Scale appropriate modelling of diffuse microbial pollution from agriculture

David M. Oliver¹*, A. Louise Heathwaite¹, Rob D. Fish¹,³, Dave R. Chadwick², Chris J. Hodgson², Michael Winter³ and Allan Butler³

¹ Centre for Sustainable Water Management, Lancaster Environment Centre, Lancaster University, Lancaster, UK, LA1 4YQ
² North Wyke Research, Okehampton, Devon, UK, EX20 2SB.
³ Centre for Rural Policy Research, Department of Politics, University of Exeter, Amory Building, Exeter, Devon, UK, EX4 6RJ.

*Corresponding author: David M. Oliver (d.m.oliver@lancaster.ac.uk). Tel: +44 (0)1524 510231, Fax: +44 (0)1524 510217.
**Abstract:** The prediction of microbial concentrations and loads in receiving waters is a key requirement for informing policy decisions in order to safeguard human health. However, modelling the fate and transfer dynamics of faecally-derived microorganisms at different spatial scales poses a considerable challenge to the research and policy community. The objective of this paper is to critically evaluate the complexities and associated uncertainties attributed to the development of models for assessing agriculturally derived microbial pollution of watercourses. A series of key issues with respect to scale appropriate modelling of diffuse microbial pollution from agriculture are presented and include: (i) appreciating inadequacies in baseline sampling to underpin model development; (ii) uncertainty in the magnitudes of microbial pollutants attributed to different faecal sources; (iii) continued development of the empirical evidence base in line with other agricultural pollutants; (iv) acknowledging the added-value of interdisciplinary working; and (v) beginning to account for economics in model development. It is argued that uncertainty in model predictions produces a space for meaningful scrutiny of the nature of evidence and assumptions underpinning model applications around which pathways towards more effective model development may ultimately emerge.

**Keywords:** diffuse pollution, end-user, faecal indicator organism, modelling, pathogen, scale, stakeholder, uncertainty

---

**I  INTRODUCTION**

Modelling the fate and transfer dynamics of faecally-derived microorganisms at different spatial scales poses a considerable challenge to the research and policy community. Current understanding of both spatial and temporal variability of faecal microorganisms in agricultural catchments is, at best, partial (Crowther *et al.*, 2003).
As researchers investigate a series of spatial scales, from replicated plot experiments investigating *E. coli* emergence in runoff through to monitoring the export of faecal bacteria from different farm operation areas, to the development of pathogen budgets at the catchment scale, it is well recognised that research comes packaged with a series of inherent uncertainties or limitations (Wagenet and Hutson, 1996; Sivapalan, 2003; Haygarth et al., 2005; Beven et al., 2006; Corwin et al., 2006).

Framed by the needs of ‘evidence-based’ policy, it is important that scientists not only understand and acknowledge these uncertainties in their work, but devise approaches for knowledge exchange that ensure that model limitations remain transparent to end-users. This requirement to acknowledge uncertainty derives, in part, from misconceptions about the capacity of models to fully capture and represent biophysical processes. Not least is the danger that, at the point of policy application, apparently ‘realistic’ constructions of the material world may be confused with reality itself (Beven, 2006a; Beven, 2007; van Wyk et al., 2008). Understanding and recognising sources of uncertainty in model structures and output is particularly pertinent for models that attempt to deal with diffuse microbial pollution from agriculture, largely because the research base (of both pathogen and faecal indicator organism (FIO) studies) upon which they are built is relatively immature in contrast to other agricultural pollutants such as phosphorus (P) and nitrogen (N) (Kay et al., 2008a).

Prediction of microbial parameter concentrations and loads in receiving waters is nonetheless essential to assist planning and policy decisions, in order to protect human health (Wilkinson et al., 1995). Modelling microbial dynamics within catchments, farms or plots can therefore facilitate potentially multi-scaled forecasting of temporal and/or spatial trends. The UK policy context for such prediction largely centres largely on the EU Water Framework Directive (WFD; CEC, 2000) and
associated directives such as the revised Bathing Waters Directive (rBWD; CEC, 2006a) and Shellfish Waters Directive (CEC, 2006b). In the US, total maximum daily loads (TMDLs) are calculated under the Clean Water Act for key contaminants causing water quality impairment. The most prominent cause of impairment of river and stream water quality in the US is excessive levels of pathogen indicator bacteria (USEPA, 2009).

Currently, there are a number of significant gaps in our understanding of faecal microbe dynamics in the environment which constrain our ability to develop models relating to microbial pollution (Kay et al., 2008b, Kay and Falconer, 2008). Whilst it is widely recognised that upscaling spatial data via modelling approaches is fraught with conceptual and methodological difficulties (Sivapalan, 2003; Beven et al., 2006; Standing et al., 2007; Haygarth et al., 2005), the purpose of this paper is not to formulate a set of principles or laws that align models of microbial dynamics in the environment derived from different spatial scales. Rather, the aim is to explore the associated uncertainties and complexities attributed to the development of models for assessing agriculturally derived microbial pollution of watercourses and to evaluate critically a series of key issues with respect to scale appropriate modelling of diffuse microbial pollution from agriculture.

