Implications of Lags in Pollution Delivery for Efficient Agricultural Waste Load Allocations and the Design of Water Quality Trading Programs

James Shortle
Professor of Agricultural and Environmental Economics
Environment and Natural Resources Institute
Penn State University
University Park
814-865-8720

David Abler
Professor of Agricultural, Environmental and Regional Economics and Demography
Department of Agricultural Economics, Sociology and Education
Penn State University
University Park

Zach Kaufman
City of Baltimore

Kate Zip
Assistant Professor of Environmental and Resource Economics
Department of Agricultural Economics, Sociology and Education
Penn State University
University Park

Abstract: Water quality markets allowing point-nonpoint trades assume that nonpoint BMPs have their full impact on emissions when implemented. However, ambient water quality responses to BMPs vary in time from a few months to many decades. We simulate emissions allocations that satisfy static and dynamic optimization rules to explore how simple static market designs compare economically and environmentally to complex multi-period designs in the case of nitrogen emissions to the Chesapeake Bay. In this case, static rules provide large cost savings compared to dynamic rules, but result in some delay in the achievement of water quality targets.

Introduction

Water quality trading is a mechanism that can improve the economic and environmental performance of water pollution controls. The case for trading is that markets can allocate emissions across alternative sources more efficiently than traditional regulatory instruments, and require less information on the part of regulators to achieve pollution targets (Horan and Shortle 2011). However, water quality trading poses a number of challenges for the design of markets that can achieve the theoretical potential of trading (Olmstead and Fisher Vanden 2013; Horan and Shortle 2011; Shortle 2013). One which is just now receiving attention, and which we explore in this research, is lags in the response of water quality conditions to pollution control activities. Consistent with textbook models of pollution trading, water quality markets are typically designed under the assumption that ambient environmental quality improvements resulting from pollution abatement efforts are fully realized within the time interval in which the abatement effort is undertaken. Thus, point sources of water pollution are allowed under existing water quality trading programs to immediately substitute credits generated by installation of agricultural best management practices (BMPs) for reductions in point source effluent. This substitution occurs as if the BMPs produce pollution reductions that are contemporaneous substitutes for point source effluent reductions. This assumption greatly simplifies market design. However, ambient quality responses to nonpoint pollution controls can vary in time from a few months to decades, making the assumption of contemporaneous substitution generally invalid.

Markets designed under the assumption of contemporaneous substitution may perform the economic functions of markets (economic efficiency) well when lags are the reality, but they will fail to achieve their ecological functions (achieving pollution reductions needed to achieve water quality goals) for some period of time depending on the length and structure of the lags. The "fix" in theory would be to design markets that allow for trading across possibly long periods of time as well as across space. But futures for trading commodities across long periods of time can be very complex to implement, expensive to operate, and may not perform well economically when the commodities are complex or there is significant uncertainty about future economic conditions and regulatory environments (Carlton 1984). In consequence, retaining the simplicity of markets designed under the assumption of contemporaneous substitution (i.e. no lags) might work out well if the cost savings from the simpler design are significant and if the delays in achieving environmental targets are acceptable.

This research is motivated by a need to understand the implications of lags in agricultural pollution for the efficiency and design of water quality markets. We are especially interested in the question of whether simple market designs under the assumption of contemporaneous substitution do reasonably well compared to complex dynamic markets that facilitate trading across time and space. For this analysis, we compare pollution control outcomes between simple and complex dynamic allocation rules in the context of nutrient pollution control for the Chesapeake Bay, and identify conditions under which simple rules may perform well. We begin with conceptual model of efficient pollution control allocations with lags to develop key concepts and the analytical framework used for this study. We then apply the models to nitrogen pollution control in the Chesapeake Bay.

Efficient Pollution Control with Lags

Market-based approaches to pollution control entail exchanges of property rights directly or indirectly in polluting emissions. Fundamental tasks in the design of markets are (a) defining the tradable commodity to which the property rights apply, (b) specifying trading rules governing exchanges of the commodity, and (c) specifying aggregation rules that limit the aggregate supply of the commodity so that market allocations of polluting emissions do not violate the overarching environmental goal served by the trading program (Horan and Shortle 2011). Lags have implications for each of these tasks.

In a simple "textbook" model of pollution permit trading without lags, the amount of pollution reaching a "receptor location" at which pollution loads are measured during the year (e.g. the mouth of the Susquehanna River on the Chesapeake Bay) depends entirely on emissions during the year. In this case, emissions within the year substitute for each source at the rate indicated by the appropriate trading ratio. With lags, the pollution reaching the receptor during the year may have been emitted during that year or in prior years. To put it another way, pollutants emitted within a year may reach the receptor during the year or in subsequent years. If the goal is to limit the amount reaching the receptor in a given year to a specific amount, the definition of the commodity, trading rules, and the aggregation rule must take account of lagged pollution delivery.

To introduce lags we begin by indexing emissions not only by source but also by the date (e.g. year) released: emissions from source i at time t are e_{it}. Some fraction of a source's emissions reaches and contributes to ambient pollution in a regulated waterbody at some future time. For simplicity, we assume the delivered fraction is not distributed over a period of time but arrives at a specific future date. The amount delivered at time $t + l$ from source i is given by $\theta_{il} e_{it}$ $(0 \leq \theta_{il} \leq 1)$, where l is the lag time between emissions and delivery. Different sources may have the same or different lag lengths, but the lag for any source is fixed.

Let $t = 0$ be the date at which "new" management strategies are implemented. At that date there is a sequence of legacy loads resulting from prior emissions, plus pollution loads from unmanaged natural sources. The legacy load plus the natural load at time t is referred to as the base load B_t. The legacy component of the base load will generally decay over time as pollution from prior emissions works through the watershed. Thus, over time B_t converges to the natural background load (including lags in that process). The pollution reaching the receiving water at any time $t \geq 0$ is the legacy load, plus the load generated since the implementation of new management strategies at time $t = 0$:

(1) $L_t = B_t + \sum_{k=0}^{t} \sum_{i=1}^{m_k} \theta_{il} e_{ik}$

m_k is the number of sources discharging at time k ($0 \leq k \leq t$). For simplicity, this number is taken to be fixed and the same in all time periods (equal to m). The second term on the RHS of (1) is the manageable component of pollution beginning in period $t = 0$. Emissions at any time $k < t$ with lag greater than t - k will appear in delivered loads subsequent to time t.[1]

Environmental policy makers impose a limit L_{max}^t on the pollution reaching the regulated receiving waters for each time t in the planning horizon. Thus, in each period t in the planning horizon, they require

(2) $B_t + \sum_{k=0}^{t} \sum_{i=1}^{m} \theta_{il} e_{ik} \leq L_{max}^t$

A minimal requirement for the goal to be feasible is that $B_t \leq L_{max}^t$. The difference between the goal and base load is the allowable level of the managed load.

[1] For example, assume two firms $i = 1,2$ with lags $l = 1,2$ respectively. The pollution reaching the receiving waters in the first three time periods, $t = 0,1,2$, are:

$$L_0 = B_0$$

$$L_1 = B_1 + \theta_{1,1} e_{1,0}$$

$$L_2 = B_2 + \theta_{1,1} e_{1,1} + \theta_{2,2} e_{2,0}$$

With lags, the costs of achieving the pollution target for any time period will be realized over multiple time periods. In consequence, efficiency in pollution control requires allocating emissions reductions over time and space to minimize the present value of the costs of achieving the goal. Denote the abatement cost for source i at time t as $c_i(e_{it})$. This function is taken to be continuous, convex, and decreasing in emissions (i.e. more pollution, less pollution abatement, less cost). The present value of the societal costs from pollution control from all sources beginning at $t = 0$ over the time horizon T is then given by

(3) $\sum_{t=0}^{T} \sum_{i=1}^{m} c_i(e_{it})\beta_t$

$\beta_t = (1+r)^{-t}$ and r is the discount rate. The Lagrange equation for the optimization problem is

(4) $J = \sum_{t=0}^{T} \sum_{i=1}^{m} c_i(e_{it})\beta_t + \sum_{t=0}^{T} \rho_t \left[L_{max}^{t} - B_t - \sum_{k=0}^{t} \sum_{i=1}^{m} \theta_{il} e_{ik}\right]$

ρ_t is the Lagrange multiplier for the environmental constraint at time t. Assuming an interior solution, the first-order necessary conditions for the optimization problem are:

(5) $\frac{\partial J}{\partial e_{it}} = \frac{\partial c_i}{\partial e_{it}} \beta_t - \rho_{t+l}\theta_{il} = 0$

For any source with lag l, emissions at time t substitute with emissions that reach the regulated water from other sources in period $t+l$. Those emissions may have occurred in time periods prior to t for sources with lag lengths greater than l, or in time periods subsequent to t for sources with lag lengths less than l. Accordingly, for any source j discharging at time $t + k \leq t + l$ $(-t \leq k \leq l)$, whose emissions also arrive at time $t+l$, equation (5) implies that an optimal allocation requires

(6) $\frac{\partial c_i}{\partial e_{it}} \frac{1}{\theta_{il}} = \frac{\partial c_j}{\partial e_{j(t+k)}} \frac{\beta_k}{\theta_{j(l-k)}}$ for any $i, j, k, and\ l, t + k \leq t + l, -t \leq k \leq l$

The left-hand side of equation (6) is the marginal cost of reduction in source i's emissions at time t divided by the proportion of the emissions that reaches the receiving waters l periods later.

With this division, the term can be interpreted as the marginal cost of reducing its contribution to ambient pollution at time $t+l$. The marginal benefit of a decrease in ambient pollution by source i at time $t+l$ due to a reduction in emissions at time t is the forgone cost of reducing pollution from other sources that contribute to ambient pollution at time $t+l$ to satisfy the ambient pollution constraint at that time. This forgone cost is given by the right-hand side of (6) for source j discharging pollution at time $t+k$ that reaches the receiving water at time $t+l$. Condition (6) indicates that in optimality, the marginal cost of abatement for source i at time t is equal to the discounted marginal abatement cost for source j at time $t+k$.[2]

Equation (6) has important implications for pollution management. Marginal abatement cost inequalities are often used as indicators of inefficiency in pollution control. In simple static models, inequalities in marginal ambient abatement costs imply that cost savings can be realized by reallocating abatement from the high marginal cost sources to lower marginal cost sources. Observations that marginal agricultural nonpoint abatement costs are often lower than those for point sources are, for example, commonly used to indicate inefficiencies in abatement allocations between point and nonpoint sources and potential gains from pollution trading. However, equation (6) implies that when there are lags, sources subject to lags should have lower relative marginal costs in an efficient allocation than sources not subject to lags. Thus, differences in marginal costs do not necessarily imply inefficiencies and gains from trade. The equation also indicates that the marginal cost differences in an efficient solution will vary with the lag length and the discount rate. When comparing a source not subject to lags to a source with lags, the

[2] Again assume two firms $i = 1,2$ with lags $l = 1,2$ respectively, condition (8) implies that to reduce emissions in time period 2, either firm 2 reduces emissions in time period 0 or firm 1 reduces emissions in period 1. At time period 0, it is cheaper to reduce the discounted emissions from firm 1 in period 1 than the emissions from firm 2 in period 0. Therefore, the firms with the shorter lags will reduce more emissions and thus have a higher MAC.

difference in marginal costs between them will be greater, the longer the lag length or the higher the discount rate.

