J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

Quantifying Background Nitrate Removal Mechanisms in an Agricultural Watershed with Contrasting Subcatchment Base-flow Concentrations

Wesley O. Zell[1] (wzell@usgs.gov) *corresponding author

Teresa B. Culver[2] (tbc4e@virginia.edu)

Ward E. Sanford[1] (wsanford@usgs.gov)

Jonathan L. Goodall[2] (goodall@virginia.edu)

[1] U.S. Geological Survey, 12201 Sunrise Valley Dr, MS 432, Reston, VA 20192

[2] Department of Civil and Environmental Engineering, University of Virginia, Thornton Hall, P.O. Box 400259, Charlottesville, VA 22904-4259

The authors declare no competing interests.

ABSTRACT

Numerous studies have documented the linkages between agricultural nitrogen loads and surface water degradation. In contrast, potential water quality improvements due to agricultural best management practices are difficult to detect because of the confounding effect of background nitrate removal rates as well as the groundwater-driven delay between land surface action and stream response. To characterize background controls on nitrate removal in two agricultural catchments we calibrated groundwater travel time distributions with subsurface environmental tracer data to quantify the lag time between historic agricultural inputs and measured base-flow nitrate. We then estimated spatially-distributed loading to the water table from nitrate measurements at monitoring wells, using machine learning techniques to extrapolate the loading to unmonitored portions of the catchment in order to subsequently estimate catchment removal controls. Multiple models agree that in-stream processes remove as much as 75% of incoming loads for one subcatchment while removing less than 20% of incoming loads for the other. The use of a spatially variable loading field did not result in meaningfully different optimized parameter estimates or model performance when compared to spatially constant loading derived directly from a county-scale agricultural nitrogen budget. While previous studies using individual well measurements have shown that subsurface denitrification due to contact with a reducing argillaceous confining unit plays an important role in nitrate removal, the catchment-scale contribution of this process is difficult to quantify given the available data. Nonetheless, the study provides a baseline characterization of nitrate transport timescales and removal mechanisms that will support future efforts to detect water quality benefits from ongoing BMP implementation.

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

1 INTRODUCTION

Numerous studies have documented the linkages between agricultural nitrogen loads and surface water degradation (Vitousek et al., 1997; Schindler and Vallentyne, 2008). The adverse effects of excess dissolved nitrate include the seasonal dissolved oxygen deficits and algal blooms that persist in many bays and estuaries despite widespread implementation of agricultural best management practices (BMPs) in upstream catchments. This persistence is due to the ongoing discharge of groundwater nitrates that have accumulated in surficial aquifers during the past century (Puckett et al., 2011). For example, in some agriculturally intensive regions of the Chesapeake Bay watershed as much as 70% of nitrate loads are delivered to the Bay or its tributaries as groundwater discharge (Lindsey et al., 2003, Ator and Denver, 2012; Sanford and Pope, 2013).

While loading reductions and water quality improvement due to BMPs have been documented at laboratory and field scales (e.g., Staver and Brinsfield, 1998), the anticipated effects of these practices are often difficult to detect at the outlets of agricultural watersheds in which they have been widely implemented (Osmond et al., 2012; Meals et al., 2010). This difficulty is in part due to the lag time between land surface action and surface water response that results from groundwater transport pathways (Sanford and Pope, 2013; Tesoriero et al., 2013; Science and Technical Advisory Committee, 2005). However, the effects of BMPs are also difficult to disentangle from other spatially and temporally distributed factors affecting in-stream loads (Gitau et al., 2010; Sutton et al., 2009). These factors, which may vary widely between catchments, include background rates of nitrogen removal at the land surface, in the aquifer, or in the receiving stream (Meals et al., 2010; Böhlke and Denver, 1995).

The objective of the study described in this paper was to differentiate and quantify long-term, catchment-integrated nitrate removal mechanisms in two adjacent agricultural headwater catchments with similar contributing land use histories but contrasting stream nitrate concentrations. While the fate and transport of agricultural nitrate has been widely investigated, there have been few if any studies that characterize long-term effects of subsurface or in-stream nitrate removal in a highly spatially variable system. For example, many studies have examined the short-term (1-5 year) groundwater-driven discharge of agricultural nitrates to headwater streams with goals of differentiating seasonally variable nitrate sources (Yevenes and Mannaerts, 2012) or identifying changes in hydrologic connectivity between uplands and discharge areas (Petry et al., 2002; Wriedt et al., 2007; Molenat et al., 2008). For the characterization of in-stream nitrate variability at these shorter time scales, it is not necessary to account for the full, multi-decadal loading history, and it is common to treat upgradient catchment nitrate as an effectively steady-state reservoir draining subject to hydrological controls (Vidon and Hill, 2004; Montreuil et al., 2010). While some studies have directly measured in-stream rates of nitrogen removal for headwater catchments (Royer et al., 2004; Vidon and Hill, 2004; Mulholland et al., 2008), questions remain about extrapolating these measurements to larger spatial and temporal scales (Boyer et al, 2002). Few groundwater studies have examined the long-term behavior of agricultural nitrogen inputs and export. Aquilina et al. (2012) used groundwater nitrate and chlorofluorocarbon (CFC) measurements to reconstruct the long-term nitrate input function for a catchment in Brittany (France); they simulated long-term in-stream nitrate concentrations at the catchment outlet

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

but assumed conservative export and did not investigate removal processes. Sanford and Pope (2013) combined groundwater travel times from a calibrated regional simulation with a regression method to estimate spatially constant nitrate removal terms for the Delmarva Peninsula (USA); however, the scale of their investigation did not allow for spatial variation of removal terms or the effects of catchment-scale hydrogeological variability. A few studies have documented the potential of catchment-scale, physics-based simulation of nitrate fate and transport through coupled landscape-groundwater-surface water systems (Conan et al., 2003; Galbiati et al., 2006; Wei et al., 2018), but these simulations are likewise limited to short time scales and subsurface linkages are not constrained by environmental tracer data.

To address these knowledge gaps we leveraged a multi-decadal record of catchment nitrate inputs and exports as well as a unique dataset of environmental tracer measurements, groundwater nitrate measurements, and in-stream nitrate measurements to simulate long-term average nitrate controls for adjacent headwater streams. We use a fully distributed, three-dimensional numerical simulation of the groundwater system to link land surface inputs to subsurface and in-stream nitrate concentrations, and we examined the significance of spatially distributed representation of catchment nitrate loading for parameter estimation and uncertainty.

2 Materials and Methods

2.A Overview of Study Site

The 61-km^2 study site (hereafter referred to as the ‘Upper Chester’ - cf. Nelson and Spies, 2013) is in Kent County, MD, and is a low-relief agricultural watershed drained by small gaining streams; the Chesterville Branch (USGS gage 1493112) and Morgan Creek (USGS gage 1493500) subcatchments are the focus of this paper (**Figure 1**). These subcatchments have similar land-use histories, soil types, and stream discharge rates but widely different in-stream nitrate levels. Water quality throughout the Upper Chester deteriorated during the last century due to agricultural intensification and elevated fertilizer inputs (**Figure 2;** cf. Böhlke and Denver, 1995). In recent years, a variety of management practices aimed at damping adverse agricultural effects and improving water quality have been implemented in the Upper Chester (Nelson and Spies, 2013). Concentrations at the Morgan Creek stream gage ranged between 2 and 3 mg NO_3-N/L for the duration of its sampling history; in contrast, concentrations at the Chesterville Branch stream gage have increased from 4-6 mg/L in the early 1990s and currently persist near 10 mg/L (**Figure 3**).

Insert Figure 1

Insert Figure 2

Insert Figure 3

Previous studies in Morgan Creek suggest several potential reasons for the disparity in mean baseflow nitrate concentration though no catchment-scale studies have integrated the available data and quantified their relative contributions. Böhlke and Denver (1995) found evidence of denitrification (elevated nitrate δ^{15}N levels, excess dissolved N_2, and indicators of pyrite reduction) due to a glauconitic confining unit that outcrops at the lower reaches of Morgan Creek (**Figure 1**). Bachman et al. (2002) observed increasing silica concentrations in a

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

downstream direction on Morgan Creek. Groundwater silica concentrations elsewhere on the Delmarva Peninsula have been shown to positively correlate with tritium-derived groundwater ages (Clune and Denver, 2012), such that increased silica in the lower reaches of Morgan Creek may indicate the dilution of agricultural nitrates with older, higher silica, nitrate-free groundwater that reaches the stream from the lower confined aquifer. Sediment cores in lower Morgan Creek show an abrupt change in the elevation of the confining unit and thus suggest a discontinuity that could allow influx of older groundwater from the deeper, confined aquifer (Bachman et al., 2002). Finally, because the Morgan Creek stream channel is downcut into the low-permeability confining unit, direct groundwater discharge through the streambed is limited, and groundwater instead emerges through seeps at the edge of a near-stream floodplain before traveling to the main channel via small rivulets and sheetflow; Duff et al. (2008) observed decreasing nitrate concentrations in the groundwater, rivulets, and stream, respectively, for lower Morgan Creek, suggesting the importance of riparian nitrogen removal.

Chesterville Branch has not been investigated with the same detail as Morgan Creek. However, the higher-permeability surficial sediments are much deeper under Chesterville Branch than Morgan Creek (Böhlke and Denver, 1995), suggesting that a higher percentage of base-flow discharge bypasses nitrate removal mechanisms in the riparian zone (Zell et al., 2018). Similar bypasses, and their importance for nitrate processing, have been noted in other agricultural systems (e.g., Tesoriero et al., 2013; Vidon and Hill, 2004).

2.B Simulation of Groundwater Flow and Nitrate Transport

In a separate study we document the development, calibration, and sensitivity/uncertainty analysis of several candidate numerical flow and transport simulation models for the Upper Chester (Zell et al., 2018). The models represent a range of plausible interpretations of the Upper Chester groundwater system and simulate steady-state subsurface flow and advective solute transport using the US Geological Survey (USGS) finite-difference code MODFLOW (Harbaugh, 2005) and its companion particle-tracking software MODPATH (Pollock, 2012). For each model, spatially variable recharge, horizontal hydraulic conductivity, anisotropy, and porosity were calibrated against groundwater levels, base-flow discharge, and more than 200 subsurface measurements of atmospherically derived age tracers.

For the present study, we selected the two best performing flow and transport models and used them to (i) generate flux-weighed travel time distributions (TTD) at nitrate monitoring locations and (ii) identify the associated contributing recharge area for each nitrate monitoring location in the study area (i.e., monitoring wells and the Chesterville Branch and Morgan Creek catchment outlets). In the remainder of the manuscript we refer to these TTD models as the nitrate transport 'base models'. The selected base models are chiefly differentiated by assumptions about base-flow indices (BFI) in the Upper Chester and are consequently labeled 'LowBFI' and 'HighBFI' (see Zell et al., 2018, for more details). Given the TTDs provided by these base models, the concentration of a solute at a monitoring location *j* may be calculated by the convolution of a TTD with the time series of solute inputs to the catchment, expressed in its discrete form as

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

$$C_j[t] = \sum_{\tau=0}^{\infty} C[x,y,t-\tau]g_j[\tau] \quad , \qquad (1)$$

where $C(x,y,t\text{-}\tau)$ is the solute (e.g., dissolved nitrate) input signal and $g_j(\tau)$ is the flux-weighted TTD of groundwater sampled at j.

