Modeling water quality effects of variable-rate nitrogen fertilizer application using DRAINMOD-NII

and

Development of a new statistical framework for evaluation of goodness-of-fit for hydrologic models

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Grace L. Wilson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advised by Dr. David Mulla

July 2018

Acknowledgments

The investigators gratefully acknowledge financial support from the Clean Water Land and Legacy Amendment Fund administered through the Minnesota Department of Agriculture.

Table of Contents

Acknowledgments i

Table of Contents ii

List of Tables vi

List of Figures vii

Introduction 1

PART I: Literature Review 3

Chapter 1. Fertilizer and water quality concerns, and Concepts used for model evaluation 4

Overview of Fertilizer Timing and Application Rate on Water Quality 4

Nitrogen fertilizer and subsurface drainage in agricultural systems 4

Improving water quality through fertilizer management practices 6

General Concepts Useful in Model Evaluations 10

Overview 10

Goodness of Fit: Methods to evaluate models 10

Uncertainty of Observed and Predicted Data 13

Methods to Account for Uncertainties 14

GOF indicators adjusted to take into account uncertainties 20

Measurement Uncertainty 20

New Framework and Method for Adjusting GOF 23

PART II: Modeling water quality effects of variable-rate nitrogen fertilizer application using DRAINMOD-NII 24

Chapter 2. Effects of Fertilizer Timing and Variable Rate N on Nitrate-N Losses from Tile Drained Corn and Soybean Rotation Simulated Using DRAINMOD-NII 25

Synopsis 25

Introduction 26
Methods 28
Experimental Site Characteristics 28
Field Management 29
DRAINMOD Simulations 30
Results and Discussion 33
Results of Model calibration and subsurface drainage 33
Observed and Predicted Nitrate Concentrations for 2016 and 2017 35
Nitrate Load for 2017 37
Comparison of nitrate loads for single, split, and VRN management 39
Change in N budgets 41
Conclusions 43
Chapter 3. Predicted Nitrate-N Loads for Fall, Spring, and VRN Fertilizer Application in Southern Minnesota 44
Synopsis 44
Introduction 45
Materials and Methods 46
DRAINMOD-NII Model Description 46
Model Calibration 46
Site Descriptions 47
Model Calibration 48
DRAINMOD Simulations 49
Regression Analysis 51
Results and Discussion 52
DRAINMOD Calibration 52

Nitrate Losses Predicted by DRAINMOD-NII ... 53

Regional Regression Equations ... 55

Comparison between Fall, Spring and Split-VRN Applications ... 57

Conclusions ... 58

PART III: Development of a new statistical framework for evaluation of goodness-of-fit for hydrologic models ... 60

Chapter 4. A new framework for assessing the goodness-of-fit of hydrologic and water quality models. Part I: Theoretical Development ... 61

Synopsis ... 61

Introduction ... 61

Framework ... 64

Accounting for Spatial and Temporal Factors ... 64

Observed Data ... 65

Predicted Values ... 67

Spatially Averaged terms ... 68

Goodness of Fit indicators: NMSE and MSE ... 70

Expected value of MSE and NMSE ... 71

Adjusted MSE and NMSE ... 75

Influence of component terms on MSE and NMSE ... 81

Expected value of MSE and NMSE not adjusted for errors ... 81

Expected value of MSE and NMSE adjusted for errors ... 83

Summary ... 86

Chapter 5. A new framework for assessing the goodness-of-fit of hydrologic and water quality models. Part II: Evaluation and Application ... 88

Synopsis ... 88

Introduction ... 89

Methods 93
Accuracy of Adjusted MSE and NMSE using Expected Value Approach 93
Results 100
Accuracy of Adjusted MSE and NMSE using Expected Value Approach 100
Illustration of Method Using a Case Study 115
Experimental Sites 115
Model Performance using adjusted and un-adjusted MSE and NMSE 116
Comparison of GOF for DRAINMOD results 117
Summary and Conclusions 120
Bibliography 121
Appendix 131
A. General Relationships for Expected Value and Variance 131
General Relationships 131
B. Full derivation of E(SSE) and E(SSTO) for separate parameter for each time and space 131
Sums of squares error: SSE_{st} 132
Sums of squares total: $SSTO_{st}$ 133
Mean Square Error and Normalized Mean Square Error 137
Expected value of (non-adjusted) MSE and NMSE 137
C. Variance of the actual adjusted and theoretically derived GOF 139
Numerator, SSE 139
Adjusted SSE, measurement error: 140
Adjusted SSE, measurement + parameter error 141
Variance of the difference between actual and theoretical 142
D. Visual Basic Code for generating Monte Carlo Simulations 143

List of Tables

Table 1: Summary of monthly NSE values for subsurface drainage and nitrate load. 33
Table 2: Fertilizer rates and timings used for DRAINMOD-NII simulations.. 51
Table 3: Summary of DRAINMOD-NII model performance...52
Table 4: Summary of key variables used for adjusting GOF statistics 69
Table 5: Field measured values of subsurface drainage from Waseca, MN.....................94
Table 6: Values of error terms used in Monte Carlo simulations. 95
Table 7: Probability that adjusted NMSE less than or equal to 0.1. 107
Table 8: Probability that adjusted MSE are less than or equal to 2.83. 112
Table 9: Annual observed and DRAINMOD predicted subsurface drainage flow at Waseca and Lamberton...118
Table 10: Adjsuted GOF statistics calculated for Waseca and for Lamberton. 119

List of Figures

Figure 1: Seasonal trends in plant N uptake and months with highest N loss potential. ..8

Figure 2: Waseca study location, drainage intensities and nitrogen treatments. 29

Figure 3: Observed and predicted monthly subsurface drainage.....................................34

Figure 4: Observed and predicted monthly subsurface drainage. 35

Figure 5: Observed and simulated nitrate concentration for the years 2016 and 2017.....36

Figure 6: Observed and simulated nitrate load and subsurface drainage for 2017. 38

Figure 7: Simulated nitrate load for single pre-plant, split, and VRN 40

Figure 8: Simulated average annual nitrogen budgets...42

Figure 9: DRAINMOD simulated nitrate load for spring fertilizer application vs. growing season precipitation 53

Figure 10: DRAINMOD simulated nitrate load for split-VRN fertilizer application vs. growing season precipitation 54

Figure 11: DRAINMOD simulated nitrate load for fall fertilizer application vs. growing season precipitation...55

Figure 12: Ln-ln plots of the regression predicted nitrate losses vs. DRAINMOD-NII predictions..56

Figure 13: Regression predicted nitrate loss for southern Minnesota under spring, fall, or split-VRN fertilizer timings 58

Figure 14: Diagram showing the relationship between terms for observed data.............67

Figure 15: Relationship between the component terms and the resulting expected value of MSE and NMSE 83

Figure 16: Adjusted MSE and NMSE 86

Figure 17: Convergence of the mean difference..101

Figure 18: Mean value from MC simulations..102

Figure 19: Coefficient of variation calculated from MC simulations. 104

Figure 20: Distribution of $\Delta NMSEi, \varepsilon + \varepsilon$...105

Figure 21: Probability that the NMSE difference is less than or equal to 0.1 108

Figure 22: Standard deviation from MC simulations 110

Figure 23: Mean difference between the NMSE adjusted using known random numbers and the expected value theory..111
Figure 24: Standard deviation from MC simulations of the MSE differnce...................113
Figure 25: Mean MSE difference 114

Introduction

In the first part of this research, the field-scale hydrologic model, DRAINMOD-NII, was used to make predictions on nitrogen losses for fields utilizing different nitrogen fertilizer management practices. DRAINMOD-NII was used to predict nitrate loads for different fertilizer application rates and timing to corn at Waseca, Lamberton, and Willmar in Southern Minnesota. The fertilizer practices simulated included a single application in the spring before planting, a split application with half applied pre-plant and half at approximately the V6 stage for corn, and split and variable rate N practice (VRN) which utilized the split timing and a lower rate based on in-season monitoring of plant N requirements. Results of simulations at all three locations were used in a regression analysis to develop equations to predict nitrate-N loses for the region more generally as a function of fertilizer timing and application rate. Model results showed that fertilizer application applied as a VRN practice has the potential to reduce nitrate loads in southern Minnesota.

In the second part of this project, a new framework was developed which would take into account uncertainty in the evaluation of hydrologic and water quality models. Model performance is evaluated using statistical goodness-of-fit (GOF) measures that compare the observed data measured in the field to model predictions. Though both the observed and predicted data contain uncertainty due to measurement error and model input parameter error, current methods of calculating GOF do not take these uncertainties into account. Here, a framework was developed that allows the observed and predicted data to be described in terms of the uncertainty due to measurement and model parameter error. This framework was used to adjusted mean square error (MSE) and normalized mean square error (NMSE) to account for errors in the observed and predicted data. Additionally, the accuracy of this approach was evaluated.

A note on the format of the dissertation: This dissertation is written in three parts. The first section, Part I (chapter 1), is a review of pertinent literature for both the effects of fertilizer timing and rate on water quality, and on the statistics for model evaluation. In Part II of the dissertation (chapters 2 and 3), the work on modeling water quality effects using DRAINMOD-NII is presented. In Part III of the dissertation (chapters 4 and 5), the work on development of a new statistical framework is presented. Figure and table numbers are continuous through the dissertation. However, equation numbers are specific to the chapter in which they are referenced.

PART I: Literature Review

Chapter 1. Fertilizer and water quality concerns, and Concepts used for model evaluation

Overview of Fertilizer Timing and Application Rate on Water Quality

Nitrogen fertilizer and subsurface drainage in agricultural systems

Nitrogen (N) is essential for crop growth, and a key to maximizing crop productivity is agricultural management that ensures adequate soil available nitrogen. In the first half of the 20th century, crop and livestock production were integrated, and nitrogen for crops was obtained primarily through fixation by forage legumes and from livestock manures (Dinnes et al. 2002). By mid-century several technological innovations, such as improved crop cultivars, increased mechanization, and introduction of chemical nitrogen fertilizers, led to increased acreage in simplified short-term annual crop rotations and a separation of crop and livestock production (Dimitri et al. 2005, Russelle et al. 2007). In intensive monoculture systems, synthetic or "chemical nitrogen" fertilizers have become the primary nitrogen input, since less "biological nitrogen" was available in the form of manure or fixed during crop rotations. Synthetic N fertilizer has become the primary nitrogen source for farmers managing intensive monoculture systems (Robertson and Vitousek 2009), including most of the farmland in the Upper Midwest.

The rise in synthetic nitrogen fertilizer use occurred simultaneously with intensification of another field management practice: installation of subsurface tile drainage systems. Subsurface tile drainage systems improve crop production and allow for more intensive management by removing excess water from the fields, allowing for timely fieldwork and sufficient aeration for crop roots (Evans and Fausey 1999). Tile drainage continues to be installed in fields; in 2002, the rate of new installation was estimated to be 100 million feet of tile drainage lines installed annually in Minnesota (Busman and Sands 2002). In addition to improving crop yields, the combination of fertilizer application and installation of extensive tile drainage systems has had a profound effect on the movement of nitrogen in Midwestern soils.

Subsurface Drainage Systems

Agricultural tile drainage systems involve the installation of permeable pipes to remove standing excess water. Up until the 1970's, these pipes were made of concrete or clay "tiles" though today most pipes are perforated polyethylene tubing (Busman and Sands 2002). Drainage systems are beneficial to crops in agricultural soils that have low water infiltration rates (such as clay soils), or in areas of flat or depressional topography. A large fraction of cropland in Minnesota fits these categories; in the heavily row-cropped, 4-million hectare Minnesota River Basin, approximately 60% of the soils are classified as poorly drained (Sands et al. 2006).

Subsurface drainage systems are typically designed to keep the height of the water table at a depth which is optimal for crop growth (Van Der Ploeg, R. R. et al. 1999). In the Upper Midwest region, this typically results in pipes placed at depths 100 to 120 cm below the soil surface (Sands et al. 2006). Water flows through drainage systems only under saturated field conditions, when there is a hydraulic potential gradient between the system and the surrounding soil (Ritzema 1994). Subsurface drainage systems are connected to surface waters via a network of drainage ditches, which carry water and associated soluble pollutants from upland farm fields into streams, rivers, and lakes.

Nitrogen Movement to Surface Waters Facilitated by Drainage Systems

Nitrogen in its nitrate form is highly water soluble and repelled by the negative charge on soil particles (Follett 2008), making it highly susceptible to transport from fields to surface waters via subsurface drainage systems (Baker et al. 1975). Nitrogen in the soil, either from applied fertilizers or in soil organic matter, is converted into nitrate-N through the processes of mineralization and nitrification. Though application of nitrogen fertilizer introduces additional nitrogen to the soil, the mineralization of soil organic matter can also be responsible for the formation of a substantial amount of nitrate-N in agricultural fields in the region (Randall and Mulla 2001, Randall and Goss 2008, Ghane et al. 2016).

Nitrate losses through subsurface drainage systems are primarily determined by the volume of water transported through the drainage system, and the nitrate-N concentration in the soil. Factors that influence these include: precipitation, crop maturity and evapotranspiration (ET) rates, soil organic matter content, and fertilizer management practices. High precipitation—especially when coupled with low ET rates—are highly correlated with nitrate losses (Randall and Mulla 2001). Randall and Goss (2008) summarize 25 studies which all show year to year variability in drainage flow and nitrate loss due to the variability in annual precipitation. Besides influencing tile flow, precipitation can also influence soil nitrate concentration through the effect of soil moisture on mineralization rates. Mineralization rates are particularly high when wet years follow very dry years, resulting in more nitrate-N in the soil and higher susceptibility to leaching (Randall 1998). Fertilizer management practices including fertilizer application rate and timing also influence soil nitrate concentrations, with higher concentrations resulting from higher fertilizer application rates (Randall and Goss 2008).

Drainage and nitrate load show a seasonal response, with higher flow and associated N loss occurring during months when there is little plant uptake of water or nitrogen. For warmer regions where fields are fallow but not frozen during the winter, the peak flow and associated N loss occurs during this fallow winter period. In Indiana, Kladivko et al. (2004) found that 81% of annual tile drain flow and 78% of N load occurs in the months November through April. For northern areas of the Upper Mississippi River basin where soils remain frozen during the winter months, low ET rates and higher drainage volumes generally coincide with the spring months of April, May, and June. In a 15-year study in southern Minnesota, these three months accounted for 71% of the annual drainage volume and 73% of the annual nitrate loss from a corn-soybean rotation (Randall 2004).

Improving water quality through fertilizer management practices

While nitrogen increases crop yields, excess nitrogen in the water supply is detrimental to fresh and marine aquatic ecosystems, contaminating drinking water and coastal areas (Schulte et al. 2006). Nitrogen from annual row crop agriculture in the Upper Midwest is a main contributor to the hypoxic zone in the Gulf of Mexico (CENR 2010). A significant portion of the nitrate-N polluting both fresh and marine waters originates from agricultural regions which utilize subsurface drainage (David et al. 2010). The Minnesota Pollution Control Agency estimates that 70% of nitrate-N found in Minnesota surface waters comes from cropland tile drainage (MPCA 2013). Excess nitrate-N in drinking water causes methemoglobinemia, a disease that affects the oxygen carrying capacity of the blood (children are particularly susceptible to this condition). The costs of reducing nitrates to safe levels for drinking can be considerable. In the heavily row-cropped region of Des Moines, IA, a $3.7 million facility was required to remove the excess nitrates that accumulate in the Mississippi River (Des Moines Water Works 2009).

There is increasing pressure on farms in the Upper Midwest to reduce nitrate-N losses. Recent guidance by the US EPA suggests that the amount of nitrate-N entering the Mississippi River from upstream farm fields would need to decrease by 45% in order to reduce the extent of the hypoxic zone in the Gulf of Mexico to a reasonable size (US EPA 2007). There are several nutrient best management practices and technologies that can be used on-farm to reduce nitrate-N losses from fields, and subsequent loadings to surface waters. One nutrient management approach to minimize N losses is to alter the rate and/or timing of N fertilizer application.

Fertilizer Timing

For northern regions, the greatest potential for nitrogen loss occurs in the spring months, when drainage rates are high and plant uptake of N is low. Theoretically, applying nitrogen fertilizer when plants have increased nitrogen demands—which occurs after the period of highest drainage and greatest leaching potential—should reduce nitrogen losses (figure 1). Typical fertilizer application often occurs well before the largest N demands of the crop occur; in southern Minnesota, a large proportion of fertilizer application occurs in fall before planting (MPCA 2013). Fall application of N fertilizer has been show to result in the highest losses of N compared to other application timings. Randall and Mulla (2001) showed that over a 6-year period for a continuous corn system at Waseca, MN, fall application of N fertilizer resulted in an average nitrate load that was 36% higher than a spring application. In a modelling study by Nangia et al. (2008), switching from fall to spring application resulted in a 9% decrease in nitrate losses.