The various scales of modelling that are addressed in this review (and associated examples of the science and policy questions) are illustrated in Figure 1. Within these distinct scales, there will be uncertainties and assumptions unique to that specific model resolution. However, there will remain common elements of uncertainty that encompass all scales of modelling (discussed herein). Indeed, one scale of modelling may well shape the uncertainties that arise at a different scale of interest (Haygarth et al., 2005).
II MICRO MODELS OF MICROBIAL DYNAMICS

Laboratory scale experiments and associated models provide the fundamental mechanistic information on which to base subsequent hypotheses and provide the rationale for inclusion of parameters within process-based models. However, when scaling from the laboratory to the field, researchers need to be aware of the limitations of laboratory-derived data resulting from controlled experimentation since these fail to embody the depth of interaction and complexity of environmental processes operating in unison in the field (Beven et al., 2006).

Models of faecal microbe dynamics at the laboratory scale often relate to the estimation of: trapping efficiency of microorganisms within soil systems (e.g. Artz et al., 2005; Smith and Badawy, 2008); microbial release from faeces (e.g. Hodgson et al., in press); attachment interactions between microbes and soil and organic particles (e.g. Oliver et al., 2007; Kuczynska et al., 2005); or die-off characteristics of microbes within different environmental matrices (e.g. Peng et al., 2008; Oliver et al., 2006; Peleg, 2003). Such studies are valuable for informing model development because they consolidate our understanding of fundamental environmental processes. The difficulty lies in scaling-up such knowledge and advances in understanding of basic behavioural traits of FIOs and potential pathogens to improve parameterisation of on-farm risk assessment tools and models suitable for policy makers. Not least among the difficulties of scaling-up is that processes dominating microbial activity within the soil core will not emerge as key drivers of catchment scale microbial dynamics. The scaling-up of processes and functions, as noted by Standing et al., (2007) requires an understanding of the connections, linkages, non-linear relationships and feedbacks inherent to the system of study.
An example of potential scaling errors is evident when deriving microbial survival curves. Survival curve models are often generated in the laboratory under defined temperatures and controlled conditions and then applied to field and farm studies (e.g. Fig 2A), but when exposed to the variable outside environment, such idealised curves may be unrepresentative of field-relevant behaviour (e.g Fig 2B). However, field-derived microbial die-off data are scarce and few studies provide quality data to enable a comparison, although recent studies in North America and New Zealand are emerging (Muirhead, 2009; Soupir et al., 2008; Sinton et al., 2007; van Kessel et al., 2007; Meays et al., 2005). Microbial die-off studies are rarely conducted on a seasonal basis, and field investigations undertaken to provide a month-by-month assessment of microbial die-off under field-relevant conditions are especially scarce (e.g. Muirhead, 2009). Field-derived data are therefore sporadic even at the most refined scale of understanding (e.g. a dung pat) and yet uncertainties will be magnified as we scale up to derive national accounts of microbial reservoirs on land (Fig 2C). In laboratory studies, the die-off profile of faecal bacteria generally follows that of first-order decline (Peleg, 2003), and this is the traditionally accepted paradigm of approximating microbial die-off, e.g. the log-linear decline profile. However, new empirical research conducted under environmental conditions has reported on the potential for re-growth of \textit{E. coli} (Fig 2B) (Muirhead, 2009; Soupir et al., 2008; Sinton et al., 2007; van Kessel et al., 2007).

Whilst laboratory studies are useful, they can, therefore, misrepresent behavioural traits encountered in the field because of their inability to accommodate variable interacting factors and associated effects in measured parameters. The lack of representation of re-growth dynamics in laboratory derived models of microbial die-off, and the general assumption of first-order decline may equate to a quantitative
error and potential underestimation of diffuse-source microbial risks to soil and water quality. However, current modelling approaches fail to account for the observed re-growth of microbial populations during the period immediately post-defecation (Kay et al., 2007a). This is problematic given that faecal bacteria are the key indicator for regulatory bathing and shellfish water quality monitoring. Understanding their abundance is important in order to quantify farm and regional sources of terrestrial inputs to ensure programmes of measures, as required under Article 11 of the WFD (CEC, 2000) are targeted effectively. To underpin improved modelling of microbial survival dynamics, high resolution empirical data under field conditions are now needed to approximate accurately the microbial re-growth phase, which has been speculated from low frequency sampling (e.g. Sinton et al., 2007; van Kessel et al., 2007). Indeed, with microbial modelling approaches adopting first-order kinetics, it remains unclear how much of an impact the ‘ignored’ re-growth phase would have on predictions of total FIO burden to land from livestock faeces.

This example serves to underline the wider point, that laboratory based models in the first instance are able to identify key factors responsible for impacting on microbial behaviour, but exhibit limitations when transferred to real-world scenarios. In this sense, current UK modelling approaches are rudimentary because of the uncertainties in the probability distribution of behavioural variables used and the reliance on potentially erroneous algorithms in pathogen fate models. However, there is an important, and more general, point herein - these uncertainties cannot simply be reduced to parameter uncertainty because they also reflect a degree of model structural uncertainty.