2.C Estimation of nitrate loading to the water table

As an initial estimate of nitrate inputs to the water table (i.e., $C(x,y,t\text{-}\tau)$ in **Equation 1**) we calculated a county-level nitrogen budget for the years 1930-2015 using Kent County (MD) agricultural data and nitrogen wet deposition data for the Maryland Eastern Shore (**Figure 2**). In the remainder of the paper we refer to this spatially constant, county-scale time series as the 'reference loading'. We then calibrated multiple sets of spatially variable loading factors that, when applied to the reference loading time series, resulted in a range of estimates of the temporally and spatially variable flux of nitrate across the water table. Loading factors were estimated by calibrating the nitrate transport model (**Equation 1**) solely against observed groundwater nitrate concentrations under different calibration scenarios that varied with respect to both the base flow and transport model as well as the weighting scheme applied to the nitrate observation dataset (**Table 1** and Supplemental Materials Section **S1**).

Insert Table 1

Due to the generally oxic character of the subsurface and the expected conservative nitrate transport from the water table to observation wells, we assumed that most groundwater nitrate concentrations in the Upper Chester provide direct information about nitrate loading to the water table (i.e., after nitrate removals such as crop export or soil denitrification) (Green et al., 2010). However, while groundwater nitrate observations in the Upper Chester are abundant when compared to many sites, monitoring well nitrate data only constrain a portion of the model domain. Each stage 1 scenario therefore included use of a Gradient Boosted Regression (GBR) method to extrapolate the calibrated water table loading factors from the monitored subdomain to the entire model area on the basis of proxy relationships with other mapped data. These candidate explanatory variables included soils and land use data derived from national-scale datasets as well as estimates of hydrologic states and system properties developed during this study. The GBR methods and results are fully described in the Supplemental Materials.

2.D Estimation of catchment nitrate removal

In the second stage of parameter estimation, we used the stage 1 scenarios as a range of possible loadings and estimated nitrate removal at the confining unit and in/near each stream by calibrating the resulting nitrate transport models against base-flow nitrate observations. Base-flow nitrate concentrations were considered to be those measurements associated with a stream discharge observation for which the separated base-flow was more than 85% of total flow (using the digital filter separation method of Arnold et al., 1995) (**Figure 3**).

As discussed in the development of the flow and transport base models, uncertainties about the spatial distribution of catchment hydraulic conductivity, porosity, and recharge propagate through to the simulated base-flow TTDs (Zell et al., 2018). Therefore, while the base-flow TTDs used in this study are relatively well-constrained by hydraulic and atmospheric tracer data, it is expected that they may be updated by conditioning them upon stream nitrate data. We

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

consequently allowed the calibration to adjust the TTDs by means of stream-specific scaling factors that may improve the simulation of base-flow nitrate trends. Note also that, for purposes of evaluating model uncertainty, the inclusion of TTD scaling parameters is a means of remedying non-conservative uncertainty estimates that result from assuming that base-flow travel times are known.

Due to correlations between stream removal and confining unit removal in Morgan Creek it was assumed that fixing the confining unit parameter value (and thus routing all available stream nitrate information to the stream removal parameters) would reduce the uncertainty on the estimated stream parameters. On the evidence of high removal efficiencies observed by Böhlke and Denver (1995) we included stream scenarios ('Fixed CU') for which the confining unit removal efficiency was assumed to be perfectly known as 0.80. Similarly, it was assumed that the use of the TTD scaling parameters – while admitting a hydrologically realistic degree of flexibility to the simulated groundwater lag times – increased uncertainty on the estimates of stream removal rates. To query this effect we included an additional calibration scenario ('No TTD Scaling') for which the stream base-flow TTD simulated by the base model was fixed (**Table 1**).

3 Results

3.A Nitrate loading to the water table

The calibrated loading scenarios collectively reproduce the relative mean groundwater nitrate concentration in each subcatchment as well as the preponderance of individual observed values (**Figure 4**), though some groundwater monitoring locations are not well-simulated by any of the scenarios. Simulated values in the Morgan Creek catchment (e.g., at well KEBE206, labeled on **Figure 4**) are much more sensitive to the calibration structure than are simulated values in Chesterville Branch; this is expected given that simulated subsurface travel times in Morgan Creek are more sensitive to assumptions about the BFI than are those in Chesterville Branch (cf. Zell et al., 2018). For both subcatchments the mean simulated value is slightly lower than the mean observed value due to two features of the stage 1 regression methodology. First, all simulated and observed groundwater nitrate concentrations were transformed by the natural log in order to prevent the measurements of very high nitrate concentrations from dominating the regression. Second, observation variance was higher (and therefore observation weights were lower) for the highest concentrations.

Insert Figure 4

3.B Catchment nitrate controls

For stage 2 of parameter estimation each stream nitrate model was driven by a different nitrate loading scenario but calibrated against an identically weighted dataset of stream nitrate observations; therefore, unlike calibration stage 1, the weighted sum of square errors (WSSE) provides a comparative measure of stream model performance (**Figure 5**). The models collectively show that stream processes in Morgan Creek remove a higher fraction of incoming loads (0.55-0.77) than do stream processes in Chesterville Branch (0.05-0.41); for the nine best performing models, the range of calibrated removal rates for Chesterville Branch is even lower (0.05-0.19), though for most cases the estimate for Chesterville Branch is highly uncertain. The large uncertainty bounds for the Chesterville Branch removal parameter reflect the relatively

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

large measurement variability for the Chesterville Branch base-flow nitrate concentrations used as calibration targets (**Figure 6a**). As measured by the WSSE, nine of the twelve resulting calibration scenarios performed similarly (i.e., with calibration WSSE <= 0.25 of the highest WSSE) in their capacity to reproduce the stream nitrate time series (**Figure 5** and **Table 1**). With the exception of Loading Scenario B, each of the LowBFI models performed better than the corresponding HighBFI model; this performance difference may corroborate the results of the earlier calibration studies, which found that the LowBFI hydrology model performed better than the HighBFI hydrology model in its simulation of water levels and atmospheric tracer transport.

Insert Figure 5

While the nitrate removal impact of the confining unit is well-demonstrated from the interpretation of individual wells in the Morgan Creek catchment (Böhlke and Denver, 1995), the catchment-level impact of the confining unit on nitrate removal is uncertain given the available data and the subsurface model used in this study (i.e., the simulated location of the confining unit and its resulting impact on the flow regime and nitrate removal). Assigning the confining unit a very high removal rate (i.e., the Fixed CU scenario) did result in the lowest estimate of in-stream/near-stream removal rates but did not meaningfully impact the model performance or the uncertainty associated with the remaining estimated parameters. In contrast, removing the TTD scaling parameters did greatly reduce the uncertainty of the estimated removal rates for both Chesterville Branch stream processing and the confining unit but at the cost of model performance.

When compared to Chesterville Branch, the higher rates of in-stream nitrate removal and the larger influence of the confining unit in Morgan Creek result in a more subdued response to changes in catchment agricultural inputs (**Figure 6a**). For the HighBFI base models, the total nitrate loads seen by both subcatchment groundwater systems are similar (**Figure 6b**). However, for the LowBFI base models, total nitrate loads to the water table are noticeably higher in the Chesterville Branch subcatchment despite the smaller surface water drainage in Chesterville Branch compared to Morgan Creek. This difference in loading is in part because the simulated groundwater divide for the LowBFI scenarios is not coincident with the surface water divide, such that recharge from the upper portions of the Morgan Creek subcatchment discharges to Chesterville Branch, thus making the contributing drainage areas for each stream more similar than would be suggested by topography alone (Zell et al., 2018). Given the modelled location of the confining unit, few Chesterville Branch flow paths that recharged after 1940 contact the reducing confining unit and there is thus negligible nitrate removal. In contrast, the confining unit removed roughly one-third of Morgan Creek loads under the Fixed CU scenario and approximately 10% of loads when averaged across multiple scenarios.

Insert Figure 6

In addition to illuminating the nitrate removal distinctions between the two catchments, the models suggest that Morgan Creek base-flow has a larger fraction of pre-agricultural water than does base-flow in Chesterville Branch, such that lower nitrate concentrations in Morgan Creek may result in part from the dilution of agricultural nitrates. For both streams, model calibration to the in-stream nitrate data shifted the base-flow TTD towards ages older than

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

those derived from the base-model calibration against subsurface tracer data (**Figure S3**); the calibrated shift is greater for Morgan Creek and results in a median age older than the median age in Chesterville Branch. As other authors have discussed (cf. Kirchner, 2006) parameters used to describe environmental systems may function as proxies for processes that are not explicitly represented in a model. In this case, in-stream nitrate data may be informing the base-flow TTD previously constrained by the subsurface tracer data; however, it is also possible that the calibrated TTD scaling factors reflect other hydrological or geochemical dynamics that would likewise delay the translation of the fertilizer purchase record to an in-stream water quality signal. For example, any subsurface retardation of nitrate relative to the non-retarded transport of atmospheric tracers would, under our conceptual model, be subsumed by adjustments to the TTD; these velocity differences could result from nitrate sorption, which is generally considered to be negligible but has been observed in some column studies (Clay et al., 2004). Similarly, TTD adjustments here could be in response to dispersive effects not evident during base model calibration nor represented by our advective assumptions....

4 Discussion and Conclusions

Quantifying background controls on nitrate transport and removal is essential for their subsequent disentanglement from water quality trends that may be due to management actions. As such, this study provides a baseline characterization of nitrate transport timescales and removal mechanisms that will support future efforts to detect water quality benefits from ongoing BMP implementation. Simulated groundwater nitrate concentrations for several sites were highly sensitive to the range of stage 1 calibration scenarios (**Figure 4**). The poorly performing groundwater sites are likely a result of local heterogeneity of either the flow system or the nitrate loading at a spatial scale not available to our discretization. For example, groundwater nitrate measurements include observations from three closely-spaced transects of 3-4 wells each that sampled shallow groundwater in the lower reach of Morgan Creek (these include wells KEBD162 and KEBD163 – cf. **Figure 4**). This small area (i.e., all observations separated by less than 300 meters) is subject to very steep nitrate gradients that cannot be reproduced by our catchment-scale model. These gradients appear to be due to a combination of converging flow paths of widely disparate ages, nitrate removal due to denitrification in the confining unit sediments, and concentrated near-stream loading that possibly originates at a dairy operation near lower Morgan Creek (Puckett et al., 2008; Bachman et al., 2002; Böhlke and Denver, 1995).

While the sensitivity of simulated groundwater nitrate to the range of stage 1 calibration scenarios may suggest the importance of multiple plausible estimates of water table loading, the loading time series derived directly from the county-scale agricultural nitrate budget, without subsequent conditioning on groundwater nitrate data, resulted in similar calibrated stream models (**Figure 5**). Thus, for purposes of estimating long-term catchment nitrate controls from annually averaged stream nitrate concentrations, the available groundwater nitrate data did not substantially affect either the estimate of total nitrate available for stream export or the distribution of those inputs across the landscape. However, upland groundwater nitrate data and the specification of spatially distributed inputs may be more important for resolving nitrate export behavior at shorter timescales or when using a transient flow and

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

transport simulation to characterize subsurface linkages. For example, several studies have shown that in-stream nitrate concentrations can be highly sensitive to the time-variable hydrological connectivity that delivers base-flow from uplands to discharge areas (Petry et al., 2002; Ocampo et al., 2006; Wriedt et al., 2006; Molenat et al., 2008). The wide sub-annual range of base-flow nitrate concentrations shown in **Figure 6a** may reflect seasonal changes in upland gradients and leachate flushing as more permeable upland or midslope areas are activated and deactivated with rising and falling water tables. These effects, as well as the temperature-dependent stream metabolism effects on nitrogen removal (Bernot et al., 2006), are not captured by our simulation of the long-term export signal.

