The Expected Value of Sample Information for Nonpoint Water Quality Management

Hwansoo Sung
Ph.D Candidate
Agricultural and Environmental Economics
Department of Agricultural Economics and Rural Sociology
308 Armsby Building
The Pennsylvania State University
University Park PA 16801
Phone: 814-863-8248

James Shortle
Professor of Agricultural and Environmental Economics
Department of Agricultural Economics and Rural Sociology
112 Armsby Building
The Pennsylvania State University
University Park PA 16801
Phone: 814-865-7657

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Long Beach, California, July 23-26, 2006

Abstract

There is considerable interest in watershed-based water quality protection. However, the approach can be highly information intensive, necessitating decisions about the types and amounts of data used to guide decisions. This study examines the Bayesian value of different types and amounts of sample information for reducing nutrient pollution in the Conestoga watershed of Pennsylvania, focusing on nitrogen from agricultural sources. Uncertainty is modeled from the perspective of a social planner seeking to maximize the economic efficiency of water quality control. A nested Monte Carlo procedure combined with an Evolutionary Optimization Strategy with Covariance Matrix Adaptation is used to compute resource allocation that optimizes the expected net benefit after updating for varying sample sizes and information types (broadly classified as pertaining to abatement costs, pollution fate and transport, and benefits of environmental protection). The results provide insights the returns from information investments to improve water quality management.

KEY TERMS: water quality management; value of sample information; Monte Carlo simulation

I. Introduction

Effective implementation of watershed-based water quality management requires significant information on pollution sources, water quality conditions, and relationships between land uses and pollution loads, and pollution loads and water quality conditions (National Research Council, 2000). This Information is typically imperfect, necessitating decision making under uncertainty. Uncertainty and the attendant risk can be diminished by investments in additional information. Yet, given scare resources for water quality management, managers must make decisions about the types and amounts of information that they collect.

One tool for guiding choices about the amounts and types of information is the value of information. Defined generally, the value of information is the increase in the expected "utility" of a decision that would derive from the collection and use of new information relative to the expected outcome achieved without new information. Value of information studies often focus on the *value of perfect information,* that is, new information that eliminates uncertainty about a specific parameter or set of parameters (Thompson and Evans, 1997; Yokota and Thompson, 2004). A recent example relevant to this study is Borisova et al (2005), who estimate the value of perfect information about the benefits and costs of water quality protection under alternative water quality policy regimes. However, casting information choices as being between imperfect prior information and perfect posterior information is both unrealistic and uneconomic. Rather, the realistic and economic choices recognize that the value of perfect information may not justify the cost of obtaining it.

A more useful concept is the *expected value of sample information* (EVSI). EVSI is the difference between the expected value of decisions selected after the collection of additional data (i.e., a sample) that is used to update the decision maker's priors, and the expected value of optimal decisions made only with prior information. While the concept is not new, applications have been limited (Yokota and Thompson, 2004). The EVSI concept is now beginning to attract attention in the environmental science literature and health economics, due in large part to advances in computing that make estimation of EVSI tractable (Dakins et al., 1996; Claxton et al, 2001; Ades et al., 2004; Yokota and Thompson, 2004).

This paper examines the value of economic and biophysical sample information for water quality management in the context of a case study of nitrogen pollution control from agricultural sources in the Conestoga River watershed (CRW) of Pennsylvania. Agriculture in the Conestoga is a major source of nutrients entering the Chesapeake Bay. The paper begins with a model of Conestoga watershed management problem. The concept of EVSI is then developed in the context of this problem, and an algorithm for computation of EVSI presented. Results and analysis are then presented.

II. Conestoga Watershed Model

The CRW is located in Lancaster County, in Southeast Pennsylvania, and within the Chesapeake Bay watershed (CBW) (Figure 1). Land use in the CRW is a highly agricultural, with 52% of the total land area in agriculture. Animal intensive agriculture in the watershed has been identified as a leading source of nutrients entering the Chesapeake Bay from Pennsylvania (Coale, 2003; Mark and Knaffke, 1998;

Susquehanna River Basin Commission, 1991). Reducing nutrient loads from agriculture is the CRW, is a major objective of water quality planners.

The objective of water quality decision making is taken to be maximizing the economic efficiency of water quality protection. Accordingly, the results obtained here are for the value of information for achieving efficient water quality protection. Three models are used to compute the expected net benefits of changes in resource allocation for water quality protection: (1) a model that determines the level of nitrogen loads to the mouth of the CRW from agricultural operations in the watershed, and correspondingly, the costs of changes in resource allocation for water quality protection; (2) a model determines the transport of nitrogen from the watershed to the Chesapeake Bay; and (3) a model of the costs of forgone ecosystem services in the Bay resulting from increased nutrient pollution.

Agricultural Production Model

The agricultural production model component is a highly simplified representation of the dominant integrated corn-dairy production system, intended to capture key policy variables and relationships. In the model, farmers produce corn and milk. Corn production requires plant nutrients (mainly nitrogen and phosphorous) and land. This study focuses on nitrogen. Nitrogen is supplied to corn land from dairy manure, and supplemented by purchases of commercial fertilizer. Corn is fed to cows and sold on the market.