Split application of nitrogen fertilizer addresses the asynchrony in timing between nitrogen application and plant uptake by applying part of the total N amount at or before planting, and part as side-dress during the growing season (fig 1). This approach delays full application of fertilizer until corn nitrogen uptake rates peak around growth stage V6-V8. Field studies have not shown consistent results with regard to the change in nitrate loss for a split application strategy vs. a single spring application. At Waseca, MN, Randall et al. (2003), found that the four year (1990-1993) totals for N lost were higher for a split (40% pre-plant/60% side-dress) compared to a single spring application, with the 4-year totals of N lost equaling 190 kg/ha for the split rate compared to 177 kg/ha for the single pre-plant. In Iowa, Jaynes (2015) found that fall application of N (196 kg N/ha) showed higher nitrate concentration and load than spring or split fertilizer applied at a slightly lower rate (168 kg N/ha). The spring application showed on average a lower N load (15.7 kg N/ha) than the split (19.5 kg N/ha), though these results were not significantly significant at p=0.05. In contrast to these results, Bakhsh et al. (2002), found a 25% reduction in N loss for a split application of N to corn compared to single spring application in Iowa.

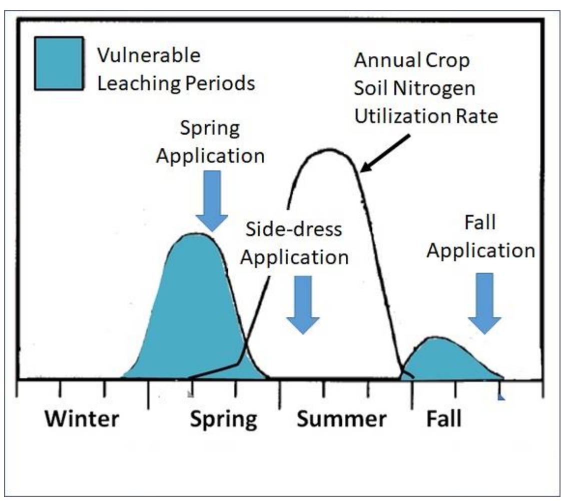

Figure 1: Seasonal trends in plant N uptake and months with highest N loss potential. Fertilizer applied in the fall shows a large temporal mismatch between plant uptake and N fertilizer application, leaving applied fertilizer in the soil during the period of greatest N loss potential. Spring application limits the temporal mismatch.

Fertilizer Rate and Spatial Variability

Reduced fertilizer rates have been shown to reduce both N load and soil concentrations. For poorly drained soils in Iowa, Jaynes et al. (2001) found about a 27% reduction in N load for a 33% reduction in fertilizer rate, with nitrate loads of 29, 35, and 48 kg N/ha for N rates of 57-67, 114-135, and 172-202 kg N/ha. Similarly, lower nitrate loads were also found in southern Minnesota, with loads of 19, 25, 59, and 120 kg/ha measured for application rates of 20, 112, 224, and 448 kg N/ha (Gast et al. 1978). For the 50% decrease in fertilizer application rate from 224 kg N/ha to 112 kg N/ha, this resulted in an almost 58% reduction in nitrate load.

Additionally, fields in the upper Midwest can display significant within-field variability in soil organic matter and nitrogen content (Mamo et al. 2003; Scharf et al. 2005), yet more than two-thirds receive blanket N application (Erickson and Widmar 2015). Addressing the spatial variability in plant nitrogen requirements in a field could result in overall lower fertilizer application rates and lower losses of N (Power et al. 2000).

Applying the Right rate and the Right Time

Split variable-rate nitrogen (VRN) fertilizer application utilizes a split timing, but bases N fertilizer application rates on spatial variations in plant N requirements, addressing both the temporal and spatial mismatches between N application and crop development. VRN utilizes methods of estimating soil or plant N content to determine how much N should be applied at side-dress, potentially reducing the overall N applied. Nitrogen management practices which focus on applying the right rate of N at the right time, such as split and variable rate nitrogen (VRN), have been proposed as a practice which could reduce N loads to surface waters (i.e. Nelson 1985). Side dressing rates can be variable based on crop sensing (Mulla 2013) or in-season modeling of plant available N. In a modeling study which compared nitrate load between split application, and split with a reduced rate, Sela et al. (2016) found a 36% decrease in N losses corresponding to a 34% decrease in side-dress N application. However, there are few other studies that have researched the effects of this fertilizer application strategy on water quality.

General Concepts Useful in Model Evaluations

Overview

Models that simulate the hydrology and chemical transport of agricultural systems are used to make predictions of the effects that different management practices, cropping systems, and climate changes have on the movement of water, transport of contaminants, and water usage. These predictions are often used as the basis for policy decisions and management practices. Therefore, it is important to have an understanding of how well the results of the model simulation match the natural processes of the system (Harmel et al. 2014). Evaluation of model performance is done by comparing simulated model results that predict different hydrologic components (subsurface drainage, runoff, streamflow, etc.) or nutrient loads, to corresponding observed data measured in situ.

In this review of literature, a summary of methods often used to measure the goodness of fit between predicted and observed values, the types of uncertainties inherent in evaluation of model accuracy, and different approaches used to account for these uncertainties.

Goodness of Fit: Methods to evaluate models

Methods for model evaluation include both graphical and statistical performance measures (Engel et al. 2007, Moriasi et al. 2015). Graphical performance measures allow for visual comparison of predicted and observed data. They are qualitative and are dependent on the judgement skills of the modeler. Statistical methods evaluate model performance as a pairwise comparison of the measured (observed) data and the corresponding model simulated (predicted) data. These statistical goodness of fit (GOF) measures allow for a more quantitative assessment of how well the model simulations match the observed data, making them more useful in taking a standardized approach to model evaluation (Legates and McCabe 1999). Goodness of fit estimates include those that assess the absolute error and those that normalize it based on the variance in the observed data.

Absolute Error

Measures of absolute error recommended for use in evaluating hydrologic model performance include root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and percent bias (PBIAS) (Legates and McCabe 1999). These are defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N}} \tag{1.1a}$$

$$MSE = \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N} \tag{1.1b}$$

$$MAE = \frac{\sum_{i=1}^{n} abs(O_i - P_i)}{N} \tag{1.1c}$$

$$PBIAS = \frac{\sum_{i=1}^{n}(O_i - P_i)}{\sum_{i=1}^{n} O_i} \times 100 \tag{1.1d}$$

Where O_i is the observed data for a specified time step, i, P_i is the corresponding predicted data, and N is the number of observations.

Normalized error

Measures of absolute error do not take into account variability in the observed data, and so can bias against datasets that show significant variability with respect to the time steps. Other measures of goodness of fit normalize the comparison by taking into account the range in the observed data. Here the *residual term* $(O_i - P_i)$ is divided by a measure of this range, which often is the variance of the observed data defined as:

$$VAR(O) = \frac{\sum_{i=1}^{n}(O_i - \bar{O})^2}{n} \tag{1.2}$$

where $\bar{O}$ is the mean value of the observed data. Measures of goodness of fit that normalize the data include the normalized mean square error (NMSE), Nash-Sutcliffe Coefficient of Efficiency (NSCE), and index of agreement (d) (Legates and McCabe 1999, Gershenfeld and Weigend 1993):

$$NMSE = \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \tag{1.3a}$$

$$NSCE = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} = 1 - NMSE \tag{1.3b}$$

$$d = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}[abs(P_i - \bar{O}) + abs(O_i - \bar{O})]^2} \tag{1.3c}$$

Squared Correlation Coefficient and Coefficient of Determination

The squared correlation coefficient (r^2), sometimes called the coefficient of determination (i.e. Legates and McCabe 1999, Moriasi et al. 2015), is also a commonly reported GOF measure, where it is defined as the square of the correlation coefficient, r:

$$\begin{aligned} r^2 &= \left(\frac{\frac{1}{n-1}\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{s_x s_y}\right)^2 \\ &= \left(\frac{\sum_{i=1}^{N}(O_i - \bar{O})(P_i - \bar{P})}{(\sum_{i=1}^{N}(O_i - \bar{O})^2)^{0.5}(\sum_{i=1}^{N}(P_i - \bar{P})^2)^{0.5}}\right)^2 \end{aligned} \tag{1.4}$$

Where s_x is the standard deviation of the observed term, and s_y is the standard deviation of the predicted term.

Though used as a recommended measure of GOF (i.e. Moriasi et al. 2015), the squared correlation coefficient can be a confusing statistic in evaluating model GOF. The correlation coefficient gives the linear fit between the observed and predicted data—how closely they fall on a straight line. In the context of model evaluation, it can be confusing to use. The r^2 statistic is commonly used for the case of a linear relationship, where r^2 is the same as the coefficient of determination (C_d):

$$C_d = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \tag{1.5}$$

The coefficient of determination is sometimes thought of as the fraction of the total variation in the dependent variable (here, this is the predicted value) “explained” by the regression equation. Because of the familiarity of r^2 with the linear case, it can be a confusing statistic to use in the hydrologic modelling context as it implies linearity. Additionally, as a metric unto itself, the square of the correlation coefficient is not more informative than other normalized measures of GOF, which have a more straight-forward calculation. Mathematically, the coefficient of determination is identical to the Nash-Sutcliffe coefficient given by equation 1.3b. The coefficient of determination is generally used when P_i's are determined by linear models, whereas the Nash-Sutcliffe coefficient is used when P_i's are determined from nonlinear, hydrologic/water quality models.

Guidelines for values of GOF indicators that indicate "acceptable" model performance can be found in Moriasi et al. (2007 and 2015). These papers recommend that models can be deemed "satisfactory" for flow for $R^2>0.6$, NSCE >0.5, and PBIAS between ± 10 to 15%. For sediment "satisfactory" ranges are $R^2>0.4$, NSCE >0.45, PBIAS between ± 20 to 30%; for nitrogen and phosphorus, $R^2>0.3$, NSCE >0.35, and PBIAS between ± 20 to 30%.

Uncertainty of Observed and Predicted Data

The GOF indicators presented in equations 1.1 and 1.3 give a performance measure assuming that the predicted and observed data are both "true" values. In reality, both the model-predicted and observed values have an associated error or uncertainty. Uncertainties in observed data often result from instrument or analysis error during data collection (Harmel et. al 2006). For model predictions, the uncertainty is due to: 1) uncertainty in model input values (not knowing the "correct" value for model inputs and parameters), and 2) errors in model structure or algorithms (how well the math used by the model represents reality).

Uncertainty in Observed Data (O_i)

Uncertainty in the observed data is typically considered a result of instrument measurement error and can be considered to be random, unless there is some known systematic instrument error. Harmel et al. (2006) reported on uncertainty quantified as a percent error in measured water quality data introduced by four procedural categories:

streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis. They concluded their paper by emphasizing the importance of treating the observed data not as single estimates but rather as ranges, noting that, "models should be expected to produce output within the uncertainty limits inherent in measured data, not to produce outputs with low deviation from measured data."

Uncertainty in Model Prediction (P_i): Model parameter uncertainty

The uncertainty in model inputs is due to incomplete knowledge of both model parameter values (those based on functional relationships, such as curve number or Mannings N), and measured input data such as soil physical properties (saturated hydraulic conductivity, porosity, etc.), or climate data. Measurement error caused by inaccuracies in instrumentation or in analysis techniques can introduce uncertainty into the measured input data. Additionally, many inputs (such as soil physical properties, crop growth parameters, climate data, etc.) exist across a range for a given modelling unit, and uncertainty is introduced by having to choose which value(s) in that range will be input into the model. Thus, uncertainty in model inputs can be introduced by instrument measurement error, but also results from the difficulty in describing the spatial and temporal variability inherent in the natural system in terms of the discrete units required for modeling (Guzman et. al 2015, Haan 1989, Mulla and Addiscott 1999, Shirmohammadi et. al 2006). The uncertainty in model inputs is generally considered to be random (Haan 1989); for example, you are equally likely to choose a K_{sat} that is an overestimate of the "true" value as you are a value that is an underestimate. Uncertainty introduced into model predicted results from these sources will be referred to as model parameter error.

Methods to Account for Uncertainties

Methods to evaluate uncertainty due to parameter error

If a single correct model input value for each parameter in a model existed and was known, a single correct model output value could be obtained. However, since all parameters are inherently random variables (model inputs can vary over a range), their values are not known with perfect certainty. Because of this, the model predicted output will also vary over a possible range of values (depending on what exact input value is used), and model prediction can be thought of as a distribution with a mean and variance (uncertainty). Monte Carlo, first-order analysis, likelihood analysis, and likelihood analysis using a Bayesian framework are all common forms of analysis of model parameter error (Haan 1977, Stow et al. 2007, Haan and Skaggs 2003), and are described in more detail below. Variations of these methods used to assess uncertainty in model predicted data are discussed in reviews by Guzman et al. (2015) and Shirmohammadi et al. (2006).

Estimation of Model Parameter Errors

Monte Carlo: With the Monte Carlo (MC) approach, information on the known variance and mean of model *parameters* is used to determine the distribution of the *model output*, including its mean and variance (Haan 1989). The first step in this approach is to determine the probability density functions for each parameter of interest, so that a range of values is generated for each parameter, K. Then, the model is run using each of these K values, for each parameter of interest, generating an output distribution based on the different values of K.

The Monte Carlo approach is a robust methodology for determining the probability distribution function (pdf) for model output, however it has some limitations. The probability density function—or at least distribution, mean, and variance—must be known for each of the parameters. Monte Carlo is also computationally intense and can be time consuming, as many simulations (for the range of K's) are needed to generate the model output distribution. Additionally, it is important to maintain correlations among parameter values. For example, many soil characteristics are correlated; if for example a change is made to soil porosity, hydraulic conductivity would also be expected to change with it. These correlations are difficult to define and therefore difficult to examine with MC techniques.

First-Order Analysis (FOA): Similar to Monte Carlo techniques, first-order analysis estimates the mean and variance of the model output (Garen et al. 1981, Harr 1987). With this method, the mean and variance of the model output is determined based on Taylor series expansion, ignoring second and higher order terms. Using this framework, the mean of the output (η) is equal to the model output that occurs using the mean value of each parameter, K:

$$E(\eta) = \eta(\mu_{K1}, \mu_{K2}, \dots \mu_{Kp}) \tag{2.1}$$

The variance of the model output is the sum of the product of the square of the sensitivity coefficient obtained using the mean value for a given K ($S^2_{\mu_{K_i}}$) and variance for each random parameter, K ($VAR(K_i)$):

$$VAR(\eta) = \sum_{i=1}^{n} S^2_{\mu_{K_i}} VAR(K_i) \tag{2.2}$$

The sensitivity coefficient for the parameter K is defined as the change in the response of the model for a change in the parameter value.

First-order analysis is a less time consuming method to estimate the mean and variance of the model output compared to Monte Carlo techniques, however, accurate results will only be achieved if the model can be reasonably approximated as being linear. For a linear model, the FOA gives an exact solution, since second and higher order terms in the Taylor series approximation for a linear model are equal to zero. However, for a non-linear model, these terms are not equal to zero, and FOA is only an approximation that becomes less reliable as K deviates more from the mean value and the model becomes more non-linear. For the case of the linear model, or when FOA can provide a good approximation of the model output mean and variance, the model mean and variance obtained using MC should be the same as the model output obtained used the mean values of K.

Likelihood Function: In the framework of model evaluation, the likelihood function is used to determine the likelihood of obtaining a set of observed values for different possible model parameter values (Sorooshian et al. 1983, Wilson and Haan 1991). The most likely parameters are those that maximize the probability of obtaining the observed data, thus minimizing the error, or difference between the observed, O_i, and predicted obtained using those parameters, $\eta(K)_i$:

$$\varepsilon_i = O_i - \eta(K)_i \tag{3.1}$$

To determine the distribution of the likelihood of different parameters (K) in minimizing the difference between the predicted and observed, first, model simulations are run with different K's. For each simulation, the error term (ε_i) is calculated. Then, the probability of each error term is calculated, which requires knowledge of the distribution of the error term. For uncorrelated, normally distributed errors with a mean of zero, the probability density function for a given parameter can be computed by:

$$f(\varepsilon_1) = \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\varepsilon_1}{\sigma_1^2}\right)^2\right),\ \ f(\varepsilon_2) = \frac{1}{\sigma_2\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\varepsilon_2}{\sigma_2^2}\right)^2\right) \tag{3.2}$$

Where $f(\varepsilon_i)$ is the probability for obtaining the first error (ε_1) for a specific K, and σ_i^2 is the variance of the errors for obtained for that K. The relationships of equation 3.2 can be readily defined for different distributions of the error

The next step is to calculate the joint probability of errors for that given parameter K, which is equal to the product of each of the probabilities from equation 3.2. For uncorrelated, normally distributed errors, this can be calculated as:

$$\begin{aligned} f(x\ /\ K) &= f(\varepsilon_1, \varepsilon_2, \dots \varepsilon_n\ /\ K) \\ &= \frac{1}{(2\pi)^{n/2}\sigma_1\sigma_2 \dots \sigma_n}\exp\left(-\sum_{i=1}^{n}\left(\frac{\varepsilon_i}{2\sigma_i^2}\right)^2\right), \end{aligned} \tag{3.3}$$

where $f(\varepsilon_1, \varepsilon_2, \dots \varepsilon_n\ /\ K)$ is the joint probability function. The likelihood distribution is created by repeating this process for each parameter, K. For a given set of errors—such as those obtained once this processes has been repeated—the likelihood function is defined as being equal to the joint probability:

$$L(K/x) = L(K\ /\ \varepsilon) = \frac{1}{(2\pi)^{n/2}\sigma_1\sigma_2 \dots \sigma_n}\exp\left(-\sum_{i=1}^{n}\left(\frac{\varepsilon_i}{2\sigma_i^2}\right)^2\right), \tag{3.4}$$

The resulting likelihood distribution is not a probability density function, but can converted to one by standardizing the likelihood function so that the area under the curve is equal to one. The parameter that gives the best model results (minimizes model error) is equal to the maximum value of the likelihood distribution. This most likely parameter can also be obtained without going through the process of calculating the joint probability function for each K, but can be calculated using maximization criteria of the joint probability function. The maximum, most likely point on the distribution can be calculated by taking the derivative of the distribution. For normally distrusted errors, the maximum of the function is defined as:

$$M = \sum_{i=1}^{n} \frac{\varepsilon_i}{\sigma_i^2} \tag{3.5}$$

The distribution of the likelihood function is dictated by the distribution of the errors. While the examples used here assume normally distributed model errors, there are other likelihood functions defined for other distribution of errors (such as double exponential or power exponential).