The current analysis cannot conclusively explain the disparate removal rates between the two catchments. However, two potential factors may be considered. Studies in the last decade (Wollheim et al., 2006, Mulholland et al., 2008; Alexander et al., 2009; Böhlke et al., 2009; Scanlon et al., 2010) have variously shown nitrate removal efficiency in smaller order streams to be a function of hydraulics (i.e., stream depth or velocity) or water quality (i.e., stream nitrate concentration). At shallow depths and low velocities, stream nitrates have longer exposure to nitrogen uptake services from in-channel biota and sediments (Alexander et al., 2009). Near-stream and in-stream hydraulics are likely of importance here; i.e., as described above, the confining unit which outcrops in the Morgan Creek lower reaches may not only account for nitrogen removal through denitrification, but also controls the seepage of base-flow discharge to the main channel in a manner that increases exposure of nitrates to biotic uptake. Furthermore, a coarse comparison of the stream velocities and associated cross-sectional flow areas (**Figure S3**) suggests that Chesterville Branch has shorter in-stream residence times due to a shorter network and higher velocities. The National Hydrography Dataset (NHDPlusV2) approximation of the Morgan Creek stream network (above the gage used in this study) is approximately 3.5 times the length of Chesterville Branch (also above the gage), such that the length-normalized removal efficiencies may be more similar than their accumulated downstream effect.

Less well understood is the evidence that nitrogen removal efficiency declines with increasing nitrate concentration. For example, Mulholland et al. (2008) found across a range of smaller order streams that increasing the stream nitrate concentration from 1.5 to 15 mg/L may reduce the nitrate removal fraction by more than half. The results of this study may reiterate questions of potential importance for management of nitrogen export from lower order streams; namely, are the low removal rates in Chesterville Branch (i) a characteristic of the natural system or (ii) a legacy of stream degradation? If the latter, might nitrate processing be improved (i.e., restored) by reducing the headwater loads? Further study is required to evaluate the relative importance of headwater loads versus loads from tributaries or base-flow discharge further downstream at the subcatchment outlet (see **Figure S4**), and whether these loads are responsible for degrading the in-stream processing capacity. These future studies could include prediction uncertainty analysis and monitoring network design that would reduce the uncertainty associated with the relative influence of confining unit and in-stream/near-stream mechanisms. But the results of this analysis suggest a greater urgency for placement of

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

BMPs in the Chesterville Branch catchment as well as continued in-stream monitoring that may detect their potential effects.

5 Acknowledgements

This study received funding by the National Science Foundation under award numbers CBET-0846244 and CCF-1451708. The authors thank Jim Tesoriero and two anonymous reviewers for their improvements to the manuscript.

6 Data Availability

The data generated during this study, including input and output files for the simulations referred to in the manuscript, are available as a USGS data release (Zell and Sanford, 2019).

7 Supplemental Material

The supplemental material for this manuscript describes the model development and calibration procedure in greater detail.

8 WORKS CITED

Alexander, R. B., Böhlke, J. K., Boyer, E. W., David, M. B., Harvey, J. W., Mulholland, P. J., Wollheim, W. M. (2009). Dynamic modeling of nitrogen losses in river networks unravels the coupled effects of hydrological and biogeochemical processes. *Biogeochemistry*, *93*(1-2), 91-116.

Aquilina, L., Vergnaud-Ayraud, V., Labasque, T., Bour, O., Molénat, J., Ruiz, L., ... & Longuevergne, L. (2012). Nitrate dynamics in agricultural catchments deduced from groundwater dating and long-term nitrate monitoring in surface-and groundwaters. *Science of the Total Environment*, *435*, 167-178.

Arnold, J. G., Allen, P. M., Muttiah, R., & Bernhardt, G. (1995). Automated base flow separation and recession analysis techniques. *Groundwater*, *33*(6), 1010-1018.

Ator, S. W., & Denver, J. M. (2012). Estimating contributions of nitrate and herbicides from groundwater to headwater streams, Northern Atlantic Coastal Plain, United States. *JAWRA Journal of the American Water Resources Association*, *48*(6), 1075-1090.

Bachman, L. J., Krantz, D. E., Böhlke, J., & Hantush, M. M. (2002). Hydrogeologic framework, ground-water geochemistry, and assessment of nitrogen yield from base flow in two agricultural watersheds, Kent County, Maryland. U.S. Environmental Protection Agency, Washington, D.C., EPA/600/R-02/008.

Bernot, M. J., Tank, J. L., Royer, T. V., & David, M. B. (2006). Nutrient uptake in streams draining agricultural catchments of the midwestern United States. *Freshwater Biology*, *51*(3), 499-509.

Böhlke, J. K., & Denver, J. M. (1995). Combined use of groundwater dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds, Atlantic Coastal Plain, Maryland. *Water Resources Research*, *31*(9), 2319-2339.

Böhlke, J. K., Antweiler, R. C., Harvey, J. W., Laursen, A. E., Smith, L. K., Smith, R. L., & Voytek, M. A. (2009). Multi-scale measurements and modeling of denitrification in streams with varying flow and nitrate concentration in the Upper Mississippi River Basin, USA. *Biogeochemistry*, *93*(1-2), 117-141.

Boyer, E. W., Alexander, R. B., Parton, W. J., Li, C., Butterbach-Bahl, K., Donner, S. D., ... & Grosso, S. J. D. (2006). Modeling denitrification in terrestrial and aquatic ecosystems at regional scales. *Ecological Applications*, *16*(6), 2123-2142.

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

Clay, D. E., Zheng, Z., Liu, Z., Clay, S. A., & Trooien, T. P. (2004). Bromide and nitrate movement through undisturbed soil columns. *Journal of Environmental Quality*, *33*(1), 338-342.

Clune, J. W., & Denver, J. M. (2012). *Residence time, chemical and isotopic analysis of nitrate in the groundwater and surface water of a small agricultural watershed in the coastal plain, Bucks Branch, Sussex County, Delaware.* U.S. Geological Survey Scientific Investigations Report 2012-5235.

Conan, C., Bouraoui, F., Turpin, N., de Marsily, G., & Bidoglio, G. (2003). Modeling flow and nitrate fate at catchment scale in Brittany (France). *Journal of Environmental Quality*, *32*(6), 2026-2032.

Duff, J. H., Tesoriero, A. J., Richardson, W. B., Strauss, E. A., & Munn, M. D. (2008). Whole-stream response to nitrate loading in three streams draining agricultural landscapes. *Journal of Environmental Quality*, *37*(3), 1133-44.

Fienen, M. N., Muffels, C. T., & Hunt, R. J. (2009). On constraining pilot point calibration with regularization in PEST. *Groundwater*, *47*(6), 835-844.

Galbiati, L., Bouraoui, F., Elorza, F. J., & Bidoglio, G. (2006). Modeling diffuse pollution loading into a Mediterranean lagoon: development and application of an integrated surface–subsurface model tool. *Ecological Modelling*, *193*(1-2), 4-18.

Gitau, M. W., Chaubey, I., Gbur, E., Pennington, J. H., & Gorham, B. (2010). Impacts of land-use change and best management practice implementation in a conservation effects assessment project watershed: Northwest Arkansas. *Journal of Soil and Water Conservation*, *65*(6), 353-368.

Green, C. T., Puckett, L. J., Böhlke, J. K., Bekins, B. A., Phillips, S. P., Kauffman, L. J., Denver, J.M., Johnson, H. M. (2008). Limited occurrence of denitrification in four shallow aquifers in agricultural areas of the United States. *Journal of Environmental Quality*, *37*(3), 994-1009.

Harbaugh, A. W. (2005). MODFLOW-2005, the US geological survey modular ground-water model: The ground-water flow process. U.S. Geological Survey Techniques and Methods 6-A16.

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, 42(3).

Lindsey, B. D., Phillips, S. W., Donnelly, C. A., Speiran, G. K., Plummer, L. N., Böhlke, J.K., Busenberg, E. (2003). Residence times and nitrate transport in ground water discharging to streams in the Chesapeake Bay Watershed. U.S. Geological Survey Water-Resources Investigations Report 2003–4035.

Meals, D. W., Dressing, S. A., & Davenport, T. E. (2010). Lag time in water quality response to best management practices: A review. *Journal of Environmental Quality*, 39(1), 85-96.

Molénat, J., Gascuel-Odoux, C., Ruiz, L., & Gruau, G. (2008). Role of water table dynamics on stream nitrate export and concentration in agricultural headwater catchment (France). *Journal of Hydrology*, *348*(3-4), 363-378.

Montreuil, O., Merot, P., & Marmonier, P. (2010). Estimation of nitrate removal by riparian wetlands and streams in agricultural catchments: effect of discharge and stream order. *Freshwater Biology*, *55*(11), 2305-2318.

Mulholland, P. J., Helton, A. M., Poole, G. C., Hall, R. O., Hamilton, S. K., Peterson, B. J., Dahm, C. N. (2008). Stream denitrification across biomes and its response to anthropogenic nitrate loading. *Nature*, 452(7184), 202-205.

Nelson, J., & Spies, P. (2013). The Upper Chester river watershed: Lessons learned from a focused, highly partnered, voluntary approach to conservation. *Journal of Soil and Water Conservation*, 68(2), 41A-44A.

Ocampo, C. J., Sivapalan, M., & Oldham, C. (2006). Hydrological connectivity of upland-riparian zones in agricultural catchments: Implications for runoff generation and nitrate transport. *Journal of Hydrology*, *331*(3-4), 643-658.

Osmond, D., Meals, D., Hoag, D., Arabi, M., Luloff, A., Jennings, G., Line, D. (2012). Improving conservation practices programming to protect water quality in agricultural watersheds: Lessons learned from the National Institute of Food and Agriculture-Conservation Effects Assessment Project. *Journal of Soil and Water Conservation*, *67*(5), 122A-127A.

Petry, J., Soulsby, C., Malcolm, I. A., & Youngson, A. F. (2002). Hydrological controls on nutrient concentrations and fluxes in agricultural catchments. *Science of the Total Environment*, *294*(1-3), 95-110.

Pollock, D. W. (2012). User guide for MODPATH version 6: A particle tracking model for MODFLOW. U.S. Geological Survey Techniques and Methods 6-A41.

Puckett, L. J., Tesoriero, A. J., & Dubrovsky, N. M. (2011). Nitrogen contamination of surficial aquifers: A growing legacy. *Environmental Science and Technology*, *45*(3), 839.

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

Puckett, L. J., Zamora, C., Essaid, H., Wilson, J. T., Johnson, H. M., Brayton, M. J., & Vogel, J. R. (2008). Transport and fate of nitrate at the ground-water/surface-water interface. *Journal of Environmental Quality*, *37*(3), 1034-50.

Royer, T. V., Tank, J. L., & David, M. B. (2004). Transport and fate of nitrate in headwater agricultural streams in Illinois. *Journal of Environmental Quality*, *33*(4), 1296-1304.

Sanford, W. E., & Pope, J. P. (2013). Quantifying groundwater's role in delaying improvements to Chesapeake Bay water quality. *Environmental Science and Technology, 47*(23), 13330-13338.

Scanlon, T. M., Ingram, S. M., & Riscassi, A. L. (2010). Terrestrial and in-stream influences on the spatial variability of nitrate in a forested headwater catchment. *Journal of Geophysical Research: Biogeosciences*, *115*(G2).

Schindler, D. W., & Vallentyne, J. R. (2008). *The Algal Bowl*. University of Alberta Press.