Corn production is modeled using a two-level constant elasticity of substitution (CES) production function with constant returns to scale (Sato, 1967; Abler and Shortle, 1992; Kawagoe et al, 1985; Thirtle, 1985; Bingswagner, 1974):

(1) $$Y = B\left(\beta A^{\gamma} + (1-\beta)(uN)^{\gamma}\right)^{\frac{1}{\gamma}}$$

where Y is corn production, u is the share of applied nitrogen (N) that utilized by plants, B is a production function parameter for watershed, $\gamma = \dfrac{\sigma - 1}{\sigma}$, σ = elasticity of substitution between A and N, and $\beta = \dfrac{e_A}{e_A + e_N}$ (e_j stands for elasticity of demand for factor j (j = A or N)).

Nitrogen is supplied to the crop from two sources: purchased fertilizer (X_P) and cattle manure (X_M). These sources are considered to be imperfect substitutes. Assuming that excretion rate per cow is v, and that number of milk cow is C_N,the nitrogen supplied to the crops is expressed as:

(2) $$N = \delta\left(X_P^b + (rvC_N)^b\right)^{\frac{1}{b}}$$

where δ is a scale parameter and b = $\dfrac{\sigma_N - 1}{\sigma_N}$, σ_N = elasticity of substitution between X_P and X_M.[1] The nitrogen per unit of manure (r) is measured by the nitrogen contained in manure when excreted, less losses affected by manure management practices of storage and field application (Rotz, 2004).

Farmers are assumed to maximize profit, and to have no collective influence on the prices of inputs and outputs except for local crop land. The profit from corn production is:

(3) $$\pi_C = P_C B\left(\beta A^{\gamma} + (1-\beta)(u\delta\left(X_P^b + (rvC_N)^b\right)^{\frac{1}{b}})^{\gamma}\right)^{\frac{1}{\gamma}} - P_f X_P - RA$$

---

[1] All the values of substitution elasticity of 'A and N' and 'X_P and X_M' are positive, in this study. This implies that two input factors in each CES function are interpreted as substitutes, rather than complements (Kemfert, 1998).

where P_C and P_f are the prices of corn and nitrogen, respectively, and R is the market rental rate of land in the watershed.

Agricultural land is reasonably assumed to be inelastic in supply in the watershed, so that policy induced changes in the demand for land can be expected to impact local rental rates for agricultural land. Land supply is modeled as

(4) $$A = sR^{\eta}$$

where s is a parameter and η is the elasticity of land supply.

The profits from dairy production (π_M) are modeled as a function of cow numbers. Specifically, the following nonlinear function is used to approximate the restricted profit function of dairy production:

(5) $$\pi_M = C_N - \tau C_N^Z$$

where τ and Z are parameters.

Nitrogen Load Model

Following Abler et al (2002), the expected annual load to the month of the watershed (NL) resulting from crop production is modeled:

(6) $$NL = w_1 P_r^2 N_C A + w_2 (P_r^2 N_C)^2 A$$

where w_1, w_2 and w_3 are coefficients, P_r is the mean annual precipitation and A is land acreage devoted to corn production. The nitrogen concentration N_C is specified as the ratio of nitrogen runoff mass (1-u)N and water volume P_rA:

(7) $$N_C = \phi \frac{(1-u)N}{P_r A}$$

where ϕ is a calibration coefficient.

The portion of the load at the mouth of the watershed that reaches the Chesapeake Bay is modeled with a constant transport coefficient Θ so that the total nitrogen load reaching the Bay (L) is expressed as:

(8) $L = \Theta NL$

Environmental Damage Costs

The economics costs of nitrogen loads in the Conestoga watershed to the Chesapeake Bay is expressed as an increasing and convex function of the total nitrogen load L to the Bay:

(9) $D(L) = \rho L^{q}$

where$D$ is economic damage, ρ is a coefficient, q is elasticity of damage function, and $\partial D/\partial L > 0, \partial^2 D/\partial L^2 > 0$.

Social Net Benefit (SNB)

Under the assumptions to this point, the social surplus from agricultural production is,

(10) $$\begin{aligned} SocialNetBenefit = E[&\pi_C(X_P, C_N, A) + \pi_M(C_N) + ER(A) \\ &- D(L(X_P, C_N, A))] \end{aligned}$$

III. Uncertainty and EVSI

The planner's uncertainty about the costs and benefits of water quality protection is modeled by treating parameters of the various functions determining SNB as random variables with known distributions. Since the estimates of these uncertain parameters vary widely, uniform distributions[2] are used to describe the prior distributions of the parameters, as done in previous works (Horan et al, 2002; Borisova et al, 2005). Uncertainty is classified into three categories: (1) the costs of changes in resource allocation for water quality protection in agriculture; (2) the effects of changes in resource allocation on nutrient loads reaching the Bay from the CRW; and (3) the environmental benefits of reduced nutrient loads.