III HILLSLOPE AND ON-FARM MODELLING OF MICROBIAL DYNAMICS

Microbial watercourse contamination can occur via both surface and subsurface flow pathways (Oliver et al., 2005) and so models operating at the hillslope scale need to
account for multiple hydrological pathways to understand or to predict contamination
events fully. However, it can be argued that the existence of multiple pathways that
operate variably in space and time lend themselves to an aspect of the diffuse
pollution problem that is, in itself, indeterminate. That said, there are examples of
mechanistic models that do attempt to couple surface and subsurface flow and
transport of bacteria on hillslopes (e.g. Kouznetsov et al., 2007). In scaling from the
laboratory to the hillslope environment, models are required to deal with new
landscape dimensions that, whilst relatively simple (relative to a catchment), allow for
parameter values to become less constrained. An increase in uncertainty within the
modelled system will also need to be recognised because of the potential for
discontinuity of flow at larger scales due to processes such as infiltration, and
through heterogeneities linked to micro-topography and macropore flow (e.g.
McGechan et al., 2008). Despite this, the hillslope-scale model of Kouznetsov et al.
(2007) estimated faecal coliform transport on slopes under simulated rainfall within
satisfactory limits (R² values of 0.69 and 0.81 for bare clay loam and bare sandy
loam plots, respectively) of the observed data. Some studies even suggest that
complex behaviour patterns can be reduced to surprisingly low variability in model
outputs (Kirchner et al., 2000; Whitehead et al., 2009). Others, using drainage water
quality from small plots that had been grazed by livestock and received slurry
applications, developed a transport model for faecal microbes in soil systems
(McGechan and Vinten, 2003; 2004). This accommodated modifications of the
MACRO model to include processes important for E. coli fate and transport and
identified bypass flow pathways as being important for E. coli transfer through
hillslope plots. However, the authors acknowledge that while the set of parameters
used for model calibration provided a good closeness-of-fit to the experimental data,
caution must be exercised due to the large number of parameters used in their
model. In other words, the parameterisation providing this level of fit between model
prediction and observed data may not be unique to the set of values chosen. This
can result in equifinality whereby the same model, with subtly different parameter
realisations, produces equally plausible representations of environmental systems
(Beven, 2006b). This can be especially true of models accounting for process
interactions that result in parameters that can affect the form and strength of those
interactions in unexpected ways.

Hillslope modelling studies, such as the examples cited, rely on knowledge of
complex processes acquired from smaller-scale research to enhance the predictive
power of resulting models. Models that operate at the farm scale can have direct
influence on land-based decisions, and integrate knowledge derived from laboratory,
field and hillslope studies. Decision support tools such as soil management plans
(Defra, 2009) are a simplistic form of such an approach to modelling. Extension of
such basic risk mapping approaches are exemplified in concepts such as the
Phosphorus Index (PI), which incorporates simplified observations into a risk
indexing (or ranking) approach for land and nutrient management (Heathwaite et al.,
2000). Explicitly, the PI ranks the relative risk of fields contributing to the transfer of P
across the landscape. Such risk indexing approaches are not yet common for FIOs
or pathogens, though the existing frameworks for nutrients may be adopted and then
modified using adaptations, such as die-off coefficients, relevant for microbiological
parameters. Goss and Richards (2008) have argued that development of a risk-
based index of the potential for pathogens from agricultural activity to impact on
water quality is required as an interim stage in the establishment of a fully
quantitative microbial risk assessment approach. They propose the need for an
inventory of confined and diffuse microbial sources on each farm in a given
catchment. The risk of receiving waters being contaminated is then dependent on the
number and size of microbial stores, functionality and extent of hydrological
pathways and any subsequent attenuation en-route from source to receptor. Whilst
the authors provide a logical conceptualisation of the agricultural system, there
remains much complexity in converting perceptual and conceptual understanding into procedural models (Krueger et al., 2007) and this is particularly true for understanding the spatial and temporal variability in hydrological pathways that are not necessarily readily or easily generalised.

Oliver et al. (under revision) have made provisional steps in developing a field-based risk indexing approach for FIOs, drawing on the current evidence base to define source, transfer and connectivity-related risk-drivers for microbial watercourse pollution at the field scale and expert judgments to assist in ranking the relative importance of those interacting risk factors. This FIO risk index operates as an easy-to-use, flexible and adaptable tool for communicating, in a visual manner, the output of actual risk assessment to real end-users e.g. farmers or ‘risk managers’. In this respect, the FIO risk index can be used as a scenario testing tool, allowing changes in potential risk to be calculated based on changes in farm management such as reducing manure application rates, changing manure application methods or preventing the application of slurry to high risk fields. Embracing the PI approach to suit microbial contaminants in agricultural systems provides an example of how scientists can develop and modify risk management approaches by learning from research conducted for one pollutant (e.g. P) and transferring concepts to other pollutants (in this case FIOs). Oliver et al. (in review) present an initial test of their risk index but warn that such on-farm evaluation is challenging and that diffuse pollution makes ‘whole farm’ validation of such risk assessment tools difficult because of uncertainties in pollutant sources and pathways.