The likelihood function differs conceptually from MC and FOA. For both MC and FOA methods, the distribution of the input parameters is used to obtain the distribution of the model output. With the likelihood function, the outcome is not obtaining the distribution of the model output, but rather the distribution for likely parameters. However, the model output distributions obtained using MC and FOA methods are due to differences in model parameters, so the approaches used by the likelihood function and that used by MC and FOA both give a distribution which shows the mean and variance in model parameter error.

Bayesian Framework: The Bayesian framework for evaluating model performance is based on Bayes theorem for discrete variables (Edwards and Haan 1989, Wilson and Haan 1991, Freer et al. 1996). Conceptually, in evaluating model performance, the Bayesian framework is very similar to the likelihood function. With this method, the results of the likelihood function are combined with previous knowledge of the distribution of parameters to obtain a new probability distribution of the likelihood of model parameters that minimize the error between observed and model predicted data.

The Bayes theorem states that for a given set of errors, the conditional probability of K (parameters) can be obtained by:

$$f(K/x) = \frac{f(x/K)F(K)}{f(x)} \tag{4.1}$$

Where:

- $f(x/K)$ is the likelihood function (described in detail in the previous section), which gives the joint probabilities of error terms for each parameter, K. The likelihood function in the Bayes theorem framework incorporates information on parameter values with site-specific data.
- $f(K)$ is the prior distribution, which incorporates prior information on parameter values using the modeler's previous experiences, physical insight and/or calibrations. This distribution is unrelated to what parameter range is used in the current analysis.
- $f(K/x)$ is the resulting distribution from the Bayes theorem, referred to as the posterior distribution.
- $f(x)$ is not a function of the parameter values, but is used to normalize the function so that the integration of the pdf is equal to 1.

The Bayesian framework is generally considered as superior to the likelihood function because it improves the estimate of parameters by also including prior information.

In order to analyze model performance using the Bayesian framework, the user needs an estimate of the prior distribution for model parameters, and the likelihood function for model errors for the current project. The prior distribution can be obtained as the product of previously known pdfs, which for uncorrelated, normal errors is of a similar form to the joint probabilities in equation 3.4:

$$f(K_0, K_1, \ldots K_p) = \frac{1}{(2\pi)^{n/2}\sigma_{K_1}\sigma_{K_2}\ldots\sigma_{K_p}} \exp\left(-\sum_{i=1}^{p}\left(\frac{K_i - \kappa_i}{2\sigma_{K_i}^2}\right)^2\right), \tag{4.2}$$

Likewise, for uncorrelated, normal errors, the likelihood function can be defined as in equation 3.4. However, more complicated prior distribution functions and likelihood functions need to be used for the case of correlated parameters, or for models that have multiple objectives, where more than one independent variables is used in estimating statistical parameters.

Generalized Likelihood Uncertainty Estimation (GLUE): The generalized likelihood uncertainty estimation (GLUE) methodology proposed by Beven and Binley (1992) is often recommended in the literature as a way to estimate uncertainty in model predictions (i.e. Yang et al 2008). Despite having utilizing the term "likelihood" in its name, it differs from the likelihood functions described earlier. Here, "likelihood" is used in a broad sense, and is specified as any measure of GOF that can be used to compare observed and predicted data. Despite being recommended as a means to estimate model error, its methodology has also been criticized. Mantovan and Todini (2004) offer criticism of the GLUE methodology and do not recommend it as a method for model evaluation. .

Model Structural/Algorithm Uncertainty

Uncertainty in model structure/algorithms occurs when the processes, assumptions, and algorithms used by the model fail to represent reality (Haan 1989, Refsgaard et al. 2006). Unlike the error introduced from model inputs, the uncertainty in model structure is not considered random, as a problem in math should consistently over or under-estimate output. Model structural error is harder to asses than model parameter error because it is non-random.

GOF indicators adjusted to take into account uncertainties

Though there has been much work on ways to quantify uncertainty in model predictions and observed data, there are fewer examples of methods that utilize these uncertainties in the GOF measures. Those approaches that have been presented in the literature rely on modifying the residual term $(O_i - P_i)$ and using the modified term in its place in the GOF equation of choice.

Measurement Uncertainty

Using the estimates of measurement uncertainty from Harmel et al. (2006), Harmel and Smith (2007) modified the residual $(O_i - P_i)$ used in GOF pairwise comparison methods so that the goodness of fit between model and observed data would take into account the measurement uncertainty of the observed data. In their method, a new modified residual takes the place of $O_i - P_i$ in GOF calculations based on two different procedures depending on whether or not the user knows the mean, variance, and shape of the observed distribution.

Assuming the distribution is unknown, the modified residual was obtained by using the expected range in error reported in Harmel et al. (2006) to create upper and lower boundaries for the observed data:

$$UO_i = O_i \pm \frac{PER \times O_i}{100} \tag{5.1}$$

Where the upper limit is the observed value (O_i) plus the product of the probable error range (PER, as a percent) and the observed data, while the lower limit is the observed value minus this product.

The modified residual is then based on whether or not the predicted value lies within these upper and lower limits for the observed data. If it does, the residual term becomes zero, indicating a perfect fit. If P_i was greater than the upper limit, the residual term would become equal to the difference between the upper limit and the predicted term; if P_i were less than the lower limit, the residual would become equal to the difference between the lower limit and P_i:

$$modified\ residual = \begin{cases} 0 & \text{If } UO_i(lower) \leq P_i \leq UO_i(upper) \\ UO_i(lower) - P_i & \text{If } P_i < UO_i(lower) \\ UO_i(upper) - P_i & \text{If } P_i > UO\_i(upper) \end{cases} \tag{5.2}$$

In the case where the user knows probability distributions for their measured data, the distribution information is used instead of the probable error range. In this case, the residual term $(O_i - P_i)$ is modified based on a correction factor (CF). The correction factor is the probability that P_i lies within the range of observed values. Information about the shape of the distribution is required for this calculation. The modified residual becomes:

$$modified\,residual = \frac{CF}{0.5} \times (O_i - P_i) \tag{5.3}$$

Measurement and Parameter Uncertainty

The method presented by Harmel and Smith (2007) only accounts for measurement error in the observed data in the GOF calculation. Following the work of Haan et al (1995), Harmel et al (2010) suggested a method by which to account for uncertainty in both the predicted and observed data in calculating goodness of fit. Similar to the work of Harmel and Smith (2007), in this method, a correction factor is applied to the residual $(O_i - P_i)$ to create a modified residual, and the modified residual would take the place of the usual residual term $(O_i - P_i)$ in GOF equations. In this approach, the modified residual is calculated as:

$$modified\,residual = CF \times (O_i - P_i) \tag{6.1}$$

The correction factor in this method is calculated based on the amount of overlap (DO_i) between the probability density functions (pdfs) of the predicted and observed data:

$$CF = 1 - DO_i \tag{6.2}$$

The degree of overlap between the observed and predicted pdfs is represented by their joint probability. Assuming the observed and predicted data are independent, DO_i can be calculated as the product of the probability that the observed data point lies within the range of the predicted pdfs max and minimum values, and the probability that the predicted data point lies within the range of the observed pdfs maximum and minimum values:

$$DO_i = \left[prob\left(O_i < P_{i,max}\right) - prob\left(O_i < P_{i,min}\right)\right]\left[prob\left(P_i < O_{i,max}\right) - prob\left(P_i < O_{i,min}\right)\right] \tag{6.3}$$

The methodology proposed by Harmel et al (2010) requires detailed knowledge of the statistical characteristics of the predicted and observed data, including the mean, variance, and shape of the resulting distributions.

New Framework and Method for Adjusting GOF

In the work presented in this dissertation, a new framework was developed that allows the observed and predicted data to be described in terms of the uncertainty due to measurement and model parameter error. This framework was used to adjust the MSE and NMSE to account for random errors in the observed and predicted data, using expected values of these random variables. Unlike the approach used by Harmel et al. (2010), the adjustment presented here considers the errors themselves in the definitions of the observed and predicted data. The theoretical framework and method are presented in chapter 4, while the method's accuracy is examined in chapter 5.

PART II: Modeling water quality effects of variable-rate nitrogen fertilizer application using DRAINMOD-NII

Chapter 2. Effects of Fertilizer Timing and Variable Rate N on Nitrate-N Losses from Tile Drained Corn and Soybean Rotation Simulated Using DRAINMOD-NII

Synopsis

Nitrogen from farm fields is a source of pollution to fresh and marine waters, and is a major player in the creation of the hypoxic zone in the Gulf of Mexico. Nutrient management practices that modify fertilizer application rates and timing to better match plant requirements have been proposed as a way to reduce nitrate nitrogen losses from farm fields and make improvements to surface water quality. In this study, the field-scale hydrologic and nitrogen simulation model DRAINMOD-NII was used to predict nitrate loads over a 14-year period (2003-2016) for different fertilizer application rates and timing to corn at Waseca in Southern Minnesota, USA. The fertilizer practices simulated included a single application in the spring before planting, a split application with half applied pre-plant and half at approximately the V6 stage for corn, and split and variable rate N practice (VRN) which utilized the split timing and a lower rate based on in-season monitoring of plant N requirements. Field trials in 2016 and 2017 of VRN application at the site were used to inform the nitrogen rates used for the VRN scenario simulated, and to validate model performance in simulating daily nitrate concentration and load for sites receiving VRN application. Nitrate concentration and load data measured for the 2016 and 2017 growing season showed high variability; predicted nitrate concentrations for this period were a reasonable approximation of measured data, but the predicted load showed larger discrepancies from the observed data. The results of this model simulation for the 14-year period showed an average 23% decrease in N load for the split application strategy compared to a single pre-plant application in the spring. Nitrate loads were further reduced when the change in timing was coupled with a decrease in N application (N applied based on plant needs), which resulted in a reduction of 46%.

Introduction

Nitrogen (N) originating from annual corn and soybean cropping systems in the Upper Mississippi river basin is a leading cause of impairment to both fresh and marine water systems (David et al. 2010, Schulte et al. 2006). Its role as a major contributor to one of the largest hypoxic zones in the world in the Gulf of Mexico (CENR 2010), has drawn intense scrutiny of N losses from agricultural lands in the Upper Midwest. Nitrogen movement from agricultural fields to waters predominately occurs in the highly water-soluble nitrate-N form, and its rapid movement to surface waters in the Upper Midwest is facilitated by the dense network of subsurface drainage systems in the region (Dinnes et al. 2002). It is estimated that nitrate-N entering the Mississippi River from upstream farm fields needs to decrease by at least 45% in order to reduce the extent of the hypoxic zone in the Gulf of Mexico (US EPA 2007).

The amount of nitrate-N moving from fields to waters is highly influenced by factors such as annual precipitation, mineralization of soil organic matter, and fertilizer management practices (Randall and Goss 2008). Nitrate loads exhibit a seasonal response, with the greatest amounts occurring during periods of little or no crop growth, when plant evapotranspiration and uptake of nitrogen are low (Randall et al. 2003). In the northern part of the Midwestern region where soils historically remain frozen through winter, the highest nitrate losses tend to occur during the months of April-June. In a 15-year study in Minnesota, this 3-month period accounted for 71% of the annual drainage volume and 73% of the annual nitrate loss from a corn-soybean rotation (Randall 2004).

Typical fertilizer management practices in the region may exacerbate potential N losses, placing excess N in the soil during the time of greatest loss potential. Nearly all fertilizer application in southern Minnesota occurs in fall before planting (MPCA 2013). Additionally, fields in the upper Midwest can display significant within-field variability in soil organic matter and nitrogen content (Mamo et al. 2003; Scharf et al. 2005), yet more than two-thirds receive blanket N application (Erickson and Widmar 2015), which can result in higher losses of N (Power et al. 2000). Nitrogen management practices that focus on applying the right rate of N at the right time, such as split and variable rate nitrogen (VRN), have been proposed as a practice could reduce N loads to surface waters (i.e. Nelson 1985).

Split application of nitrogen fertilizer addresses the asynchrony in timing between nitrogen application and plant uptake by applying part of the total N amount at or before planting, and part as side-dress during the growing season. This approach delays full application of fertilizer until corn nitrogen uptake rates peak around growth stage V6-V8. An application strategy that utilizes this split timing, along with VRN addresses both the temporal and spatial mismatches between N application and crop development; VRN utilizes the same timing application strategy as the split application, but bases N fertilizer application rates on spatial variations in plant N requirements. VRN utilizes methods of estimating soil or plant N content to determine how much N should be applied at side-dress, potentially reducing the overall N applied. Side dressing rates can be variable based on crop sensing (Mulla 2013) or in-season modeling of plant available N.

There have been few studies on the effect of both timing and rate of fertilizer application on water quality. Here, we use the field-scale hydrologic simulation model DRAINMOD-NII to simulate the effect of fertilizer application rate and timing on nitrate loads from subsurface drainage over a 14-year period for rain-fed corn and soybean, at the Waseca Experiment Station, located in southern Minnesota. Field trials of split-VRN application at the site in 2016 and 2017were used to validate model's performance for this practice, and inform the VRN application rates used in model simulation. Fertilizer rate and timing treatments simulated for the 14-year period included: 1) single application pre-plant (PP) using N rates based on Minnesota Extension recommendations; 2) split application of that N amount, with 50% applied pre-plant, and 50% applied as side-dress; and 3) a split-variable rate nitrogen (VRN) application, where the application was split 50/50 PP and side-dress, as well as using a reduced nitrogen rate based on optical sensor measurements of N need.

Methods

Experimental Site Characteristics

The study site is located in southern Minnesota at the University of Minnesota Southern Research and Outreach Center, at Waseca, Minnesota. The drainage experiment site at Waseca consists of nine plots ranging in size from 0.8 to 2.4ha. Plots 1-5 have drainage intensities of 13 mm d^{-1} (conventional or "low" intensity), while plots 6-9 have a higher drainage rate of 51 mm d^{-1} (high intensity) (fig. 2). The different drainage intensities were achieved using different combinations of drain depth and spacing, and were calculated using the Hooghoudt equation. For drains with a 1.2 m depth, drain spacing's are 12 and 24 m for the 1.2 m depth, while for a 0.9 m depth spacing's are 9 and 18 m for the 0.9 m depth. Each treatment of similar drain depth/spacing/drainage intensity has 1 or 2 replicate plots. More detail on the drainage design can be found in Sands et al. 2008.

Soils and topography found at the site are typical for south-central Minnesota. The major soil types at the site include Webster silty clay loam (fine-loamy, mixed, superactive, mesic Typic Endoaquolls) and Nicollet clay loam (fine-loamy, mixed, superactive, mesic Aquic Hapludolls). Average annual precipitation is approximately 889 mm; precipitation during the growing season (May through September) is approximately 533 mm (averages are the 30-year normal, 1982-2010). Plots were outfitted with tipping buckets to provide a continuous measurement of subsurface drainage flow rates for the years 2003-2008. Nitrogen concentrations were recorded on an event basis for the years 2003 to 2008. Nitrogen load data for the years 2003-2008 set were estimated using the Army Corps of Engineers model, FLUX, (Walker 1996) using the daily flow and sampled nitrogen concentration measurements. In 2016 and 2017, nitrate concentration was recorded as a grab sample on a weekly basis. In 2017, instantaneous flow measurements collected as grab samples at the time of concentration sampling were used to generate an estimate of the daily load.