The planner's uncertainty about producers' costs of pollution control is modeled as a problem of asymmetric information where producers know their costs of compliance when making production decisions, but the planner does not know. The planner can only form an expectation of the farmers' responses to the policy, without perfect information. The elasticities of substitution in the nested corn production function (σ and σ_N), the elasticity of land supply (η), and the nitrogen utilization rate (u) are uncertain. So too are the manure per cow (v) and nitrogen per unit of manure (r). The prior distributions for these parameters are based on prior empirical studies (see Table 1). For the calibration of parameters in the dairy profit function (τ and Z), this study uses data on net income per cow of dairy farm in Pennsylvania (PFB MSC Services (2005)), along with baseline herd size.

Uncertainty about the movement of nitrogen from the watershed to the Bay is modeled by treating the transport coefficient, Θ, as a random variable with a known distribution. The prior for this distribution is based on Carmichael and Evans (2000). In

[2] Uniform distribution is a conventional choice for representing prior distribution of an uncertain parameter (Lawrence, 1999).

addition, annual precipitation P_r is normal distributed with a mean and variance based on precipitation data for the Lancaster County, PA.[3]

Uncertainty in the damage cost function is modeled by treating damage exponent (q) and damage coefficient (ρ) as random variables. The range of the values of these parameters is set such that the optimal expected load with updated information is in the range of 60-80% of the baseline load.

The Expected Value of Sample Information

The above 'Social Net Benefit (SNB)' function can be expressed in the simplest term as:

$$\text{(11)} \qquad \mathrm{SNB}\,(\mathrm{X}, \theta) = \pi(\mathrm{X}, \chi) - \mathrm{D}(\mathrm{X}, \psi)$$

where π(.) represents economic profits from agricultural resources used in agricultural production (X), D(.) is economic damage from the Conestoga nitrogen loads to the Chesapeake Bay, and θ (= $\begin{bmatrix} \chi \\ --- \\ \psi \end{bmatrix}$) is the vector of uncertain parameters with defined prior probability distributions of economic profit (χ) and economic damage (ψ).

The optimal resource allocation in agricultural production given the current (prior) information is:

$$\text{(12)} \qquad \mathrm{SNB_B} = \underset{X}{Max} \int_\theta SNB(X,\theta)P(\theta)d\theta$$

where P(θ) is a multi-variate prior probability distribution based on current information. Any data collected (D_I) would be used to update the uncertainty about the true underlying value of θ. Given a particular simulated data set D_I, a revised decision is made by

[3] For the estimation of the precipitation, county-level data of annual total precipitation of 73 years (1931-2004) are taken from National Climatic Data Center (NCDC).

evaluating each decision strategy in turn and then choosing the one with the highest expected SNB. Thus, expected SNB under the new information can be written as

$$(13)\qquad \mathrm{SNB_N} = \underset{X}{Max} \int_{\theta} SNB(X,\theta) P_P(\theta \mid D_I) d\theta$$

where $P_P(.)$ is pre-posterior probability density of updated θ.

Since it is not known what the result of the collection of data D_I will be, the distribution of possible results of the data D_I must be averaged. Thus, overall expected SNB of the proposed data collection is given by:

$$(14)\qquad \mathrm{SNB_N^A} = \int_{\mathrm{D_I}} \left\lfloor \underset{X}{Max} \int_{\theta} SNB(X,\theta) P_P(\theta \mid D_I) d\theta \right\rfloor P_D(D_I) d\mathrm{D_I}$$

where $P_D(.)$ is probability density function of data set D_I. This expression clearly shows the nested expectation procedure. The outer expectation relates to the variety of possible results of the proposed data collection. The inner expectation relates to the evaluation of the decision model under remaining uncertainty having obtained proposed data. In the inner circle, the uncertain parameters are sampled from their pre-posterior distributions $P_P(\theta_i|D_{Ii})$ for $\theta_i \in \theta$.

Finally, EVSI is defined as the difference between the expected value of decision made after a data of parameters of interest D_I have been collected (14) and expected value of a decision made with baseline information (12):

$$(15)\qquad \mathrm{EVSI} = \mathrm{SNB_N^A} - \mathrm{SNB_B}$$

$$= \int_{\mathrm{D_I}} \left\lfloor \underset{X}{Max} \int_{\theta} SNB(X,\theta) P_P(\theta \mid D_I) d\theta \right\rfloor P_D(D_I) d\mathrm{D_I} - \underset{X}{Max} \int_{\theta} SNB(X,\theta) P(\theta) d\theta$$

This two-step update procedure is implemented numerically using a "*nested inner Monte Carlo integration*' (Ades et al., 2004). In Figure 3, the corresponding flow chart for the above computation of EVSI is presented. The most important part of undertaking the above EVSI analysis is synthesizing the existing prior evidence with the simulated data to form a simulated posterior probability distribution for the parameter of interest. This process of updating the probability distribution $P(\theta)$ given new data D_I makes EVSI inherently Bayesian.

