Effects of agricultural land management changes on surface water quality: a review of meso-scale catchment research

Melland, A.R[1,2]., Fenton, O[3*]., Jordan, P[4,2].

[1] National Centre for Engineering in Agriculture, University of Southern Queensland, Toowoomba, Qld 4350, Australia.

[2] Agricultural Catchments Programme, Teagasc, Johnstown Castle, Wexford, Ireland.

[3] Teagasc, Environmental Research Centre, Wexford, Ireland

[4] School of Geography and Environmental Sciences, Ulster University, Coleraine, N. Ireland.

*Corresponding author: owen.fenton@teagasc.ie

Short title

Outcomes of water quality mitigation practices in agriculture

Abstract

Measuring the environmental impacts of agricultural practice is critical for policy formulation and review, including policies implemented to improve water quality. Here, studies that measured such impacts in surface waters of hydrologically diverse meso-scale catchments (1-100 km^2) were reviewed. Positive water quality effects were measured in 17 out of 25 reviewed studies. Successful farm practices included improved landscape engineering, improved crop management and reductions in farming intensity. Positive effects occurred from 1 to 10 years after the measures were implemented, with the response time broadly increasing with catchment size. However, it took from 4 to 20 years to

confidently detect the effects. Policy makers and scientists should account for these hydrological and biogeochemical time lags when setting policy and planning monitoring in meso-scale catchments. To successfully measure policy effects, rates of practice change should also be measured with targeted water quality parameters.

Keywords

Mitigation, measures, BMP, agriculture, management practice, water quality, catchment, nitrogen, phosphorus

Highlights

- In reviewed catchments, it took 1 to 10 years for policies to have a measurable effect on water quality
- Positive mitigation effects on surface water quality took 4-20 years to measure
- Time lags explain why positive effects aren't always evident within governance cycles

1. Introduction

Agricultural management practices that can effectively mitigate against on and off-farm surface water quality degradation have been demonstrated at field (**Smith et al., 2001; Melland et al., 2016;**), hillslope (**Freebairn et al., 2009;** Sousa et al., 2013) and micro catchment scales (**McDowell et al., 2009**; Melland et al., 2014; Tomer et al., 2014). In contrast, the effectiveness of farm practice change for water quality improvement at larger scales is less clear (Fenton et al., 2011; Vero et al., 2017). Policy makers need to be informed about the spatial and temporal links between field-scale land management and national-scale water quality in order to develop appropriate policies, to justify expenditure on policy implementation and to promote policy implementation (Roberts et al., 2014; Minella et al., 2008; Collins et al., 2008). **Herein, we review the outcomes of studies that have directly measured impacts of agricultural mitigation measures in medium, or meso-scale, catchments (1-100 km^2, incorporating 1st – 3rd order streams and representing a scale between farm and river basin scales) over the last 20 years.** We use this scale to incorporate the scale of statutory water quality monitoring in rivers while also the link between farm scale and catchment.

Such meso-scale studies are limited in the literature due to the challenging and resource intensive nature of this type of study (Melland et al., 2014). The challenges include the uncertainty in cause-effect relationships due to the complexity of hydrological, climatic, biogeochemical and anthropogenic processes occurring in time and space, and this often results in insufficient collection of water quality and land management information (Cherry et al., 2008). These constraints are compounded by the long periods of time that are normally needed to identify trends and account for time lags in water quality response to, and implementation of, mitigation measures (Meals et al., 2010; Spooner et al., 1987).

When considering hydrological and biogeochemical time lags for nitrogen (N, longer residence times associated with mainly subsurface losses) and phosphorus (P, lower residence times associated with mainly surface losses) within meso-catchments it may not always be possible to document residence times or give detailed data pertaining to e.g. redox conditions. Furthermore, P losses also occur via groundwater and N losses along surface pathways. For the purposes of the present study, permeability, with respect to the soil-subsoil-bedrock continuum, was used as a guide to establish which pathway dominates (Table S1-2). Such a proxy, although not quantitative, can assign dominant pathways of loss, attenuation capacity and highlight if receiving surface waterbodies are dominated by flows derived from surface or groundwater (Fealy et al., 2010). For example in meso-scale catchments (Mellander et al., 2014) dominated by imperfect or poorly drained soils the dominant loss pathway will be through surface and shallow subsurface pathways (e.g. lateral movement of infiltrating and shallow groundwater due to low permeability layers such as fragipans or artificial drainage systems) (McDaniel et al., 2008; Shore et al., 2013). In well or excessively-drained equivalents subsurface pathways will dominate but the hydrogeochemistry of the system may vary in terms of dissolved oxygen, electrical conductivity and bacterial energy source availability which in turn may attenuate or enhance nutrient flows via those subsurface pathways. For example, McAleer et al. (2017) examined two well-drained catchments with contrasting