When attempting to conceptualise whole-farm systems, socio-economic drivers of risk are often not explicitly accounted for within model building exercises. Chadwick et al. (2008) outlined the case for an interdisciplinary approach for assessing microbial watercourse risks from livestock farming and the rationale for inclusion of a
series of natural, social and economical risk factors in the development of a cross-
disciplinary toolkit for reducing FIO loss from land to water has been detailed by
Oliver et al. (2009). In this toolkit approach, four key overarching risk criteria were
identified as dictating FIO loss from grassland farm enterprises, namely:
accumulating microbial burden to land; social and economical obstacles to taking
action; landscape transfer capacity; and farm infrastructure. The design of the toolkit
was such that end-users were able to interpret the interplay and importance of
different farm scale elements in order to identify where mitigation strategies would be
most effective with respect to reducing FIO loss from the farm enterprise to water.
Similarly, researchers in New Zealand have developed whole farm models for
agricultural systems with integrated economic sub-models (Monaghan et al., 2008).
This latter approach not only recognised that economics must be an integral
component of modelling frameworks tasked with optimising the adoption of mitigation
options on-farm, but was also designed to prioritise mitigation options using a holistic
approach for a suite of contaminants, including FIOs. Others have attempted to
account for uncertainty in the costs and effectiveness of measures to limit microbial
pollution of watercourses using a model based on statistical analysis and expert
judgement (Brouwer and de Blois, 2008). The interaction between environmental and
economic uncertainty was complex and driven by the array of direct and indirect
costs attributed to different actors in the catchment. Alternative strategies for
modelling the economic impact of agricultural pollution mitigation are exemplified by
‘cost curve’ approaches as evidenced for phosphorus research (e.g. Haygarth et al.,
2009) whereby an analysis of the potential effectiveness of mitigation measures (e.g.
the reduction of mass of P transferred) and potential cost (in terms of GB pounds
sterling £) to the farming industry is undertaken. With continued development of the
FIO and pathogen research evidence base, such approaches are likely to have
considerable relevance for microbial pollutants, too.
The need for holistic models of multi-pollutants is now critical because they reflect important agricultural issues such as pollution swapping and ecosystem service swapping (Stevens and Quinton, 2009a; 2009b). Emerging research is now proposing the need for new model frameworks that can explore the effectiveness of diffuse pollutant mitigation options for multiple pollutants (including those designed for pathogens). Undoubtedly, such tools are high on the agenda for policy-makers who are required to identify which mitigation options are likely to reduce the target pollutant without increasing impact from others (Chadwick et al., 2006; Cuttle et al., 2007; Stevens and Quinton, 2009a).

IV MODELLING CATCHMENT MICROBIAL DYNAMICS

European legislation in the form of the rBWD and the Shellfish Waters Directive have driven an emerging imperative for the development of catchment models capable of predicting FIO fluxes (Kay et al., 2008b). However, basic information on microbial dynamics in different catchment compartments is often inadequate (Kay et al., 2007a). As a result, models use extrapolated data, often in coarse formats with associated assumptions and simplifications averaged across the spatial resolution of the dataset. This can, as a consequence, produce uncertain model output. While FIO based modelling has developed, it must be recognised that new information, in the form of catchment scale FIO export coefficients, is required for continued improvements in the modelling of diffuse microbial pollution (Kay et al., 2008a).

Insert Fig 3 here

Catchment scale models provide a means of identifying key regions of microbial water quality concern within a hydrological boundary. They also help elucidate critical broad-scale land areas contributing pollutants to watercourses within catchments and particular ‘at-risk’ reaches of drainage systems. Finer scale
information is often aggregated into a defined spatial area, such as a 1km$^2$ grid cell to
create an overall catchment map of distributed spatial risk (Fig 3 A, B, C,
respectively). Some researchers have attempted to include both source and transfer
components into catchment models of microbial pollution. For example, Tian et al.
(2002) developed a model of watershed microbial contaminants on grazed farmlands
and represented the catchment via a series of cells, within which three process
interactions were combined to represent: (i) bacterial inputs from grazing animals; (ii)
removal of bacteria by diffusion; and (iii) bacterial reduction due to attenuation. This
strategy is complementary to the farm scale approach outlined by Goss and Richards
(2008) in their conceptualisation of a pathogen index via compartmentalising an
agricultural system. Similarly, Collins and Rutherford (2004) describe a dynamic
process based model that was developed to predict *E. coli* concentrations in streams.
This model structure was based on a mass balance approach conserving the ‘mass’
of *E. coli*, and was not unlike that of the earlier model described by Tian et al. (2002).
The former model was not developed to derive accurate, high frequency predictions,
but rather, to assess the relative impact of mitigation measures at a catchment scale.
In fact, this work attempted to link catchment and in-stream processes for a more
holistic catchment model (as advocated by Jamieson et al. 2004). Flow accumulation
maps were derived within a GIS to calculate the number of upslope cells that
potentially contributed to surface runoff. The identification of dominant flow paths for
incorporation in hydrological models approximating contaminant transfer has been
acknowledged in other research (Quinn et al., 2004), and enables a stronger
weighting to be given to bacterial delivery in cases of dominant flow. However, the
approach adopted by Collins and Rutherford (2004) was limited because they did not
account for hydrological connectivity from land to water. Instead each cell within their
rasterized model structure was assigned a water yield which was routed as a direct
input to the nearest stretch of stream. They also highlighted other sources of
uncertainty, in that many of the assumptions upon which the model was based were
not validated. For example, a detailed history of grazing was used to calculate *E. coli* burden received by each field in a catchment on a daily basis. This can introduce a series of uncertainties because assumptions will be made with regard to which fields are grazed at any one time and also livestock grazing densities, faecal excretion rates, shedding rates of bacteria from different livestock types, dry matter content of faecal deposits, and microbial die-off coefficients for different livestock faeces and manures. All of these are likely to accommodate a particular range of error, which itself may also exhibit temporal variability.