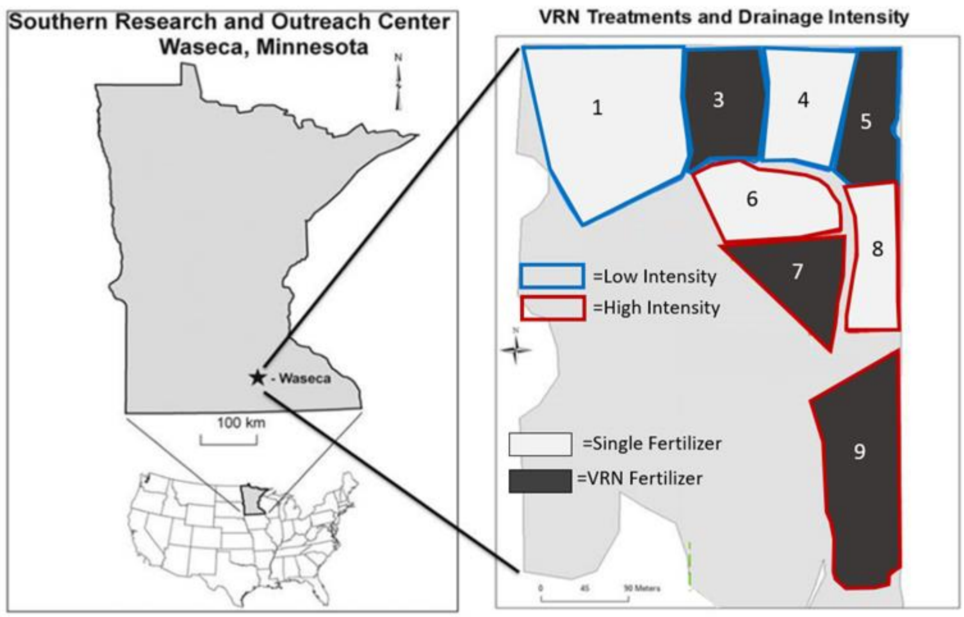

Figure 2: Waseca study location, drainage intensities and nitrogen treatments.

Field Management

Corn (*Zea mays* L.) and soybean (*Glycine max* L.) were grown in rotation for four years, starting with soybean in 2003, with corn following corn for the fifth year (a soy-corn-soy-corn-corn rotation). For the period 2003-2008, anhydrous ammonia was applied for corn pre-plant in April at a rate of 142 kg N/ha in 2004, 134 kg N/ha in 2006, and 185 kg N/ha in 2007. No spring nitrogen fertilizer was applied during the soybean years. An application of NPK fertilizer (nitrogen, phosphorus and potassium) was applied following soybean harvest in the fall, at a rate of 66/168/224, 0/0/224, and 27/112/134 kg/ha in 2003, 2005, and 2008 respectively.

In the 2016 growing season, the plots were divided into single pre-plant (PP) and VRN fertilizer treatments. The experimental design was completely randomized with two replicates within each drainage intensity. In 2016, the single pre-plant nitrogen treatment consisted of applying the economically optimum nitrogen rate (EONR of 151 kg/ha based on Minnesota Extension Services recommendations) to maize plots 1, 4, 6 and 8 one day before planting, while the in-season variable rate treatment consisted of a rate of 50 kg/ha of nitrogen applied to Subfields 3, 5, 7 and 9 before planting, and the remaining 46 to 56 kg/ha applied at growth stage V6. In 2017, fertilizer was applied at a slightly higher rate to second year maize plots for the single PP application at a rate of 180 kg N/ha, while the VRN application received a slightly lower amount, with 60 kg N/ha applied at planting, and an additional 80 kg N/ha applied at the V6 stage. Pre-plant fertilizer application occurred on April 30 in 2016, and May 11 in 2017; while the split-application was applied June 20 in both 2016 and 2017. The VRN side-dress application rate was derived from CropCircle Green NDVI sensing at V6 based on two reference zones, each with nitrogen response plots used for confirming the EONR.

DRAINMOD Simulations

Model Description

The hydrologic model DRAINMOD-NII was used to simulate drainage, nitrate concentration and loads, and soil nitrogen processes for the drainage plots. DRAINMOD is a process-based, distributed, field-scale model, developed to describe the hydrology of poorly or artificially drained lands (Skaggs et al. 2012). It conducts water balances on hourly and daily time scales, and predicts hydrologic parameters including infiltration, runoff, evapotranspiration, seepage, water table depth, and subsurface drainage on a daily, monthly, and annual time step. The companion model, DRAINMOD N-II predicts nitrogen transport and transformation processes in the vadose zone, predicting processes such as mineralization, immobilization, nitrification, as well as nitrogen flows out of the system in the form of plant uptake, denitrification, volatilization, and losses from surface runoff and subsurface drainage (Youssef et al. 2005).

Initial Model Parameterization

Initial input parameters describing the soil water and chemical properties of the Waseca plots were based on values determined by Luo et al. (2010). Soil samples taken from the site were used to determine model inputs including: soil particle distributions, soil water characteristic curves, and saturated hydraulic conductivity. Hydraulic conductivity showed great variability in both the lab measured samples and county soil survey data (Luo et al., 2010) and was used as a calibration parameter. Other parameters calibrated included: maximum surface storage (1 cm), Kirkham’s depth for flow to drains (0.5 cm), actual distance to impermeable layer (200 cm), initial depth to water table (150 cm), critical ice content above which infiltration stops (0.1 cm^3/cm^3), and the snow melt coefficient (4 mm/dd °C). Soil organic matter (4.7-6.2%) was estimated in each plot using kriging interpolation based on soil sampling data from a 0.4 ha grid taken in the spring of 2016 before planting. Calibration was preformed manually. Climatic data measured at the Waseca experiment station site for the years 2003-2016 was used as model input, including: daily maximum and minimum temperatures, and hourly precipitation. Simulations were run with a 5-year soybean/corn rotation starting with soybean in 2003 and alternating with corn for 4 years, with the 5th year as corn following corn

Model Calibration for years 2003-2008

DRAINMOD was calibrated for subsurface drainage flow and nitrate load for the years 2003-2008, corresponding to period of continuous observed drainage and nitrate load measurements. Model calibration was done for one plot within each depth/spacing treatment (plots 3, 4, 6, and 7), and validated for the remaining plots (1, 5, 8, and 9). Ammonium fertilizer was applied only during corn years, 7 days before planting, and incorporated at a depth of 15 cm. For model calibration, fertilizer was applied at a rate of 135 kg N/ha. Model performance was evaluated and by calculating the monthly Nash-Sutcliffe Coefficient of Efficiency (NSE; Moriasi et al. 2007, defined as:

$$NSCE = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \qquad (1)$$

Where O_i is the observed data for a specified time step, i, P_i is the corresponding predicted data, N is the number of observations, and $\bar{O}$ is the mean value of the observed data.

Model Simulations for 2016 and 2017

Following calibration, the model was run for a 5-year period (2013-2017) to compare measured data to DRAINMOD-predicted nitrate concentration and load for plots treated with VRN fertilizer application vs. the single spring pre-plant application. The 5-year simulation utilized a soy-corn-soy-corn-corn rotation. For all plots, fertilizer was applied during the corn years only, and for the first corn year, fertilizer was applied as a single spring application at a rate of 150 kg N/ha, similar to field practices.

DRAINMOD-NII does not allow for different fertilizer rates within the field, and so fields simulated using the model are treated as homogeneous with regard to soils and management practices. To account for spatial variability in fertilizer application in model simulation, the simulated VRN scenario used the average application rate from the 2016 and 2017 field work. Plots 1, 4, 6, and 8 were simulated using the single PP fertilizer management, with a single application of ammonium N-fertilizer at a rate of 150 kg N/ha, 7 days before planting. Plots 3, 5, 7, and 9 were used to simulate the VRN fertilizer application. The VRN fertilizer management scheme used an application rate of 100 kg N/ha, with half (50 kg N/ha) applied 7 days before planting, with the remaining half (50 kg N/ha) surface applied 60 days after planting, as had been applied during the 2016 growing season.

The model simulated results for nitrate concentration (2016 and 2017) and load (for 2017 only) were compared against the observed data measured in 2016 and 2017. The observed concentration data were collected as grab samples, which were assumed to be representative of daily conditions. The daily load values for 2017 were calculated assuming the instantaneous flow rates measured were representative of daily conditions. These daily observed values were compared against the average of the simulated data for a seven-day period around the measurement date; the average of the predicted value for the day that the observed data were collected and all values predicted within the window of plus and minus three days of that.

Model Simulation of Improved N Management Practices

DRAINMOD-NII was then run over a 14-year period (2003-2016) for the 5-year soy-corn-soy-corn-corn rotation using three N treatments for every year corn was grown. The three N treatments included: 1) a single pre-plant application in the spring, 2) split-VRN application, and 3) split application. Plots 1, 4, 6, and 8 were simulated using the single PP fertilizer management, with a single application of ammonium N-fertilizer at a rate of 150 kg N/ha, 7 days before planting. Plots used in the model simulation of the alternative-N practices were based on those that received the VRN fertilizer in the 2016 and 2017 field experiments, and included plots 3, 5, 7, and 9. The VRN fertilizer management scheme used an application rate of 100 kg N/ha, with half (50kg N/ha) applied 7 days before planting, with the remaining half (50 kg N/ha) surface applied 60 days after planting. The split application treatment was modeled using the same plots as the VRN treatment (plots 3, 5, 7, and 9), and same application rate as the single PP (150 kg N/ha), with half of the fertilizer (75 kg N/ha) applied 7 days before planting, with the remaining 75 kg N/ha applied 60 days after planting.

Results and Discussion

Results of Model calibration and subsurface drainage

Model performance for the calibration dataset (years 2003-2008), is presented in Table 1 as NSE values based on comparing observed and predicted monthly drainage or nitrate load. Generally, a monthly NSE value of 0.5 or greater is considered acceptable while a value of 0.7 or greater is considered good (Moriasi et al. 2007; Skaggs et al. 2012). Most NSE values are acceptable to good, with the exception being Subfield five. The poor results for Subfield five are likely due to instrument malfunction, resulting in no data collection for the year 2004. Unusually low flow for the year 2008 also indicates possible instrument malfunction that year as well. (fig. 3b) Nitrate load has slightly lower NSE values (but still acceptable) than for drainage.

Table 1: Summary of monthly NSE values for subsurface drainage and nitrate load for Subfields (plots) during calibration and validation (years 2003-2008). Subfields 3, 4, 6, and 7 were used for calibration, while Subfields 1, 5, 8, and 9 were used to validate model performance.

	Monthly NSE Value							
	Plot 1	Plot 3	Plot 4	Plot 5	Plot 6	Plot 7	Plot8	Plot 9
Drainage	0.35	0.56	0.49	-0.65	0.5	0.5	0.66	0.72
Nitrate Load	0.4	0.54	0.55	-1.7	0.6	0.4	0.35	0.7

Figures 3 and 4 show the monthly subsurface drainage predicted by DRAINMOD as well as the observed values for each plot. DRAINMOD predictions are shown with the triangle symbol, while the observed data utilizes a circle. Since DRAINMOD subsurface drainage was calibrated for average hydrologic site conditions for each drainage treatment, model predictions are the same for replicate plots within a treatment. Model predictions match the observed data well for most years, with the exception of the year 2006, where model predictions consistently underestimate flow for all plots.

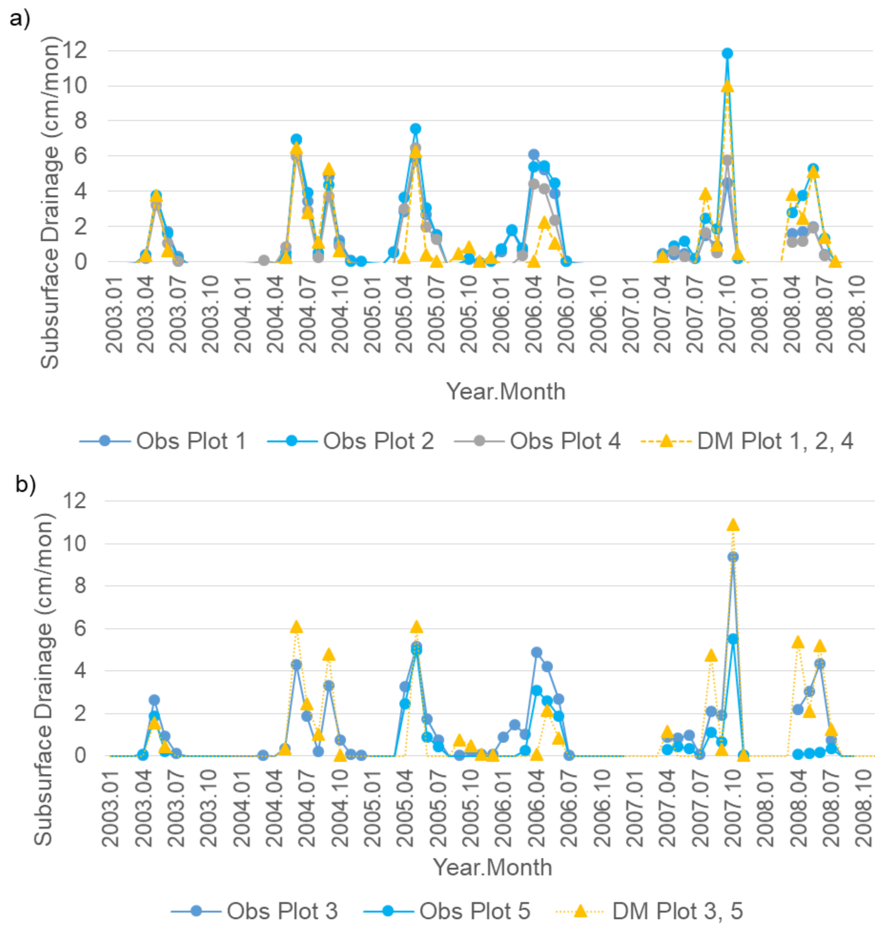

Figure 3: Observed and predicted monthly subsurface drainage for plots with 1.3cm/d drainage intensity. Replicate plots for treatment 1 (drainage depth = 120 cm, spacing = 2400 cm) are shown in figure 3a, and for treatment 2 (drainage depth = 90cm, spacing 1800 cm) in figure 3b.

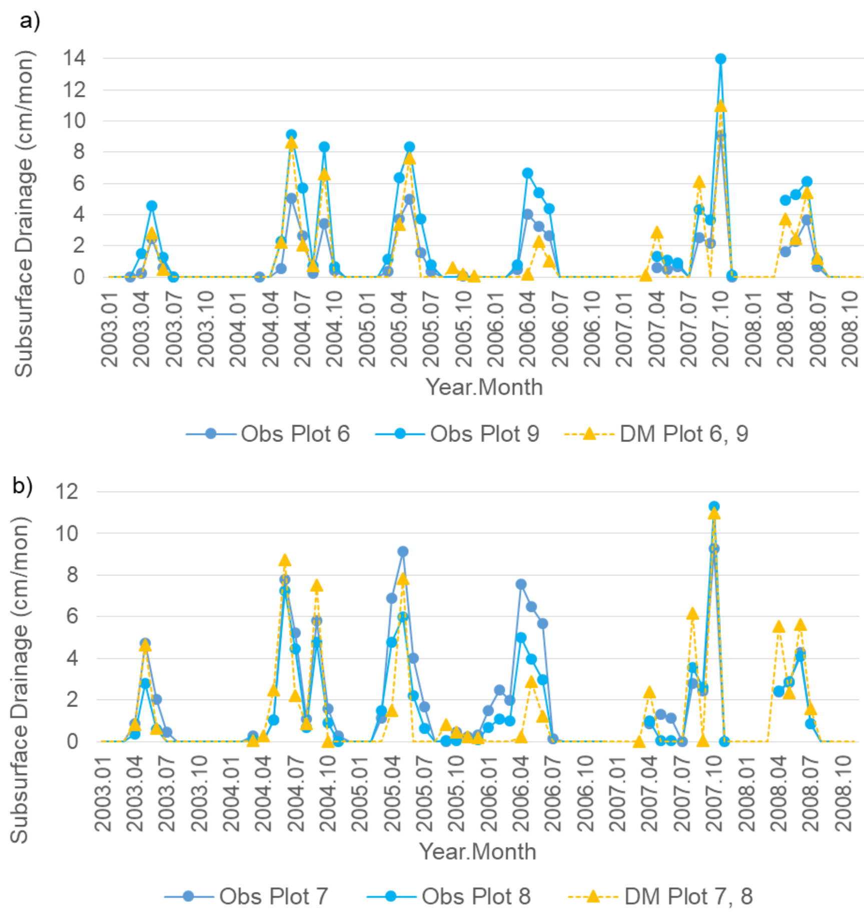

Figure 4: Observed and predicted monthly subsurface drainage for plots with 1.3cm/d drainage intensity. Replicate plots for treatment 3 (drainage depth = 90 cm, spacing = 900 cm) are shown in figure 4a, and for treatment 4 (drainage depth =120 cm, spacing = 1200 cm) in figure 4b.

Comparison of Observed and Predicted Nitrate Concentrations for 2016 and 2017

The average of the observed and predicted nitrate concentrations for Subfields treated with single uniform and VRN side-dress fertilizer applications in 2016 and 2017 are summarized in figure 5. For both the 2016 and 2017 growing seasons, the measured concentrations were slightly higher than those simulated by DRAINMOD. In 2016, the average daily nitrate concentration for the observed and simulated data in the single treatment Subfields equaled 13.2 and 12.1 mg/L, respectively. For the VRN plots, the observed and simulated concentrations for the observed and simulated data were 11.5 and 7.7 mg/L, respectively. In 2017, the nitrate concentration for the observed and simulated data in the single treatment was 11.1 and 7.6 mg/L, while the VRN treatments were 9.1 and 6.4 mg/L respectively.

In 2016, the simulated VRN Subfields showed on average a much greater percent change in nitrate concentration than the single application Subfields, with a 36% percent reduction predicted by the model, while only a 13% reduction was measured in the field (fig. 5a). In 2017, the simulated VRN Subfields showed a smaller reduction in nitrate concentration than the observed value, with a 10% reduction predicted by the model, while an 18% reduction was measured in the field (fig. 5b).