subsurface lithologies (slate versus sandstone). Physical factors, including agronomy, watertable elevation and soil-subsoil-bedrock permeability, all influenced the hydrogeochemical signature of the aquifers. Stream nitrate (NO_3^-) load was 32% lower in the sandstone catchment even though agronomic nitrogen (N) inputs were substantially higher than the slate catchment. Therefore, the dominance of surface or groundwater pathways within a catchment and the residence time and geochemistry associated with these pathways must be considered when assessing the efficacy of practice(s) on water quality. In terms of N and biogeochemical lags, soil organic N in the source zone is influenced by the source zone NO_3^- concentration, legacy organic N depletion rate constant, mean annual recharge, soil saturation and soil porosity (Van Meter et al., 2015; Ascott et al., 2017 (defined as NO_3^- storage in the Vadose zone)). Outside of the source zone the transformation rate of NO_3^- in the subsurface is important e.g. the denitrification rate in subsoil, subsoil-bedrock interface and in bedrock (Jahangir et al., 2013). In terms of dissolved reactive P it is the chemistry of the soil-subsoil-bedrock continuum and the redox conditions that cause retention or mobilisation of P (Daly et al., 2017). In terms of the subsurface hydrological time lags, which involve mainly dissolved forms of N and P in the unsaturated and saturated zone, parameters such as residence time from the sampling point to the catchment outlet, the physical properties of the underlying aquifer and the overall hydraulic gradient pushing this migration is important (Van Meter et al., 2015; Vero et al., 2017). Further complications to conceptual models of nutrient transport can be encountered in groundwater-dominated karst environments where the concentration, load and residence times across different subsurface pathways (conduit versus different fracture sizes) can vary greatly as

demonstrated by Fenton et al. (2017) using high resolution loadagraph separation techniques. Acknowledging these conceptual complexities, studies included in the scope of the present review were those that directly measured chemical and/or biological water quality responses in surface water (lakes or rivers) to agricultural practices in meso-scale catchments.

2. Materials and methods

Studies of single, paired and multiple catchments were reviewed, with the latter being included in the review only if the median size of catchment was meso-scale. For each study, a combination of qualitative and quantitative analyses was conducted.

Quantitative analyses included assessments of the response time, the measurement time, the measurement lag (Fig. 1) and the implementation lag. These were defined as:

- **Response time was the number of years from when a threshold or maximum rate of implementation of a practice was reported or inferred to have been achieved, to when a (significant) effect on water quality was deduced to have occurred.**

- **Measurement time was the number of years taken to measure a statistically (or physically) significant water quality response to an agricultural practice and unless otherwise reported, was taken as the total length of the**

measurement period. This was usually longer than the response time because the initiation of significant water quality effects or trends was only evident or convincing once a longer time series of data was collected. The measurement time was not defined as the sum of the other terms, rather the implementation lag was defined as finishing when the response and measurement times began.

- **Measurement lag was the difference between the response time and the measurement time. Measurement lags reflect the extra time needed to measure water quality indicators in order to separate signals/responses from environmental noise and in many cases reflected a period of measurement required before a practice change occurred in order to establish a baseline. In contrast, the response time only started once full/threshold implementation of the practice change was complete.**

- **Implementation lag was the number of years between the reported or inferred initiation of practice change and when a maximum or threshold rate of implementation was reported, or inferred, to have been achieved.**

Qualitative analyses included summaries of:

- o monitoring approaches used

- classifications of effects on water quality indicators as positive, neutral or negative
- classification of positive effects according to the type of hydrological transport pathway most influencing the response of the water quality indicator
- classification of positive effects according to the type of water quality indicator as chemical (N, P, suspended sediment (SS)) or biological (diatom, macroinvertebrate, macrophyte)
- classification of drivers of practice change as mostly voluntary, mostly incentivized for research collaboration or mostly mandatory
- reasons why effects were not measurable
- reasons why negative effects occurred
- soil, geology and hydrological flow pathways and residence times

3. Results and Discussion

3.1 Monitoring approaches

Twenty-four studies from across Europe, USA, New Zealand and Brazil that matched the scope of the review were identified (Table S2). Within these, 46 different experimental approaches were used to measure the effect of agricultural practice on chemical and/or biological water quality (Table S2). The number of approaches exceeded the number of studies because multiple approaches were combined within single studies to optimise the potential for detecting significant effects and causal relationships. The most commonly used approach was a time series of data in which water quality was measured at various temporal resolutions **before and after** a significant change in agricultural practice **(e.g. Jaynes et al. 2004; Bishop et al.,**

2005; Makarewicz et al., 2009; Sutton et al., 2009). Sometimes these studies included measurements during the period of practice change (**Bishop et al., 2005**). The 'before' phase established similarity between paired catchments before practices were applied in one or more of the catchments. **For example, Jaynes et al. (2014) (Fig 1.) showed when best management practices (BMP) were implemented to limit NO_3- leaching to tile-drains, the 'response time' for effects to occur was 1.5 years. However, 3.5 years of data were needed to measure the statistically significant downward trend in NO_3- concentration (i.e. the measurement time). The statistical effect was measured by comparing water quality between catchments with and without, and before and after, practices changed.** Time-series approaches also included measurement over a period of **gradual change** in management practice (e.g. Schilling et al., 2006; Kronvang et al., 2008) and measurement over a period that was considered to reflect a **baseline** condition (i.e. when negligible change in practice was assumed to have occurred) (Wall et al., 2011). Baseline studies sometimes later became 'gradual change' or 'before and after' studies if the monitoring remained in place over a period of significant practice change. For example, five dairy catchments in New Zealand were monitored to identify baseline water quality and management practices over 3-5 years (Monaghan et al., 2009) and then later, trends over time were analysed in comparison with gradual changes in practice (Wilcock et al., 2013). Often the focus of the temporal baseline studies was to identify effects across a spatial gradient and/or through cause-effect linkages.