Thus, predicting the *E. coli* burden deposited on land from a defined number of livestock each shedding $X \times 10^Y$ cells in $Z$ g faeces per day and dying off at a constant rate of $A$ cfu day$^{-1}$ (despite UV variations and temperature fluctuations on diurnal timesteps), would produce a highly regular model response (e.g. Fig 4A). In fact, the bacterial burden would increase to a point of equilibrium whereby exponential die-off would balance fresh faecal additions and whereupon the removal of livestock at the end of the grazing season would result in the *E. coli* burden declining exponentially. Such scenarios have been used by others (e.g. Oliver *et al.*, 2009) and provide an indication of the magnitude of microbial burden deposited on land rather than a truly representative model of real-world scenarios given the potential for environmental variables to impact on microbe behaviour. A more realistic scenario would be one whereby the day-to-day shedding of *E. coli* by cattle is correlated as $W$ (where $W$ is a value $< 1.0$, i.e. variability in the excretion of cells will occur), die-off coefficients are assumed to accommodate a degree of error (e.g. $X\%$ from the average values chosen as first-order approximations) due to interacting environmental variables, and shedding may be $+\text{ or } - Z$ orders of magnitude around the expected mean shedding rate, for example. The output of the simplistic ‘crisp’ model shown in Figure 4A is reproduced in Figure 4B but with an incorporated day-to-day shedding correlation of 0.7, a 33% error in die-off coefficient used and a $+\text{ or } -$
one order of magnitude uncertainty associated with microbial load of faeces. Figure 4B also includes uncertainty boundaries around the predicted data showing the 5th and 95th percentile values for modelled predictions to provide a more realistic estimate of *E. coli* burden on land.

The inclusion of microbial die-off in models of catchment microbial dynamics can prove important but is, nevertheless, debated in the literature. It is a biological trait that differentiates microbial models from other agricultural pollutant models, and in one catchment scale model, the output consistently over predicted faecal coliform delivery to watercourses because it did not account for enhanced die-off resulting from detrimental environmental factors, such as sunlight (Fraser *et al.*, 1998). In contrast, other research has suggested that catchment models can perform well but be simplified by excluding die-off parameters within model structures (Haydon and Deletic, 2007).

Insert Fig 4 here

Jamieson *et al.* (2004) provided a synopsis of current catchment scale models of FIO pollution and identified that large scale models tend to exclude subsurface flow contributions, instead accommodating only overland flow as primary hydrological pathway of microbe transfer. This is interesting, and contrasts with the hillslope scale models discussed earlier whereby attempts are made to include a variety of pathways and accommodate more detail at smaller spatial areas. On hillslope plots, the role of matrix flows transferring water and contaminants into throughflow pathways is more important than at the catchment scale where the importance of overland flow may override other pathways. The key point is that dominant processes change as we explore different scales. However, Haydon and Deletic (2006) report on the development of a coupled pathogen-hydrological catchment model which
makes an attempt to estimate pathogen discharges from large catchments. The authors detail two conceptual models. First, a description of surface and subsurface pathogen transport processes by means of wash-off and loss equations are incorporated and coupled with an existing hydrological model that predicts flows. Second, a basic model accommodating only surface transport processes was coupled with a stormflow and baseflow separation model. Evaluation of the models using three catchments suggested that the more complex model performed better with regard to the observed data, but as discussed earlier, caution should be exercised with complex models due to the potential for equifinality, arising from multi-parameter models (Beven, 2006b).

Simpler approaches to catchment modelling do exist. Crowther et al. (2001) described a desk approach considered to be a cost effective exploratory management tool for non-compliance of beaches in the UK. This functioned through a regression modelling approach and was used to provide preliminary insight into factors impacting on FIO concentrations at coastal locations. While the use of coarse scale data is rarely adequate, Crowther et al., (2003) found that a high level of explained variance of high flow FIO concentrations were obtained in multiple regression models even when using only basic data on land use type; thus, successful predictive models can be developed without detailed information on stocking density, grazing patterns and animal waste applications. Similarly, Kay et al. (2005), in the Ribble catchment in North West England, focused on land use relationships with FIO concentrations in surface waters of the catchment. This study adopted a stepwise regression modelling approach to examine relationships between GIS datasets such as elevation, land cover and digital map data and microbial water quality determined via a targeted monitoring programme. The results produced statistically significant models to predict geometric base and high flow FIO concentrations from land use variables. More recently, FIO budgets for the same
catchment have been derived to provide an assessment of source apportionment in the Ribble basin (Stapleton et al., 2008) building on the previous land cover model developed by Kay et al. (2005). Regression can be a useful tool for identifying pollution ‘hotspots’ because of the ability to identify differences between observed and predicted data and to explore spatial patterns across catchments. Critically, these simple approaches work well through an appreciation of the limitations of coarse scale data and its integration with complementary ground-truthed data.

Additional benefits have been described through the use of multiple models to simulate catchment scale microbial fate and transfer, whereby improved prediction capacity (and therefore incorporated fate and transfer processes) can be identified for the catchment under investigation by comparing a range of model outputs (Chin et al., 2009). While the focus of this discussion has been on diffuse microbial pollution from agriculture, it is important to acknowledge that, at subcatchment and catchment scales, there is a need to incorporate non-agricultural sources of FIOs such as waste water treatment plants and combined sewer overflows into model design (Petersen et al., 2009). This can then lead to further debate relating to source apportionment of faecal microorganisms detected in the environment.