A natural form of a market trading pollutants with lags would be one in which polluters could buy or sell forward. A farmer installing BMPs today that would not have impacts until sometime in the future could contract to supply credits to a source seeking credits at the future date. In theory, a perfectly competitive futures market could achieve a dynamically efficient allocation. However, forward markets designed for long periods of time can be expensive to set up and operate, as well as complex for participants (e.g., Carlton 1984). While we do not attempt to estimate these set-up and operating costs, it would only make sense to incur them if the dynamically efficient allocation achieved significantly better results than a static market design in terms of allocating costs of pollution control across sources and time. Below we ask whether the results of a static market design might be reasonably acceptable compared to the results of the dynamically efficient allocation. In a nutshell, do we need to account for lags in market design? We explore this question using a model of nitrogen pollution control from point source and agricultural nonpoint source pollution in the Chesapeake Bay as a case study.

Lags and Efficient Load Allocations for the Chesapeake Bay

Reducing nutrient and sediment pollution of the Chesapeake Bay has been a major policy goal of the US Environmental Protection Agency and states in the Bay's watershed since the early 1980s. Limited progress led the EPA to issue a total maximum daily load (TMDL) for the Bay in December 2010. The TMDL, which is discussed extensively elsewhere (see Shortle et al. 2014) specifies load (applicable to nonpoint sources) and waste load (applicable to point sources) limits to be achieved by 2025 for nitrogen, phosphorus, and sediment across all polluting sectors (agriculture, wastewater treatment plants, regulated and unregulated urban and suburban runoff, septic, forest, and air deposition). Regulated jurisdictions are the portions of the states included in the watershed (Delaware, Maryland, New York, Pennsylvania, Virginia, West Virginia), and the District of Columbia. The District of Columbia is not included in our analysis as it does not have any agricultural lands.

The TMDL load and waste load limits take into account the relative effectiveness of pollution load reductions on water quality conditions in the Bay, the relative amounts of pollution from alternative sources and locations, and other factors. They do not take into account the relative cost-effectiveness of alternative sources and locations explicitly. Further, Watershed Implementation Plans (WIPs) which describe the practices the states plan to implement to meet their TMDL limits were not developed to be cost-minimizing (Shortle et al. 2013). The inattention to cost-effectiveness suggests there may be considerable potential for cost-savings from implementation of strategies, including Bay-wide trading, to enhance cost-effectiveness (Kaufman et al. 2014). Further, the TMDL allocations are based on the EPA's Chesapeake Bay Watershed Model, which does not account for lags in delivery (US EPA 2010). The model simulations predict pollution load reductions in the Bay as steady-state, long-run responses to pollution control activities in the Bay watershed. In other words, when a best management practice (BMP) is implemented, the model credits that jurisdiction with the full nutrient reduction benefit associated with that practice (STAC 2013).

While lags are now of increasing interest to water quality managers, there is no comprehensive understanding of lag lengths across the Chesapeake Bay Watershed, including how they are affected by hydrogeomorphic conditions and BMP types and locations. However, their potential significance is suggested by the approximate lag length ranges by pollutant and transport type for the Bay shown in Table 1.

For this research we consider two allocations of point source and agricultural nonpoint nitrogen pollution control to meet load limits in the Bay watershed. [3] One allocation is a dynamically optimal allocation (DOA) which is computed by solving a dynamic optimization model with lags for the Bay watershed consistent with the conceptual model we presented above. The second is a static optimal allocation (SOA) that is computed by solving a conventional static cost minimization model with limits on steady loads rather than actual annual loads. Comparisons of the two allocations will provide insight about the implications of lags for efficient allocations across source types and locations, and about the relative merits of simple static versus complex dynamic markets.

Dynamic Optimal Allocation (DOA)

Our dynamic and static allocation models make significant use of parameters and relationships in the Phase 5.3.2 of the EPA's Chesapeake Bay Watershed Model (CBWM). Consistent with the CBWM, we subdivide the Bay's watershed into geographic management units for modeling agricultural abatement. Phase 5.3.2 of the CBWM subdivides that watershed into approximately 2500 land-river segments(USEPA 2010).[4] Utilization of this large number of

[3] We apply the models to phosphorus pollution control as well, with similar results as the nitrogen case. The phosphorus model results are available upon request.

[4] CBWM is a complex hydrological model that simulates the movement of nutrients over and through the land utilizing monitoring data from rivers and streams, hydrogeomorphology, climate, and other land-based characteristics. The CBWM divides the Bay watershed into land-river segments, which are used for modeling the relationship between land uses and pollution loads. The land-river segments result from the intersection of the CBWM's land segmentation and river segmentation.

segments would result in significant computational difficulties for multi-period nonlinear dynamic optimization. We therefore simplify by aggregating the land-river segments into a smaller number (10) of "bins" differentiated by CBWM delivery factors that indicate the proportion of the nitrogen pollution load that originates in a segment that actually reaches the Bay (analogous to the parameter θ_{il} in the conceptual model presented above). Other things equal, the marginal costs of pollution abatement delivered to the Bay vary inversely with the delivery factors, i.e., the lower the delivery factor for an upstream source, the higher the marginal cost of reducing pollution delivered to the Bay from that source. Spatial variations in delivery factors are a key determinant of spatial variations in marginal abatement costs of pollution loads delivered to the Bay and therefore a key determinant of efficient allocations of abatement across the watershed (Shortle et al. 2013).

Agricultural emissions from the bins are modeled as the product of agricultural nonpoint source edge-of-segment (EOS) emissions per acre within the bin multiplied by the acres within the bin. [5] EOS emissions are the proportion of emissions from agricultural sources that reach receiving waters in their associated land-river segment. Lags are modeled within bins by assigning lag lengths to proportions of the land areas within the bins. As noted previously, there is little comprehensive understanding of lag lengths across the Chesapeake Bay. We therefore utilize scenarios to explore the implications of lag lengths and their spatial distribution.

Nonpoint source emissions per acre at time t from bin b with lag length l are denoted r_{tbl}.

[5] The model excludes direct discharges from regulated Concentrated Animal Feeding Operations (CAFOs), which account for a small percentage of nitrogen and phosphorus loads delivered to the Bay. The primary source of pollutants from CAFOs is manure that is spread on cropland or pastureland (Chesapeake Bay Program 2010). These loads are considered nonpoint pollution and are included in the load reduction data utilized for this analysis. The model represents industrial and municipal discharges using a single point source.

The fraction of these emissions subsequently delivered to the Bay is given by the delivery factor θ_b applicable to all acres in bin. Bin acres, given by A_b, are distributed across possible lag lengths, with the proportion in source b with lag length *l* given by p_{bl}. Thus, the deliverable emissions at time *t* from bin b with lag length *l* are given by $\theta_b r_{tbl} A_b p_{bl}$. Land areas for the 10 bins, differentiated by delivery factors, are calculated using 2011 baseline land use data from the CBWM. We use 2011 as beginning of our planning period since it is the first year in which the Bay TMDL was in force.

Point source emissions are modeled as a single source. The pounds of point-source emissions delivered at time t are denoted e_t. Since point sources are not lagged, emissions reductions implemented at time t are realized immediately. Point source data are in terms of delivered reductions, so a delivery factor is not necessary for the aggregate point source.

As above, let $t = 0$ be the date (the year 2010) from which we implement "new" management strategies. At that date there is a sequence of legacy loads resulting from prior emissions, plus pollution loads from unmanaged natural sources, B_t. Pollution reaching the receiving water at any time $t \geq 0$ is the base load, plus the load generated since the implementation of new management strategies*:*

(7) $L_t = B_t + \sum_{b=1}^{10} \sum_{k=0}^{t} \sum_{i=0}^{l} p_{bi} \theta_b r_{kbi} A_b + e_t$

The second and third terms on the right-hand side of equation (7) are the manageable components of pollution, beginning in period $t = 0$. For simplicity, we ignore the base load and focus only on the manageable components.

We impose a limit L_t^{max} on the manageable flow of pollution reaching the Bay in each time period. For each time period t, these limits require

(8) $\sum_{b=1}^{10} \sum_{k=0}^{t} \sum_{i=0}^{l} p_{bi} \theta_b r_{kbi} A_b + e_t \leq L_t^{max}$

A dynamically efficient allocation minimizes the present value of the costs of achieving the load limits. The abatement cost for source b at time t with lag length l is given by $c_b(r_{tbl})$. The abatement cost for the point source at time t is given by $c(e_t)$. The present value of the societal costs from pollution control from all sources beginning at $t = 0$ over the planning horizon T is then given by

(9) $\sum_{t=0}^{T}\sum_{b=1}^{10}\sum_{i=0}^{l} p_{bi}A_b c_b(r_{tbi})\beta_t + \sum_{t=0}^{T} c(e_t)\beta_t$

$\beta_t = (1 + r)^{-t}$ and r is the discount rate. The plan obtained by minimizing equation (9) subject to the *T+1* load constraints given by (8) is referred to as the dynamically optimal allocation (DOA). The DOA will consist of emissions paths of length *T+1* for the acres of each assigned lag length in each of the ten bins plus a point source emissions path of *T+1* periods. Equation (6) implies that in the DOA, the discounted marginal cost of abatement for delivered emissions at time t is equal across the point source as well as all NPS sources b and lag lengths *l*, normalized by bin acreage and delivery factor. In other words, for any two NPS sources b_1 and b_2, and for the point source, optimality requires that the following hold in all time periods:

(10) $$\frac{c'_{b_1}(r_{tb_1l})\beta_t}{\theta_{b_1}A_{b_1}} = \frac{c'_{b_2}(r_{tb_2l})\beta_t}{\theta_{b_2}A_{b_2}} = c'(e_t)$$

This condition is interpreted as was equation (6) above.