For optimization, the objective function (SNB) is highly nonlinear with respect to the control variables. Accordingly, SNB is optimized by heuristic method, rather than by direct optimization procedure. In this research, the Evolution Strategy with Covariance Matrix Adaptation (a heuristic optimization algorithm; Hansen and Kern, 2004) is used for the optimization. CMA-ES is a local search algorithm that only few best solution candidates are selected in each generation. So, the algorithm exploits search space more effectively, especially in large population of candidate solutions. This nested Monte Carlo simulation is performed using Matlab 7.0.

Bayesian Updating

The prior distributions for uncertain parameters are described mainly by uniform distributions which are conventional choices for representing a high degree of uncertainty (Lawrence, 1999). Such prior distributions are combined with the new simulation data using Bayesian updating process. In the absence of new data from the fields, this study generates new data such that the likelihoods for the proposed data are conjugate with the prior uniform distributions, and that the posterior have revised means and smaller variances than priors. Due to space limitations, such new information of the uncertain parameters is not presented in detail here.

In this study, the prior uniform distributions are transformed into normal distributions to mathematically simplify the Bayesian updating process as done by Chung (2004). Transforming from original uniform distribution of range (a, b), the mean (μ_0) and the variance (σ_0) of normal distribution are derived as

$$(16)\qquad \begin{aligned} \mu_0 &= \frac{1}{2}(a+b) \\ \sigma_0^2 &= \frac{1}{12}(b-a)^2 \end{aligned}$$

Combining this transformed normal distribution with the Bayes' theorem, the posterior distribution of sample size m (P^U), is derived as the normal distribution with the following mean (μ_u) and variance (σ_u):

$$(17)\qquad \begin{aligned} \mu_u &= \frac{\overline{X}(\sigma_0^2) + \mu_0(\sigma^2/m)}{\sigma_0^2 + (\sigma^2/m)} \\ \sigma_u^2 &= \frac{(\sigma_0^2)\times(\sigma^2/m)}{\sigma_0^2 + (\sigma^2/m)} \end{aligned}$$

where $\overline{X}$ is a sample mean and σ is a sample variance. At last, drawing from this updated posterior distributions, P^U, the Monte Carlo simulation is performed to derive ex ante posterior expected SNBs.

Sampling Plans

Four alternative sampling strategies are applied, for which expected values of sample information are compared. The first strategy is that each sample collects data on all the eleven uncertain parameters. In practice, it is, however, impossible for social planner to get into the study design of all uncertainty[4] because of huge cost of such

[4] All the samples of uncertain parameters are simultaneously increased to examine EVSI.

research. Rather, it is more reasonable to concentrate on a particular study design on a subset of all the uncertain parameters.[5]

Three alternative sampling strategies rather than the one with the samples of all the uncertain parameters, are presented in this study: 1) the set of uncertain parameters determining pollution control costs, 2) the set of pollution loading and transport parameter, and 3) the pair of parameters in the damage cost function that determine benefits of a given load reductions.

Samples varying in size from 20 to 7,000[6] are randomly generated for the above alternative sampling plans. This sample data is combined with the corresponding priors by the Bayesian inference to derive the EVSI of each sample size.

IV. Results

The computed values of EVSI for four information structures are presented in Table 2 and Figure 2. The entries are simply the difference in expected social net benefits between the given information structure and the baseline information structure. Additional information collection improves the EVSI for all information structures. For instance, the EVSI increases from \$0.769 million for a sample of 20 to \$3.16 million for a sample of 7,000 in the case of sampling of all parameter information. As shown in Table 2 and Figure 2, there are extensive increases in EVSI as the sample size is increased for a domain of small sample sizes. This implies that a small investment in research in can have a large payoff increasing the efficiency of water quality protection

[5] The derived EVSI from such particular study design is referred to "partial EVSI" (Brennan and Kharroubi, 2005).
[6] When sample of all the eleven uncertain parameters is simultaneously collected, estimates of EVSI are found to be invariant with respect to the sample sizes in the simulation over the sample size 7,000.

for a small scale research. Such gains are diminishing as research design reaches a certain level of sample size (around 1,000). This result is, of course, contingent on the quality of the prior information, assumptions about functional forms, and assumptions about the posterior distributions of the unknown parameters. In this study, a sample of 7,000 essentially eliminates uncertainty about the parameters considered. We note that EVSI is an upper bound on what should be spent on data collection: it does not take into account the costs of data collection. On the other hand, our EVSI is also a lower bound on the value of the information, because it does not include spillovers of this information. For instance, since nitrogen loadings from the Conestoga watershed are spilled over so many places beyond the Chesapeake Bay, EVSI would be underestimated in this study where the economic damage of nitrogen loadings is confined only to the Bay.

The improvement of EVSI also depends on the type of information collected. Sample collection of all the eleven uncertain parameters improves the EVSI at the highest level. Meanwhile, the value of obtaining additive information of the parameters of control cost turns out to be greater than those of obtaining better information of 'load and transport' or 'damage cost', but it is small in comparison to obtaining information of all uncertain parameters.

The additional important finding is that there are positive correlations among values of information of sampling strategies. As shown in Table 3, the differences in EVSI between information structure of all uncertain parameters and a sum of alternative three sample strategies are positive for all the sample sizes. These results imply that there are some interactions effects on the improvement of EVSI among three separate strategies.