New methodologies and improvements in microbial source tracking (MST) sciences can feed into models in an attempt to improve, or constrain, uncertainties of microbial origin (e.g. Simpson et al., 2002; Meays et al., 2004; Stapleton et al., 2007). Of course, MST only provides a ‘snap-shot’ of the source contribution at the time of sampling and this should be borne in mind because contributions from varying sources may alter considerably within a matter of hours. Some improvements in the reliability of faecal coliform TMDL modelling have been reported through integration of MST (Hyer and Moyer, 2004). In this study, MST allowed for an apportionment of the contribution of different animals, but the whole process made the authors appreciate the variability (> 4 or 5 orders of magnitude) in FIO density per animal.
type source, and raised awareness of the ease by which a source could be under or
over represented within the model. Clearly, MST has the potential to enhance model
development, but is also in itself subject to associated uncertainties. Currently, it
offers a complementary component to aid a suite of catchment wide monitoring and
modelling approaches (Reischer et al., 2008).

V EVIDENCE BASE LIMITATIONS

Limitations in the modelling capacity of diffuse microbial pollution from agriculture can
arise from the use of nationally available datasets (e.g. land use data; Castilla and
Hay, 2007) as well as inadequacies in the empirical evidence base due to the
relatively immature understanding of mechanistic processes of diffuse microbial
pollution. A series of emerging key issues are outline below, stemming from this
assessment of scale appropriate modelling of diffuse microbial pollution from
agriculture.

Issue 1: Inadequate baseline sampling to underpin model development

Inadequacies in national datasets of routine baseline microbial water quality (see Kay
et al., 2007b) hinder the ability to validate models across catchments and this is
problematic in the context of UK microbial modelling research. The problem is
exacerbated because routine baseline monitoring is not seen in a favourable light
with regard to funding bodies: it is not considered as ‘cutting edge’, yet remains
important for underpinning credible model development (ERFF, 2008). Science can
only progress in terms of modelling so long as data are there to support, or indeed
reject, predictions (Beven, 2007). The quality and availability of microbial data at any
scale will play a considerable role in governing model development. For example,
detailed mechanistic models of microbial pollution of coastal waters will be
inappropriate in a data scarce catchment (Silgram et al., 2008). Beven (2007)
comments that modelling should be treated as a learning process, but that learning facilitated through modelling mechanisms can only proceed with continued data assimilation. In the UK (and elsewhere), there is a need to begin forming reliable databases of microbial water quality data. At present, the scarcity of microbial water quality data in standardised formats is highly problematic.

Issue 2: The sources and magnitudes of FIOs and pathogens are relatively unknown

Problems of modelling faecal microorganisms in catchments relate to the plethora of sources within the environment and their variability in space and time (Dorner et al., 2007). Some believe that the greatest uncertainty in model predictions can be associated with partitioning of contaminants between various catchment related sources (Silgram et al., 2008). FIOs can originate from livestock, from humans via sewerage overflows or from leaking septic tanks. In some instances, farmers spread septic tank waste to land, or in worst-case scenarios, do not even have infrastructure in place for a septic tank to hold domestic waste water. Additionally, the congregation of birds in key locations such as trees that line river corridors, or coastal structures such as piers, can contribute to the deterioration of river and bathing water quality through their direct deposition of faeces into water. Wild animals will also provide a further source of FIOs to the landscape – the true magnitude of which is under researched, though some have reported on microbial quality of wilderness areas (Derlet, 2008). Overall, therefore, the relative magnitudes of all FIO and potential pathogen sources are relatively unknown (Hyer and Moyer, 2004). This is even before we begin to understand comprehensively the die-off characteristics of FIOs sourced from different species and deposited in different faecal matrices in the environment.
Technologies such as MST are beginning to offer improved approaches that can help ascertain sources in catchments, but catchments are rarely compartmentalised into neat urban and rural regions. In addition, there is the need for better characterisation of faecal inputs to land (both volume and distribution), but this is difficult to gauge at the catchment scale (Tate et al., 2003). This contrasts with the farm scale, whereby specific livestock numbers can be determined (through projects embedding within farm enterprises with farmer co-operation and participation) and assumptions can be made with respect to microbial shedding. However, even then, we often have a lack of data to reflect and to capture spatial (and temporal) distributions of livestock at the farm scale. At the catchment scale, the distribution of livestock is more difficult to ascertain, as are manure applications to land in space and time. Confidentiality issues restrict individual farm details being released across catchment areas, and data are generally only accessible from the agricultural census which itself is confounded by the fact that data are collected at the scale of the farm holding but results are used at various aggregate levels. Moving beyond FIOs, the complexity increases when concerned with specific pathogens. As noted by Ferguson et al. (2007), reliance on FIOs may lead to either an under or over estimation of pathogen risks because FIOs and pathogens exhibit different fate and transport characteristics in the environment.