We assume a constant-elasticity cost function for point source emission reductions. Given limited data on agricultural abatement costs within bins, we assume that those costs are homogeneous across agricultural lands within, but not across bins, and specify the costs per acre as constant elasticity cost functions. Let e_t^b be the point source 2011 baseline emissions and r_{tbl}^b be the 2011 baseline nonpoint source (NPS) emissions per acre. The NPS per acre and PS abatement cost functions are respectively

(11) $$c_b(r_{tbl}) = \frac{\alpha_b}{1+\gamma_b}\left(r_{tbl}^b - r_{tbl}\right)^{1+\gamma_b}$$

and

(12) $$c\ (e_t) = \frac{\alpha}{1+y}\left(e_t^b - e_t\right)^{1+y}$$

where $\alpha_b > 0, \gamma_b > 0, \alpha > 0,$ and $y > 0$. These functions are continuous, convex, and decreasing in emissions (i.e. more pollution, less pollution abatement, less cost). The marginal costs of emissions are negative because more emissions mean fewer reductions: $c_b'(r_{tbl}) = -\alpha_b\left(r_{tbl}^b - r_{tbl}\right)^{\gamma_b} \leq 0$ and $c'(e_t) = -\alpha\left(e_t^b - e_t\right)^{y} \leq 0$. NPS and PS baseline loadings for the 10 bins, differentiated by delivery factors, are calculated using 2011 baseline loading data from the CBWM.

We utilize a discount rate of 7% per Office of Management and Budget requirements (OMB Circular No. A-94). The optimal terminal time T should in principle be infinite (i.e. to manage pollution optimally for the rest of time). However, to simplify computation, and without harm to our results, we use finite terminal times based on the longest lags in the model. The cost functions we use are independent of time. In consequence, if the load limits are fixed over time, or become fixed within an amount of time less than the longest lag, then the optimal allocation will converge to a steady-state by a time T equal to the longest lag length. The optimal allocations in the steady state would simply repeat the optimal allocation at time T. We vary T with alternative scenarios for lag lengths. We refer to the passage of time between 0 and the achievement of the steady state at time T as the *implementation phase*, during which emissions are reduced from baseline levels and move toward the steady state levels.

Static Optimal Allocation (SOA)

We define a static optimal allocation (SOA) as one with constant point and nonpoint emissions that minimizes the periodic cost of achieving the load limits in a steady state. Formally for the *T* period planning horizon, the SOA is obtained as a sequence of optimizations with the following structure in any period t:

(13) minimize$\sum_{b=1}^{10} \sum_{i=0}^{l} p_{bl} A_b c_b(r_{bti}) + c(e_t)$ subject to

(14) $\sum_{b=1}^{10} \sum_{i=0}^{l} \theta_b r_{bti} p_{bl} A_b + e_t \leq L_t^{max}$

The optimality condition for the SOA is the same as equation (12) but with $\beta_t = 1$. SOA cost minimization for any annual limit would allocate emissions across all NPS sources b and the point source to equalize the marginal costs of abating delivered pollution subject to the resulting allocation achieving the annual limit in a steady state. Because the SOA does not account for the time path of delivered load reductions but only requires that the limits be satisfied in a steady state, the path of delivered emissions for any annual plan would be in excess of the corresponding annual load limits until the lags are resolved. If the limit is invariant over time, then the SOA is obtained with a single optimization.

Model Scenarios and Parameters

Comparison of pollution control outcomes between simple and complex dynamic allocation rules in the context of nutrient pollution control requires (1) defining the bins by delivery factors within each bin, (2) calculating marginal abatement cost (MACs) curves for each bin, (3) defining the lag lengths for each bin, (4) setting the load limits (L_t^{max}) in each time period, and (5) making assumptions about adjustment costs.

Delivery Factors

The land area of and the baseline loadings for the 10 bins differentiated by delivery factors are specified utilizing 2011 baseline land use and loading data from the CBWM. Bin 1 includes all acreage in land-river segments with delivery factors $0 \leq df \leq 0.1$, Bin 2 includes all acreage in land-river segments with delivery factors $0.1 < df \leq 0.2$, Bin 3 includes all acreage in land-river segments with delivery factors $0.2 < df \leq 0.3$, and so on. Each bin is assigned a single delivery factor at the midpoint of the bin range. Therefore, Bin 1 acreage is assigned a delivery factor of 0.05, Bin 2 acreage a delivery factor of 0.15, Bin 3 a delivery factor of 0.25, and so on. The 2011 baseline data are based on CBWM runs including nutrient reduction benefits from all BMPs credited within the model as of or before the baseline year. Delivery factors are the proportion of edge of segment (EOS) emissions that reach the Bay. As noted above, EOS emissions are the proportion of emissions from agricultural sources that reach receiving waters in their associated land-river segment. The CBWM includes delivery factors for every land-river segment and pollutant.

Delivery factors differ by pollutant and vary with distance to the Bay and various hydrogeomorphic and topographic characteristics of each land-river segment. The distribution of land-river segments into bins for nitrogen is shown in Figure 1. In general, the greater the distance from the Bay, the lower the delivery factor. However, this is not always the case. For example, delivery factors for nitrogen are greater for land-river segments along the Susquehanna River in Pennsylvania than for some land-river segments in Maryland.

Marginal Abatement Costs

Once agricultural acreage was subdivided into bins, we utilized nutrient reduction data from Shortle et al. (2013) to facilitate calculation of marginal abatement cost (MAC) curves for each bin. In Shortle et al. (2013), cost-effective BMP portfolios were calculated by estimating nutrient reductions associated with each BMP within each land-river segment. Costs for each BMP/land-river segment combination were calculated along with their associated MACs. Each BMP/land-river segment combination was sorted from low to high based on MAC by jurisdiction, resulting in MAC curves for each Bay jurisdiction. MACs serve as a guidepost for efficient implementation. Ideally, the lowest MAC practices are installed first, and so on until load allocations are met. Using this data, we allocated nitrogen reductions at the BMP/land-river segment scale to their respective bins. This enabled the construction of MAC curves for each bin. Units for reductions were EOS pounds per acre since delivery factors were included as a parameter in the model. We also normalized reductions by bin acreage. Thus, our MAC units were dollars per EOS pound reduced per acre.

To facilitate optimization, cost function parameters were estimated to provide smooth approximations of the empirical nitrogen MAC curves for each bin based on the Shortle et al. (2013) results. Each bin includes many prohibitively expensive practices that can lead to cost parameter estimates that do not approximate MAC curves well over most of the range of costs. In the interest of improving the fit of our estimated functions, we capped maximum reductions within each bin. We did so by either (1) capping maximum reductions at 90% of total reductions or (2) capping maximum reductions at the maximum MAC for point sources. Maximum reductions were capped at whichever method resulted in the lowest maximum MAC for that bin, and cost parameters were calculated without the prohibitively expensive practices above that maximum. The 90% reduction threshold resulted in the lowest maximum MAC for all bins except bins 1 and 2 for nitrogen. These bins have extremely low delivery factors, thus deflating their MAC equivalent to the maximum MAC for point sources. This reduced the MAC associated with the least efficient practice included in the analysis for each bin substantially and improved the fit of our cost functions. The functions consistently underestimate actual costs reflected by the data at low MAC levels and overestimated actual costs at higher MAC levels, but this error was deemed acceptable as it is consistent across all bins. Furthermore, the value of comparing a steady-state solution to a lagged solution remains despite this error. Point source cost functions were estimated in an identical fashion using data from Ribaudo (2013), with nitrogen reductions capped at approximately 90% to eliminate prohibitively costly reductions and to improve the fit of the cost functions.

Lag Lengths

As noted previously, there are little comprehensive data on lag lengths across the Chesapeake Bay. To explore the implications of lag lengths, we consider scenarios in which lags for nitrogen range from 0 to 19 years, and other scenarios in which lags for nitrogen range from 0 to 39 years. This results in models that are 20 and 40 periods long. The terminal time for the dynamic optimization was chosen such that all lags were resolved in the final period.

The lag lengths and acreages available for reductions in delivered pollution loads differ in any given period. For example, in period 0, only delivered load reductions from the point source and NPS acreage with no lag are possible. In period 1, additional delivered load reductions are possible from NPS practices implemented in period 0 with lag length 1. In period 2, additional reductions are possible from NPS practices implemented in period 0 with lag length 2 and from period 1 with lag length 1. For the 20 period models, effort undertaken in period 0 for acres with lag length 19 would not affect water quality until time period 19. For 40 period models, effort undertaken in period 0 for acres with lag length 39 would not affect water quality until time period 39.

Lag Length–Delivery Factor Correlations

A fundamental question is how lag lengths may interact with delivery factors to determine efficient allocations. Since there is no spatially-specific data source on lag lengths in the Chesapeake Bay Watershed, we consider scenarios with alternative interactions. Our scenarios model both a strong correlation and a weak correlation between lag length and delivery factors. The strong correlation scenario assumes that lower lag lengths are associated with agricultural acreage that has a higher delivery factor, and vice versa. We assigned a limited number of lag lengths to each bin, with acreage subdivided equally among all lag lengths in the bin. Therefore, if Bin 1 includes lag lengths of 17, 18, and 19 years, each lag length is allocated 1/3 of agricultural acres in Bin 1. The lag lengths assigned to each bin for the 20 and 40 period models in the strong correlation scenario are shown in Table 2. The weak correlation scenario assumes no correlation between lag lengths and delivery factors. Every lag length is included in each of the ten bins, with acreage once again subdivided equally among lag lengths. Therefore, 1/20 of agricultural acreage in each bin is assigned to each lag length for the 20 period model while 1/40 of agricultural acreage in each bin is assigned to each lag length for the 40 period model.

Constant versus Decreasing Load Limits

Static models of pollution trading typically assume the pollution load limit to be constant over time. The presence of lags leads to interesting questions about the dynamic structure of the load limit. There are many possibilities. Lags limit feasible reductions in pollution in the short term but not in the long term. Lags can also increase short term costs by limiting the contemporaneous substitution of lower cost sources for relatively higher cost point sources to achieve target load. Both of these facts suggest time varying load limits that become increasingly restrictive over time. Concern for cost would suggest beginning with modest load limits in the short term but increasing stringency over time. Concern for water quality gains would suggest stronger limits in the short run and again increasing stringency over time as the feasible reductions increase. The difference in costs and water quality gains between these two cases would depend on how tight the constraint is initially and how fast it increases over time.