Multiple catchments were often used to evaluate water quality impacts of farm practice by using **paired** catchments where water quality from catchments without agricultural practice change was compared to those where change occurred over the same time series **(e.g. Jaynes et al. (2004) or Lemke et al. (2011)).** The paired catchments not up or downstream of each other had similar physical (e.g. soils, geology and/or topography) and climatic characteristics. Water quality was also monitored in two or more catchments with a **gradient** of differing physical characteristics but with similar (Melland et al., 2012) or dissimilar (Yates et al., 2007) practices. Spatially nested (i.e. longitudinally connected) catchments were also used to compare water quality **up and downstream** of a farm practice (Inamdar et al., 2002).

Identifying causal relationships between practices and water quality was an indirect outcome of most of the temporal and spatial analytical approaches and was the primary objective of the final category of studies. In these experimental designs, two or more components of the DPSIR framework were measured. This framework describes causal interactions between society and the environment and is used to identify the drivers of practice (D), the pressures (P) practices place on water quality, the state (S) of the water quality, the resulting impact (I) on values of the water, and the policy response (R) to the impact (IMPRESS, 2002). Biophysical links between the measured components were considered to be largely non-contestable or were modelled (Monaghan et al., 2007; Kronvang et al., 2008; Kyllmar et al., 2006; Wall et al., 2011).

3.2 Response times for positive effects on water quality

Positive effects on one or more water quality indicators were measured in 17 of the 25 studies reviewed. These positive effects occurred 1-10 years after practices were implemented (Fig **2**). In contrast, 4-20 years were needed to detect the positive effects on water quality (Fig **2**). The measurement lag (time between the effect occurring and the effect being measured to have occurred) ranged from 1-18 years. Not all studies had time lag information as they did not study a time element (e.g. spatial comparisons without temporal information on practice change) or there was no major practice change over the monitored timeframe (e.g. baselines studies).

Both the *response time* and the *measurement time* broadly increased with increasing median catchment size in each study (Fig **3)**. There was also a tendency for the response time to increase as the travel-time of the pathway of pollutant flow increased. For example, sediment and P transport, which occurs predominantly via the overland flow pathway had opportunities to be remediated quickly, whereas NO_3^- leached via subsurface flow pathways took longer to remediate. **Despite variation in flow travel times between catchments, the linear regression correlation coefficients between catchments size and *response time* was 0.43 (*P*<0.05), and between catchment size and *measurement time* was 0.36 (*P*<0.05), with one outlier removed for each regression.**

I*mplementation lag times* ranged from 0.5 to 14 years and tended to increase with catchment size up to about 20 km^2 (Fig 4). There was no clear association between *implementation lag time* and the policy mechanism that was used to facilitate practice change in the studied catchments. For example, when practice change was mandatory, the time between initiation of practice change and threshold or maximum rates of implementation of the practice was often longer than in cases where the practice was voluntarily adopted. Data on temporal and spatial nutrient sources and management were generally scarce in comparison with water quality data despite their importance in identifying cause-effect relationships.

3.3 Effective practices

The practices that resulted in mitigation of one or more water quality indicators were usually combinations of measures that addressed nutrient or pollutant sources, pathways, delivery and impact. Structural and cultural measures applied in smaller catchments (0.4 to 5.9 km^2), significantly reduced concentrations and fluxes of P, N and SS two years after implementation in a catchment of Lake Conesus in New York state, USA (Makarewicz et al., 2009). The measures included substantial changes to the intensity of farming including a reduction in dairy farm stocking rates and converting cropped land to perennial alfalfa. In Southern Brazil, sediment yields decreased mainly due to reduced runoff after introduction of minimum tillage and increased crop cover (Minella et al., 2008). Elsewhere water management also played a key role in mitigation. For example, reduced P concentrations and loads to the Everglades wetlands, South Florida, USA were mainly due to better management of irrigation and rainfall drainage water (Daroub et al., 2011). In this and other mitigation programs success was also attributed to high spatial rates of implementation of the effective practices (Yates et al., 2007).

Biological water quality indicators were less frequently monitored (five of the studies) than chemical and hydrological indicators and, where measured, effects on biological indicators were more often neutral than positive. Improved macro-invertebrate indicators were observed, however, after 20 years of practice change in Canada (Yates et al., 2007). Also, in Lake Consensus, USA a reduction in algal biomass was observed within three years of erosion control, stream fencing, nutrient management, crop and grazing rotations and reduced land use intensity in contributing catchments

(Makarewicz et al., 2009). A five-year period of measurement was considered too short for a positive macro-invertebrate response in a catchment in New Zealand where dairy shed-effluent was spread to land rather than discharged to streams (Wilcock et al., 2009). Diatoms (unicellular algae) are sensitive to small changes in chemical water quality and a positive effect on diatom assemblage was observed within 10 years of riparian vegetation, stream fencing, farm yard and manure management in Delaware, USA (Gabel et al., 2012). Ecological restoration of surface water bodies due to agricultural practice can be delayed due to hydrological and biogeochemical time lags along subsurface flow pathways (Vero et al., 2017), and by processes such as sediment storage and remobilisation within streams and rivers (Hamilton et al., 2012).