Issue 3: Gaps in the empirical evidence base

There is a need for the continued and targeted growth of the empirical evidence base of diffuse microbial pollution to underpin model development for all scales. At the same time, researchers must balance the generation of new data with model requirements. Difficulties do exist in constructing large scale microbial water quality models because of the limited availability of a fundamental empirical evidence base. These problems are magnified when dealing with specific pathogens rather than indicator bacteria, because there is even less robust spatial and temporal data to
support model development. This reflects the increased costs, time and skills required to enumerate specific pathogens in contrast to FIOs, and also nutrients. Furthermore, the natural spatial variability of microbial contaminants is very high, so that when coupled with costs of analysing for pathogens in the environment, it can raise some key problems for modelling (Pachepsky et al., 2006). Researchers in Australia have attempted to develop process based deterministic catchment scale models (Ferguson et al., 2004, 2005, 2007) but emphasize that application of their model is constrained by the availability of appropriate data and the series of built-in assumptions inherent to such large scale budget models. This reiterates the need to take a precautionary approach when considering highly complex deterministic models. Naturally, researchers are keen to accommodate more scientific understanding into model development, but this can lead to models with many parameter values that are difficult to measure, and thus result in model equifinality as described in Sections III and IV. Models which integrate a large number of parameter values are at risk of providing output which is highly uncertain because we cannot necessarily measure input data in situations representing real circumstances (Beven, 2007). The use of poorly defined parameter ranges within models, or parameter abuse (Hamilton, 2007), must be borne in mind as we continue to develop models of diffuse microbial pollution. One strategy we do have to help consolidate evidence base limitations is to accommodate expert data (e.g. Fish et al., 2009; Giannetti et al., 2009) within model development. Many modelling approaches do make use of expert-derived weightings and coefficients. However, accommodating surrogate measures through expert data is an experimental and novel area of research. Thus, the technique raises it own particular uncertainties, not least how to reconcile heterogeneous understandings of the problem at hand (Fish et al., 2009). Developing a standardized set of techniques for generating these surrogate measures is, perhaps, a necessary next step (e.g. EA, 2008).
Undoubtedly, conceptual frameworks of models are constrained by existing
knowledge and data, but this applies to most models of diffuse pollution. While
embedding expert knowledge into decision support tools may allow for vulnerability to
error from the expertly-derived coefficients, precautionary steps can be taken to
appreciate how influential the uncertainties which arise may be. For example, as
detailed by both Fish et al., (2009) and Haygarth et al. (in press), it is wise to
undertake an uncertainty analysis to determine the responsiveness of modelled
output to changes in individual, expert-derived coefficients.

Issue 4: Appreciating the added-value of interdisciplinary working
Interdisciplinary modelling frameworks that recognise the importance of integrating
science and social science (and indeed multiple-pollutants) represent an important
shift for more holistic models at farm and catchment scales. Such approaches are
beginning to appreciate the importance of coupling socio-economics with landscape
characteristics and environmental processes, but also for increased integration at the
interface between scientific disciplines and policy (Macleod et al., 2008). This is
clearly important for scenario modelling of mitigation options designed to reduce
microbial impairment of watercourses; different socio-economic circumstances will
favour or hinder the adoption of new management approaches on-farm. However,
the scales of operation of different disciplines do not always align in a convenient
fashion (Stevens et al., 2007). Social processes and structures often imply ‘farm
scale’ risk, but insight is often qualitative, site specific and indirect. This can raise
interesting debates as to whether this is a problem arising from the nature of social-
science data or alternatively an inappropriate approach to modelling the system.
While therefore presenting a challenge for the modelling community, interdisciplinary
modelling frameworks must now be seen as a significant step forward. By situating
models within their wider social and economic context, we can begin to modify and
accommodate variations in, for example, grazing routines. These are generally
represented equally throughout model spatial scales and yet fail to represent farmer
behaviour in the sense that some farmers will adopt best management practices
whereas others fail to embrace successful land and water stewardship (LaWare and
Rifai, 2006). Incorporating such complexity in models will not be easy, but by
beginning to acknowledge and truly account for the added-value of interdisciplinary
approaches within model design, we will make considerable progress.

Issue 5: The importance of economics

Models embracing both an appreciation of the inherent uncertainties associated with
environmental microbial dynamics and complexities relating to economics of
mitigation and health impacts are currently limited but are likely to be developed and
used increasingly for modelling management scenarios. Some research has
attempted to assess FIOs impacts on human health based on non-market
approaches to economic valuation such as “willingness to pay” methods (Johnson et
al. 2008), and has thus modelled the value of health benefits to society of reducing
FIO loading to bathing waters. This also serves to reinforce the need for
interdisciplinary approaches to managing microbial risks from agriculture as outlined
in Issue 4. The ‘cost curve’ approach (e.g. Haygarth et al., 2009) provides a
preliminary framework for identification of the key nutrient mitigation methods, and
illustrates areas of research where we currently lack knowledge. At the present time,
research in the area of FIO fate, transfer and mitigation is immature with respect to P
and N but the ‘cost curve’ methodology provides a template that could be applied to
FIO (or pathogen) mitigation options as further empirical evidence emerges. It is
worth emphasising that socio-economic drivers impact on all scales and aspects of
modelling, whether it be determining how farm debt impacts on management
decisions or budgeting for the number of samples to analyse in the laboratory for
robust model evaluation.
V1 CONCLUSIONS

Haygarth et al. (2005) recognise that we need to appreciate the relative strengths and weaknesses of both small scale laboratory studies and the complexity of unreplicated large scale research. The majority of the key issues outlined in this review for modelling microbial pollution from agriculture tend, at first glance, to apply to larger scale modelling approaches. This perhaps reflects the increased complexity associated with coarse scale modelling. But underneath the obvious linkages between the ‘issues’ and ‘scales’ discussed, there lie an array of inter-related points of interest that cut across many scales (e.g. socio-economic impacts in the laboratory through to the farm and catchment, multi-scale interdisciplinary working).

Advances in modelling approaches have been made and include, for example, the development of user-friendly graphic interfaces, stakeholder participation (Newham et al., 2006) and use of GIS layers and databases. However, basic approaches have changed little. In contrast, it is our understanding of processes through combined modelling and experimentation which has advanced, and so up-to-date knowledge and data need to be continually incorporated into existing and developing model structures. Models provide frameworks within which future research agendas can be prioritised. This should not translate into an absolute need to parameterise model frameworks with all available knowledge and understanding gained from smaller scale studies. Simplistic tools capable of gauging relative implications of particular scenarios are equally as useful (Strauss et al., 2007).