We consider two types of limits. One is a conventional constant limit that is invariant over time. The other is increasingly restrictive. We selected the constant limit at a level that could be achieved without agricultural load reductions when point source reductions are maximized. This level was chosen to ensure that the model was feasible in early periods. This load limit resulted in total delivered reductions of approximately 24.5 percent across all sources. The decreasing load limit scenario begins at this level but becomes tighter by 0.5 percent each period through period 19 in both the 20 and 40 period models. Therefore, the load limit in period 0 reflected total reductions of ~24.5 percent, ~25 percent in period 1, ~25.5 percent in period 2, and so on until a maximum total reduction of ~34 percent was reached in period 19. These reductions more closely mirror those required by the TMDL. For example, the TMDP requires total agricultural reductions relative to the 2011 baseline of approximately 34 percent for nitrogen. The load limit remained constant at total reductions of ~34 percent in periods 20 through 39 in the 40 period model. This facilitated comparison between the 20 and 40 time period models since required reductions increased at the same rate in periods 0 through 19 for both models.

Static Model Scenarios

SOA emissions and costs are independent of lag lengths and correlations, although SOA delivered loads do vary with lag lengths and correlations. Table 3 summarizes the combination of criteria for each model run, along with the abbreviations, in parenthesis, that are used to refer to them. Both the SOA and DOA models were coded in GAMS and solved using the Path NLP Solver.

Results

The dynamic optimal allocations (DOA) and the static optimal allocations (SOA) are compared in terms of (1) the total present value of implementation costs, (2) time paths of nitrogen reductions, (3) the distribution of reductions between nonpoint source and point source pollution, and (4) time path of undiscounted marginal abatement costs.

The Total Present Value of Implementation Costs

Table 4 presents the total present value costs for the implementation phases of the static optimal (SOA) and dynamic optimal (DOA) allocations for nitrogen. These are the costs of achieving steady state outcomes (i.e., the implementation phase) for both the SOA and DOA, with the number of periods equal to the longest lag length plus one. As noted previously, pollution control costs will continue to be realized beyond the final period of this implementation phase.

First, the present value of costs of the implementation phase of the 40Y models are necessarily greater than those for the 20Y models due simply to differences in the planning horizons considered. The present value of the costs of achieving load reduction targets during the first 20 periods of the 40 period scenarios are included for comparison. The costs of achieving load reductions in the first 20 periods of the 40Y models are 24% (WC) to 33% (SC) greater than the costs of 20Y models. This difference stems from the fact that longer lags require that emissions reductions from relatively high cost point sources must be relied on for longer periods of time to achieve targeted water quality reductions.

Second, the implementation phases of DOA are substantially more costly (78% to 134% in the 20Y model and 118% to 171% in the 40Y model) than the implementation phase of the SOA. The sign of the difference is expected because the DOA are required to satisfy the load limit in each period whereas the SOA satisfy the load limit only in the steady state. The magnitude of the difference is, however, noteworthy, implying significant potential cost savings from reduced complexity. Of course, those costs savings must be weighed against the environmental losses from delayed achievement of the environmental goal.

Third, the costs of the DOA are significantly (24% in both the 20Y and 40Y models) less when there is a strong correlation (SC) between lag lengths and delivery ratios. The explanation for this result is that SC concentrates low lag lengths in bins where EOS load reductions have the largest impact on loads delivered to the Bay compared to weak correlation (WC). One way to understand this result focuses on marginal costs. Other things equal, a SC implies that bins with shorter (longer) lag lengths are also bins with relatively lower (higher) marginal abatement costs for delivered pollution. Bins with lower marginal abatement costs for delivered pollution are preferred for cost-minimization without lags. Adding lags strengthens this preference when SC exists. But the preference for high delivery ratios is offset to some degree in the WC case. Essentially, the longer lags drive up the discounted marginal costs of high delivery ratio bins relative to the SC scenario. A second way to understand this result is that WC between delivery ratios and lag lengths results in fewer available reductions in high delivery ratio bins in early periods. Having a lower delivery ratio drives up marginal abatement costs for delivered loads, resulting in higher costs for models with WC between delivery ratios and lag lengths.

Fourth, costs increase across all models with a decreasing target load compared to the constant target due to the higher levels of required reductions after the first period. DOA costs are again significantly higher than those of the SOA, with WC costs exceeding those for SC scenarios. For models with decreasing target loads, DOA costs are 88% and 141% higher than SOA costs for SC and WC scenarios, respectively. DOA SC costs are 28% higher than WC costs.

Time Paths of Nitrogen Reductions

The time paths of nitrogen reductions are indicated in Figure 2 for the 20Y model and Figure 3 for the 40Y models. In the DOA, the overall level of nitrogen reductions jumps up to the regulatory target at the beginning of the planning period and remains at that level throughout. Since it meets the regulatory target in all periods, the DOA is not included in Figures 2 or 3.

In the SOA, pollution control effort is implemented in the first period, but these actions are not adequate to achieve the target reduction until achievement of the steady state at the end of the implementation phase. The PS delivered load across the implementation phase is constant at the SOA steady state level. The NPS load declines over time as lags are resolved, reaching the steady state NPS allocation at the end of the implementation phase. The percentage of required reductions attained in the first period is approximately 55 percent for both the SC and WC models, though the time path of reductions differs significantly between the two. Reductions in the SC model increase more rapidly in earlier periods since shorter lag lengths are concentrated in bins with higher delivery ratios. In contrast, the WC model increases in a more linear manner, reflective of the even distribution of lag lengths across all bins. Midway through the implementation phase (period 9 for the 20Y model and period 19 for the 40Y model), the SC model has met approximately 91 percent of required load reductions while the WC model has attained approximately 76 percent of required reductions.

The time path of reductions in the SOA varies little with the decreasing versus the constant target levels. From Figure 4, it is apparent that the correlation between lag lengths and delivery ratios drives differences in the time path of SOA reductions, though the percentage of required reductions attained is slightly less for decreasing load scenarios. SOA reductions reach 88% and 74% of the decreasing load target after period 19 for SC and WC scenarios, respectively. These reductions are just short of the 91% and 76% of reductions reached for the constant target scenarios (40Y model).

The Distribution of Reductions between Nonpoint Source and Point Source Pollution

Figures 5 and 6 show the PS/NPS emissions split for SC and WC 20Y models, respectively. Figures 7 and 8 display the PS/NPS split for SC and WC 40Y models, respectively. In both the 20Y and 40Y models, the SOA allocates 40 percent of reductions to PS (and 60% percent of reductions to the NPS), in the steady state. Practices consistent with this allocation are implemented at the beginning of the implementation phase, but the split is not realized until the steady state is achieved. Figures 5, 6, 7, and 8 reflect this, as NPS reductions increase gradually to the steady state level of 60 percent by the final period. PS reductions are constant at the steady state SOA level. The DOA relies heavily on the PS reductions in early periods, with the percentage of reductions allocated to NPS gradually increasing through time. This pattern is required to achieve the constant load reduction target as very few NPS reductions are available in early periods due to the impacts of lags on NPS.

The steady state NPS and PS reduction percentages for the DOA approximate those of the SOA, though with somewhat greater reliance on PS (44% DOA 20Y and 48% DOA 40Y versus 40% SOA) than NPS in the SC steady state. WC steady state SOA reduction percentages are identical to those of the SC model, though NPS reductions increase in a more linear fashion. The WC exhibits even greater DOA reliance on PS reductions, with steady state reductions of 48% in the 20Y model and 54% in the 40Y model. This greater reliance on the PS than NPS even in the steady state is the result of the differences in the optimality conditions for the SOA and DOA models, which, as noted previously in the development of the conceptual model, imply higher marginal abatement costs for delivered pollution from lagged NPS than for unlagged PS.

Consistent with our discussion of the effects of correlation between lags and delivery ratios on costs, the PS/NPS split is tilted more towards NPS than PS in the model with SC than in the model with WC: weak correlation between delivery ratios and lag lengths results in fewer available reductions in high delivery ratio bins in early periods. NPS reduction percentages increase more rapidly in earlier periods in the SC model but more slowly in later periods, while NPS percentage reduction increases under WC are more or less linear over time. The NPS percentage reductions in the final period are always higher for models with SC than those with WC. This is due to the uneven distribution of short lag lengths to bins with high delivery ratios in the SC model and the even distribution of lag lengths in the WC model. SC concentrates shorter lag length reductions in bins where the marginal cost of reductions tends to be lower due to higher delivery ratios.

The NPS/PS split is little influenced by the type of target. The DOA relies heavily on PS reductions in early periods, with the percentage of reductions allocated to NPS gradually increasing through time. The steady state NPS and PS reduction percentages for the DOA approximate those of the SOA, though with somewhat greater reliance on PS, particularly for WC scenarios. In the final period for 40Y models, the SOA allocates 55 and 45 percent of reductions to NPS and PS, respectively. The DOA initially allocates 16% to NPS for the SC scenario and 13% to NPS for the WC scenario. These allocations increase to 48 and 41 percent, respectively, in the final period.

Time Path of Undiscounted Marginal Abatement Costs

The time paths of undiscounted marginal abatement costs are presented in Figures 9 and 10, for the 20Y and 40Y models, respectively, and also Table 5. The SOA costs are constant across time at about $5.50/lb. for both the 20Y and 40Y models. The DOA costs start at about $15/lb. (20Y) and $16.30/lb. (40Y) in the SC model and $16.60/lb. (20Y) and $17.30/lb. (40Y) in WC model and decline over time to steady state values of about $6/lb. (20Y) and $7/lb. (40Y) in the SC model and $7/lb. (20Y) and $8.40/lb. (40Y) in the WC model.

The larger initial marginal cost reflects the limited capacity of NPS to provide delivered loads reductions, thus requiring large reductions from the PS that increase PS marginal costs. Increased availability of NPS load reductions over time allows increases in PS effluents and accompanying decreases in marginal costs. Undiscounted marginal costs for the SC model are lower than those for the WC model in every time period and decrease more rapidly in initial periods. This is consistent with our discussion of the implications of concentrating shorter lag lengths in high delivery ratio bins in the SC case.

Figure 11 presents undiscounted marginal costs for nitrogen with decreasing targets. Undiscounted marginal costs for the SOA gradually increase as the environmental target decreases through period 19, from about $5.50/lb. to about $10.50/lb. They remain constant after period 19 as required reductions remain constant thereafter at about 34%. In the DOA, decreasing load targets imply higher marginal costs while greater availability of NPS reductions through time implies lower marginal costs. In the WC scenario, marginal costs increase rapidly through period 19, primarily due to the fact that PS reductions are maximized. As a result, all new reductions must come from NPS sources. In contrast, the increased availability of NPS reductions initially yields lower costs in the SC model, with costs increasing in later periods after short lag reductions are exhausted.