3.4 Negative and immeasurable effects

Changes in agricultural practices have in many cases resulted in no measureable improvements to water quality at the meso-scale, even after up to 15 years of monitoring (Table S2). In many cases, a mixed response to practice change occurred. For example, in the Waiokura catchment in New Zealand, positive effects on phosphorus, suspended sediment (dissolved reactive phosphorus (DRP), total P (TP) and SS **flux** declined by 25-40%), and faecal indicators were measured, whereas there was no measured change in stream macroinvertebrate indicators and the N **flux** response was negative (Wilcock et al., 2009). **Both surface and subsurface flows transport farm pollutants in the catchment. The negative N flux effect was explained by higher N leaching losses owing to higher N fertiliser and supplementary feed inputs to the catchment over the period of measurement, whereas the positive effects were realised via mitigation of surface flow pathways. The neutral effect on stream macroinvertebrates was attributed to the**

short timeframe of the study (5 years), a lack of upstream sources of sensitive species for recolonisation and because high water temperature did not limit the invertebrate communities prior to stream habitat restoration

The reasons that improved water quality was not measured in these studies included: limitations of the monitoring method, the time-frame of monitoring being too short to account for hydrological and/or biogeochemical time lags, the effect of the practice was small compared with background effects or counteracting processes, and/or the measures were potentially ineffective for the pollutant of concern.

No catchment monitoring approach (e.g. paired, before/after, or linking cause-effect approaches) consistently failed to observe practice effects at the meso- scale. However, measurement uncertainty in every approach limited the ability to measure the impacts of practices. The uncertainty inherent in most nutrient flux measurements, but particularly where there was a lack of high flow water quality data, was a limitation (e.g. Iital et al., 2008; O'Donnell et al., 2012). Further to this, data on temporal and spatial land management, such as nutrient source use, were generally sparse in comparison with water quality data, and were often insufficient for identifying cause-effect relationships. Other studies identified that nested scales of monitoring were needed to link cause and effect (Iital et al., 2008), and others suspected that major step changes in effect had potentially occurred before monitoring had begun (Bechmann, et al., 2008).

Some cases, proved difficult to verify whether a lack of effect was a result of ineffective measures, or because time lags for improvement of water quality were longer than the monitored period (Bergfur et al., 2012). The time lag in response is affected by the rate of change or degree of impact of a certain measure. The smaller the rate of change, the longer the time needed to detect an improvement in water quality against the backdrop of inter-annual variation (Bechmann et al., 2008). For example, at least 20 years of monitoring was estimated to be needed in order to detect a 25% decrease in atrazine flux in streamflow from a 73 km^2 cropped catchment in Northcentral Missouri, USA (O'Donnell et al., 2012).

In many cases, the potential to measure improvement in one or more water quality indicators was limited by the counteracting impact of a few management events (such as an untimely manure application or cattle accessing a stream) (Makarewicz et al., 2009;Wilcock et al., 2007), or weather events (Inamdar et al., 2002;Wilcock et al., 2009). For example, the degree of impact of a reduction in soil plant-available P levels in two Irish 5 km^2 catchments over five years was too small to measurably reduce high flow P concentrations in the stream. Instead wet years and seasons led to an increase in stream P concentrations (Campbell et al., 2015). Elsewhere, implementation rates were too low for a positive effect to be measured against background influences. For example, in the Upper Snake/Rock Creek catchments in Idaho, USA, SS fluxes did not reduce due to the sprinkler irrigation technology that was introduced because water quality was dominated by the influence of the remaining area of land under furrow irrigation (Bjorneberg et al., 2008).

In some cases the practices implemented were ineffective for the water quality indicator measured (or vice-versa). For example seven years of monitoring failed to identify any changes in NO_3-, total P, dissolved reactive P, SS or flow after implementation of grassed waterways, stream buffers, and strip-tillage in a catchment of the Mackinaw River, central Illinois, USA. The best management practices (BMP) were designed to control surface losses but were ineffective at mitigating the large percentages of the total loads that were lost via subsurface tile drains (Lemke et al., 2011).

4 Summary of implications for catchment scientists and policy makers

The review highlighted that to measure water quality change in meso-scale catchments, scientists should account for long times lags, from four to 20 years, when designing measurement programs. Long term (c.a. 30 year) studies of water quality are used in the USA in a network of agricultural catchments (Long Term Agroecosystem Research (LTAR)) to allow for time lags and to measure slow changes (Bartuska et al., 2015). However, securing continuous funding for long-term studies remains a challenge.

To enhance the scientific information and knowledge that meso-scale catchment studies generate, five outcomes of agricultural practice change should be explored or predicted. Firstly, the studies should highlight practice change scenarios that are likely to be ineffective for certain parameters e.g. NO_3^- versus P, or indeed for losses along certain pathways. Retro-fitting the correct measure(s) to site specific losses along known pathways that are based on site specific knowledge can have a positive effect on water quality. For example Tomer et al. (2014) describe how riparian re-vegetation at local to basin scale was encouraged to improve stream water quality after studies found that a large amount of SS in the streams was from stream bank rather than field erosion.