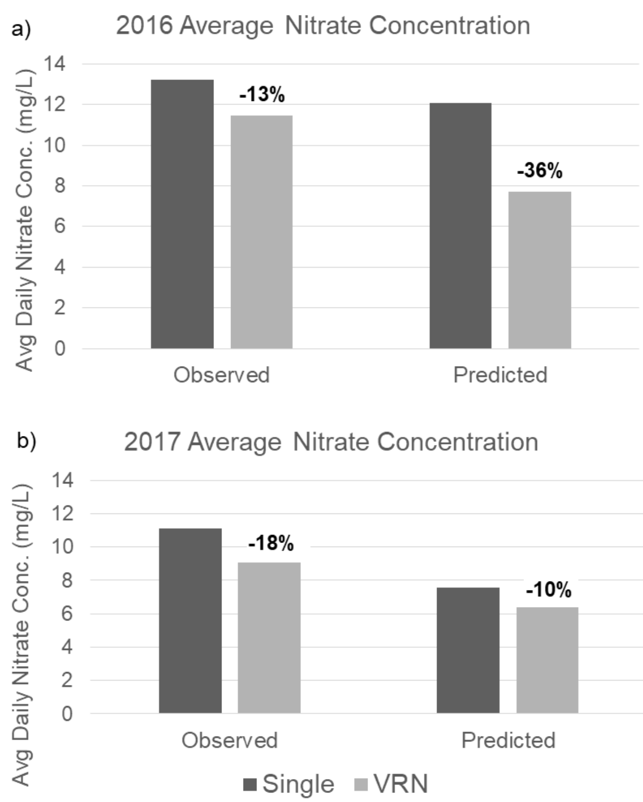

Figure 5: Observed and simulated nitrate concentration for the years 2016 (fig. 5a) and 2017 (fig. 5b). The observed data are an average of the individual grab samples taken in each Subfield and the predicted data are the 7-day average of model predictions around the measurement date.

Though modeled and simulated nitrate concentration values are not identical, the model results show reasonable approximation of field measurements. The observed data were taken as individual grab samples, and may not be representative of conditions for the entire day. If stormflow samples had been collected, observed values would likely have been decreased, since nitrate concentrations generally decrease during stormflow. In contrast, predicted data are simulated as the average daily value—which could explain the lower concentration values in the predicted data.

Nitrate Load for 2017

The average observed daily nitrate load is larger than the predicted nitrate load for 2017 (fig. 6a). For the single application plots, the average of the observed daily values was 0.11 kg N/ha, while for the predicted data it was 0.05 kg N/ha. For the VRN plots, the average of the observed daily values was 0.15 kg N/ha; for the predicted data it was 0.04 kg N/ha. Besides showing larger daily nitrate loads, the trend in the observed data for the single application vs. VRN is opposite of the trend seen in the measured data. The observed nitrate load for the daily values shows that plots using VRN fertilizer have a 32% increase in nitrate load compared to the single application plots. This differs from the model results, which indicate an 11% reduction for the VRN treatment corresponding to the days of field observation (fig. 6a).

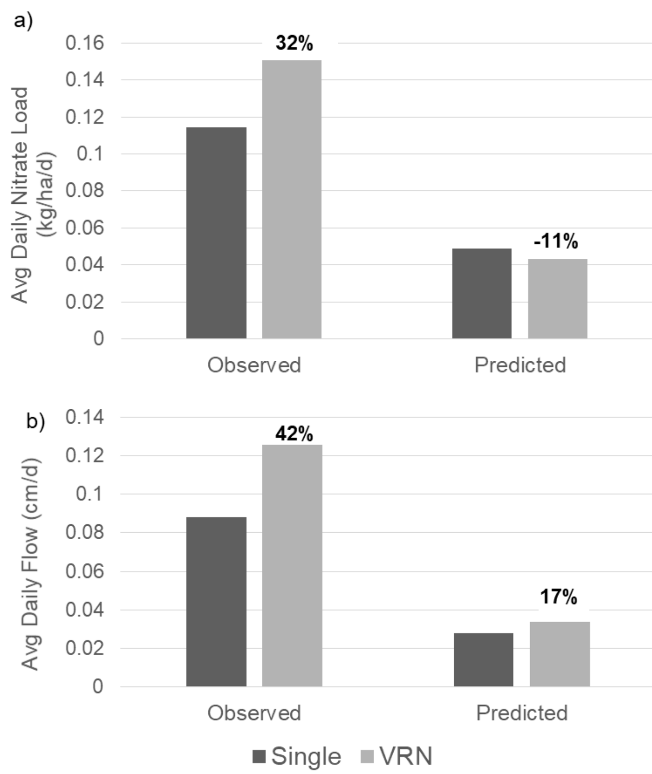

Figure 6: Observed and simulated nitrate load (a) and subsurface drainage (b) for 2017. Nitrate load and subsurface flow are shown as an average daily value for all plots in each treatment.

For the observed data, the plots receiving VRN fertilizer application show higher nitrate load than the single application plots due to higher subsurface flow. Nitrate load is a function of subsurface flow and nitrate concentration (Randall and Goss 2008). From figure 5b, the observed nitrate concentration for the plots receiving VRN fertilizer application was less than that receiving the single application. However, the observed average daily subsurface drainage for the plots receiving VRN fertilizer application is 43% greater than those receiving the single fertilizer application, with a daily flow rate of 0.125 cm/day for the VRN plots compared to 0.088 cm/day for the plots receiving the single fertilizer application (fig. 6b). The predicted subsurface flow is also higher for the plots receiving the single fertilizer application, though the model predicts less flow for these days overall compared to the observed data.

Comparison of nitrate loads for single, split, and VRN management

For the 14 years of model simulations using one of three fertilizer application treatments for all corn years, nitrate loads from subsurface drainage were reduced every year of model simulation for both the split and VRN applications compared to the single PP. The annual nitrate load was generally related to precipitation, with higher loads generally predicted for all strategies during wetter years than dry years (fig. 7).

The results of model simulation over the full 14-year period (2003-2016) showed an average annual reduction in nitrate load for both alternative N practices compared to single uniform PP application of N. The average nitrate load for Subfields simulated with the split application had a 23% reduction in nitrate-N-load compared to those in the single PP treatment, while the VRN treatment had a 46% reduction in nitrate-N load compared to the PP treatment (fig. 7b).

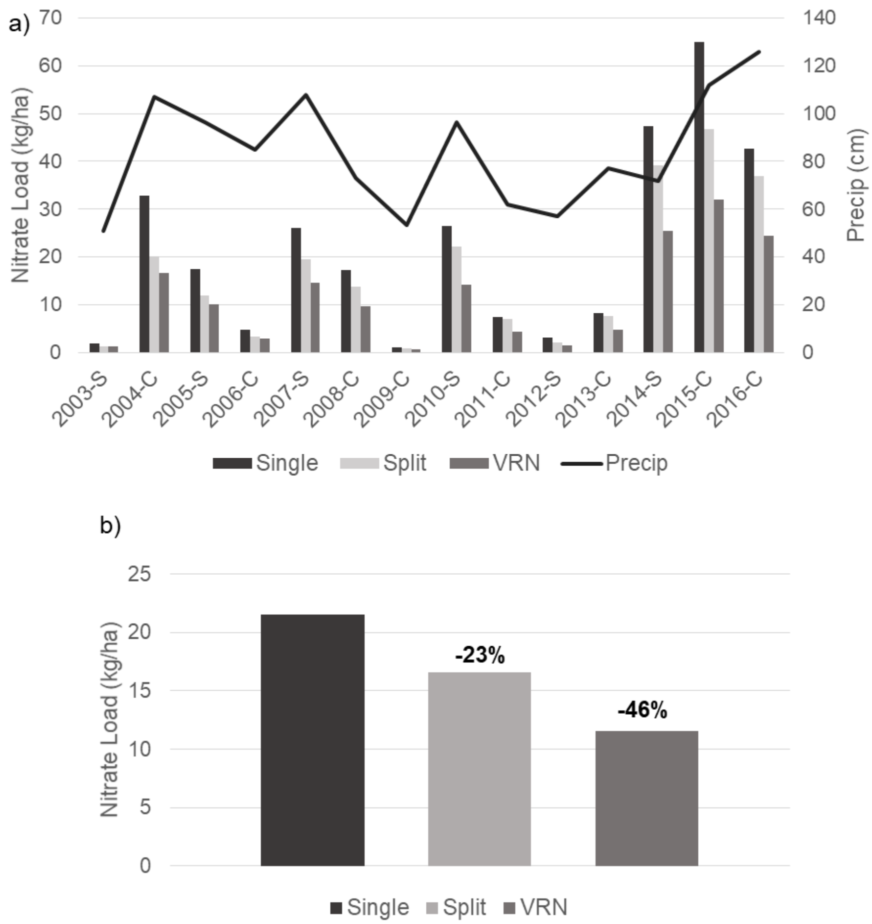

Figure 7: Simulated nitrate load in subsurface drainage for single pre-plant, split, and VRN for: a) each year of model simulation, and b) averaged over the 14-year period.

The reduction of 23% for the split application found here is similar to 25% reduction in N leaching loss reported by Bakhsh et al. (2002) for a split application of N to corn compared to single spring application for a field-study in Iowa. However, studies have not consistently shown that split application reduces nitrate-N load compared to single application in the spring. At Waseca, MN, Randall et al. (2003), found that the 4-year (1990-1993) totals for N lost were higher for a split (40% PP/60% side-dress) compared to a single spring application, with the 4 year totals of N lost equaling 190 kg/ha for the split rate compared to 177 kg/ha for the single pre-plant.

It is difficult to assess the accuracy of the predicted 46% average reduction in N load for the VRN treatment since there are no other studies in this region that both change the rate and timing of N. Many studies have shown reductions in both N load and concentration for reduced N fertilizer rates (Gast et al., 1978; Kladivko et al., 2004), with Jaynes et al. (2001) finding nitrate loads of 29, 35, and 48 kg N/ha for N rates of 57-67, 114-135, and 172-202 kg N/ha (about a 27% reduction in N load for a 33% reduction in fertilizer). In a modeling study using Adapt-N which compared nitrate load between split application, and split with a reduced rate, Sela et al. (2016) found a 36% decrease in N losses corresponding to a 34% decrease in side-dress N application. This difference in N load for split and split plus the reduced rate is similar to the difference here between DRAINMOD predicted N-load for the split and VRN treatments, where a 33% reduction in total N application resulted in a 30% reduction in N load.

The predicted 46% decrease in N load for the VRN treatment was achieved by assuming a continuous 33% reduction in fertilizer rates for all years compared to the single spring application. This reduction in fertilizer application for the VRN treatment was based on in-season monitoring of plant N requirements over two years. It is likely that there was residual soil nitrogen left over from fertilizer applied in the years prior to the VRN field experiment, and this may have allowed the VRN application to be lower during those two years than what could be expected over a longer time frame, such as the 14 years simulated here. Therefore, the 46% reduction for VRN over the 14-year period may be an overestimate, unless fertilizer application rates based on in-season crop N requirements would show a consistent 33% reduction compared to the standard application rate.

Change in N budgets

For all simulation years, DRAINMOD results show an overall reduction in nitrogen added to, lost from, and stored in the soil system for both alternative N management practices compared the single PP uniform N application (fig. 8).

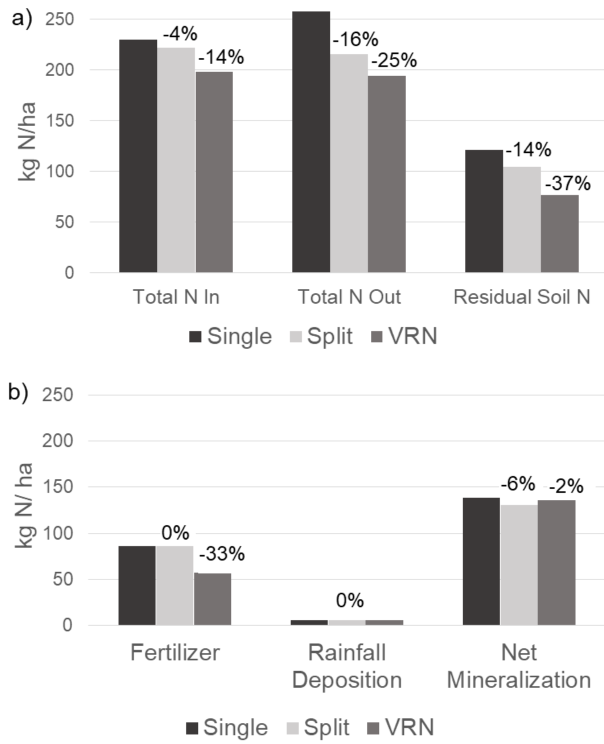

Figure 8: Simulated average annual nitrogen budgets averaged for Subfields in each treatment (single, split, and VRN) over the 14-year period, including a) total soil nitrogen additions, losses, and residual soil N, and b) sources of nitrogen into the system. In b), the sum of the residual soil nitrogen and total nitrogen out of the system does not equal the total nitrogen into the system. This is because the DRAINMOD simulations used an initial nitrate concentration in the top of the soil profile of 5 mg/L.

It is not surprising that there is a reduction in nitrogen inputs to the system for the VRN compared to the single application, since the VRN involves applying 33% less fertilizer than the single PP application (fig. 8b). The split application, however, uses the same amount of fertilizer as the single pre-plant, but still shows an overall reduction of nitrogen inputs. In DRAINMOD-NII, besides fertilizer application, nitrogen can enter into the system as the result of two processes: 1) rainfall deposition (which was small and independent of N fertilizer treatment), and 2) mineralization of soil organic matter. The reduction in nitrogen into the system for the split application simulated by the model is clearly a result of a reduction in mineralization (fig. 8b)—an interesting result as it implies that in the model simulation at least, shifting the timing of fertilizer application reduced the amount of SOM mineralized.

The average annual residual soil nitrogen decreased for both the split and VRN treatments compared to the single PP, but there was a larger reduction in soil residual N for VRN compared to split. This makes sense for the VRN because there is less fertilizer applied. The overall decrease in nitrogen lost from the system is explained both by the reduction of nitrogen inputs, and also by the reduction in soil storage, which reduced the amount of N available in the soil.

Conclusions

Nutrient management practices which modify fertilizer application rates and timing to better match plant requirements have been proposed as a way to reduce nitrate nitrogen losses from farm fields and make improvements to surface water quality. In this study, the field-scale hydrologic and nitrogen simulation model DRAINMOD-NII was used to predict nitrate loads over a 14 year period (2003-2016) for different fertilizer application rates and timing to corn at Waseca in southern Minnesota, USA. Field trials in 2016 and 2017 of VRN application at the Waseca AERF site were used to inform the nitrogen rates used for the VRN scenario simulated, and to validate model performance in simulating daily nitrate concentration for Subfields receiving VRN application. Comparison of DRAINMOD model results for nitrate concentration data measured for the 2016 and 2017 growing season showed good agreement between measured and predicted nitrate concentrations. Alternative nitrogen fertilizer application strategies that change the timing or rate and timing of fertilizer to more closely match plant nitrogen requirements reduced nitrate-N loads from corn and soybean cropping systems. For a 14 year simulation, model results for split side-dress N application showed a 23% average annual reduction in N load compared to a single pre-plant application, while VRN side-dress application showed a 46% reduction. The larger decrease in the average annual nitrate load for the VRN application, which utilized a reduced fertilizer rate in addition to split side-dress fertilizer application timing, indicates that while changing the timing of fertilizer application reduces N load, changing the rate has an important impact as well. Model simulations showed that changes in nitrogen application timing also reduced the amount of mineralized SOM, and reduced residual soil nitrogen.

Chapter 3. Predicted Nitrate-N Loads for Fall, Spring, and VRN Fertilizer Application in Southern Minnesota

Synopsis

Nitrate nitrogen (N) from tile-drained agricultural fields is a source of pollution to fresh and marine waters. Sensor-based technologies that allow for in-season monitoring of crop nitrogen requirements may represent a way to reduce nitrate-N losses by allowing for fertilizer application on a more precise spatial and temporal resolution. However, little research has been done to determine its effectiveness in reducing nitrate-N losses. In this study, the field-scale hydrologic and nitrogen simulation model DRAINMOD-NII was used to estimate nitrate-N loads for fertilizer applied to corn given different application rates and application timing approaches for three sites in southern Minnesota: Waseca, Willmar, and Lamberton. The results of these simulations were used in a regression analysis to develop equations to predict nitrate-N loses for the region more generally as a function of fertilizer timing and application rate. Fertilizer timing treatments used in model simulation included fall, spring, and variable rate nitrogen (VRN) applications. The VRN application involved half of the total fertilizer rate applied in the spring before planting, with the remaining half applied at approximately corn V6 growth stage. The results of the DRAINMOD simulations showed the highest nitrate loads occurred for all fall application rates, with VRN having the lowest predicted N losses. An exponential regression model which used fertilizer application rate and growing season precipitation as the dependent variables had good agreement with DRAINMOD results, with R^2 of 0.46, 0.49, and 0.52 for spring, split-VRN, and fall application, respectively. The regression model showed that the variation in the DRAINMOD predicted nitrate-N load for the different application timing treatments was highly related to annual precipitation and fertilizer application rates. The nitrogen loss predicted by the regression model for a growing season precipitation of 60 cm and fertilizer applied at a rate of 100 kg N/ha was 2.37 kg N/ha and 3.88 kg N/ha less than the fall application for spring and VRN applications, a 16% and 27% reduction, respectively. Increasing the fertilizer application rate to 180 kg N/ha, nitrogen losses were reduced by 5.85 kg/ha and 9.55 kg/ha for the spring and VRN applications compared to fall application, a 23% and 39% reduction, respectively.