The existence of such correlations among three strategies may suggest that research design on sampling strategies should be organized to aggregate information of all the possible uncertain parameters across ‘control cost’, ‘loading and transport of nutrients’, and ‘damage cost of pollution’, rather than to concentrate on specific study on the particular parameters.

It should be noted that these findings cannot be generally applied to the water quality management of other regional agricultural watersheds in the U.S. Given various regional characteristics of economic, biophysical and other attributes, the findings of this case study are contingent on the empirical model which is calibrated reflecting the specific characteristics of the Conestoga watershed.

V. **Conclusions**

This paper examines the expected value of sample information for alternative types of information for controlling water pollution from agriculture in the Conestoga watershed, and the gains from more intensive sampling (Table 2 and Figure 2). The results show that EVSI is increasing but at a diminishing rate as the sample size increases. The analysis also shows that information on control cost has the greatest impact on EVSI among alternative types of information.

Another important finding of the analysis is that the expected value of sampling strategy involving data collection of all uncertain parameters is greater than sum of expected values of three alternative sampling plans. This suggests that data collected about one particular study design on a subset of uncertain parameters may also provide some evidence about another subset of uncertain parameters. It is therefore important to implement such correlations between values of alternative sample information into the systemic sampling design.

References

Abler, D, J. Shortle, J. Carmichael, and R. Horan (2001) Climate change, agriculture, and water quality in the Chesapeake bay region. *Climatic Change,* 55(3), 339-359.

Ades, A.E., G. Lu and K. Claxton (2004) Expected value of sample information calculations in medical decision modeling. *Medical Decision Making*, 24, 207-227.

Arrow, K., H. Chenery, B. Minhas and R. Solow (1961) Capital-labor substitution and economic efficiency. *Review of Economics and Statistics*, 43(3), 225-250.

Baumol, W.J. and W.E. Oates (1979) *Economics, Environmental Policy, and the Quality of Life*. Englewood Cliffs, NJ: Prentice Hall.

Baumol, W.J. and W.E. Oates (1988) *The Theory of Environmental Policy*. New York: Cambridge University Press.

Borisova, T., J. Shortle, R.D. Horan and D. Abler (2005) The value of information for water quality management. *Water Resources Research*, 41(6).

Brennan, A. and S.A. Kharroubi (2005) Expected value of sample information in survival trials using Bayesian approximation with the Weibull proportional hazard model. Discussion paper, University of Sheffield.

Buchanan, J.M. and W.C. Stubblebine (1962) Externality. *Economica*, 29(116), 371-384.

Carmichael, J. and B. Evans (2000) A reduced form model of non-point pollution loading for regional scale analysis. working paper, Pennsylvania State Univ.

Chung, T.H., Y. Mohamed and S. AbouRizk (2004) Simulation input updating using Bayesian techniques. *Proceedings of the 2004 Winter Simulation Conference*, R.G. Ingalls, M.D. Rossetti, J.S. Smith and B.A. Peters, eds.

Claxton, K., P. J. Neumann, S. Araki and M.C. Weinstein (2001) Bayesian value-of-information analysis: an application to a policy model of Alzheimer's disease. *International Journal of Technology Assessment in Health Care*, 17, 38-55.

Coale, F.T. (2003) Agricultural best management practices for reducing nitrogen loading from the Conestoga River watershed. Interim project report submitted to Environmental Defense. Highland, Maryland.

Daskins, M.E., J.E. Toll, M.J. Small and K.P. Brand (1996) Risk-based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information. *Risk Analysis*, 16(1).

Floyd, J.E. (1965) The effects of farm price supports on the returns to land and labor in agriculture. *The Journal of Political Economy*, 73(2), 148-158.

Hansen, N. (2005) The CMA Evolution Strategy: A Tutorial.

Hansen, N. and S. Kern (2004) Evaluating the CMA Evolution Strategy on Multimodal Test Functions. *Eighth International Conference on Parallel Problem Solving from Nature PPSN VIII, Proceedings*, 282-291, Berlin: Springer.

Henley, N., J.F. Shogren, and B. White (1997) *Environmental Economics: In Theory and Practice*, Oxford, New York.

Horan, R., J. Shortle, and D. Abler (2002) Point-nonpoint nutrient trading in the Susquehanna Susquehanna river basin. *Water Resources Research*, 38(5).

Horan, R. and J. Shortle (2001) Environmental Instruments for Agriculture. In:

Environmental Policies for Agricultural Pollution Control / Shortle J. S. and D. Abler (eds.) CABI Publishing, Oxon (UK), NY (USA).

Kao, J.J. and H.J. Hong (1996) NPS model parameter uncertainty analysis for an off-stream reservoir. *Journal of the American Water Resources Association*, 32(5), 1067-1079.

Kemfert, C. (1998) Estimated substitution elasticities of a nested CES production function approach for Germany. *Energy Economics*, 20, 249-264.

Lawrence, D.B. (1999) *The Economic Value of Information*. Springer-Verlag.