Secondly, where practice change improves water quality, the degree to which water quality targets are likely to be achieved should be explored. In the studies reviewed, water quality targets were rarely attained. Thirdly, the temporal and spatial scale of effectiveness of a practice change scenario should be estimated because the monitoring period and location of monitoring needed depends on the parameter or indicator of improved water quality. A fourth science-related recommendation is that the potential for pollution swapping should be examined (Stevens et al., 2009). For example, Weaver et al. (2014) identified that for sandy catchments dominated by subsurface nutrient flows, riparian fencing and vegetation was likely to decrease sediment, but increase the proportion of bioavailable P, entering waterways. It is likely that modelling, and not just direct measurement, will be needed for some of these predictions.

The richness of information generated by meso-scale water quality impact studies could also be enhanced by explicit, rather than implicit, evaluation of cause and effect links between practices and water quality, and by statistically robust analysis of *response times* and *measurement times*. These require actual changes in land management practice to be measured at spatial and temporal frequencies suited to the water quality indicator of interest and are critical pieces of information needed to evaluate the effectiveness of practice changes (Tomer et al., 2014).

A final recommendation to the science community is that to provide sufficient information for balanced decisions about changing practices, the ratio of costs to benefits of implementing practice changes should be calculated (e.g. Fezzi et al., 2010;Mausbach et al., 2004). Direct measurement of costs and benefits of mitigation measures at meso-scale is challenging but possible (Roberts et al., 2012; Stoeckl et al., 2014), and is increasingly being conducted in monitoring and research.

Five policy related messages emerged from the review. Firstly, **based on the studies included in the present review that conformed to the inclusion rules imposed, results** suggest that policy makers should account for the likely time frames of 1-10 years for positive effects to occur after threshold level implementation of agricultural practice change when setting expectations and planning policy implementation and evaluation. **Other catchments may have even longer timeframes due to catchment characteristics such as hydrological flow residence times.** A second finding was that in most catchments where beneficial effects of mitigation practices were successfully measured, combinations of practices, rather than single practices, had been implemented. These practices addressed more than one of the sources, pathways, delivery or impact of the nutrient or pollutant. Moreover, positive effects were often associated with a reduction in agricultural land use intensity, rather than with a change in practice within an existing land use or farming system. Policy makers should also be mindful of the notion that improved water quality does not always lead to water quality standards being met.

A third finding for setting agricultural policy for surface water quality improvement is the critical importance of matching practice change measures with the specific water quality problem, and ensuring that measured water quality indicators are biophysically connected with the implemented practices in space and time. The choice of indicator of system quality or change can influence assessments of whether mitigation measures have been successful or otherwise (Lillebo et al., 2007). Some indicators may not be affected by specific mitigation measures. For example, in catchments and seasons where point sources have a large influence on low-flow river nutrient concentrations, the effect of implementing measures to mitigate diffuse nutrient inputs to rivers may not be detected by water quality indicators monitored during low or ambient flow (Jordan et al., 2012). In short, water quality mitigation practices need to be implemented to a threshold level, and indicators of water quality improvement need to be measured, in the right place and at the right time.

A fourth policy finding from the review was that ‘no measurable effect’ of implemented practices was a common outcome for the water quality indicators measured. Reasons for a lack of measurable effect include some manageable limitations such as insufficient monitoring time scales to account for hydrological and/or biogeochemical delays, insufficient collection of ‘source’ information and uncertainty in flux measurement (the latter being estimated as up to 11% (Harmel et al., 2006) or 45% (Melland et al., 2012) under ideal conditions). Monitoring programs should be designed and refined as much as possible to eliminate these management limitations and therefore increase the likelihood of effects being measured. Ineffective practice change scenarios can then be identified and used to inform policy-making cycles, as per the DPSIR framework.

A fifth consideration for setting policy is knowledge of threshold rates of practice change required to effect a change in water quality. Threshold BMP rates were not often discerned or articulated by the studies reviewed. However, Yates et al. (2007) found some streams exhibit a threshold effect whereby some measured improvements show sharp rather than continuous changes and identified that implementation lags occur as a function of the area, temporal rate, and the magnitude of practice change in a catchment. Schilling et al. (2006) also showed that water quality improvements within a monitored period increased as the catchment size decreased and attributed this to the proportion of catchment area across which the BMP had been implemented increasing with decreasing catchment size. However, maintaining threshold levels of BMP implementation with increasing catchment size will not always result in water quality improvement because the dominant processes causing poor water quality can change with scale. For example, Wilson et al. (2014) found that, as the sediment transport pathway length increased with increasing catchment size (4-198 km^2), the proportion of sediment delivered from eroded fields decreased (due to reduced surface hydrological connectivity (e.g. Sherriff et al., 2016)) relative to that eroded from channel banks. In this scenario, a different suite of erosion mitigation measures would be appropriate at changing scales. This demonstrates that, similar to the need to identify catchment-specific suites of practice changes, it is likely that threshold rates of practice implementation will also be catchment-specific (Tomer et al., 2011).

Lastly, whilst mandatory changes may have been expected to shorten the practice implementation lag, there was no apparent link between implementation lags, or measurement lags and the main practice change implementation approach across the catchment. This highlights that implementation of cultural and structural changes to farms and farm practices takes time, even where measures are mandatory (Kronvang et al., 2008). Case by case analysis would be required to identify any potential for improved adoption via better selection of policy mechanism to achieve threshold implementation rates.