Science and Technical Advisory Committee (2005). *Understanding Lag Times Affecting the Improvement of Water Quality in the Chesapeake Bay: A Report from the Chesapeake Bay Program Scientific and Technical Advisory Committee*. Science and Technical Advisory Committee Publication 5-001. Accessed August 8, 2019 at http://www.chesapeake.org/pubs/lagtimereport.pdf.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database. Available online at http://sdmdataaccess.nrcs.usda.gov/.

Staver, K. W., & Brinsfield, R. B. (1998). Using cereal grain winter cover crops to reduce groundwater nitrate contamination in the Mid-Atlantic Coastal Plain. *Journal of Soil and Water Conservation*, *53*(3), 230-240.

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, & Soil Pollution*, *199*(1), 353-369.

Tesoriero, A.J., Duff, J.H., Saad, D.A., Spahr, N.E. and Wolock, D.M., 2013. Vulnerability of streams to legacy nitrate sources. *Environmental Science and Technology*, v. 47, 3623-3629.

U.S. Geological Survey, 2016, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed January 29, 2016, at https://doi.org/10.5066/F7P55KJN.

Wei, X., Bailey, R. T., Records, R. M., Wible, T. C., & Arabi, M. (2018). Comprehensive simulation of nitrate transport in coupled surface-subsurface hydrologic systems using the linked SWAT-MODFLOW-RT3D model. *Environmental* Modelling *& Software*.

Wollheim, W. M., Vörösmarty, C. J., Peterson, B. J., Seitzinger, S. P., & Hopkinson, C. S. (2006). Relationship between river size and nutrient removal. *Geophysical Research Letters*, *33*(6).

Wriedt, G., Spindler, J., Neef, T., Meißner, R., & Rode, M. (2007). Groundwater dynamics and channel activity as major controls of in-stream nitrate concentrations in a lowland catchment system? *Journal of Hydrology*, *343*(3-4), 154-168.

Vidon, P. G., & Hill, A. R. (2004). Landscape controls on nitrate removal in stream riparian zones. *Water Resources Research*, *40*(3).

Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., Schlesinger, W.H., and Tilman, D. G. (1997). Human alteration of the global nitrogen cycle: sources and consequences. *Ecological Applications*, *7*(3), 737-750.

Yevenes, M. A., and Mannaerts, C. M. (2012). Untangling hydrological pathways and nitrate sources by chemical appraisal in a stream network of a reservoir catchment. *Hydrology and Earth System Sciences*, *16*(3), 787-799.

Zell, W. O., Culver, T. B., & Sanford, W. E. (2018). Prediction uncertainty and data worth assessment for groundwater transport times in an agricultural catchment. *Journal of Hydrology*, *561*, 1019-1036.

Zell, W.O., and Sanford, W.E. (2019). MODFLOW-2005 and MODPATH6 models used to simulate groundwater flow and nitrate transport in two tributaries to the Upper Chester River, Maryland. U.S. Geological Survey data release. https://doi.org/10.5066/P9VWY11M

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

9 FIGURE CAPTIONS

Figure 1. Upper Chester study area. The heavy black line delineates the model domain.

Figure 2. (a) Crop acreage, (b) agricultural nitrogen inputs and exports, and (c) estimated nitrate concentrations for agricultural recharge in Kent County, MD. See the Supplemental Materials for a complete description of the input and export datasets and the calculation of the recharging nitrate time series. The high loading scenario is the rate calculated by restricting the county-scale mass of recharging nitrate to only reported corn acreage, as implemented in Equation S3 and used in this study. The low loading scenario is the rate calculated by distributing the recharging nitrate load to the sum of reported corn, soybean and wheat acreage and shown here only for purposes of comparison.

Figure 3. Observed stream nitrate concentrations at the (a) Morgan Creek and (b) Chesterville Branch gages (see **Figure 1** for gage locations). Crosses show those observations determined to have occurred under base-flow conditions and used to formulate calibration targets for this study; hollow circles show observations determined to have occurred under event flow conditions. See Figure 6 for time periods of data collection. Stream discharge and nitrate concentrations downloaded from the National Water Information System (NWIS; U.S. Geological Survey, 2016).

Figure 4. Simulated vs. observed groundwater nitrate concentrations for (a) upland and (b) riparian locations for the spatially-distributed nitrate loading scenarios estimated during stage 1 of model calibration. Each vertical line shows the range of nitrate values simulated by the multiple calibration scenarios for a single point in space and time (recalling that some observation locations have multiple measurements).

Figure 5. Model performance and estimated values for nitrate removal parameters for the Stage 2 calibration scenarios. The ‘Fixed CU’ and ‘No TTD Scaling’ scenarios are described in the text; see the Supplemental Material

for full description of the remaining scenarios. Error bars express +/- two standard deviations, calculated by PEST++ using Schur's complement (cf. Fienen et al., 2010).

Figure 6. **Simulated stream nitrate.** The shading in (a) shows the range of concentrations simulated by the nine models with the lowest WSSE (see **Table 1**); markers in (a) show the annually-averaged stream concentrations used as calibration targets; the error bars for each marker show the range of base-flow nitrate concentrations from which the annual average was calculated. Error bars without an accompanying marker show data acquired after model development and not used in calibration. The shaded and hatched regions in (b) are computed from the mean of the nine models with the lowest WSSE (see **Figure 5**); the dashed line in (b) show the simulated results of the single Fixed CU scenario.

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

10 TABLES

Table 1. Model Scenarios. WSSE = Weighted sum of squared errors calculated during stream model calibration. Model performance rank is 1 (best) to 12 (worst) and is discussed in the Results section, below.

Stage 1: Nitrate Loading to Water Table			Stage 2: Nitrate Removal		
Base Model	**Loading Scenario Name**	**Groundwater NO_3 Weighting Scheme**	**Stream Scenario Name**	**Fixed Parameters**	**Model Rank (WSSE)**
Low BFI	LowBFI Reference	[No additional calibration; spatially-constant loading derived from county data]	LowBFI Reference		3 (138)
	LowBFI A	Standard error of measurement	LowBFI A		1 (116)
	LowBFI B	Natural log of standard error of measurement	LowBFI B		12 (1225)
	LowBFI Mean	[No additional calibration; each pixel in the loading field equal to the mean of the A and B scenarios]	LowBFI Mean		2 (120)
			LowBFI FixedCU	Confining Unit Removal Fraction = 0.80	4 (172)
			LowBFI No TTD Scaling	TTD Scale Factor = 1	10 (482)
High BFI	HighBFI Reference	[No additional calibration; spatially-constant loading derived from county data]	HighBFI Reference		8 (257)
	HighBFI A	Standard error of measurement	HighBFI A		5 (190)
	HighBFI B	Natural log of standard error of measurement	HighBFI B		7 (234)
	HighBFI Mean	[No additional calibration; each pixel in the loading field equal to the mean of the A and B scenarios]	HighBFI Mean		6 (197)
			HighBFI FixedCU	Confining Unit Removal Fraction = 0.80	9 (302)
			HighBFI No TTD Scaling	TTD Scale Factor = 1	11 (901)

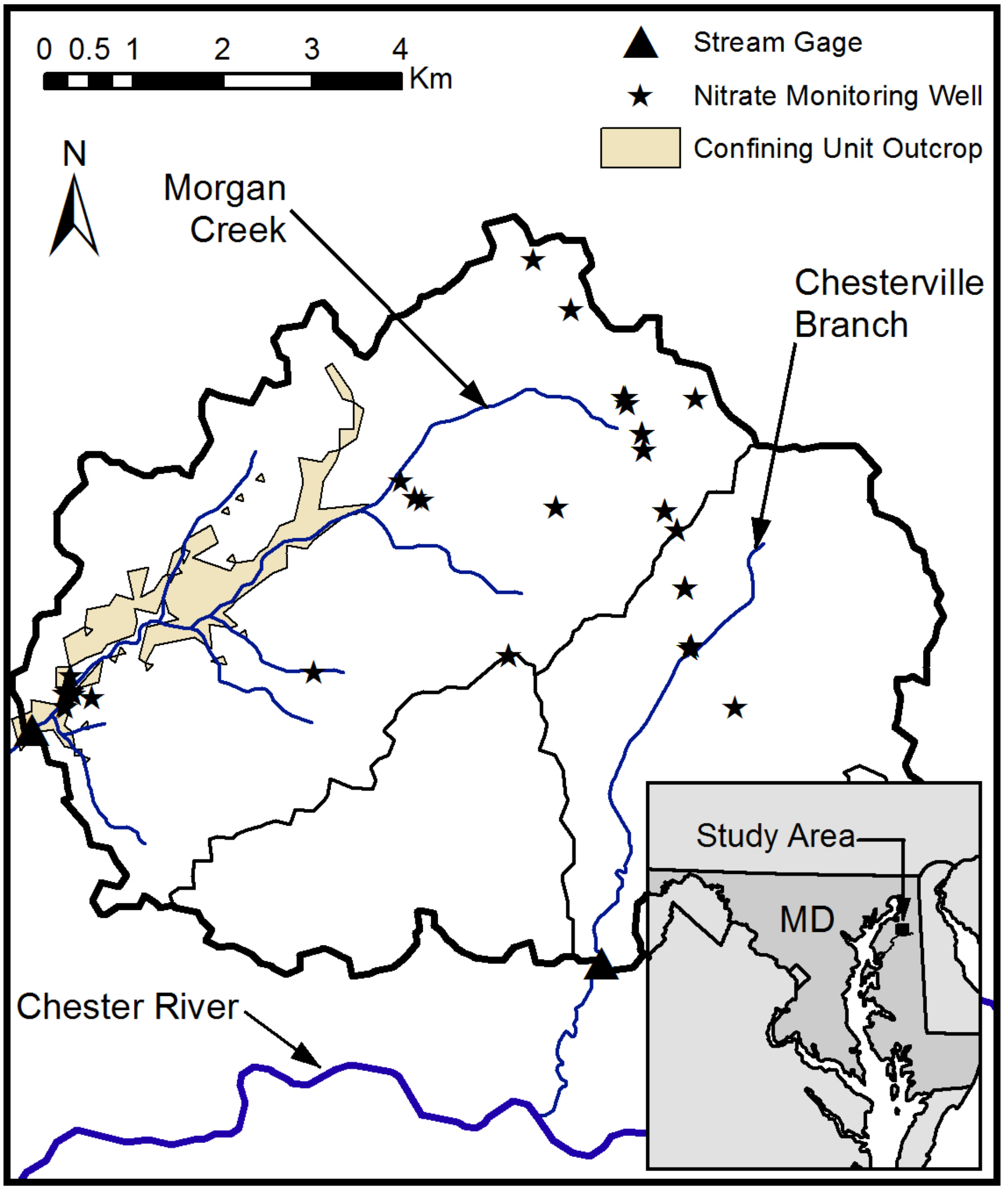

Figure 1. Upper Chester study area. The heavy black line delineates the model domain.