Introduction

Nitrogen management practices which focus on applying the right rate at the right time have been proposed to reduce N loads to surface waters for a number of years (i.e. Nelson, 1985). Split-variable rate nitrogen fertilizer (VRN) application addresses both the temporal and spatial mismatches that occur with single uniform pre-plant applications. In this management, part of the total N amount is applied at or before planting, and part as side-dress during the growing season, delaying full application of fertilizer until corn growth stage V6-V8. Side dressing rates are based on spatial variations in plant N requirements, which can be variable based on crop sensing (Mulla 2013) or in-season modeling of plant available N, reducing over-application of fertilizer on areas with high soil nitrogen content.

Though split-VRN fertilizer application has the potential to reduce nitrate-N loads to surface waters from agricultural fields, there has been little research on how it compares to conventional single, uniform application in this regard. In this study, we used the hydrologic and nitrogen simulation model -NII to predict nitrate-N losses for fields managed with a single spring or fall fertilizer application, or split-VRN applications for three sites in southern Minnesota: Waseca, Lamberton, and Willmar. Simulations were done for various nitrogen rate applications and over several years of climate data. The results of these simulations were used to develop regional regression equations that could be used to estimate nitrate-N loss for fall, spring, and VRN application for southern Minnesota.

Materials and Methods

DRAINMOD-NII Model Description

The hydrologic model DRAINMOD-NII was used to simulate subsurface drainage and nitrate losses at three sites in southern Minnesota: Waseca, Willmar, and Lamberton. DRAINMOD is a process-based, distributed, field-scale model, developed to describe the hydrology of poorly or artificially drained lands (Skaggs et al., 2012). It conducts water balances on hourly and daily time scales, and predicts hydrologic parameters including infiltration, runoff, evapotranspiration, seepage, water table depth, and subsurface drainage on a daily, monthly, and annual time step. The companion model, DRAINMOD-NII predicts nitrogen transport and transformation processes in the vadose zone, including processes such as mineralization, immobilization, nitrification, as well as nitrogen flows out of the system in the form of plant uptake, denitrification, volatilization, and losses from surface runoff and subsurface drainage (Youssef et al., 2005).

Model Calibration

DRAINMOD-NII was calibrated for measured subsurface drainage flow and nitrate-N loads at each of the three sites. Drainage flow and nitrate losses over multi-year periods had been measured at Waseca, Willmar and Lamberton as part of other research projects. The measured datasets were used to calibrate the model at each site for subsurface drainage and nitrate losses. Initial DRAINMOD-NII input parameters for simulations of all three sites were based on values reported by Luo et al. (2010), with additional calibration of parameters at each location to ensure good fit between the model-simulated and measured datasets. Calibration was performed manually for all locations.

Model performance was evaluated by calculating the Nash-Sutcliffe Efficiency (NSE) coefficients and percent bias (PBIAS) (Legates and McCabe 1999), which are defined as:

$$NSE = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \tag{1}$$

$$PBIAS = \frac{\sum_{i=1}^{n}(O_i - P_i)}{\sum_{i=1}^{n} O_i} \times 100 \tag{2}$$

Where O_i is the observed (measured) data, $\bar{O}$ is the mean of the observed data, and P_i is the corresponding model-predicted value.

Site Descriptions

Waseca

Subsurface drainage and nitrate-N load data measured at the University of Minnesota Southern Research an Outreach Center, in Waseca, MN, were used in DRAINMOD calibration. The Waseca site consists of nine drainage experimental Subfields ranging in size from 0.8 to 2.4ha. Half of the Subfields have drainage intensities of 13 mm d^{-1} (conventional or "low" intensity), while the other half have a higher drainage rate of 51 mm d^{-1} (high intensity). More detail on the drainage design can be found in Sands et al. (2008). Corn (*Zea mays* L.) and soybean (*Glycine max* L.) were grown in rotation for four years, starting with soybean in 2003, with corn following corn for the fifth year (a soy-corn-soy-corn-corn rotation). The major soil types at the site include Webster silty clay loam (fine-loamy, mixed, superactive, mesic Typic Endoaquolls) and Nicollet clay loam (fine-loamy, mixed, superactive, mesic Aquic Hapludolls).

Average annual precipitation at the Waseca site is approximately 900 mm; precipitation during the growing season (May through September) is approximately 533 mm (averages are the 30-year normal, 1981-2010; NOAA, 2016). Plots were outfitted with tipping buckets to provide a continuous measurement of subsurface drainage flow rates for the years 2003-2008. Nitrogen concentrations were recorded on an event basis for the years 2003 to 2008. Nitrogen load data for the years 2003-2008 were estimated using the Army Corps of Engineers model, FLUX (Walker 1996), using the daily flow and sampled nitrogen concentration measurements.

Willmar

The subsurface drainage and nitrate loads used to calibrate DRAINMOD for Willmar were obtained from a study by Ghane et al. (2016). This work took place at Goran's Discovery Farm, a privately owned farm located just southeast of the town of Willmar, MN. In their work, daily subsurface flow and nitrate load data were collected for the period 2007-2013 from a 50-ha field with subsurface drainage installed at an average depth of 1.2 m with 24 m spacing. Located in this field was also a surface inlet that connected to the subsurface drainage system. The dominant soil type in the field was poorly drained Canisteo (fine-loamy, mixed, calcareous, mesic Typic Endoaquolls), and Harps loam (fine-loamy, mixed, mesic Typic Calciaquolls). Average annual precipitation is approximately 748 mm; precipitation during the growing season is approximately 510 mm. More detailed information on soil measurements, field measurement of drainage flow and nitrate load, and field management can be found in Ghane et al. (2016).

Lamberton

Measured drainage and nitrate loads for the years 1990-1998 were obtained from plot studies at the University of Minnesota Southwest Research and Outreach Center (SROC) in Lamberton, MN and used for DRAINMOD calibration. Plots at this site were planted in a corn-soybean rotation, and had dimensions of 13.7 m by 15.24 m, with subsurface tile drainage placed at a depth of 1.2 m, and spacing of approximately 27 m (personal communication, Jeff Strock, University of MN Extension). Soils at SROC in Lamberton are predominately poorly draining Webster silty clay loam (fine-loamy, mixed, superactive, mesic Typic Endoaquolls), moderately well-drained Normania loam (fine-loamy, mixed, superactive, mesic Aquic Hapludolls), and well-drained Ves loam (fine-loamy, mixed, superactive, mesic Calcic Hapludolls) (Oquist et al., 2006). Average annual precipitation is approximately 709 mm; precipitation during the growing season (May through September) is approximately 497 mm.

Model Calibration

Waseca

DRAINMOD was calibrated for the years 2003-2008, corresponding to period of continuous observed drainage and nitrate load measurements. Model calibration was done for one Subfield within each depth/spacing treatment (Subfields 3, 4, 6, and 7), and validated for the remaining Subfields (1, 5, 8, and 9). For calibration, ammonium fertilizer was applied only during corn years, 7 days before planting, at a rate of 135 kg N/ha. Climatic data measured at the Waseca experiment station site for the years 2003-2016 was used as model input and for calculating daily potential evapotranspiration

(PET) using the Penman-Monteith method. Measured climatic data used as model input included hourly precipitation and daily maximum and minimum temperatures, while daily temperatures, relative humidity, solar radiation, and wind speed were used to calculate PET. Parameters calibrated included: saturated hydraulic conductivity, maximum surface storage, Kirkham's depth for flow to drains, actual distance to impermeable layer, initial depth to water table, critical ice content above which infiltration stops, the snow melt coefficient, and soil organic matter.

Willmar

DRAINMOD-NII was calibrated for the years 2009-2013. The model was calibrated based on the crop rotation described in Ghane et al. (2016), which for the years 2009-2013 was corn-soy-corn-corn-corn. For calibration simulations, fertilizer was applied as ammonium at a rate of 160 kg N/ha in the fall prior to planting corn. Hourly precipitation data from the Willmar airport, the nearest weather station, was used as model input. All other climate data—including daily maximum and minimum temperatures, relative humidity, solar radiation, and wind speed—were obtained from the Litchfield airport, located approximately 48 km east of Willmar. Daily PET was calculated using the Penman-Monteith method. Parameters calibrated included: optimum temperature for denitrification, ET factors for the months of April, May, and June, and saturated hydraulic conductivity.

Lamberton

DRAINMOD-NII was calibrated for the years 1990-1998, for a corn/soybean rotation, starting with soybean in 1990. Fertilizer was applied during corn years as ammonium at a rate of 150 kg N/ha, a rate typical for the region. Climate input data, including hourly precipitation, daily maximum and minimum temperatures, and daily PET, were obtained from Gary Sands' group at the University of Minnesota. Parameters changed during calibration included: saturated hydraulic conductivity, and ET factors for the months of April, May, and June.

DRAINMOD Simulations

Following calibration, DRAINMOD-NII was run for all three locations for a fall, spring, and split-VRN fertilizer application. For Waseca, simulations were only done for two of the overall nine Subfields (# 4 and 6), representing both high and low intensity drainage. These simulations utilized a corn-soybean rotation for all locations.

In order to capture the nitrate-loss response with regard to fertilizer application rate, several fertilizer application rates were used in the simulations. For the spring application simulations, nitrogen fertilizer was applied as ammonium in the spring 7 days before planting corn (occurring in late April), at rates of: 10, 100, 150, 160, and 200 kg N/ha. For the split-VRN application, half the total N fertilizer was applied 7 days before planting, and the remaining half applied at approximately the V6 stage for corn, 60 days after planting (mid-June). The application rates used for the split-VRN applications were: 10, 100, 150, and 200 kg N/ha. For the fall application, fertilizer was applied 120 days after soybean was planted (mid-October), at rates of 10, 100, 150, and 200 kg N/ha.

The rates of 150 and 160 kg N/ha chosen for these simulations are similar to those used by farmers in the region (Bierman et al., 2011). However, the rate of 160 kg N/ha was not used in the split-VRN or fall simulations because the spring simulations showed little difference in the application of 150 compared to 160 kg N/ha. The application rate of 100 kg N/ha was equal to the fertilizer rate applied to split-VRN fields in a 2016 field experiment at Waseca. The upper and lower amounts (10 and 200 kg N/ha) were arbitrarily chosen to ensure a large range of possible nitrogen fertilizer rates. Simulation rate and timing for each site are summarized in table 2.

Table 2: Fertilizer rates and timings used for DRAINMOD-NII simulations. Fertilizer applied in the fall occurred after soybean; in DRAINMOD, this was set as 120 days after soybean was planted.

Location	Fertilizer Application Timing		Total Fertilizer Rate (kg N/ha)	Years Used in Simulation
Waseca (Subfields 4 and 6)	Spring	7 days before planting corn	10, 100, 150, 160, 200	2003-2016
	VRN	½ of total 7 d before planting, ½ of total 60 d after planting corn	10, 100, 150, 200	
	Fall	120 days after planting soybean	10, 100, 150, 200	
Willmar	Spring	7 days before planting corn	10, 100, 150, 160, 200	2008-2013
	VRN	½ of total 7 d before planting, ½ of total 60 d after planting corn	10, 100, 150, 200	
	Fall	120 days after planting soybean	10, 100, 150, 200	
Lamberton	Spring	7 days before planting corn	10, 100, 150, 160, 200	1995-2005
	VRN	½ of total 7 d before planting, ½ of total 60 d after planting corn	10, 100, 150, 200	
	Fall	120 days after planting soybean	10, 100, 150, 200	

Regression Analysis

The results of the DRAINMOD-NII simulations at each of the three study locations were combined in a regression analysis to predict the nitrate loss for spring, fall, and split-VRN fertilizer application. The regression analysis used climate data (including annual and seasonal temperature and precipitation), and fertilizer application rates as potential independent variables.

Results and Discussion

DRAINMOD Calibration

Model performance for predicting subsurface drainage and nitrate losses following calibration are shown in table 3. For drainage flow, model performance is considered satisfactory for a NSE greater than 0.50, or PBIAS between +/- 10 to 15%. For nitrogen loss, satisfactory model performance is for NSE greater than 0.35, and PBIAS between +/- 20 to 30% (Moriasi et al., 2015). Unlike the NSE, PBIAS does not take into account variability in the observed data, and so can bias against datasets that show significant variability. Variability in the observed measurements for drainage at both Lamberton and Willmar, and nitrate load at Lamberton, are likely the cause of low PBIAS values for drainage, though the NSE results are good. Nitrate load at Willmar showed a smaller PBIAS because predicted data showed equal magnitude positive and negative errors compared to the observed data.

Table 3: Summary of DRAINMOD-NII model performance for predicting subsurface drainage and nitrate losses following calibration. For Waseca, model performance was evaluated using monthly NSE. Model calibration results are shown here only for Subfields 4 and 6 (those used in the model simulation for the different rate treatments). For Lamberton and Willmar, model performance was evaluated using the annual NSE and PBIAS, based on availability of measured data.

Model Goodness-of-Fit for Lamberton, Willmar, and Waseca						
	Lamberton		Willmar		Waseca—Subfield 4	Waseca—Subfield 6
	NSE	PBIAS	NSE	PBIAS	NSE	NSE
Drainage	0.63	24%	0.73	20%	0.49	0.5
Nitrate Loss	0.50	-21%	0.42	-16%	0.55	0.6

Nitrate Losses Predicted by DRAINMOD-NII

The results of DRAINMOD prediction for each of the nitrogen rate treatments are shown in figure 9 for the spring application, figure 10 for the split-VRN application, and figure 11 for fall application. In general, nitrate losses are greatest for fall applications, and least for split-VRN applications, with spring application nitrate losses falling between these two alternative practices. For all application timings, nitrate loss is related to growing season precipitation, with greater losses corresponding to greater growing season precipitation (months of April-September). Nitrate losses are also related to the application rate, with greater rates showing greater nitrate losses. This trend is most obvious with higher precipitation rates. During dry years, there is less of a difference in nitrate losses for the different application rates.

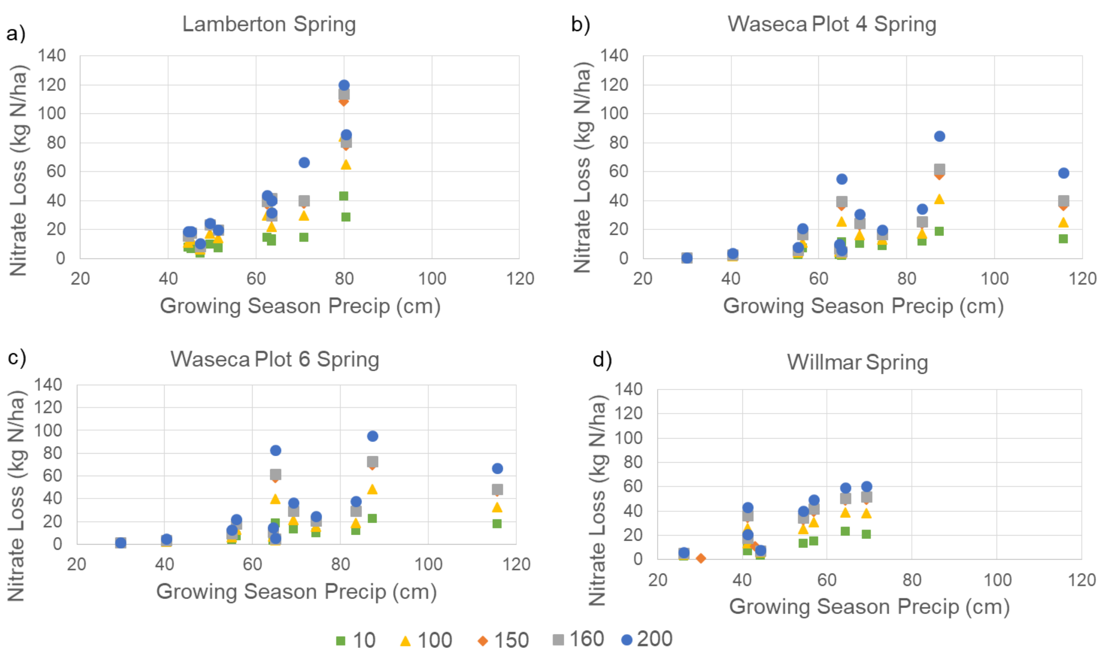

Figure 9: DRAINMOD simulated nitrate load for spring fertilizer application vs. growing season precipitation for a) Lamberton, b) Subfield 4 at Waseca, c) Subfield 6 at Waseca, and d) Willmar. Fertilizer application rates are shown in the legend with units of kg N/ha.