Other limitations to measuring improved water quality included extreme weather or management events, uncertain stream nutrient flux measurements, a scarcity of practice change information and insufficiently long monitoring programs. Meso-scale catchment studies intending to measure the effectiveness of policies need to measure the right water quality parameters, the implementation rates of policy in time and space, and the studies require sufficient time, up to 20 years **(based on the studies herein but this could be longer elsewhere)**, for effects to occur and for trends to be measured.

Acknowledgements

Time to conduct this review was provided by the authors' institutions and by the Irish Department of Agriculture Food and the Marine.

References

Ascott, M.J., Gooddy, D.C., Wang, L., Stuart, M.E., Lewis, M.A., Ward, R.S., Binley, A.M, 2017. Global patterns of nitrate storage in the vadose zone. Nature Communications 8, 1-7.

Bechmann, M., Deelstra, J., Stålnacke, P., Eggestad, H.O., Øygarden, L., Pengerud, A., 2008. Monitoring catchment scale agricultural pollution in Norway: policy instruments, implementation of mitigation methods and trends in nutrient and sediment losses. Environmental Science & Policy 11, 102-114.

Bergfur, J., Demars, B.O.L., Stutter, M.I., Langan, S.J., Friberg, N., 2012. The Tarland Catchment Initiative and Its Effect on Stream Water Quality and Macroinvertebrate Indices. Journal of Environmental Quality41, 314-321.

Bishop, P.L., Hively, W.D., Stedinger, J.R., Rafferty, M.R., Lojpersberger, J.L., Bloomfield, J.A., 2005. Multivariate Analysis of Paired Watershed Data to Evaluate Agricultural Best Management Practice Effects on Stream Water Phosphorus. Journal of Environmental Quality 34, 1087-1101.

Bjorneberg, D.L., Westermann, D.T., Nelson, N.O., Kendrick, J.H., 2008. Conservation practice effectiveness in the irrigated Upper Snake River/Rock Creek watershed. Journal of Soil and Water Conservation 63, 487-495.

Daly, K., Tuohy, P., Peyton, D., Wall, D.P., Fenton, O, 2017. Field soil and ditch sediment phosphorus dynamics from twoartificially drained fields on poorly drained soils. Agriculture Water Management 192. 115-125.

Daroub, S.H., Van Horn, S., Lang, T.A., Diaz, O.A., 2011. Best Management Practices and Long-Term Water Quality Trends in the Everglades Agricultural Area. Crit Rev Env Sci Tec 41, 608-632.

Fealy, R.M., Buckley, C., Mechan, S., Melland, A., Mellander, P-E., Shortle, G., Wall, D., Jordan, P., 2010. The Irish Catchment Programme; catchment selection using spatial multi-criteria decision analysis. Soil Use and Management 26, 225-236.

Fenton, O., Schulte, R.P.O., Jordan, P., Lalor, S.T.J., Richards, K.G., 2011. Time lag: a methodology for the estimation of vertical and horizontal travel and flushing timescales to nitrate threshold concentrations in Irish aquifers. Environment Science & Policy 14, 419-431.

Fenton, O., Mellander, P-E., Daly, K., Wall, D.P., Jahangir, M.M.R., Jordan, P., Hennessey, D., Huebsch, M., Blum, P., Vero, S., Richards, K.G., 2017. Integrated assessment of agricultural nutrient pressures and legacies in karst landscapes. Agriculture Ecosystems & Environment 239, 246-256.

Freebairn, D.M., Wockner, G.H., Hamilton, N.A., Rowland, P., 2009. Impact of soil conditions on hydrology and water quality for a brown clay in the north-eastern cereal zone of Australia. Australian Journal of Soil Research 47, 389-402. 10.1071/sr07054

Gabel, K.W., Wehr, J.D., Truhn, K.M., 2012. Assessment of the effectiveness of best management practices for streams draining agricultural landscapes using diatoms and macroinvertebrates. Hydrobiologia 680, 247-264.

Huebsch, M., Horan, B., Blum, P., Richards, K.G., Grant, J., Fenton, O., 2013. Impact of local weather conditions and agronomic practices on groundwater nitrogen content in a karst aquifer on an intensive dairy farm in Southern Ireland. Agriculture Ecosystems & Environment 179, 187–199.

Iital, A., Pachel, K., Deelstra, J., 2008. Monitoring of diffuse pollution from agriculture to support implementation of the WFD and the Nitrate Directive in Estonia. Environmental Science & Policy 11, 185-193.

Inamdar, S.P., Mostaghimi, S., Cook, M.N., Brannan, K.M., McClellen, P.W., 2002. A long-term, watershed scale evaluation of the impacts of animal waste BMP's on indicator bacteria concentratons. Journal of the American Water Resources Association 38, 819-833.

Jahangir, M.M.R., Johnston, P., Barrett, M., Khalil, M., Groffman, P., Boeckx, P., Fenton, O., Murphy, J., Richards, K.G., 2013. Denitrification and indirect N2O emissions in groundwater: hydrologic and biogeochemical influences. J. Contam. Hydrol. 152, 70–81.

Jaynes, D.B., Dinnes, D.L., Meek, D.W., Karlen, D.L., Cambardella, C.A., Colvin, T.S., 2004. Using the Late Spring Nitrate Test to Reduce Nitrate Loss within a Watershed. Journal of Environmental Quality 33, 669-677.