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

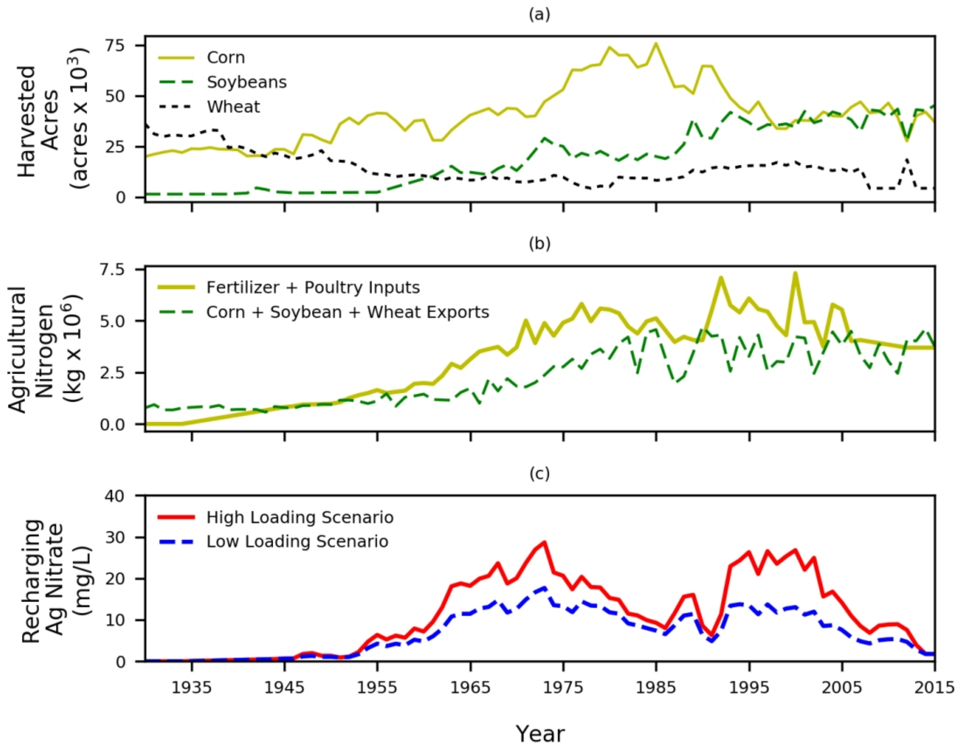

Figure 2. (a) Crop acreage, (b) agricultural nitrogen inputs and exports, and (c) estimated nitrate concentrations for agricultural recharge in Kent County, MD. See the Supplemental Materials for a complete description of the input and export datasets and the calculation of the recharging nitrate time series. The high loading scenario is the rate calculated by restricting the county-scale mass of recharging nitrate to only reported corn acreage, as implemented in Equation S3 and used in this study. The low loading scenario is the rate calculated by distributing the recharging nitrate load to the sum of reported corn, soybean and wheat acreage and shown here only for purposes of comparison.

127x101mm (300 x 300 DPI)

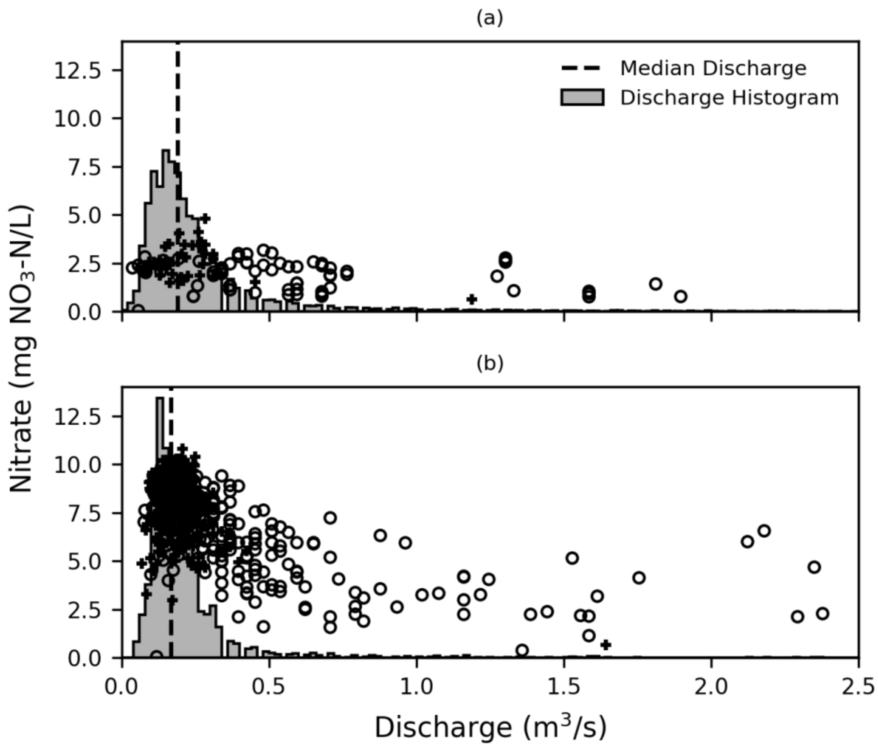

Figure 3. Observed stream nitrate concentrations at the (a) Morgan Creek and (b) Chesterville Branch gages (see Figure 1 for gage locations). Crosses show those observations determined to have occurred under base-flow conditions and used to formulate calibration targets for this study; hollow circles show observations determined to have occurred under event flow conditions See Figure 6 for time periods of data collection. Stream discharge and nitrate concentrations downloaded from the National Water Information System (NWIS; U.S. Geological Survey, 2016).

127x101mm (300 x 300 DPI)

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

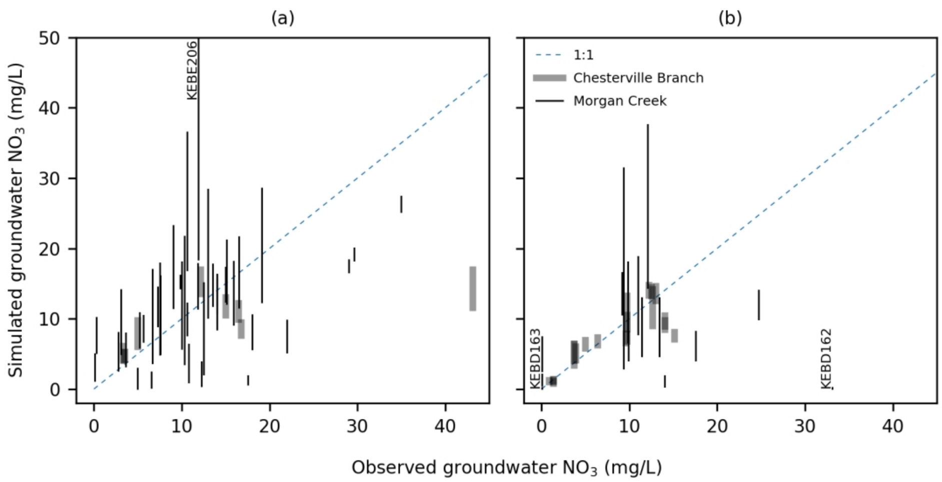

Figure 4. Simulated vs. observed groundwater nitrate concentrations for (a) upland and (b) riparian locations for the spatially-distributed nitrate loading scenarios estimated during stage 1 of model calibration. Each vertical line shows the range of nitrate values simulated by the multiple calibration scenarios for a single point in space and time (recalling that some observation locations have multiple measurements).

158x88mm (300 x 300 DPI)

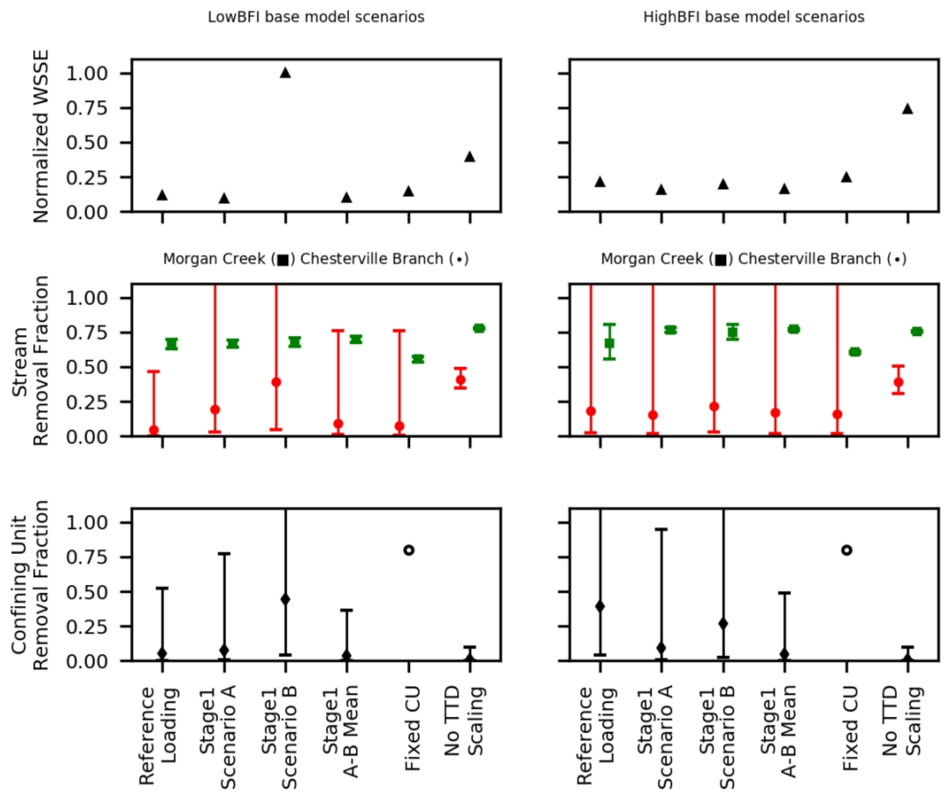

Figure 5. Model performance and estimated values for nitrate removal parameters for the Stage 2 calibration scenarios. The 'Fixed CU' and 'No TTD Scaling' scenarios are described in the text; see the Supplemental Material for full description of the remaining scenarios. Error bars express +/- two standard deviations, calculated by PEST++ using Schur's complement (cf. Fienen et al., 2010).

107x88mm (300 x 300 DPI)

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

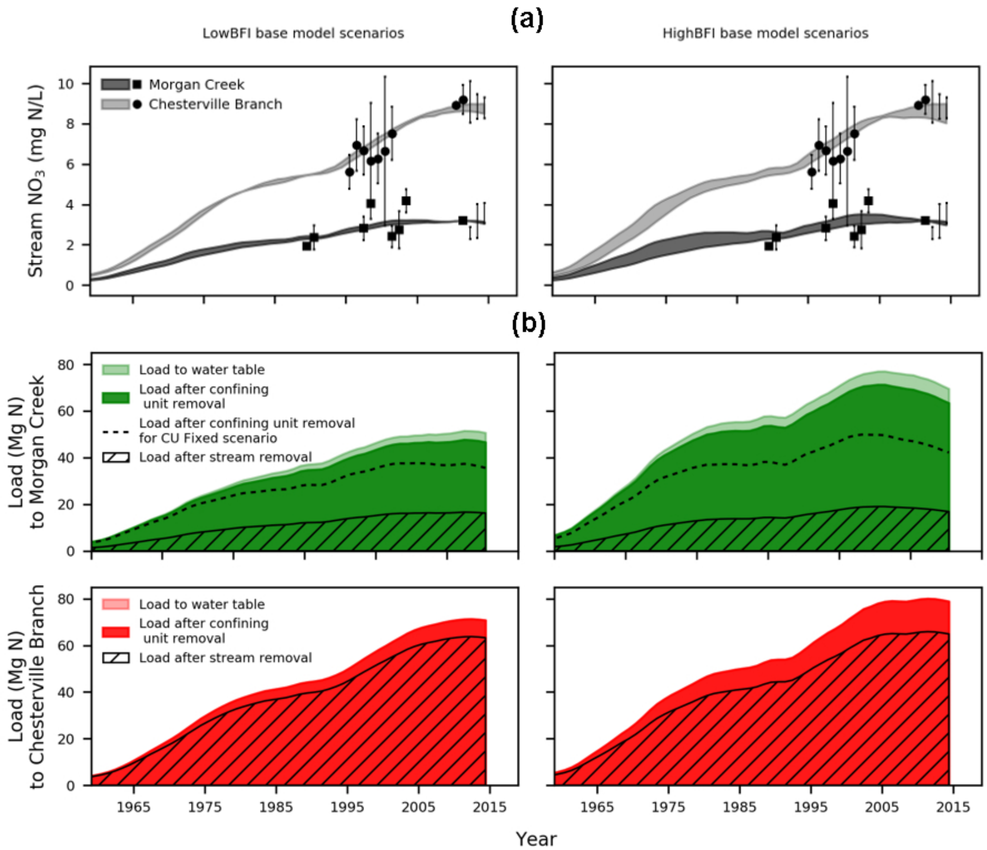

Figure 6. Simulated stream nitrate. The shading in (a) shows the range of concentrations simulated by the nine models with the lowest WSSE (see Table 1); markers in (a) show the annually-averaged stream concentrations used as calibration targets; the error bars for each marker show the range of base-flow nitrate concentrations from which the annual average was calculated. Error bars without an accompanying marker show data acquired after model development and not used in calibration. The shaded and hatched regions in (b) are computed from the mean of the nine models with the lowest WSSE (see Figure 5); the dashed line in (b) show the simulated results of the single Fixed CU scenario.