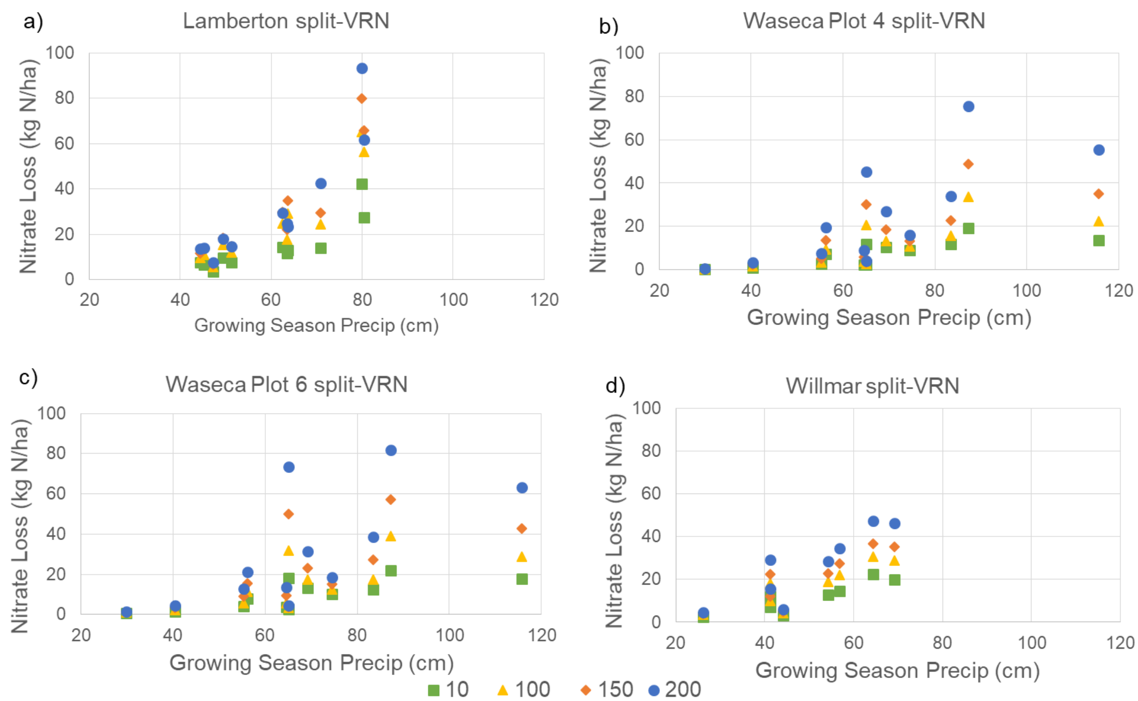

Figure 10: DRAINMOD simulated nitrate load for split-VRN fertilizer application vs. growing season precipitation for a) Lamberton, b) Subfield 4 at Waseca, c) Subfield 6 at Waseca, and d) Willmar. Fertilizer application rates are shown in the legend with units of kg N/ha.

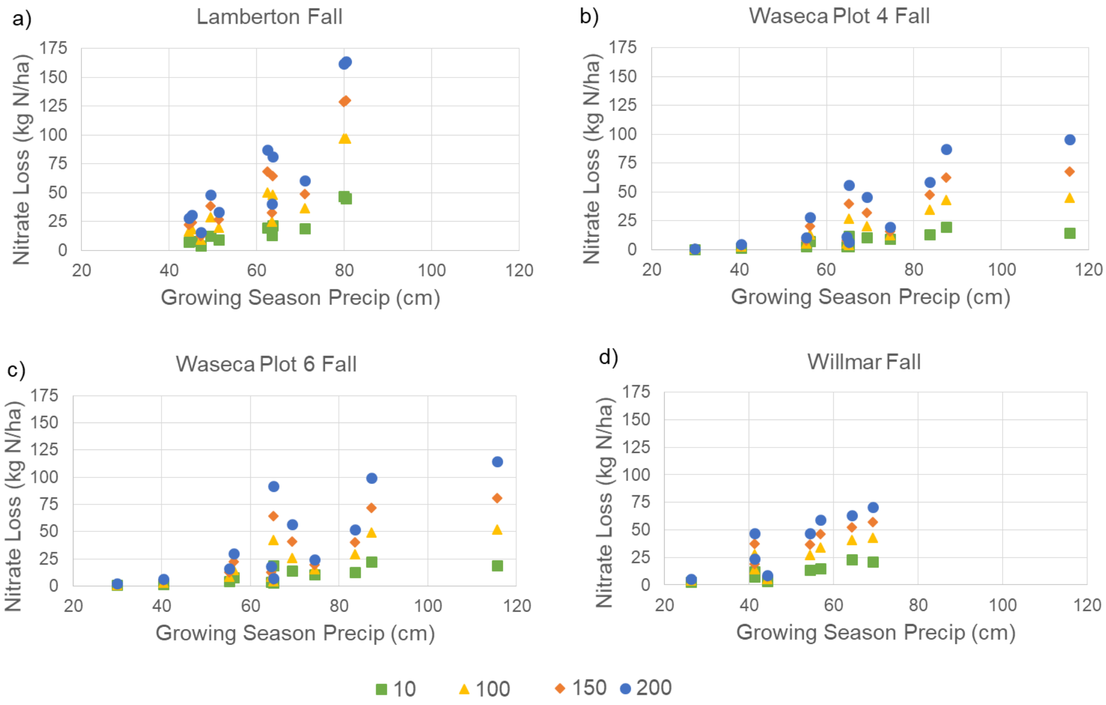

Figure 11: DRAINMOD simulated nitrate load for fall fertilizer application vs. growing season precipitation for a) Lamberton, b) Subfield 4 at Waseca, c) Subfield 6 at Waseca, and d) Willmar. Fertilizer application rates are shown in the legend with units of kg N/ha.

Regional Regression Equations

The nitrate loss results from the DRAINMOD-NII simulations for a corn-soybean rotation at Waseca, Willmar, and Lamberton, were best predicted by an exponential equation that used the fertilizer rate applied during corn years and growing season precipitation as the independent (x) variables. Equations best predicting nitrate loss (y) for spring, split-VRN, and fall application timing are shown in equations 3, 4, and 5, respectively:

$$y_{spring} = exp(-0.32022 + 0.005867x_1 + 0.03667x_2) \quad (3)$$

$$y_{VRN} = exp(-0.38152 + 0.004852x_1 + 0.0371x_2) \quad (4)$$

$$y_{fall} = exp(-0.32192 + 0.006959x_1 + 0.037938x_2) \quad (5)$$

Where x_1 is equal to the rate of nitrogen fertilizer applied (in kg N/ha), and x_2 is equal to the precipitation for the months April through September (in cm). The above equations meet standard statistical assumptions.

The spring-application predictive equation had an R^2 value equal to 0.46, split-VRN had an R^2 value equal to 0.49, and fall R^2 value equal to 0.52. The fit of DRAINMOD and regression predictions are shown in figure 12.

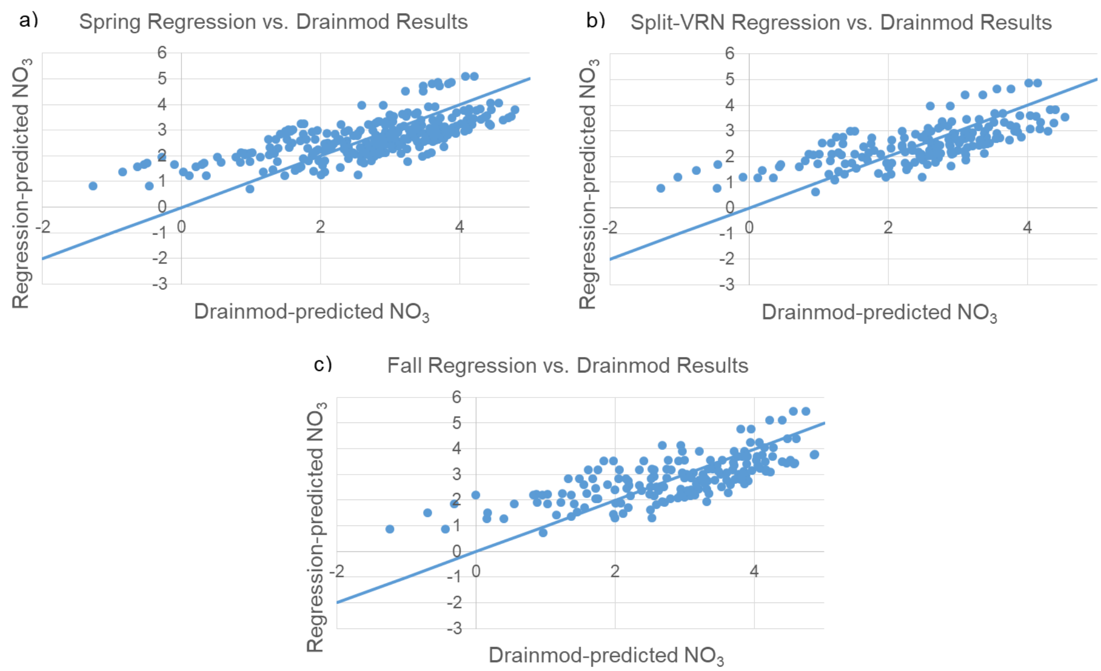

Figure 12: Ln-ln plots of the results for regression predicted nitrate losses at Waseca, Willmar and Lamberton compared to original DRAINMOD-NII simulated data for: a) spring application, b) split-VRN application, and c) fall application. A natural log (ln) scale was used to achieve the best fit of the residuals. The straight line indicates a 1:1 ratio between DRAINMOD and regression predictions, and a perfect fit between the regression and DRAINMOD results.

The exponential form of the regression equations show in equations 3 through 5 indicate that nitrate load will continue to increase with increasing precipitation. These equations use precipitation as a surrogate measure of percolation, where higher precipitation results in higher percolation and higher flow volumes from subsurface drains and higher nitrate loads. This is generally a valid assumption, however, this is not always the case. Rainfall intensity, water table depth, and landscape slope can all influence percolation and resulting water flow to subsurface drains.

Comparison between Fall, Spring and Split-VRN Applications

Using equations 3, 4, and 5, the predicted nitrate losses for southern Minnesota for spring, split-VRN, and fall application under a low (100 kg N/ha) and high (180 kg N/ha) fertilizer rate are shown in figure 13. Higher N application rates result in larger predicted nitrate loads for all timings (spring, split-VRN, and fall) compared to lower application rates. For a given precipitation and fertilizer rate, the split-VRN application has the smallest predicted nitrate loss, with fall application having the largest, though the low fall fertilizer rate has nearly the same nitrate load as the high VRN application. As precipitation increases, there is an increase in nitrate losses. For example, for a growing season precipitation of 60 cm under the low (100 kg N/ha) fertilizer application rate, the spring application results is 2.37 kg/ha less nitrogen lost than the fall application (a 17% reduction), and the split-VRN application results in 3.9 kg/ha less nitrogen than fall (27% reduction). Increasing the fertilizer application rate to 180 kg N/ha, the reductions in N losses become larger. For the same growing season precipitation of 60 cm with the high fertilizer application rate, the spring application results in 5.8 kg/ha less nitrogen lost than the fall application (a 24% reduction), while the VRN application results in 9.6 kg/ha less N lost (a 39% reduction).

As precipitation increases, the difference between the "best-case" (low fertilizer rate, split-VRN application), and "worst-case" (high fertilizer rate, fall application) scenarios increases. For a growing season precipitation of 60 cm, the "best-case" scenario of split-VRN, applied at a rate of 100 kg N/ha results in 14 kg N/ha less nitrate lost than the "worst-case" scenario of fall application at a rate of 180 kg N/ha (a 58% reduction). At 100 cm of precipitation, the "best-case" split-VRN, low rate application results in 67 kg N/ha less field nitrate losses compared to the "worst-case" fall, high rate application (59% reduction).

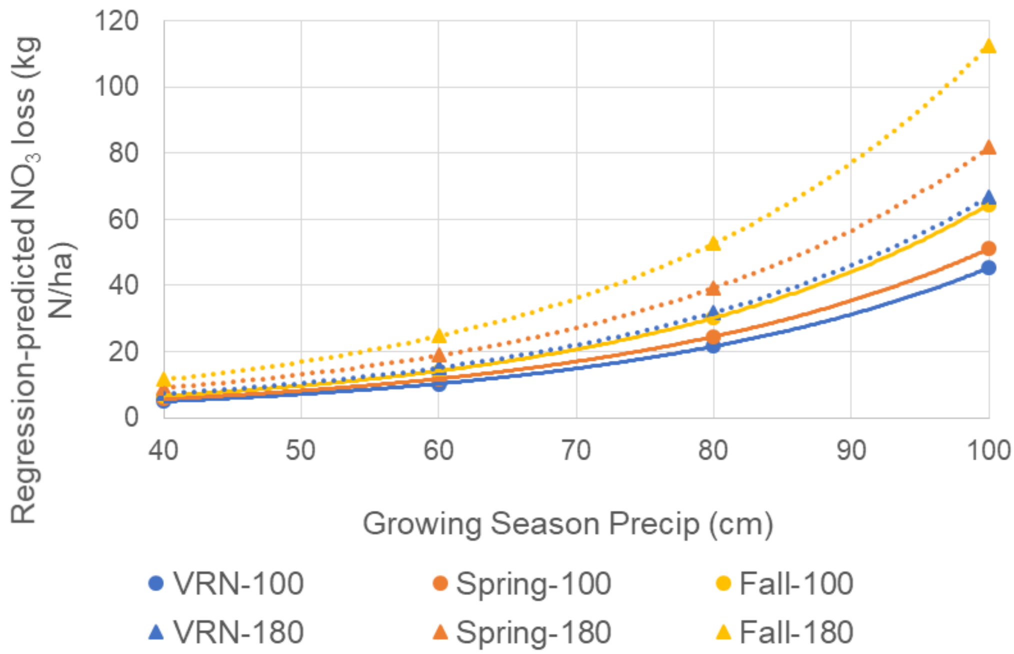

Figure 13: Regression predicted nitrate loss for southern Minnesota under spring, fall, or split-VRN fertilizer timings. Shown here are the results for a low application rate of 100 kg N/ha (VRN-100, Fall-100, and Spring-100), and a high application rate of 180 kg N/ha (VRN-180, Fall-180, and Spring-180).

Conclusions

In this study, the field scale hydrologic and nitrogen simulation model DRAINMOD-NII was used to estimate nitrate-N loads for fertilizer applied to corn given different application rates and application timing approaches for three sites in southern Minnesota: Waseca, Willmar, and Lamberton. The results of these simulation were used in a regression analysis to develop equations to predict nitrate-N loses for the region more generally as a function of fertilizer timing and application rate, as well as growing season precipitation. The results of the DRAINMOD simulations showed the highest nitrate loads occurred for all fall application rates, with VRN having the lowest predicted N losses. An exponential regression model which used fertilizer application rate and growing season precipitation as the dependent variables had good agreement with DRAINMOD results, with R^2 values of 0.46, 0.49, and 0.52 for spring, split-VRN, and fall application, respectively. The regression model showed that the variation in the

DRAINMOD predicted nitrate-N load for the different application timing treatments was highly related to annual precipitation and fertilizer application rates. At 60 cm growing season precipitation, nitrate-N losses for VRN application of 100 kg N/ha were 58% smaller than losses for a fall application of 180 kg N/ha. At 100 cm precipitation, nitrate-N losses for VRN application of 100 kg N/ha were 60% smaller than losses for a fall application of 180 kg N/ha. These results show that implementation of VRN side-dress applications based on remote sensing of crop N status can result in significant improvements in water quality.

PART III: Development of a new statistical framework for evaluation of goodness-of-fit for hydrologic models

Chapter 4. A new framework for assessing the goodness-of-fit of hydrologic and water quality models. Part I: Theoretical Development

Synopsis

Hydrologic model performance is typically determined by comparing the fit between model predictions and field-measured observations. However, the current framework used in statistical measures of goodness-of-fit do not allow for consideration of the uncertainty inherent in both the observed and predicted data in determining the fit between the observed and predicted data. In this paper, a framework is developed that allows the observed and predicted data to be described in terms of the uncertainty due to measurement and model parameter error. This framework is used to adjusted mean square error (MSE) and normalized mean square error (NMSE) to account for errors in the observed and predicted data. The adjusted MSE and NMSE derived here allow for a more robust indication of model performance than methods commonly used, supporting greater more confidence in model performance, and the accuracy of model results.

Introduction

Models that simulate the hydrology and chemical transport of agricultural systems are used to make predictions on the effects that different management practices, cropping systems, and climate changes have on the movement of water, transport of contaminants, and water usage. These predictions are often used as the basis for policy decisions and management practices. Therefore, it is important to have an understanding of how well the results of the model simulation match the natural processes of the system (Harmel et al. 2014). Recommended methods to evaluate model performance include both graphical and statistical measures that compare the model predictions to corresponding observed data measured in-situ (Engel et al. 2007, Moriasi et al. 2015). Though graphical evaluation allows for visual comparison of predicted and observed data, it is a qualitative approach that requires the judgement of the modeler. Statistical goodness of fit (GOF) measures allow for a more quantitative assessment of how well the model simulations match the observed data, making them more useful in taking a standardized approach to model evaluation (Legates and McCabe 1999).