Summary and Conclusions

This research is motivated by a need to understand the implications of lags in agricultural pollution for the efficiency and design of water quality markets. Of special interest is whether simple market designs under the assumption of contemporaneous substitution (i.e., no lags) do reasonably well compared to complex dynamic markets that facilitate trading across time and space to address lags.

This paper presents conceptual models of efficient pollution control allocations with and without lags to develop key concepts and the analytical framework used for this study. It also presents static and dynamic simulation models developed to compute static optimal allocations and dynamic optimal allocations for nitrogen pollution in the Chesapeake Bay. Given uncertainty about lag lengths and their spatial distributions, scenarios are used to gain insight about the implications of these factors. These scenarios vary the longest lag lengths in the model and the spatial correlation between lag lengths and pollution delivery coefficients. Scenarios are also used to explore implications of the time structure of pollution load targets (constant or decreasing).

The simulations indicate that length lags, the time structure of the load target, and the spatial structure of lags can have significant impacts on the costs of pollution control and the distribution of load reductions between point and nonpoint sources over time. These factors interact, so the specific implications depend on the specific relationships.

Simple static allocation rules that ignore lags allocate pollution load reductions to sources based on their marginal costs of reducing delivered pollution without regard to the timing of the reduction. Static optimal allocations (SOA) equalize the marginal costs of delivered pollution reductions assuming BMPs are fully effective in the period implemented. For a constant pollution target requiring about a 24.5% reduction in the pollutants, an SOA allocates approximately 60 and 40 percent of reductions to NPS and the PS, respectively. These allocations would only be realized after pollutants from sources with the longest lags have been delivered. Prior to that time, actual emissions reductions will not meet the load target, and the point source contribution to the actual pollution load reduction will exceed the SOA "planned" amount, while the nonpoint contribution to the actual allocation will fall short of the SOA "planned" amount.

Dynamic allocation rules take account of lags in distributing pollution across sources and time. Dynamic optimal allocations (DOA) equalize the discounted marginal costs of delivered pollution load reductions taking into account the timing of delivery. For the constant pollution load reduction targets, DOA simulations rely on point source reductions to fully achieve the target in the initial years, but substitute lower cost nonpoint reductions for point source reductions over time until a steady state allocation is achieved. The steady state NPS and PS reduction percentages for the DOA approximate to varying degree those of the SOA, though with somewhat greater reliance on PS than NPS. This greater reliance is a result of differences in DOA and SOA optimality rules. Simulations demonstrate that the initial and steady state allocations depend on lag lengths, lag-length-delivery ratio correlations, and pollutant type (reflecting differences in marginal abatement costs). They would also depend on discount rates but we limited our analysis to the OMB rate of 7%.

SOA rules save significant costs in all scenarios compared to dynamic allocation rules. This suggests significant economic merit from simpler static market designs than complex dynamic designs. The downside is that SOA rules delay achievement of water quality goals. However, we find that goal achievement under SOA rules is reasonably fast. In early periods, SOA reductions reach about 55 percent of the nitrogen load limit. These percentages increase through time, with about 90% of required nitrogen reductions being met by the strong correlation model in periods 9 and 19 for the 20 and 40 period models, respectively. About 75% of required nitrogen reductions are met by the weak correlation models in those periods. Nearly 100% of reductions are met by the final period for all scenarios. This implies that if the ultimate goal is to achieve the long term health of the Chesapeake Bay, the static solution fairs as well as cost-effective lagged solutions by the time that all lags are resolved. Of course, this assumes that factors affecting pollution in the Chesapeake Bay Watershed such as land use distribution and population remain constant through time.

Importantly, the SOA is not inconsistent with the implementation of the Chesapeake Bay TMDL. In the constant load target scenarios, the SOA requires immediate implementation of the control practices needed to achieve the targeted load reductions in the steady state. Delays in pollution reductions are the result of lags, not implementation. The SOA with the decreasing load target also requires immediate implementation of required practices, but in this case, the requirements increase through the implementation period.

In sum, the results of this study suggest that the emphasis on point sources for initial pollution control that has characterized nutrient pollution control is consistent with dynamically efficient allocations. The results also suggest that lags would lead to a greater allocation to point sources than nonpoint sources in the long run in a dynamically-efficient allocation as compared to a static-efficient allocation. Finally, our results suggest that static-efficient allocations save money and may perform reasonably well environmentally. If so, this would suggest that simpler market designs that do not explicitly account for lags may be appropriate for water quality trading.

References

Carlton, D. W. 1984. "Futures Markets: Their Purpose, Their History, Their Growth, Their Successes and Failures." *Journal of Futures Markets* 4(3): 237-271.

Chesapeake Bay Program. 2010. *ChesapeakeStat*. Available at http://stat.chesapeakebay.net/ (accessed July 2014).

Fisher-Vanden, K.; Olmstead, S. 2013. "Moving Pollution Trading from Air to Water: Potential, Problems, and Prognosis.: *Journal of Economic Perspectives* 27(1): 147-171.

Horan, R. D.; Shortle, J. S. 2011. "Economic and Ecological Rules for Water Quality Trading." *Journal of the American Water Resources Association* 47(1): 59-69.

Kaufman, Z., D. Abler, J. Shortle, J. Harper, J. Hamlett, and P. Feather (2014). "Agricultural Costs of the Chesapeake Bay Total Maximum Daily Load." *Environmental Science & Technology* 48:14131-14138.

Montgomery, W. D. 1972. "Markets in Licenses and Efficient Pollution Control Programs." *Journal of Economic Theory* 5(3): 395-418.

Nguyen, N. P.; Shortle, J. S.; Reed, P. M.; Nguyen, T. T. 2013. "Water Quality Trading with Asymmetric Information, Uncertainty and Transaction Costs: A Stochastic Agent-based Model." *Resource and Energy Economics* 35(1): 60-90.

Office of Management and Budget 1992. Circular A-94, *Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs*.

Ribaudo, M. 2013. *Encouraging Reductions in Nonpoint Source Pollution through Point-Nonpoint Trading: The Roles of Baseline Choice and Practice Subsidies*. Unpublished.

Shortle, J.; Kaufman, Z.; Abler, D.; Harper, J.; Hamlett, J.; Royer, M. 2013. "Building Capacity to Analyze the Economic Impacts of Nutrient Trading and Other Policy Approaches for

Reducing Agriculture's Nutrient Discharge into the Chesapeake Bay Watershed." Report to USDA, Washington, DC.

Shortle, J. 2013. Economics and Environmental Markets: Lessons from Water-Quality Trading. *Agricultural and Resource Economics Review*, 42(1), 57-74.

Shortle, J. 2012. *Water Quality Trading in Agriculture.* Report to the Organization for Economic Cooperation and Development, Directorate for Trade and Agriculture. OECD, Paris.

STAC (Chesapeake Bay Program Scientific and Technical Advisory Committee) 2013. "Incorporating Lag-Times into the Chesapeake Bay Program." STAC Publication Number #13-004. Edgewater, MD.

USEPA (U.S. Environmental Protection Agency) 2010. "Chesapeake Bay Phase 5.3 Community Watershed Model." EPA 903S10002 - CBP/TRS-303-10. USEPA, Chesapeake Bay Program Office, Annapolis, MD.

Tables

Table 1: Lag Length Ranges for the Chesapeake Bay

		Lags (Years)	
Pollutant Type	Transport Mode	Source to Stream	Stream to Bay
Nitrogen	Ground*	5-30	1 – 3
	Surface	<1 – 5	1 – 5
Phosphorus	Ground	5 – 30	5 – 50
	Surface*	<1 – 5	5 – 100
Sediment	Surface*	<1 – 5	5 –>100

*Major mode

Source: Chesapeake Bay Science and Technical Advisory Committee (2010)

Table 2: Subset of Lag Lengths Assigned to Each Bin for Strong Correlation Scenario

Bin	20 Period Model: Lag Lengths 0 - 19	40 Period Model: Lag Lengths 0 - 39
1 (df = 0.05)	17 – 19	35 – 39
2 (df = 0.15)	15 – 17	31 – 35
3 (df = 0.25)	13 – 15	27 – 31
4 (df = 0.35)	11 – 13	23 – 27
5 (df = 0.45)	9 – 11	19 – 23
6 (df = 0.55)	7 – 9	15 – 19
7 (df = 0.65)	5 – 7	11 – 15
8 (df = 0.75)	3 – 5	7 – 11
9 (df = 0.85)	1 – 3	3 – 7
10 (df = 0.95)	0 – 1	0 – 3

Note: df = delivery factor

Table 3: Dynamic Models Run

Number of Time Periods/Lag Lengths		Lag Length – Delivery Factor Correlation		Load Limit	
20	40	Weak (W)	Strong (S)	Steady State (SS)	Decreasing (D)
X		X		X	
X		X			X
X			X	X	
X			X		X
	X	X		X	
	X	X			X
	X		X	X	
	X		X		X
	X	X		X	
	X	X			X
	X		X	X	
	X		X		X

Table 4: Nitrogen Total Present Value Costs

Model	**Costs: Periods 0-39 (Billions)**	**Costs: Periods 0-19**
	20 Years – Constant Load Target	
DOA: Strong Correlation (SC)	n/a	$1.41
DOA: Weak Correlation (WC)	n/a	$1.85
SOA	n/a	$.79
	40 Years – Constant Load Target	
DOA: Strong Correlation (SC)	$2.16	$1.88
DOA: Weak Correlation (WC)	$2.68	$2.30
SOA	$.99	$.79
	40 Years – Decreasing Load Target	
DOA: Strong Correlation (SC)	$3.46	$2.7
DOA: Weak Correlation (WC)	$4.43	$3.41
SOA	$1.84	$1.27
	40 Years – Constant Load Target	
DOA: Strong Correlation (SC)	$3.14	$2.62
DOA: Weak Correlation (WC)	$3.43	$2.83
	40 Years – Decreasing Load Target	
DOA: Strong Correlation(SC)	$3.64	$2.85
DOA: Weak Correlation (WC)	$4.43	$3.41

Table 5: Undiscounted Nitrogen Marginal Costs

			Starting MACs	Steady state MACs
DOA	20Y	SC	$15.00	$6.00
		WC	$16.60	$7.00
DOA	40Y	SC	$16.30	$7.00
		WC	$17.30	$8.40
SOA	(constant across time)		$5.50	$5.50