158x135mm (144 x 144 DPI)

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

Supplemental Materials for:

Quantifying Background Nitrate Removal Mechanisms in an Agricultural Watershed with Contrasting Subcatchment Base-flow Concentrations

Wesley O. Zell[1] (wzell@usgs.gov) *corresponding author

Teresa B. Culver[2] (tbc4e@virginia.edu)

Ward E. Sanford[1] (wsanford@usgs.gov)

Jonathan L. Goodall[2] (goodall@virginia.edu)

[1] U.S. Geological Survey, 12201 Sunrise Valley Dr, MS 432, Reston, VA 20192

[2] Department of Civil and Environmental Engineering, University of Virginia, Thornton Hall, P.O. Box 400259, Charlottesville, VA 22904-4259

This supplemental materials section contains 20 pages, including 3 tables and 5 figures.

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

S1 Calculation of nitrate inputs to the water table

S1.A Estimation of nitrate loading to the water table in monitored portions of the catchment (Calibration Stage 1a)

Nitrate loading to the water table was first estimated by calibrating the nitrate transport model solely against observed groundwater nitrate concentrations. Groundwater nitrate observations were downloaded from the U.S. Geological Survey (USGS) National Water Information System (NWIS; U.S. Geological Survey, 2016) and aggregated into annually averaged concentrations at each observation well; in total, 213 subsurface nitrate measurements at 52 different wells were aggregated into 89 subsurface calibration targets that date between 1988 and 2004 (see **Figure 1** in manuscript for locations of nitrate monitoring wells).

In order to allow spatial variation of water table nitrate loading, we parameterized the inputs as a two-dimensional (2D) set of loading factors. Loading factors were estimated using an evenly spaced grid of pilot points, with pilot point separation approximating the length scale of the smallest agricultural fields. The 2D loading field of nitrate inputs was then interpolated from the loading factors using ordinary kriging implemented with the PEST utilities PPKFAC and FAC2REAL (Doherty, 2015). Parameter estimation was performed using PEST ++ (Welter et al., 2015) with singular value decomposition (SVD) and preferred-value Tikhonov regularization. Briefly stated, these regularization devices make a highly parameterized inverse-modeling problem well-posed and avoid over-fitting by constraining a parameter value to some prior estimate unless the calibration data provide a compelling reason for that estimate to change; cf. Fienen et al. (2009) for more detailed description of this parameter estimation methodology.

For the years 1930-2015, spatially and temporarily variable nitrate inputs were calculated by multiplying the interpolated loading factors by a reference load derived from agricultural and wet deposition components for that year; agricultural loading prior to 1930 was assumed to be zero. The reference load for each year consisted of an agricultural portion and a wet deposition portion. The agricultural portion of the reference load for each year was derived from county-level nitrogen budgets, with the total mass of agricultural nitrogen available for recharge in year *i* calculated as

$$N_{tot,in,i} = N_{ag,in,i} - N_{ag,out,i} \qquad \textbf{(Eq. S1)}$$

where $N_{ag,in}$ is the agricultural nitrogen inputs and $N_{ag,out}$ is the agricultural nitrogen exports. Historical county-level agricultural nitrogen inputs were derived from estimated and reported inorganic fertilizer sales (Alexander and Smith, 1990; Gronberg and Spahr, 2012; Brakebill and Gronberg, 2017) and estimates of poultry manure production (Sanford and Pope, 2013) (see **Figure 3** in manuscript). Historical county-level agricultural nitrogen exports were derived from the annual production of corn, soybeans, and wheat as published by the National Agricultural Statistics Service (NASS). The amount produced of each crop was converted to mass nitrogen by assuming the nitrogen content of harvested crops to be 0.9, 3.8, and 1.5 pounds nitrogen per bushel for corn, soybeans, and wheat, respectively (Murrell, 2008). In the Mid-Atlantic as much as 65-75% of the nitrogen content of soybeans can be due to atmospheric fixation and not to fertilizer inputs (http://extension.udel.edu/factsheets/nitrogen-management-for-soybean); for our mass balance calculations we consequently adjusted the nitrogen content coefficient of soybeans to 1.1 pounds nitrogen exported per bushel. The nitrogen within

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

reported harvested silage (which is not reported for the full period of record) was assumed to remain in the catchment and thus be available for leaching.

For each year, county-level estimates of the residual nitrogen available after crop uptake (Eq. 2) were converted to an areal loading rate that was applied within the model domain. Corn receives a much higher fraction of total fertilizer inputs for a given year than other crops (Hancock and Brayton, 2006), such that the ratio of fertilizer sales to harvested corn acreage provides a provisional estimate of the areal loading rate for those areas where it was applied. We consequently calculated the reference rate for year *i* as

$$Rate_{Ref,i} = \frac{N_{tot,in,i}}{Area\ Corn_i} + Rate_{atm,i} \qquad \textbf{(Eq. S2)}$$

where $N_{tot,in,i}$ is the county-level mass of nitrogen remaining after crop export for year *i* (**Eq. S1**), $Area\ Corn_i$ is the county-level area of harvested corn for the year I, and $Rate_{atm,i}$ is rate of annual wet deposition. Rates of nitrate wet deposition were obtained from the National Atmospheric Deposition Program monitoring site in Wye, Maryland, approximately 30 miles southeast of the study site (data downloaded from http://nadp.sws.uiuc.edu on 6/4/2015). Wet deposition data were available from 1983-2006. We assumed zero wet deposition for years prior to 1935; for years between 1935 and 1983 we used a linear interpolation to estimate annual wet deposition rates.

Several factors govern the delivery of excess nitrates to the water table and their transport through the subsurface to discharge locations. For example, multiple researchers have shown the particular sensitivity of leachate concentrations to precipitation patterns, as rainfall deficits during the growing season reduce crop uptake efficiencies and increase pools of excess nitrate

(Burt et al., 2008), while large rainfall amounts post-harvest accelerate nitrate flushing from the root zone to the water table (Staver and Brinsfield, 1998). In order to account for the dampening of the nitrate input signal that likely occurs at a given location through delays and mixing in the root zone and unsaturated zone (as well as similar dampening that would occur due to crop rotation – cf. Hancock and Brayton, 2006), we transformed the time series calculated with **Equation S2** using a 3-year moving average.

Note that in the case of an input signal that is constant across the landscape at a given time (as is effectively true, e.g., of age tracers that recharge from the atmosphere), the contributing recharge area for each monitoring location is unimportant. However, even in a majority agricultural catchment, the sources of nitrate may vary dramatically across the landscape, such that for purposes of simulating the nitrate concentration at monitoring wells the input signal must be specified with respect to both time and space. We generated distinct estimated distributions of loading factors by performing the Stage 1 calibration with both the LowBFI and HighBFI transport base models. In addition, we used two different weighting schemes to determine the relative importance of the different groundwater nitrate calibration targets during the optimization. Briefly stated, the weight for each annually averaged groundwater nitrate observation was initially calculated from the standard error of measurement (weighting scenario A) for all measurements within a given year (i.e., for the distribution of measurements that were aggregated into the annual average). Because this weighting scheme resulted in a large disparity of weights and a relatively small number of observations dominating the regression we also considered an alternative weighting scheme in which we reduced the range

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

of weights by using a natural log transform and enforcing a minimum weight (weighting scenario B). We therefore generated four separate water table loading scenarios for the monitored portion of the catchment. After extrapolation of the loading field from the monitored to the unmonitored portions of the catchment we generated additional loading scenarios by calculating – for each base model – the spatially distributed means of the two weighting scenarios (**Table S1**).

Table S1. Water table loading scenarios generated from Stage 1 of parameter estimation.

Scenario Name	Base Model	Groundwater NO_3 Weighting Scheme	Additional Transformations
LowBFI Scenario A	LowBFI	Standard error of measurement	--
LowBFI Scenario B	LowBFI	Natural log of scenario A weights	--
LowBFI Scenario A-B Mean	LowBFI	--	Spatially distributed field with each point equal to the mean of Distributed_A and Distributed_B scenarios
HighBFI Scenario A	HighBFI	Standard error of measurement	--
HighBFI Scenario B	HighBFI	Natural log of scenario A weights	--
HighBFI Scenario A-B Mean	HighBFI	--	Spatially distributed field with each point equal to the mean of A and B scenarios

S1.B Extrapolation of the estimated water table loading from the monitored portion to the unmonitored portions of the catchment (Calibration Stage 1b)

Following the Stage 1a estimation of nitrate loading factors we differentiated (i) the portions of the estimated loading field that were well-constrained by the groundwater data from (ii) those portions that were not well-constrained and thus required some other mechanism for estimating the loading factor. We defined the ‘monitored’ portion of the landscape as those areas for which the post-calibration reduction in parameter uncertainty (i.e., compared to the pre-calibration parameter uncertainty) for the nitrate loading factor was at

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

least 10%, where the post-calibration parameter uncertainty is derived via linearized Bayesian methods,

$$\Sigma_{\theta,post} = \Sigma_{\theta} - \Sigma_{\theta} \boldsymbol{J}^{T} \left[\boldsymbol{J} \Sigma_{\theta} \boldsymbol{J}^{T} + \Sigma_{\varepsilon} \right]^{-1} \boldsymbol{J} \Sigma_{\theta} \quad \textbf{(Eq. S3)}$$

where $\Sigma_{\Theta,post}$ is the post-calibration parameter covariance, Σ_{Θ} is the covariance matrix of prior parameter probability distribution (here derived from the estimated bounds on parameter values); $\boldsymbol{J}$ is the Jacobian matrix of observation sensitivity to parameter perturbations; and Σ_{ε} is the covariance matrix of simulation error and measurement error. For this study Σ_{ε} is defined as a diagonal matrix populated by the inverse of observation weights (all off-diagonal elements are equal to 0).