Jordan, P., Melland, A.R., Mellander, P.E., Shortle, G., Wall, D., 2012. The seasonality of phosphorus transfers from land to water: Implications for trophic impacts and policy evaluation. Science of the Total Environment 434, 101-109.

Kronvang, B., Andersen, H.E., Børgesen, C., Dalgaard, T., Larsen, S.E., Bøgestrand, J., Blicher-Mathiasen, G., 2008. Effects of policy measures implemented in Denmark on nitrogen pollution of the aquatic environment. Environmental Science & Policy 11, 144-152.

Kyllmar, K., Carlsson, C., Gustafson, A., Ulen, B., Johnsson, H., 2006. Nutrient discharge from small agricultural catchments in Sweden characterisation and trends. Agriculture, Ecosystems & Environment 115, 15-26.

Lemke, A.M., Kirkham, K.G., Lindenbaum, T.T., Herbert, M.E., Tear, T.H., Perry, W.L., Herkert, J.R., 2011. Evaluating Agricultural Best Management Practices in Tile-Drained Subwatersheds of the Mackinaw River, Illinois. Journal of Environmental Quality 40, 1215-1228.

Makarewicz, J.C., Lewis, T.W., Bosch, I., Noll, M.R., Herendeen, N., Simon, R.D., Zollweg, J., Vodacek, A., 2009. The impact of agricultural best management practices on downstream systems: Soil loss and nutrient chemistry and flux to Conesus Lake, New York, USA. Journal of Great Lakes Research 35, 23-36.

McAleer, E.B., Coxon, C.E., Richards, K.G., Jahangir, M.M.R., Grant, J., Mellander, P-E., 2017. Groundwater nitrate reduction versus dissolved gas production: A tale of two catchments. Science of the Total Environment, 586, 372-389.

McDaniel., P.A., Regan., M.P., Brooks, E., Boll, J., Barndt, S., Falen, A., Young, S.K., Hammel, J.E., 2008. Linking fragipans, perched water tables, and catchment-scale hydrological processes. Catena 73, 166-173.

McDowell, R.W., Nash, D., George, A., Wang, Q.J., Duncan, R., 2009. Approaches for quantifying and managing diffuse phosphorus exports at the farm/small catchment scale. Journal of Environmental Quality 38, 1968–1980.

Melland, A.R., Mellander, P.-E., Murphy, P.N.C., Wall, D.P., Mechan, S., Shine, O., Shortle, G., Jordan, P., 2012. Stream water quality in intensive cereal cropping catchments with regulated nutrient management. Environmental Science & Policy 24, 58-70.

Melland, A.R., Jordan, P., Murphy, P.N.C., Mellander, P.-E., Buckley, C., Shortle, G., 2014. Land Use: Catchment Management. In: Van Alfen, N.K. (Ed.), Encyclopedia of Agriculture and Food Systems. Elsevier, San Diego, pp. 98-113.

Melland, A.R., Silburn, D.M., McHugh, A.D., Fillols, E., Rojas-Ponce, S., Baillie, C., Lewis, S., 2016. Spot Spraying Reduces Herbicide Concentrations in Runoff. Journal of Agricultural and Food Chemistry 64, 4009-4020. 10.1021/acs.jafc.5b03688

Mellander, P.-E., Melland, A., Murphy, P.,Wall, D., Shortle, G., Jordan, P., 2014. Coupling of surface water and groundwater nitrate-N dynamics in two permeable agricultural catchments. J. Agric. Sci. 152, 107–124.

Minella, J.P.G., Walling, D.E., Merten, G.H., 2008. Combining sediment source tracing techniques with traditional monitoring to assess the impact of improved land management on catchment sediment yields. Journal of Hydrology 348, 546-563.

Monaghan, R.M., Carey, P.L., Wilcock, R.J., Drewry, J.J., Houlbrooke, D.J., Quinn, J.M., Thorrold, B.S., 2009. Linkages between land management activities and stream water quality in a border dyke-irrigated pastoral catchment. Agriculture, Ecosystems & Environment 129, 201-211.

Monaghan, R.M., Wilcock, R.J., Smith, L.C., Tikkisetty, B., Thorrold, B.S., Costall, D., 2007. Linkages between land management activities and water quality in an intensively farmed catchment in southern New Zealand. Agriculture, Ecosystems & Environment 118, 211-222.

O'Donnell, T.K., 2012. Assessing Watershed Transport of Atrazine and Nitrate to Evaluate Conservation Practice Effects and Advise Future Monitoring Strategies. Environmental Management 49, 267-284.

Schilling, K.E., Spooner, J., 2006. Effects of Watershed-Scale Land Use Change on Stream Nitrate Concentrations. Journal of Environmental Quality 35, 2132-2145.

Shore, M., Murphy, P.N.C., Jordan, P., Mellander, P-E., Kelly-Quinn, M., Cushen, M., Mechan, S., Shine, O., Melland, A.R., 2013. Evaluation of a surface hydrological connectivity index in agricultural catchments. Environmental Modelling & Software 47, 7-15.