Lu, Y., C.R. Camp and E.J. Sadler (2004) Efficient allocations of irrigation water and nitrogen fertilizer in corn production. *Journal of Sustainable Agriculture*, 24(4), 97-111.

Marks, R. and Knuffke, R. (1998) America's animal factories: how states fail to prevent pollution from livestock waste. *NRDC Reports*, Natural Resources Defense Council (www.nrdc.org).

MidWest Plan Service (MWPS) (2001) Manure Storages. *Manure Management System Series*. MWPS-18, Section 2. MidWest Plan Service, Iowa State University, Ames, Iowa 5011-3080.

National Agricultural Statistics Service (NASS) (2004) *Pennsylvania Agricultural Statistics 2003-2004*. Pennsylvania Statistical Office, Harrisburg, PA.

National Research Council (NRC) (2000) *Assessing the TMDL Approach to Water Quality Management*. National Academy Press, Washington, DC.

National Research Council (NRC) (1999) *New Strategies for America's Watersheds*. Washington, DC: National Academy Press.

PFB MSC Services (2005) *2005 Dairy Herd Analysis: Pennsylvania and Northeast Statistics*. PFB Members' Service Corporation.

Pizer, W. (1997) Prices vs. quantities revisited: the case of climate change. Discussion Paper 98-02, Resources for the Future. Washington D.C.

Randall, A. (1987) *Resource Economics*. 2nd ed. John Wiley & Son.

Ribaud M. (2001) Non-Point Source Pollution Control Policies in the USA. in J.S. Shortle and D.G. Abler (Ed), *Environmental Policies for Agricultural Pollution Control*. CABI Publishing, Oxon, UK and NY, USA.

Ribaudo, M.O., R.D. Horan, and M.E. Smith (1999) *Economics of Water Quality Protection from Nonpoint Sources: Theory and Practice*. Agricultural Economic Report Number 782. USDA. http://www.ers.usda.gov/publications/aer782/ (last accessed April 24, 2006)

Ribaudo, M.O., and J.S. Shortle (2001) Estimating Benefits and Costs of Pollution Control Policies, in J.S. Shortle and D.G. Abler (Ed), *Environmental Policies for Agricultural Pollution Control*. CABI Publishing, Oxon, UK and NY, USA.

Rotz, C.A. (2004) Management to reduce nitrogen losses in animal production. *Journal of Animal Science*, 82.

Shagam, S.D. (2006) Outlook for U.S. dairy. *Agricultural Outlook Forum 2006*, World Agricultural Outlook Board, USDA.

Shortle, J. S. and R. D. Horan (2001) The Economics of Non-Point Pollution Control. *Journal of Economic Surveys*, 15(3), 255-289.

Shortle, J.S., J. Winsten, H. Lin, P. Reed, J. Hamlet, B. Evans, and D. Beegle (2005)

Developing multi-scale performance indicators of agricultural nonpoint source pollution control to achieve watershed level water quality goals. working paper, Pennsylvania State University.

Susquehanna River Basin Commission (1991) *Loads and Yields of Nutrients and Suspended Sediment Transported in the Susquehanna River Basin*, 1985-1989.

Thompson, K. M. and J. S. Evans (1997) The value of improved national exposure information for perchloroethylen: a case study for dry cleaners. *Risk Analysis*, 17, 253-271.

US Environmental Protection Agency (2004) *Total Maximum Daily Loads*. http://www.epa.gov/owow/tmdl/.

Valiela, I., G. Collins, J. Kremer, K. Lajtha, M. Geist, M. Seely, J. Brawley, and C. H. Sham (1997) Nitrogen loadig from coastal watersheds to receiving estuaries: new method and application. *Ecological Applications*, 7(2), 358-380.

Verite, R. and L. Delaby (2000) Relation between nutrition, performances and nitrogen excretion in dairy cows. *Annals of Zootechnology*, 49, 217-230.

Weitzman, M.L. (1974) Prices vs. quantities. *The Review of Economic Studies* 41(4), 477-491.

Yokota, F. and K.M. Thompson (2004) Value of Information Analysis in Environmental Health Risk Management Decisions: Past, Present, and Future. *Risk Analysis,* 24(3), 635-650.

Figure 1. The Conestoga Watershed within the Chesapeake Bay Drainage

Figure 2. Social Net Benefits of Sample Information (Million Dollars)