Figures

Figure 1: Distribution of Land-River Segments into Nitrogen Bins

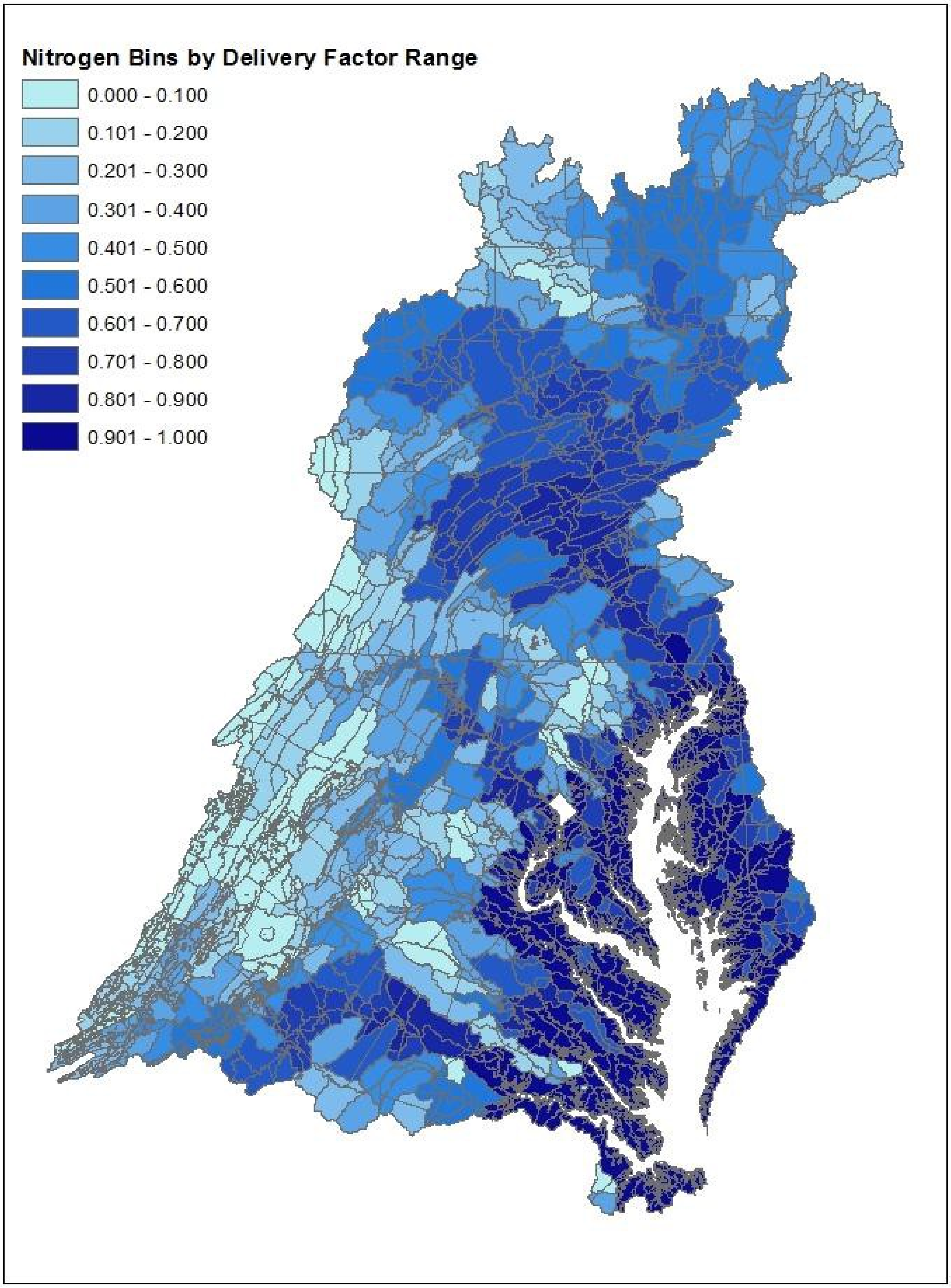

Figure 2: SOA Nitrogen Reductions: T = 20, Constant Load Target

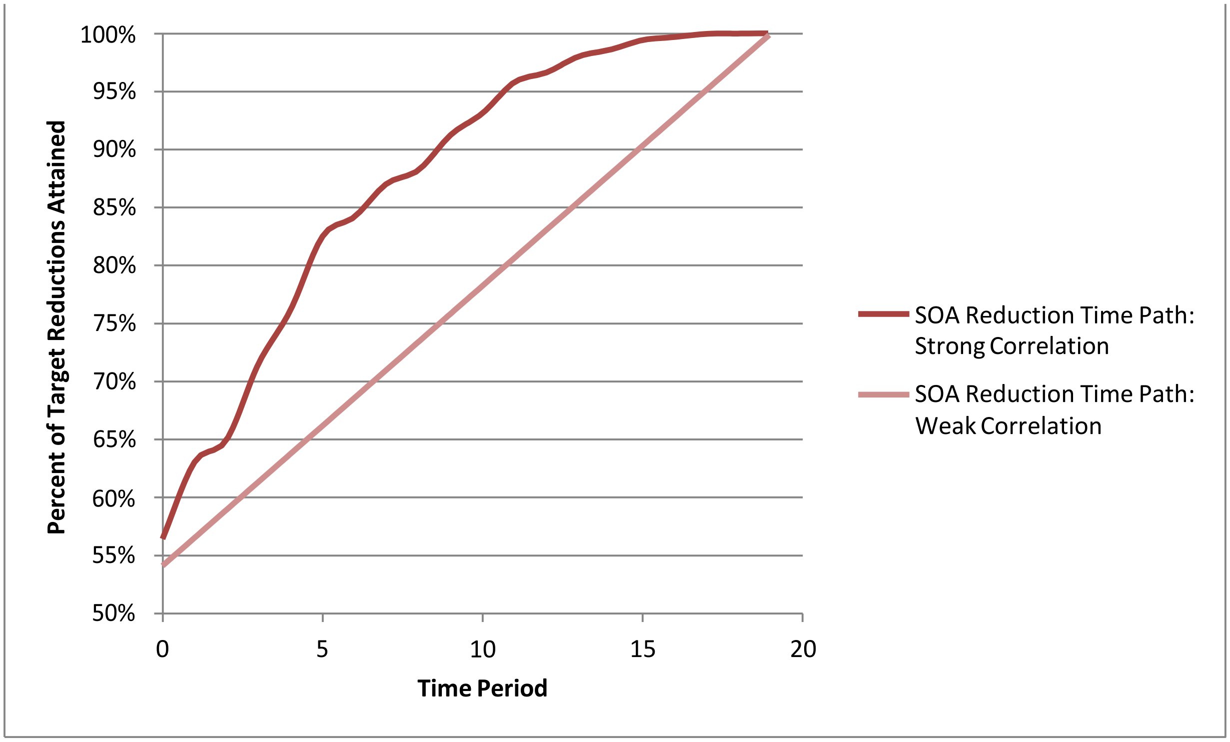

Figure 3: SOA Nitrogen Reductions: T = 40, Constant Target

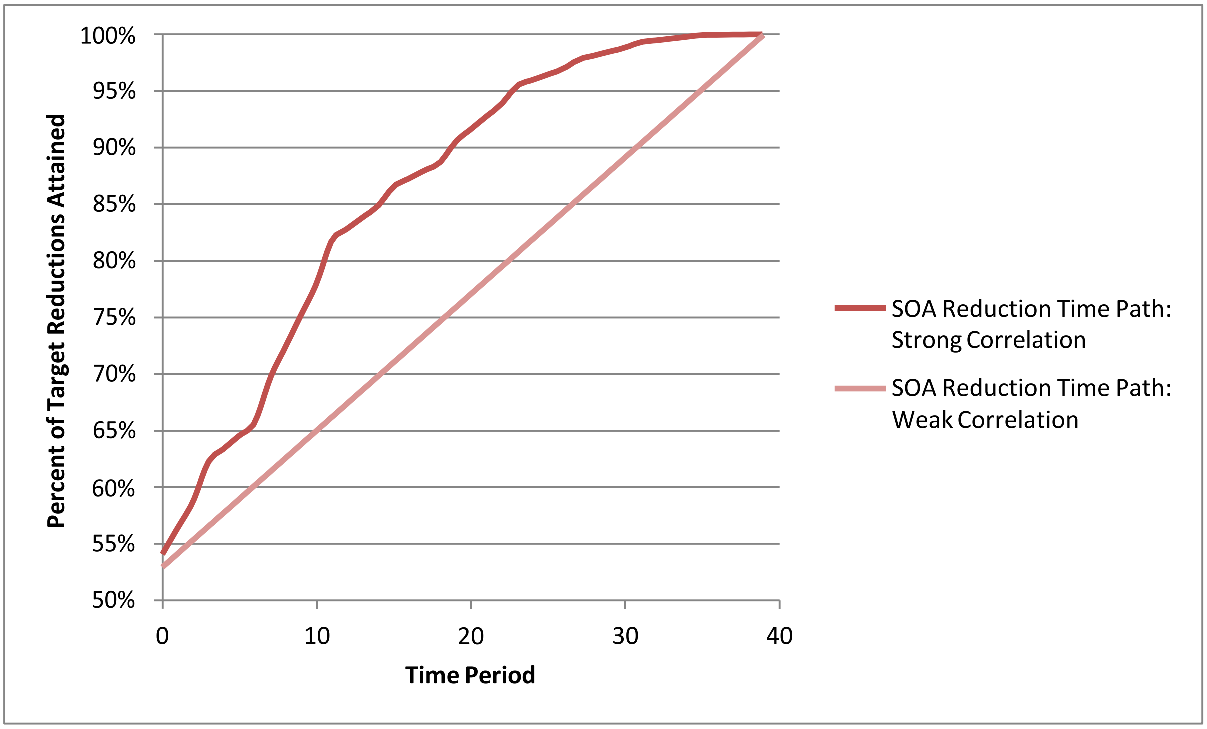

Figure 4: SOA Nitrogen Reductions: T = 40, Constant vs. Decreasing Target

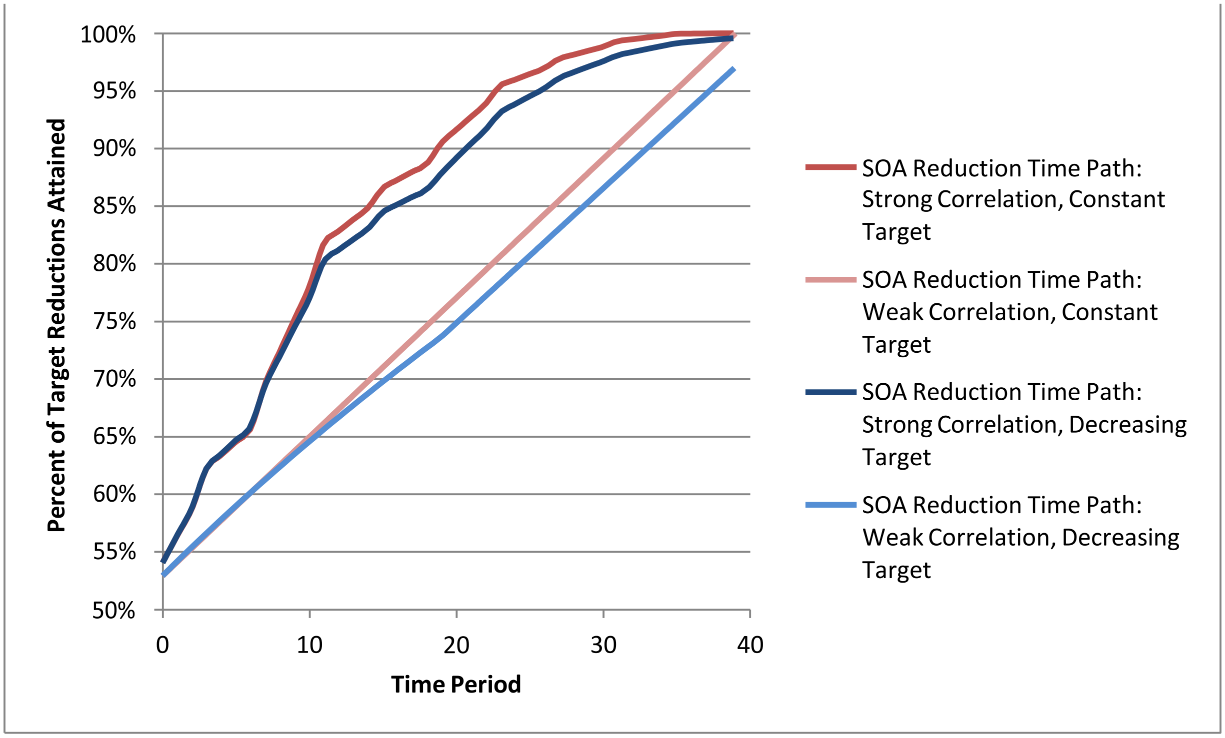