Uncertainty in the modelling of microbial pollution from agriculture is, perhaps, an interesting challenge for scientists but an inconvenient truth for policy makers. Our interest in uncertainty primarily derives from the need to move from laboratory to field to estuary scales for effective risk-based decision-making using appropriate modelling approaches. For the end-users of decision support tools, recognising
uncertainty introduces new problems of new complexions, for with this term doubt is introduced. In other words, models that are considered ‘uncertain’ may also be considered ‘inadequate’ and this may well prove problematic in circumstances where mitigation options based on model runs imply the investment of time and capital. Even so, in a wider sense, we suggest that uncertainty produces a space for meaningful scrutiny of the nature of evidence and assumptions underpinning model applications. From this, more effective model development may ultimately emerge. As such, we need to develop effective ways of communicating and debating uncertainty between different users of scientific research, from model programmers through to end-users of the emerging software and tools. This means, in short, understanding “users” of models as active participants, experts even, in fostering the acceptability of models at the point of their application. Engaging with uncertainty is a necessary political reality and a healthy component of constructing models of microbial dynamics at all scales. In fact, it may yet constitute a paradigm shift.

Acknowledgements
This work was funded as part of project RES-224-25-0086 by the Rural Economy and Land Use (RELU) programme who receive financial support from the Biotechnology and Biological Sciences Research Council (BBSRC), the Natural Environment Research Council (NERC), the Economic and Social Research Council (ESRC), the Department for Environment, Food and Rural Affairs (Defra) and the Scottish Government. The authors are also grateful for the valuable contribution of participants at the ‘Dealing with Uncertain Data’ workshop, held at Lancaster University in March 2007 and for the useful reviewer comments that helped improve the manuscript.

REFERENCES


McGechan, M. B., and Vinten, A. J. A. 2003: Simulation of transport through soil of 
E. coli derived from livestock slurry using the MACRO model. *Soil Use and 
Management* 19, 321-330.

model for assessment of best management practices for livestock farming. 
*Biosystems Engineering* 99, 292-303.

Meays, C. L., Broersma, K., Nordin, R., and Mazumder, A. 2005: Survival of 
*Escherichia coli* in beef cattle fecal pats under different levels of solar exposure. 

Meays, C. L., Broersma, K., Nordin, R., and Mazumder, A. 2004: Source tracking 
fecal bacteria in water: a critical review of current methods. *Journal of Environmental 
Management* 73, 71-79.

Monaghan, R. M., de Klein, C. A. M., and 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.

Muirhead, R. W. 2009: Soil and faecal material reservoirs of *Escherichia coli* in a 

participation in modelling for integrated catchment assessment and management: an 

Oliver, D. M., Fish, R. D., Hodgson, C. J., Heathwaite, A. L., Chadwick, D. R., and 
Winter, M. 2009: A cross-disciplinary toolkit to assess the risk of faecal indicator loss 
from grassland farm systems to surface waters. *Agriculture, Ecosystems and 
Environment* 129, 401-412.

Oliver, D. M., Page, T., Heathwaite, A. L., Hodgson, C. J., Chadwick, D. R., Fish, 
R. D., and Winter, M. under revision: Development and testing of a risk indexing 
framework to determine field-scale critical source areas of faecal bacteria on 
grassland. *Environmental Modelling and Software*, accepted with revision.

Oliver, D. M., Clegg, C. D., Heathwaite, A. L., and Haygarth, P. M. 2007: 
Preferential attachment of *Escherichia coli* to different particle size fractions of an 

Oliver, D. M., Haygarth, P. M., Clegg, C. D., and Heathwaite, A. L. 2006: 
Differential E. coli die-off patterns associated with agricultural matrices. 
*Environmental Science and Technology* 40, 5710-5716.

of *Escherichia coli* to water from drained and undrained grassland after grazing. 
*Journal of Environmental Quality* 34, 918-925.

Pachepsky, Y.A., Sadeghi, A. M., Bradford, S. A., Shelton, D. R., Guber, A. K., 
and Dao, T. 2006: Transport and fate of manure-borne pathogens: modeling 
perspective. *Agricultural Water Management* 86, 81-92


Sivapalan, M. 2003: Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection? *Hydrological Processes* 17, 1037-1041.


**Figure captions**

**Figure 1:** Distinct scales of microbial modelling research and associated examples of the science and policy questions to be addressed at those scales.

**Fig 2:** A first order die-off at the single pat scale and scaled up to the farm scale; **B** dynamics of re-growth and errors relative to first order die-off (dashed line); **C** potential re-growth variability (y) at the pat scale and implications when scaled for farm and regional burden.

**Fig 3:** The aggregation of finer scale datasets and knowledge into defined spatial resolution (1km² grids) for modelling and representation at the catchment scale.

**Figure 4:** **A**, predictions of *E. coli* numbers input to land from grazing cattle during a typical grazing season (April to October) using defined single parameter values in the model code for microbial die-off, shedding and excretion; **B**, alternative predictions of *E. coli* numbers input to land for the same time period using multi-parameter values for microbial die-off, shedding and excretion. Solid line shows 50\textsuperscript{th} percentile and dashed lines show 5\textsuperscript{th} and 95\textsuperscript{th} percentile values for modelled predictions.