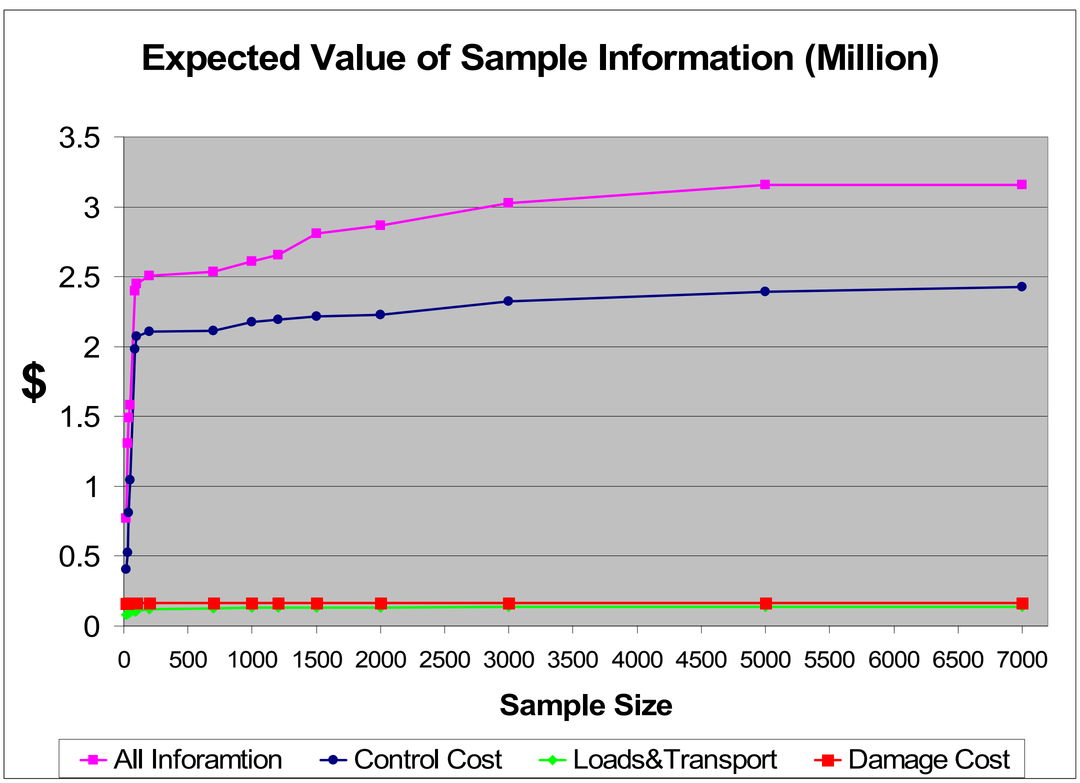

Figure 3. Flow Chart of Simulation for Sample Information Analysis

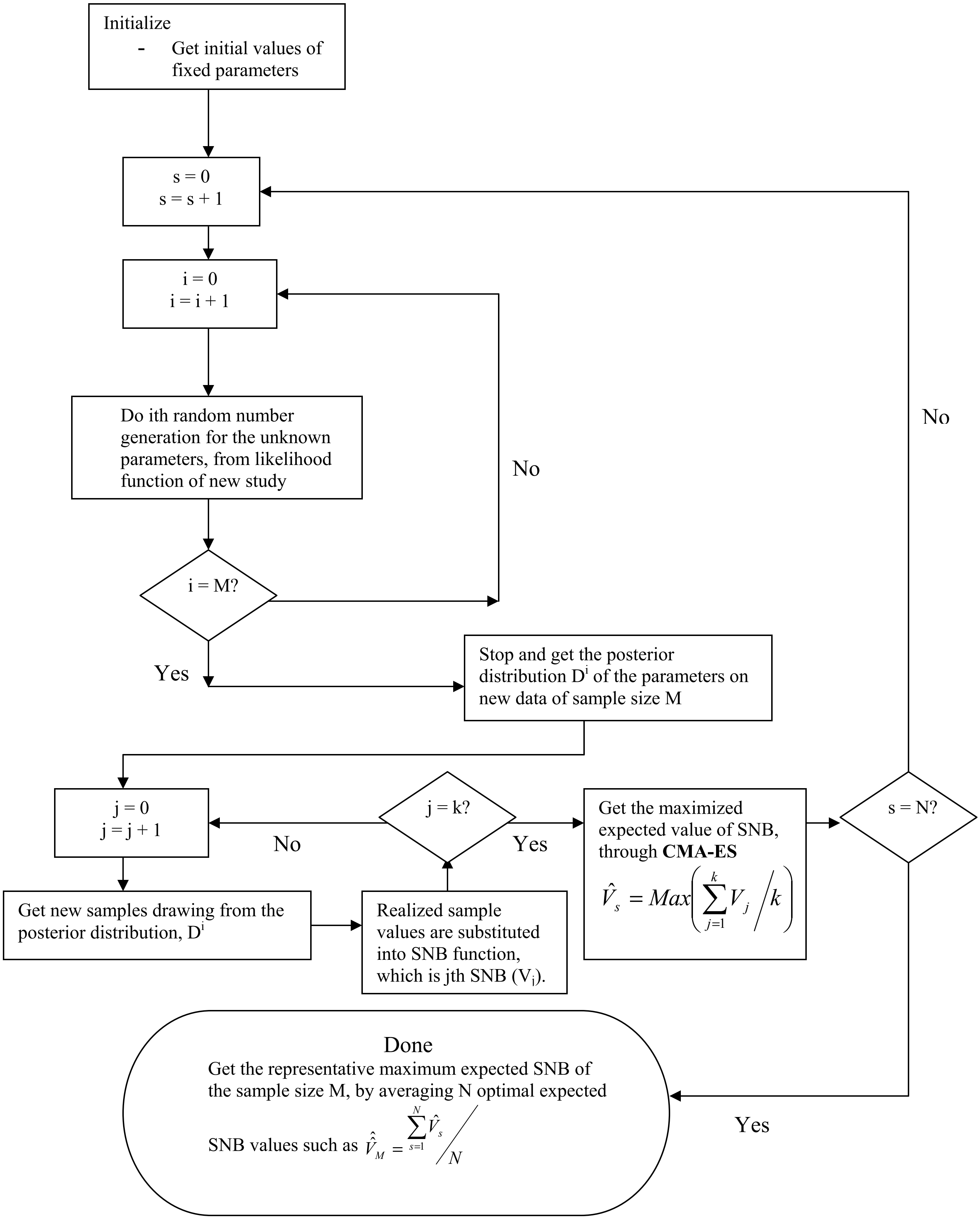

Table 1. Parameter Distributions

1.1. Economic Model for Agricultural Nonpoint Sources

Parameters	Notation	Mean	Variance	Sources
Production function parameter	B	0.6243	NA	The value is derived from the baseline information
Land cost share	S_L	0.1561	NA	USDA(2000)
Fertilizer cost share	S_N	0.2561	NA	USDA(2000)
Elasticity of substitution between land and fertilizer	σ	1.25	0.025	Binswanger (1974); Chambers and Vasavada (1983); Fernandez-Cornejo (1992); Hertel (1989); Kawagoe et al (1985); Ray (1982); Thirtle (1985)
Utilization rate	u	0.7	0.0033	Keeney (1982); Peterson and Frye (1989); NRC(1993)
Elasticity of substitution between inorganic fertilizer and manure	σ_N	2.355	0.4921	Bilgic et al (2002)
Nitrogen application parameter	δ	0.1208	NA	The value is derived from the baseline information
Loss rate of nitrogen per unit of manure	1-r	0.47	0.046875	MWPS (2001)
Nitrogen excretion rate per cow	v	120 (KgN)	300 (KgN)	Verite and Delaby (2000)
Elasticity of land supply	η	0.3	0.0075	Chavas and Holt (1990); Holt (1990); Lee and Helmberger (1985); Tegene et al (1988)
Land supply function parameter	s	43603	----	The value is derived from the baseline information
Measure of returns to scale of milk production	Z	0.7353	2.43×10^{-4}	PEB MSC Service (2005)

1.2. Nitrogen Loading and Transport

Parameters	Notation	Mean	Variance	Sources/Justification for values
Transport coefficient for the Conestoga Watershed	Θ	0.73	0.11	Smith et al (1997)
Load regression coefficient 1	w_1	3.2345×10^{-9}	NA	Abler et al (2002)