Figure 5: Distribution of NPS and PS reductions: T = 20, Strong Correlation, Constant Target

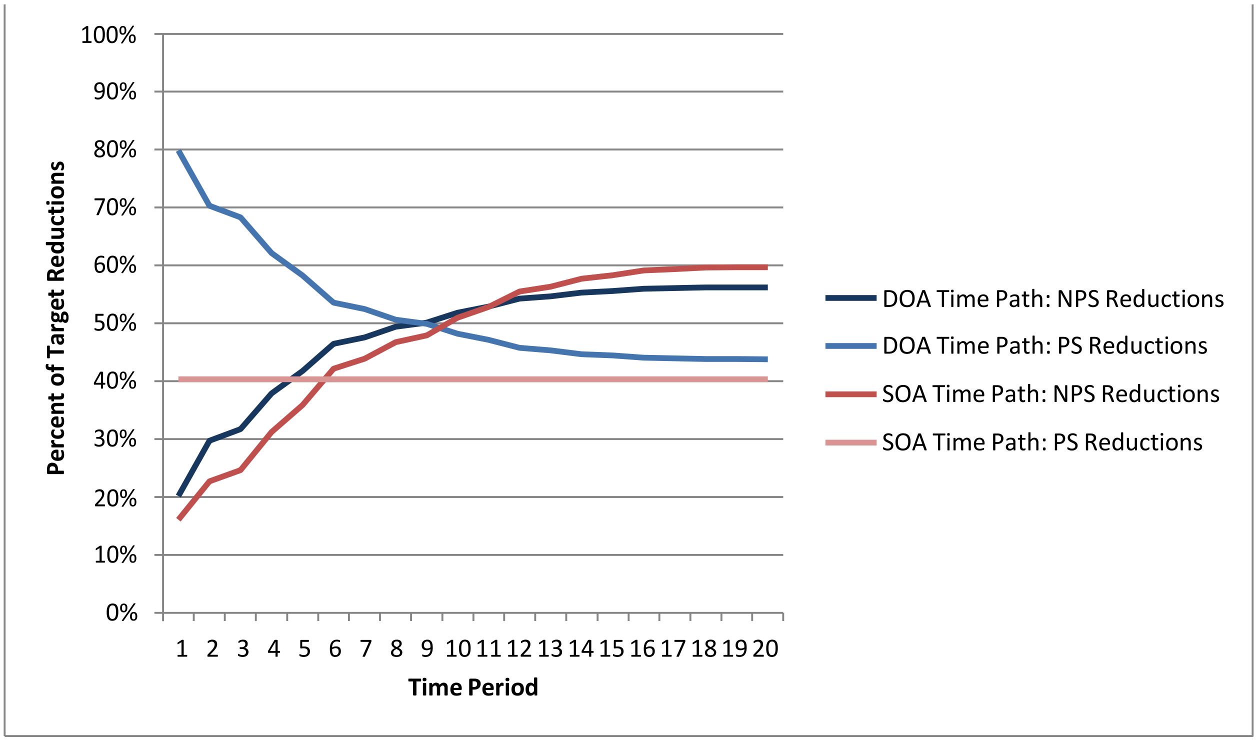

Figure 6: Distribution of NPS and PS reductions: T = 20, Weak Correlation, Constant Target

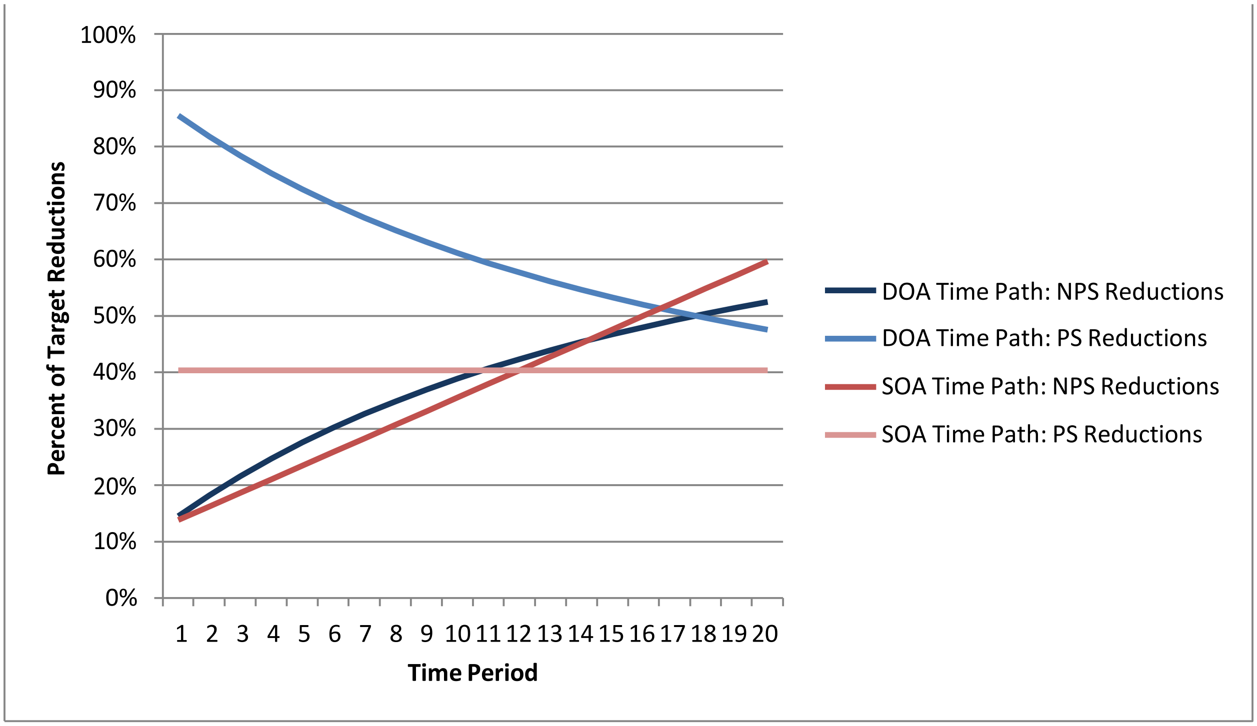

Figure 7: Distribution of NPS and PS reductions: T = 40, Strong Correlation, Constant Target

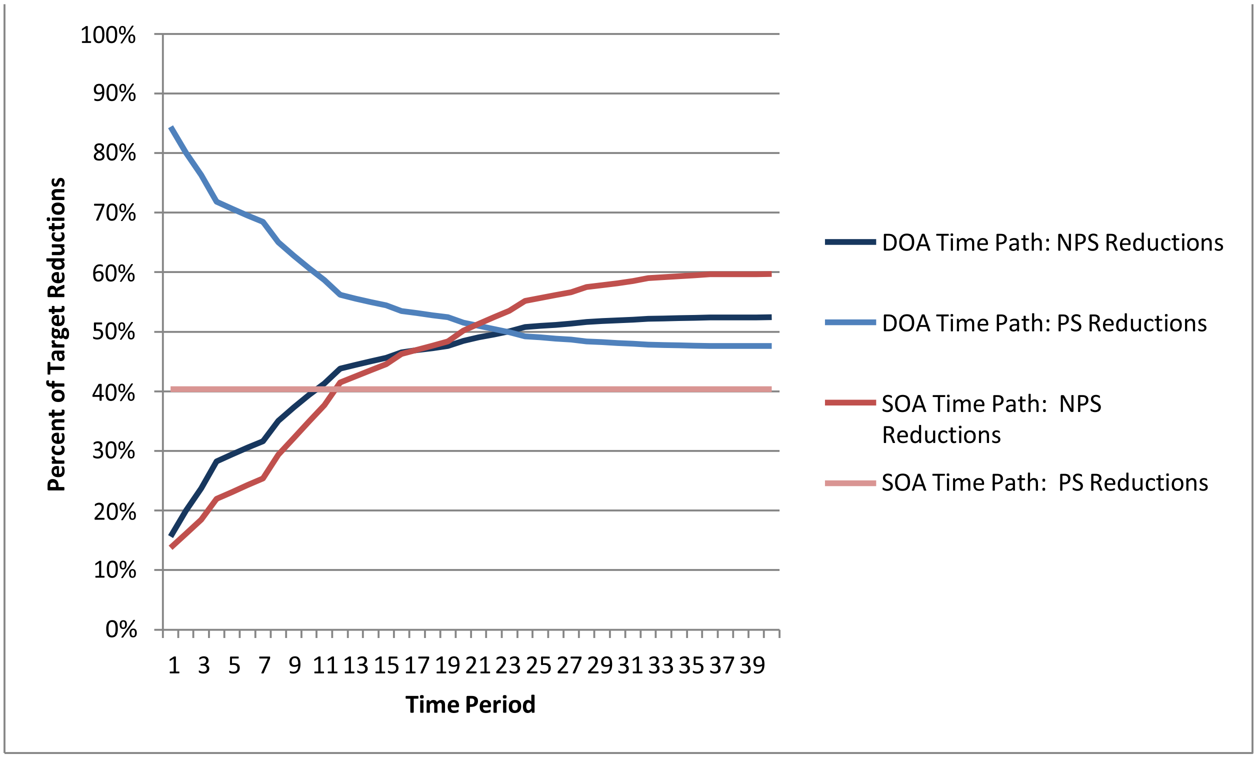

Figure 8: Distribution of NPS and PS reductions: T = 40, Weak Correlation, Constant Target