In order to extrapolate the water table nitrate loading to the entire simulation domain, we used a Gradient Boosted Regression (GBR; implemented with the Python scikit-learn library: Pedregosa et al., 2011) to develop an empirical relationship between several candidate variables (**Table S2**) and the nitrate loading rate estimated in Stage 1. (Note that for clarity, in the manuscript and in the remainder of the Supplemental Material we use the term 'GBR-estimated', 'GBR-based', etc., to refer to the empirical relationship between candidate variables and the nitrate loading derived with the GBR; we reserve the general term 'modeled' to refer to simulation of nitrate transport described above). Each of the candidate variables listed in **Table S2** is mapped for the entire simulation domain and is thus a potential predictor of nitrate loading for those areas where no nitrate loading data (i.e., groundwater nitrate data) exists. Most of the mapped variables are derived from national-scale datasets [e.g., Cropland Data Layer (USDA, 2014), National Land Cover Dataset, and Soil Survey Geographic Database]. Two datasets (spatially distributed porosity and recharge) were estimated for the model domain during the flow and transport model development described in the companion paper (Zell et al., 2018). Two additional datasets were derived for this study. First, high resolution Maryland light detection and radar (lidar) elevation data allows identification of field-scale topographic depressions that are the result of drained wetlands. These former wetlands, referred to as 'Delmarva Bays', are often characterized by higher organic content and, therefore, potentially higher rates of soil denitrification (Ator et al., 2012). We consequently generated a map of Delmarva Bays in the model area to use as input for the machine-learning extrapolation. Second, a large commercial nursery in the headwaters of Chesterville Branch is not clearly represented in land use datasets and was consequently mapped and included as a potential explanation of nitrate loading to the water table.

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

Table S2. Mapped variables used as candidate explanatory variables for the GBR-estimated nitrate input function.

Variable Name	Description	Source
CDL_X	Land use category, where X = year	Cropland Data Layer
DelmarvaBays	Topographic indication of drained wetland	This study
Mean_DEM	Mean elevation	Maryland Lidar Dataset
NLCD_X	Land use category, where X = year	National Land Cover Dataset
Nursery	Outline of nursery in Chesterville Branch headwaters	This study
Porosity	Porosity estimated during calibration of flow and transport base model	Zell et al. (2018)
Recharge	Recharge estimated during calibration of flow and transport base model	Zell et al. (2018)
SSURGO_aws_X_wta	Available soil water storage, where X = depth of soil compartment (cm)	Soil Survey Geographic Database (SSURGO)
SSURGO_drclassdcd	Soil drainage class, dominant condition	
SSURGO_drclasswettest	Soil drainage class, wettest condition	
SSURGO_hydclprs	Soil hydric classification	
SSURGO_hydgrpdcd	Soil hydrologic group	
SSURGO_pondfreqprs	Ponding frequency	
SSURGO_wtdepannmin	Water table depth, annual minimum	
SSURGO_wtdpaprjunmin	Water table depth, summer minimum	

The GBR-based extrapolation was implemented as follows. At the conclusion of Stage 1a, the model cells in the monitored subdomain were randomly assigned to a training dataset (75% of monitored model cells) and a testing dataset (25% of monitored model cells). The training dataset was used with a 10-fold cross-validation to identify the GBR hyperparameters (e.g., number of trees, tree depth, minimum samples per leaf) that minimized predictive error; the testing dataset was reserved to evaluate the performance of the GBR after the optimal hyperparameters were identified. Finally, the tuned GBR was used to assign a nitrate loading factor to each grid cell in the unmonitored portion of the model domain.

S1.C Examination of estimated nitrate inputs to the water table

The choice of base model and the associated differences in the subsurface flow and transport regime has some effect on the transmission of groundwater nitrate information from the water table to the monitoring wells (**Figure S1**). For example, for the scenario B weighting scheme, the estimated loading field that resulted from the LowBFI scenario had more point locations for which water table inputs approach five times the county-averaged rates (cf. **Figure 3** in the manuscript). These additional point locations of high loading were predominately located in the Chesterville Branch subcatchment, resulting in the higher estimates of loading to that stream (cf. Figure 5 in the manuscript).

While the isolated extreme values visible in **Figure S1** may be considered problematic when using a highly parameterized approach to estimate a field that is expected to be smoothly varying (e.g., hydraulic conductivity), the same is not necessarily the case for the field-scale differentiation of agricultural inputs that we are simulating here. It is additionally important to note that the maximum displayed loading factors are points in the interpolated loading field rather than a field-scale average loading factor. Thus, given the little information available to describe the spatial distribution of loading through time, the heterogeneity and point magnitudes suggested by the various Stage 1 scenarios are not implausible.

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

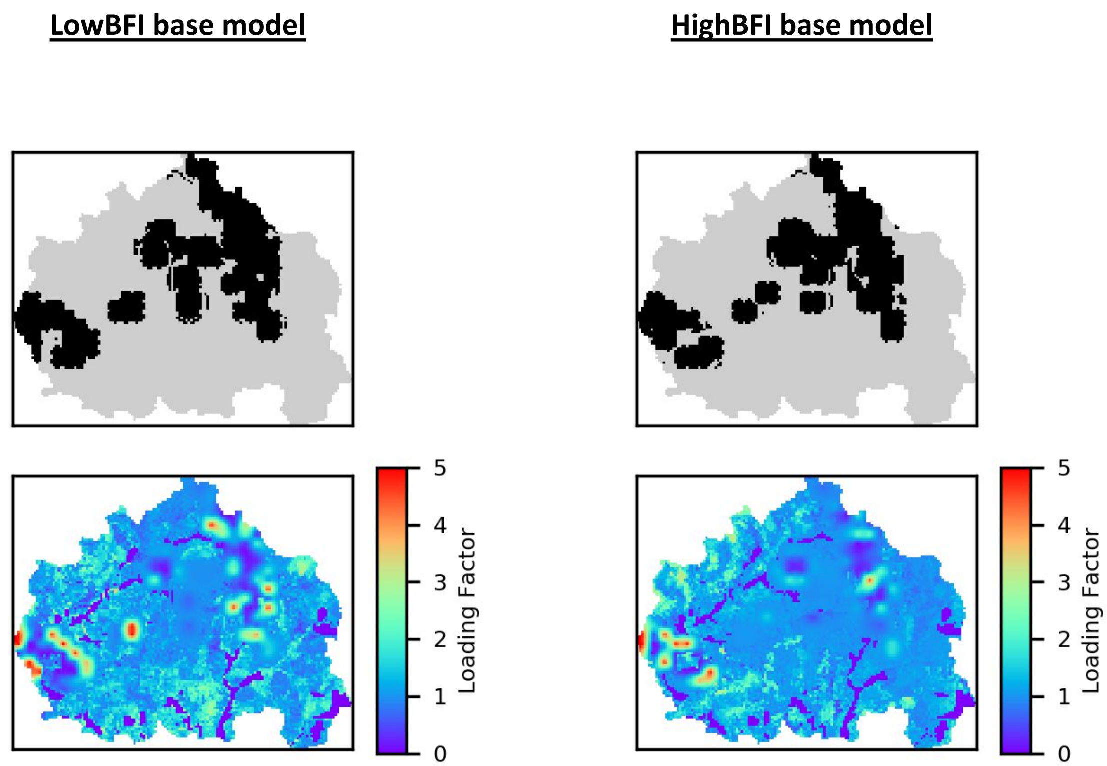

Figure S1. Outline of monitored area (top panels) and water table loading factors (bottom panels) for stage 1 Scenario B.

S1.D Examination of relationships between mapped variables and nitrate loading derived using gradient boosted regression classifier

For the GBR estimators used to extrapolate the spatial distribution of loading from the monitored to the unmonitored areas, the training R^2 ranged from 0.98 to 1.00 and the testing R^2 ranged from 0.60 – 0.72 (**Table S3**). It is important to here emphasize that the purpose of the GBR was not to predict the nitrate loading at any single location in the model domain but rather to represent the likely variance in loading across the landscape without making a priori assumptions about how that variance should be expressed through system parameterization. Representing this variance is in turn important for investigating its effect, if any, on the simulation of the nitrate mass flux seen at the catchment outlet.

Table S3. Testing R^2 values for GBR extrapolation of water table loading factors from monitored to unmonitored portions of model domain.

	LowBFI	**HighBFI**
Weighting scenario A	0.67	0.60
Weighting scenario B	0.70	0.72

While it is outside the scope of this study to fully explore the GBR-formulated relationships between the mapped variables and the water table loading factors, it is interesting to note that the GBR estimator found properties of the flow and transport system (namely, the distribution of recharge and porosity that were estimated during the transport model calibration, as well as topography) more important than other potential explanatory variables (**Figure S2**). Of

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

secondary importance was land use information derived from the Cropland Data Layer, while the Soil Survey Geographic Database (SSURGO) data played effectively no role in the estimator. Note, however, that cropland may have played a larger role in the regression if all years were aggregated into a single dataset rather than left distributed. The relative insignificance of the SSURGO variables may support our assumption of conservative nitrate behavior, since we would expect that any impacts of soils properties on the inferred loading would be due to interception and removal process (e.g., soil denitrification) rather than variations in the applied loading itself. However, because the GBR detected much more information in the land use data than the soils data, we may assume that the estimated loading factors more represent what was applied at the land surface than what was removed before reaching the water table or en route to an observation location.

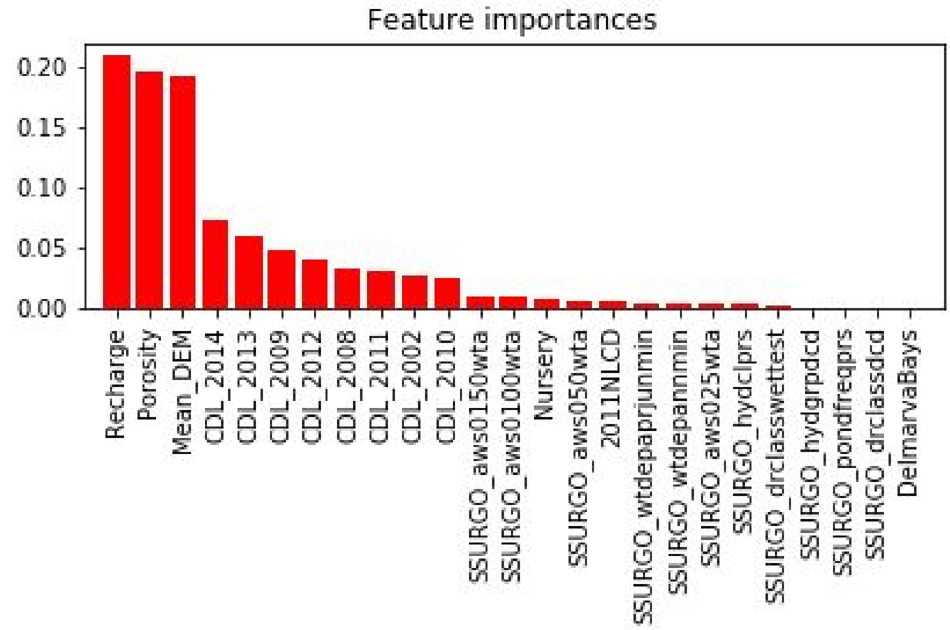

Figure S2. GBR-identified importance of explanatory variables to the estimation of nitrate loading factors. See **Table S2** for definition of variables.