Load regression coefficient 2	w_2	8.8673×10^{-17}	NA	Abler et al (2002)
Constant scale factor of nitrogen concentration in runoff	ϕ	148,953.99	NA	Abler et al (2002)
Precipitation	P_r	41.48 (inches)	66.247 (inches)	Based on the time series data of the Lancaster County, PA (NCDC)

1.3. Economic Damages from Pollution

Parameters	Notation	Mean	Variance	Sources/Justification for values
Damage exponent	q	2	0.1089	Calibrated in the range of load reduction targets set by the Chesapeake Bay Agreement (2000)
Damage coefficient	ρ	0.0202	1.3209×10^{-4}	

Table 2. Expected Value of Sample Information (EVSI) for Sample Size m (Million Dollars)

Sample Size	All Information	Control Cost (a)	Load&Transport (b)	Damage Cost (c)
20	0.769	0.408	0.0791	0.1575
30	1.308	0.527	0.0832	0.1618
40	1.493	0.809	0.097	0.1628
50	1.582	1.044	0.0998	0.1632
90	2.399	1.984	0.1026	0.1637
100	2.447	2.0717	0.1093	0.164
200	2.504	2.109	0.1226	0.1642
700	2.534	2.115	0.1282	0.1644
1000	2.611	2.173	0.1302	0.1645
1200	2.654	2.19	0.1313	0.1646
1500	2.81	2.217	0.1331	0.1646
2000	2.868	2.229	0.134	0.1646
3000	3.028	2.322	0.1345	0.1646
5000	3.155	2.393	0.135	0.1647
7000	3.16	2.425	0.1362	0.1647

Table 3. Interactions among Three Sampling Strategies (Million Dollars)

Sample Size	All Information	Sum of EVSI of Three Sampling Strategies (a+b+c)	Interactions among Three Sampling Strategies
20	0.769	0.6446	0.1244
30	1.308	0.772	0.536
40	1.493	1.0688	0.4242
50	1.582	1.307	0.275
90	2.399	2.2503	0.1487
100	2.447	2.345	0.102
200	2.504	2.3958	0.1082
700	2.534	2.4076	0.1264
1000	2.611	2.4677	0.1433
1200	2.654	2.4859	0.1681
1500	2.81	2.5147	0.2953
2000	2.868	2.5276	0.3404
3000	3.028	2.6211	0.4069
5000	3.155	2.6927	0.4623
7000	3.16	2.7259	0.4341