Sherriff, S.C., Rowan, J.S., Fenton,O., Jordan, P., Melland, A.R., Mellander, P-E., hUallachain, D., 2016. Storm Event Suspended Sediment-Discharge Hysteresis and Controls in Agricultural Watersheds: Implications for Watershed Scale Sediment Management. Environment Science and Technology 50,1769-1778

Smith, K.A., Jackson, D.R., Withers, P.J.A., 2001. Nutrient losses by surface run-off following the application of organic manures to arable land. 1. Phosphorus. Environmental Pollution 112, 53-60

Sousa, M.R., Jones, J.P., Frind, E.O., Rudolph, D.L., 2013. A simple method to assess unsaturated zone time lag in the travel time from ground surface to receptor. Journal of Contaminant Hydrology 144, 138–151.

Sutton, A.J., Fisher, T.R., Gustafson, A.B., 2009. Historical Changes in Water Quality at German Branch in the Choptank River Basin. Water Air and Soil Pollution 199, 353-369.

Tomer, M.D., Locke, M.A., 2011. The challenge of documenting water quality benefits of conservation practices: A review of USDA-ARS's conservation effects assessment project watershed studies. Water Science & Technology 64, 300-310.

Tomer, M.D., Sadler, E.J., Lizotte, R.E., Bryant, R.B., Potter, T.L., Moore, M.T., Veith, T.L., Baffault, C., Locke, M.A., Walbridge, M.R, 2014. A decade of conservation effects assessment research by the USDA agricultural research service: progress overview and future outlook. Journal of Soil Water Conservation 69, 365-373

Van Meter, K., Basu, N., 2015. Catchment legacies and time lags: a parsimonious watershed model to predict the effects of legacy storage on nitrogen export. PlosOne https://doi.org/10.1371/journal.pone.0125971

Vero, S.E., Healy, M.G., Henry, T., Creamer, R.E., Ibrahim, T.G., Richards, K.G., Mellander, P.-E., McDonald, N.T., Fenton, O., 2017. A framework for determining unsaturated zone water quality time lags at catchment scale. Agri. Eco. Environ. 236, 234–242.

Wall, D., Jordan, P., Melland, A.R., Mellander, P.E., Buckley, C., Reaney, S.M., Shortle, G., 2011. Using the nutrient transfer continuum concept to evaluate the European Union Nitrates Directive National Action Programme. Environmental Science & Policy 14, 664-674.

Wilcock, R.J., Monaghan, R.M., Quinn, J.M., Srinivasan, M.S., Houlbrooke, D.J., Duncan, M.J., Wright-Stow, A.E., Scarsbrook, M.R., 2013. Trends in water quality of five dairy farming streams in response to adoption of best practice and benefits of long-term monitoring at the catchment scale. Marine and Freshwater Research 64, 401–412

Wilcock, R.J., Betteridge, K., Shearman, D., Fowles, C.R., Scarsbrook, M.R., Thorrold, B.S., Costall, D., 2009. Riparian protection and on-farm best management practices for restoration of a lowland stream in an intensive dairy farming catchment: a case study. New Zealand Journal of Marine and Freshwater Research 43, 803-818.

Wilcock, R.J., Monaghan, R.M., Thorrold, B.S., Meredith, A.S., Betteridge, K., Duncan, M.J., 2007. Land-water interactions in five contrasting dairying

catchments: issues and solutions. Land Use and Water Resources Research 7, 2.1-2.10.
Yates, A.G., Bailey, R.C., Schwindt, J.A., 2007. Effectiveness of best management practices in improving stream ecosystem quality. Hydrobiologia 583, 331-344.

Captions for Tables

Table S1. Summary of catchment characteristics to inform residence time.

Table S2. Summary of water quality effects of agricultural management practices measured at meso-catchment scales.

Captions for Figures

Fig 1. A 4-year time series of the difference in nitrate concentration between paired treated (4.0 km^2) and untreated (4.9 km^2) subcatchments of the Walnut Creek catchment in Iowa, USA (Jaynes et al. (2004)). A 5-year period of measurements taken to establish similarity between the paired catchments before the practices were implemented is not shown. The practices were assumed to be fully implemented by 1/97 and the practice implementation lag was assumed to be 1 year prior to this. The original figure was modified to highlight when the practices had a significant effect on nitrate concentrations (response time, 2 y post-practice change), the measurement time (5 y pre-BMP plus 4 y post-BMP) and the measurement lag (9 years less 2 years) that were calculated for this review.

Fig 2. A comparison of catchment size and main water flow pathway against positive water quality response and measurement times, where bar length = measurement lag time, left extent of bar = response time, right extent of bar = measurement time. The response time is the period of time for a significant change in water quality to occur and the measurement time is the period of time needed to measure water quality to identify that a significant response had occurred. Water quality indicators are also annotated as biol. (biological indicator), N (nitrogen species), NH4 (ammonium only), P (phosphorus species) and SS (suspended sediment). The transport pathway contributing most to the state of the water quality indicator is represented as surface (grey bars), subsurface tile drains (unshaded bars) or subsurface/groundwater (black bars).

Fig. 3 Response time (years, solid symbols (●), R^2 0.43, $P<0.05$) for positive effects with a single outlier, a very fast response in a large catchment in New Zealand, removed and the measurement time (years, open symbols (○), R^2 0.36, $P<0.05$) for positive effects with a single outlier, a very slow response in the Everglades, U.S.A, removed. Linear lines of best fit and correlation coefficients are also shown.

Fig 4. Agricultural practice implementation lag time (years) for catchments of increasing size where practice change is mostly voluntary (grey bars), incentivised for the purpose of research (unshaded bars) or mostly mandatory (black bars).