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

S2 EFFECT OF CALIBRATED TRAVEL TIME SCALING FACTOR ON BASE-FLOW TRAVEL TIME DISTRIBUTIONS (DISCUSSED IN RESULTS SECTION OF MANUSCRIPT)

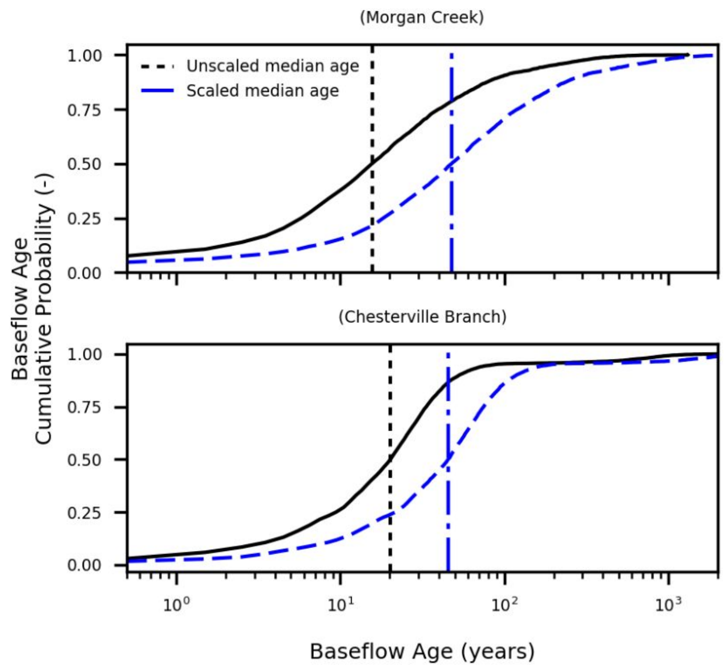

Figure S3. The base-flow age empirical cumulative distribution function (ECDF), unadjusted (solid line) and with the TTD scaling factors estimated during the LowBFI Scenario A calibration scenario (dashed line). The vertical lines show the simulated mean base-flow age for the unadjusted (dotted line) and scaled (dash-dot line) TTDs.

S3 FURTHER DISCUSSION OF STREAM NETWORK CHARACTERISTICS AS POTENTIAL DRIVERS OF CONTRASTING STREAM NITRATE REMOVAL EFFICIENCIES

The Morgan Creek riparian zone is thickly wooded, with tree debris common in the stream channel (Duff et al., 2008). As described in the manuscript, the confining unit which outcrops at the lower reaches may not only account for substantial nitrogen removal through denitrification, but also controls the manner in which discharge enters the main channel. While

Chesterville Branch has not been characterized with the same detail, it is expected that base-flow discharge to Chesterville Branch is via upwelling through the sandy bed sediments with presumably lower denitrification potential, bypassing the riparian zone processing that is an important control in Morgan Creek. This bypass has been observed in other agricultural catchments (Tesoriero et al., 2013). The organic content of the Chesterville Branch bed sediments,and the associated denitrification potential of those sediments (cf. Gu et al., 2008) is not known. Furthermore, a coarse comparison of the stream velocities and associated cross-sectional flow areas (**Figure S4**) suggests that Chesterville Branch has shorter in-stream residence times due to a shorter stream length (**Figure 1** in the manuscript) and higher velocities.

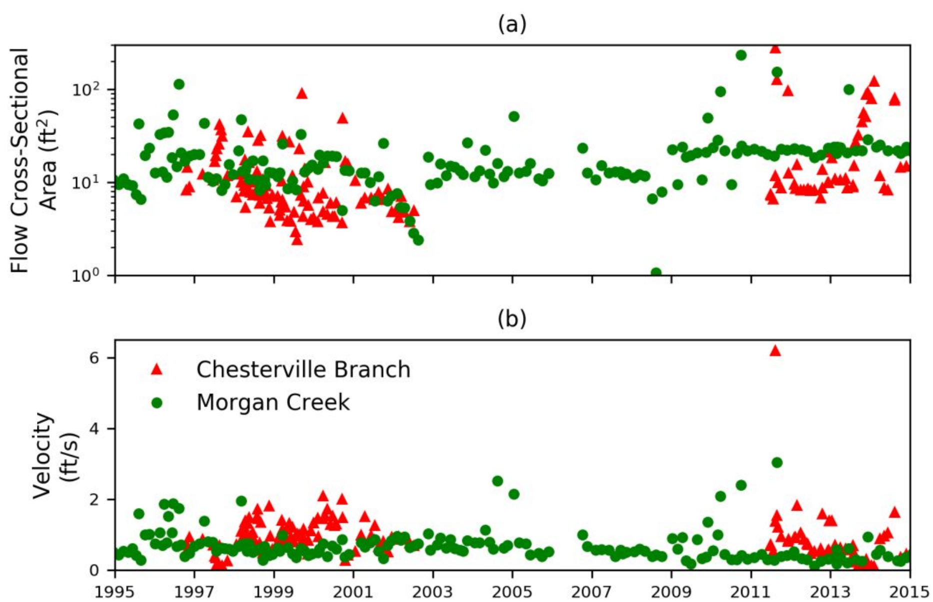

Figure S4. Flow characteristics measured at the Morgan Creek and Chesterville Branch stream gages. Each marker represents a field measurement. See **Figure 1** in the manuscript for locations.

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

Finally, evidence from a small set of synoptic studies suggests that Chesterville Branch headwater concentrations have historically been much higher than headwater concentrations in Morgan Creek (**Figure S5**). These conclusions are likewise tentative because of the few spatially distributed snapshots that include both Morgan Creek and Chesterville Branch but are consistent with the observations and conclusions of Bohlke and Denver (1995). In the early 1990s (i.e., at the time at which the stream networks were simultaneously sampled) surficial aquifer nitrate concentrations in each catchment had nitrate concentrations of 10-20 mg NO_3-N/L for observation wells near the upstream-most site in both catchments. However, Morgan Creek headwater concentrations were substantially lower than aquifer concentrations, while Chesterville Branch headwater concentrations were not.

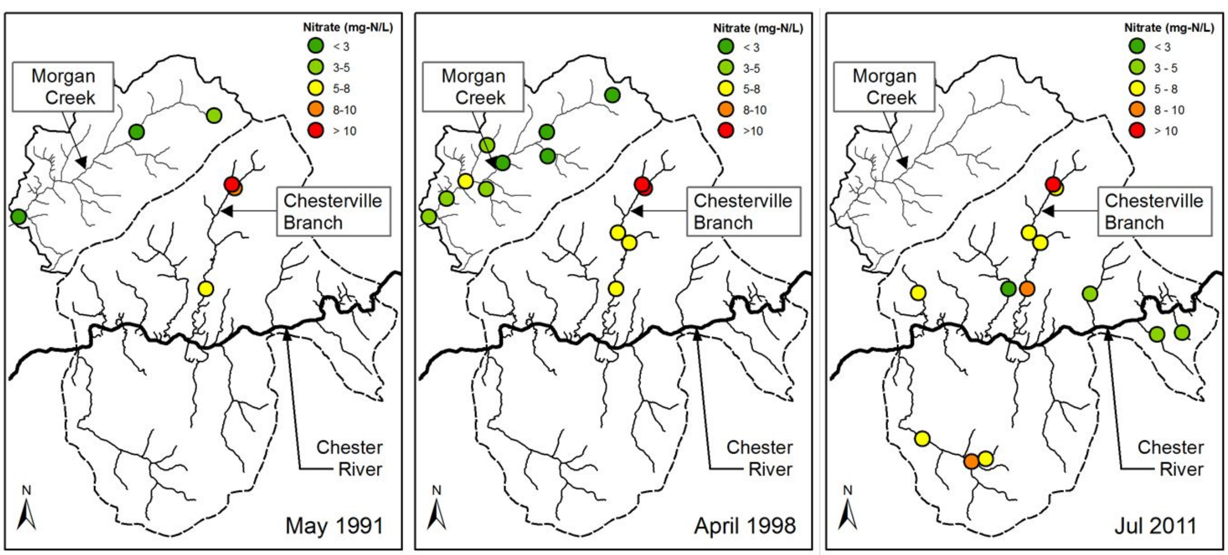

Figure S5. Base-flow stream nitrate concentrations from synoptic surface water sampling in Morgan Creek and Chesterville Branch.

S4 Works Cited

Alexander, R. B., & Smith, R. A. (1990). County-Level estimates of nitrogen and phosphorus fertilizer use in the United States, 1945 to 1985. U.S. Geological Survey Open File Report 90-130. https://doi.org/10.3133/ofr90130

Ator, S. W., & Denver, J. M. (2012). Estimating contributions of nitrate and herbicides from groundwater to headwater streams, Northern Atlantic Coastal Plain, United States. *Journal of the American Water Resources Association*, 48(6), 1075-1090.

Böhlke, J. K., & Denver, J. M. (1995). Combined use of groundwater dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds, Atlantic Coastal Plain, Maryland. *Water Resources Research*, 31(9), 2319-2339.

Brakebill, J. W., & Gronberg, J. M. (2017). County-Level Estimates of Nitrogen and Phosphorus from Commercial Fertilizer for the Conterminous United States, 1987-2012: U.S. Geological Survey data release, https://doi.org/10.5066/F7H41PKX.

Doherty, J. (2015). PEST, Model-independent parameter estimation: user manual, 5th edn.(and addendum to the PEST manual). Watermark, Brisbane, Australia. Available at www. pesthomepage.org.

Duff, J. H., Tesoriero, A. J., Richardson, W. B., Strauss, E. A., & Munn, M. D. (2008). Whole-stream response to nitrate loading in three streams draining agricultural landscapes. *Journal of Environmental Quality*, 37(3), 1133-44.

Fienen, M. N., Muffels, C. T., & Hunt, R. J. (2009). On constraining pilot point calibration with regularization in PEST. *Groundwater*, 47(6), 835-844.

Gronberg, J. M., & Spahr, N. E. (2012). County-level estimates of nitrogen and phosphorus from commercial fertilizer for the conterminous United States, 1987-2006. U.S. Geological Survey Scientific Investigations Report 2012-5207.

Gu, C., Hornberger, G. M., Herman, J. S., & Mills, A. L. (2008). Influence of stream-groundwater interactions in the streambed sediments on NO_3^- flux to a low-relief coastal stream. *Water Resources Research*, 44(11).

Hancock, T. C., & Brayton, M. J. (2006). Environmental setting of the Morgan Creek Basin, Maryland, 2002-04. US Geological Survey Open File Report 2006-1151.

Murrell, T.S. (2008). Measuring Nutrient Removal, Calculating Nutrient Budgets. *Soil Science: Step-by-step Field Analysis*.

J. Environ. Qual. Accepted Paper, posted 10/16/2019. doi:10.2134/jeq2018.11.0408

Pedregosa et al (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* (12) 2825-2830.

Sanford, W. E., & Pope, J. P. (2013). Quantifying groundwater's role in delaying improvements to Chesapeake Bay water quality. *Environmental Science and Technology*, 47(23), 13330-13338.

Tesoriero, A.J., Duff, J.H., Saad, D.A., Spahr, N.E. and Wolock, D.M., 2013. Vulnerability of streams to legacy nitrate sources. *Environmental Science and Technology*, v. 47, 3623-3629.

USDA National Agricultural Statistics Service Cropland Data Layer. (2013). Published crop-specific data layer [Online]. Available at https://nassgeodata.gmu.edu/CropScape/ (accessed March 2014). USDA-NASS, Washington, DC.

U.S. Geological Survey, 2016, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed January 29, 2016, at https://doi.org/10.5066/F7P55KJN.

Vinten, A. J. A., Smith, K. A., Burt, T. P., Heathwaite, A. L., & Trudgill, S. T. (1993). Nitrogen cycling in agricultural soils. Nitrate: processes, patterns and management, 39-73.

Welter, D. E., White, J. T., Hunt, R. J., & Doherty, J. E. (2015). Approaches in highly parameterized inversion—PEST++ Version 3, a Parameter ESTimation and uncertainty analysis software suite optimized for large environmental models. U.S. Geological Survey Techniques and Methods 7-C12. https://doi.org/10.3133/tm7C12

Zell, W. O., Culver, T. B., & Sanford, W. E. (2018). Prediction uncertainty and data worth assessment for groundwater transport times in an agricultural catchment. *Journal of Hydrology*, *561*, 1019-1036.