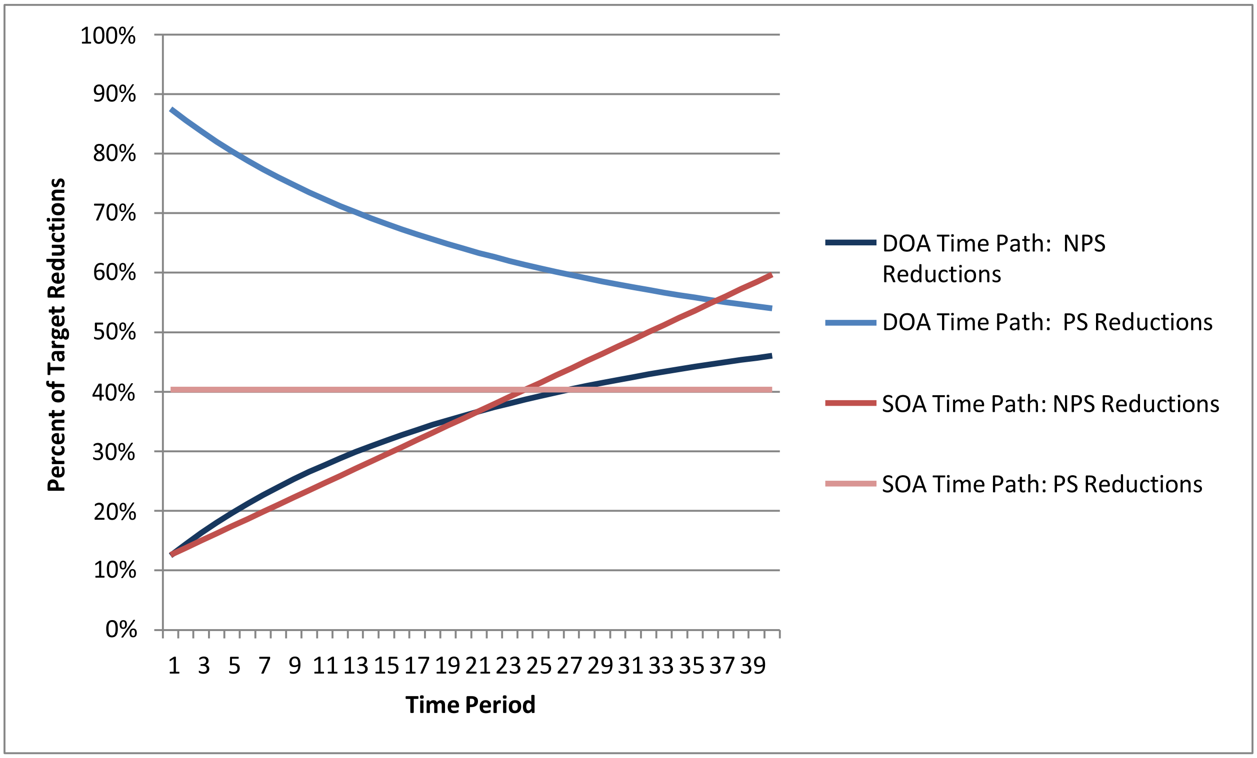

Figure 9: Undiscounted Nitrogen Marginal Costs: T = 20, No Adjustment Costs, Constant Target

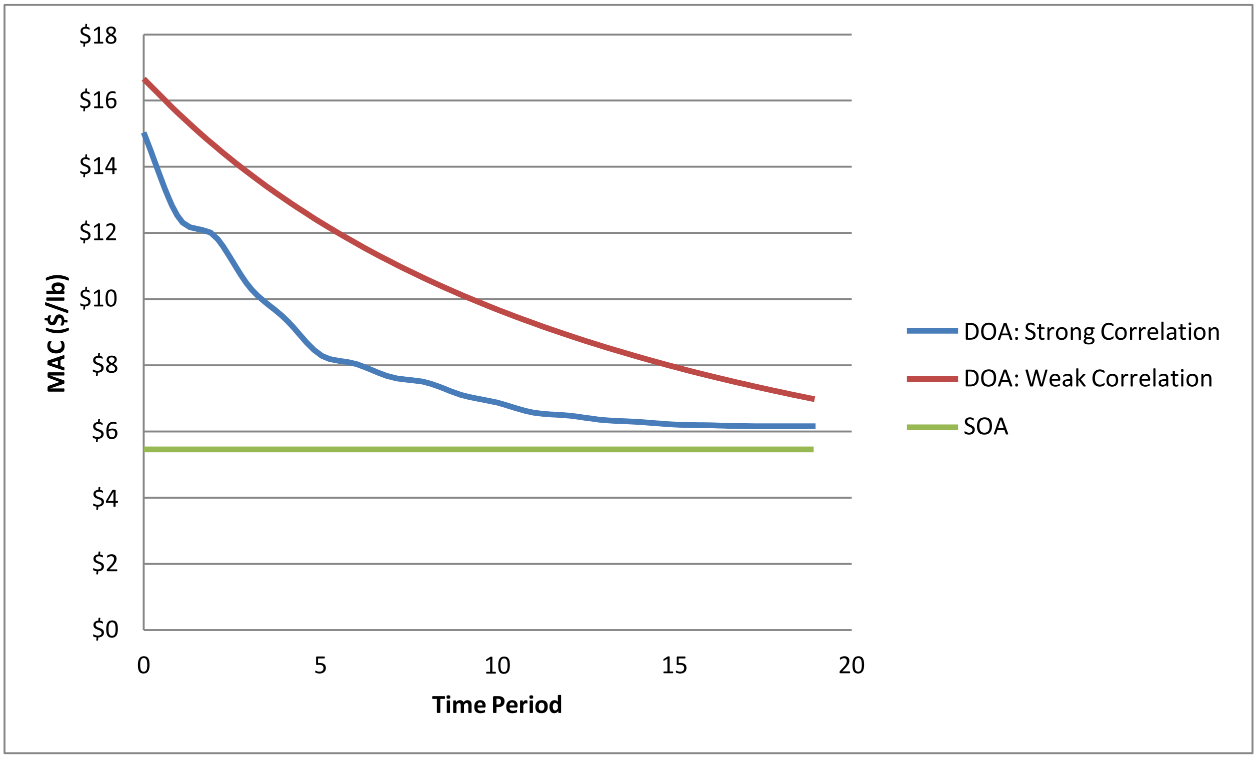

Figure 10: Undiscounted Nitrogen Marginal Costs: T = 40, Constant Target

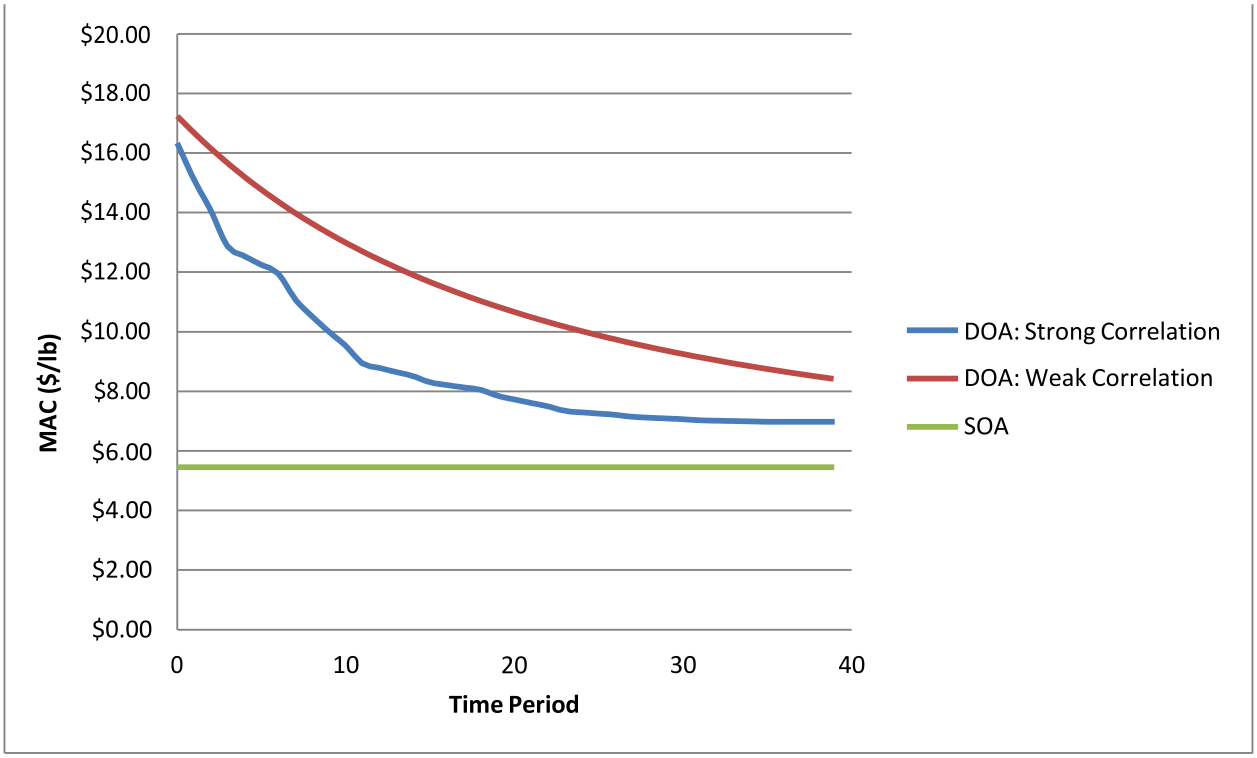

Figure 11: Undiscounted Nitrogen Marginal Costs: T = 40 Period, No Adjustment Costs, Decreasing Target