Common statistical GOF measures include absolute measures, such as mean square error (MSE) and percent bias (PBIAS), and values normalized against the range in the observed data, such as normalized mean square error (NMSE), or the Nash-Sutcliffe coefficient of efficiency (NSCE) (Legates and McCabe 1999, Gershenfeld and Weigend 1993). These measures are defined as:

$$PBIAS = \frac{\sum_{i=1}^{n}(O_i - P_i)}{\sum_{i=1}^{n} O_i} \times 100 \tag{1.1a}$$

$$MSE = \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N} \tag{1.1b}$$

$$NMSE = \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \tag{1.1c}$$

$$NSCE = 1 - NMSE = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \tag{1.1d}$$

where O_i is the observed data for a specified time step, i, P_i is the corresponding model predicted data, N is the number of observations, and $\bar{O}$ is the mean of the observed data.

The goodness-of-fit statistics of equations 1.1a-d give an indication of model performance. However, they inherently assume that each predicted and observed data point are both defined by a single, "true," value, and do not take into account the uncertainty that exists for both the model predictions and observed values. Observed values are assumed to reflect the actual response of the system. However, if the observations are unreliable because of errors in measurements or other factors, then adjustments are needed in the GOF. Likewise, model input parameters can vary over a range of values, and removal of the model parameter uncertainty from the predicted data allows for analysis of the accuracy of modeling algorithms. Uncertainties in observed data often result from instrument or analysis error during data collection (Harmel et. al

2006). If the observed data are taken as a representative response for a set of input conditions, then additional uncertainties in this response are caused by random sampling errors. Errors in model predictions result from: 1) uncertainty in model input values, and 2) errors in model structure or algorithms. Uncertainty in model input parameters occurs because of incomplete knowledge of model parameter values. This gap in knowledge is possible because of uncertainty in tabular values (such as curve number or Mannings N), errors in measured input data such as soil physical properties (saturated hydraulic conductivity, porosity, etc.), or spatial limitation of parameters such as climate data (Guzman et. al 2015, Haan et. al. 1989, Mulla and Addiscott 1999, Shirmohammadi et. al 2006). Errors due to model algorithms occur when the algorithms don't represent reality.

For predicted data, separating the magnitude of the error caused by incorrect modeling algorithms from improper parameter values is difficult, though methods for estimating model input parameter uncertainty have been widely discussed in the literature. Monte Carlo, first-order analysis, likelihood analysis, and likelihood analysis using a Bayesian framework are all common forms of analysis of model parameter error. These methods and specific variations thereof are discussed in more detail in reviews by Shirmohammadi et al (2006) and Guzman et al (2015). Estimates of errors in the observed data due to measurement error for measured stream flow and nutrient data has been summarized by Harmel et al (2006). The uncertainty in these measures were quantified as a percent error in measured flow and water quality data introduced by four procedural categories: streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis.

A more useful way of evaluating model GOF is to take into account uncertainties in the observed and predicted data. Harmel and Smith (2007) modified the residual ($O_i - P_i$) used in GOF pairwise comparison methods to take into account the measurement uncertainty in the observed data. In this method, the modified residual was calculated based on whether or not the predicted value was within an upper and lower limit of the observed data. They used the estimates of measurement uncertainty from Harmel et al (2006). Following the work of Haan et al. 1995, Harmel et al (2010) suggested a method by which to account for uncertainty in both the predicted and observed data in calculating goodness of fit. Similar to the work of Harmel and Smith (2007), in this method, a correction factor was applied to the residual to create a modified residual, which was calculated based on the amount of overlap between the probability density functions (pdfs) of the predicted and observed data. Both of these approaches requires modeler judgement of the similarity in the spread of the observed and predicted data, and knowledge of the probability density function is required for each observed and predicted value.

In the work presented here, a new framework and methodology for modifying goodness-of-fit measures is proposed that accounts for uncertainties in both the observed and predicted data in the residual term. This framework also allows for variability in time and space to be considered in model GOF calculations. An adjusted GOF indicator which takes into account the uncertainty in the observed and predicted data is derived which requires only the knowledge of the mean and variance of the uncertainty values. The adjusted GOF indicators derived here are an adjusted mean square error (MSE), and normalized mean square error (NMSE), and they account for the role of: 1) random errors in the observed data caused by measurement error, and 2) random errors in the observed data along with random errors in model input parameters.

Framework

Accounting for Spatial and Temporal Factors

For many modelling applications, a model simulation is developed based on parameters and conditions for a given site (plot, watershed, etc.), resulting in one model prediction corresponding to each observed value (such as in equations. 1.1a-d). However, there are some modelling applications where this 1-to-1 comparison can be too restrictive. For example, when designing experiments, it is common to account for natural spatial and temporal variabilities by utilizing replicates. Modelers often lump spatially different areas together into a single modeling unit to simplify the system or because parameter uncertainty is too large to justify individual parameter sets for each site. Although this may not be a common scenario, the framework developed here is flexible to allow such a possibility if required by the user.

In this framework, the observed and predicted data are defined with a spatial component (s) as well as the traditional temporal component (t). The observed and predicted data are defined as O_{st} and P_{st} respectively, the subscript denoting that it can be a function of both space (s) and time (t), rather than just the single factor ("i" in equations 1.1a-d). The summary of key variables is shown in table 4.

Observed Data

The notation used for the observed value for site, s, at time, t, is equal to the population value, or true value, for that site and time (μ_{st}), plus random error due to measurement or sampling error (ε_{st}):

$$O_{st} = \mu_{st} + \varepsilon_{st} \tag{2.1.1}$$

For a situation where there are multiple sites (as in the case with replicate field plots) and multiple times, the overall mean for the population is defined as μ_{**}. For a single site with multiple observations corresponding to different times, the temporally averaged population mean is defined as μ_{s*}; for a single time but multiple sites, the spatially averaged population mean with regard to sites is defined as μ_{*t} (fig. 14).

The population mean for a given site and time, μ_{st}, is a function of spatial and temporal factors and relates to the statistical characteristics of the larger population via a spatial factor, δ, and a temporal factor, η. For a given site and time, these factors are defined as:

$$\delta_{st} = \mu_{st} - \mu_{*t} \ \ and \ \eta_{st} = \mu_{st} - \mu_{s*} \tag{2.1.2}$$

where δ_{st} is the spatial factor that differentiates an individual site from the overall mean for sites within a given time, and likewise η_{st} is the temporal factor that differentiates in from the overall mean for all sites (fig. 14).

When considering the spatial and temporal from the larger population, these factors are also defined as:

$$\begin{aligned} \delta_{s*} = \mu_{s*} - \mu_{**} &= (\mu_{st} - \eta_{st}) - \mu_{**} \quad and \quad \eta_{*t} = \mu_{*t} - \mu_{**} \\ &= (\mu_{st} - \delta_{st}) - \mu_{**} \end{aligned} \tag{2.1.3}$$

where δ_{s*} is the spatial factor that differentiates the mean of all the sites from the overall mean, and η_{*t} is the temporal factor that differentiates the mean of all the times from the overall mean (fig. 14).

The difference between δ_{st} and δ_{s*} is equal to the co-factor term, $(\delta\eta)_{st}$

$$\begin{aligned}(\delta\eta)_{st} = \delta_{st} - \delta_{s*} &= \mu_{st} - (\mu_{s*} - \mu_{**}) - (\mu_{*t} - \mu_{**}) - \mu_{**} \\ &= (\mu_{st} - \mu_{*t}) - (\mu_{s*} - \mu_{**})\end{aligned} \tag{2.1.4}$$

This definition requires that the difference between the overall temporal factor (η_{*t}) and the individual (η_{st}) is the same as the difference between the spatial factors.

$$\delta_{s*} - \delta_{st} = \eta_{*t} - \eta_{st} \tag{2.1.5}$$

For the rest of the development, it is assumed here that the individual and overall factors are the same value (i.e. $\delta_{s*} = \delta_{st}, \eta_{*t} = \eta_{st}$), with the result that equation 2.1.5 is equal to zero.

The definition for the observed data is then written as:

$$O_{st} = \mu_{st} + \varepsilon_{st} = \mu_{**} + \delta_{s*} + \eta_{*t} + \varepsilon_{st} \tag{2.1.6}$$

The overall mean of the observed data for all sites and time points is defined as $\bar{O}_{**}$.

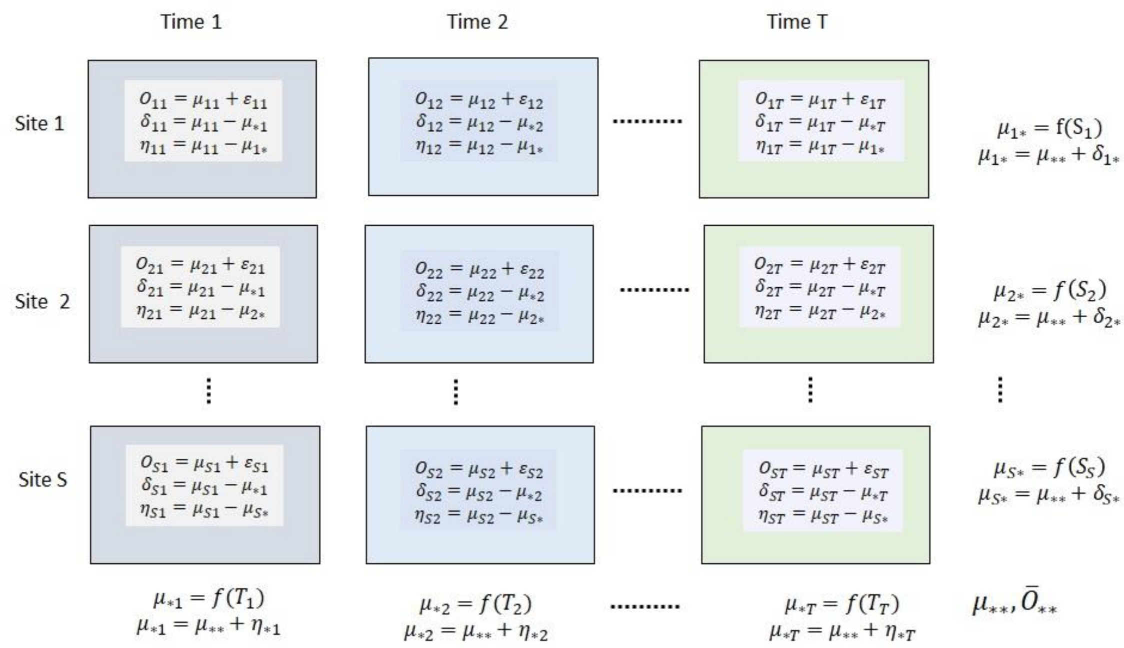

Figure 14: Diagram showing the relationship between terms for the observed data. The temporally averaged mean for a given site for all times (μ_{s*}) is a function of the sites ($f(S_S)$); while the spatially averaged mean for a given time (μ_{*t}) for all sites is a function of time ($f(T_T)$).

Predicted Values

The predicted values are defined as a function of the perfect model prediction obtained using the correct model inputs, $\tilde{P}_{st}$, with a random component resulting from using incorrect model inputs ($\tilde{\varepsilon}_{st}$). The predicted data is then defined as

$$P_{st} = \tilde{P}_{st} + \tilde{\varepsilon}_{st} \tag{2.2.1}$$

Errors in the predicted values can also occur from imperfect modeling algorithms. This type of error is non-random. For a set of correct model inputs and perfect modeling algorithms, the predicted value is equal to the observed data mean, such that:

$$P_{st} = \tilde{P}_{st} = \mu_{st} \tag{2.2.2}$$

The error caused by imperfect modeling algorithms (ρ) is defined as the difference between the observed mean and the prediction with correct parameters:

$$\rho_{st} = \mu_{st} - \tilde{P}_{st} \tag{2.2.3}$$

Spatially Averaged terms

For the case where the modeler has "lumped" spatially different areas together into a single modeling unit, the predicted value results from the single set of "average" soil and other spatial parameters, though the values can still vary with time. In this spatially averaged scenario, the observed data remains defined as: $O_{st} = \mu_{st} + \varepsilon_{st}$. This representation allows the observed mean (μ_{st}) to vary among sites while maintaining that the primary goal of the model simulation is to predict the overall spatial mean (μ_{*t}). For the set of correct model inputs and perfect modeling algorithms, the predicted data is equal to the spatially averaged observed data mean, such that:

$$P_{*t} = \tilde{P}_{*t} = \mu_{*t} \tag{2.2.4}$$

The predicted data without correct model inputs is defined as:

$$P_{*t} = \tilde{P}_{*t} + \tilde{\varepsilon}_{*t} \tag{2.2.5}$$

Where $\tilde{P}_{*t}$ and $\tilde{\varepsilon}_{*t}$ are the model prediction with no parameter error, and the value of the parameter error, respectively, for the spatially averaged model scenario.

The error caused by imperfect modeling algorithms is again defined as the different between the observed mean and the prediction with correct parameters. For spatially-averaged condition, it is:

$$\rho_{*t} = \mu_{*t} - \tilde{P}_{*t} \tag{2.2.6}$$

Table 4: Summary of key variables used in the framework for adjusting GOF statistics. Variables are summarized for the case when there is one observation for each space and time feature, and also for the case when sites are lumped spatially.

Scale	Variable	Definition
Parameter set for each site and time	$P_{st} = \tilde{P}_{st} + \tilde{\varepsilon}_{st}$	Simulated/predicted value for site s and time t
	$O_{st} = \mu_{st} + \varepsilon_{st}$	Observed value for site s and time t
	$\tilde{P}_{st}$	Predicted value with correct model parameters
	μ_{st}	Population mean for observed data for site s and time t. $\mu_{st} = \mu_{**} + \delta_{s*} + \eta_{*t}$
	$\tilde{\varepsilon}_{st}$	Random error in model prediction obtained by using the wrong model inputs
	ε_{st}	Random error in observed data, defined here as due to measurement error
	$\rho_{st} = \mu_{st} - \tilde{P}_{st}$	Error in model prediction due to use of imperfect modeling algorithms
Time and averaged spatial parameters	$P_{*t} = \tilde{P}_{*t} + \tilde{\varepsilon}_{*t}$	Simulated/predicted value at time t for spatially averaged parameters
	$O_{st} = \mu_{st} + \varepsilon_{st}$	Observed value for site s and time t
	$\tilde{P}_{*t}$	Predicted value with correct model parameters
	μ_{*t}	Spatially averaged population mean for observed data for time t
	$\tilde{\varepsilon}_{*t}$	Random error in model prediction obtained by using the wrong model inputs
	$\rho_{*t} = \mu_{*t} - \tilde{P}_{*t}$	Error in model prediction due to use of imperfect modeling algorithms

Goodness of Fit indicators: NMSE and MSE

For the most general case where there is an observation for each site and time, the goodness-of-fit indicators, MSE and NMSE, are defined as:

$$MSE = \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{N} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2}{ST} = \frac{SSE_{st}}{ST} \tag{3.1.1}$$

$$NMSE = \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2}{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}_{**}\right)^2} = \frac{SSE_{st}}{SSTO_{st}} \tag{3.1.2}$$

Where the numerator for both MSE and NMSE is equal to the sum of squares of error (SSE); the denominator for MSE is equal to the total number of observations (product of the number of sites, S, and time points, T); the denominator for NMSE is equal to the sums of squares of total (SSTO); and the subscript "st" denotes the relationships being a general form based on statistical characteristics for each site and time.

Using the relationships for O_{st} and P_{st} defined in table 4, the difference between the observed and predicted data, and the difference between the observed and its mean, are defined as:

$$\begin{aligned} SSE_{ST} &= \sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2 \\ &= \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} + \varepsilon_{ij} - \tilde{P}_{ij} - \tilde{\varepsilon}_{ij}\right)^2 \\ &= \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\rho_{ij} + \varepsilon_{ij} - \tilde{\varepsilon}_{ij}\right)^2 \end{aligned} \tag{3.2.1}$$

$$\begin{aligned} SSTO_{ST} &= \sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}_{**}\right)^2 = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}}{ST}\right)^2 \\ &= \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} + \varepsilon_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}}{ST}\right)^2 \end{aligned} \tag{3.2.2}$$

Expected value of MSE and NMSE

The definitions of SSE and SSTO in equations 3.2.1 and 3.2.2 contain random variables, and therefore MSE and NMSE are random variables as well. As shown later, the expected values of these random variables can be defined from the statistical characteristics (such as mean and variance) of the observed and predicted random errors. By using the rule of expected values (Casella and Berger 2002) the expected value of MSE and NMSE are defined as:

$$E(MSE_{ST}) = \frac{E(SSE_{ST})}{ST} \tag{3.3.1}$$

$$E(NMSE_{st}) = E\left(\frac{SSE_{ST}}{SSTO_{ST}}\right) \tag{3.3.2}$$

Taylor series for ratio of 2 random variables: Determining the expected value of MSE is relatively straightforward because it is only a function of the random variable SSE. However, the expected value of NMSE is more complicated because it is a ratio of two random variables. The expected value for the ratio of two random variables can be obtained from a Taylor series expansion (Casella and Berger 2002):

$$E\left(\frac{\mathrm{x}}{\mathrm{y}}\right) = \frac{\mathrm{E(x)}}{\mathrm{E(y)}} + \left(\frac{E(x)}{E(y)^3}\right)VAR(y) - \frac{COV(x,y)}{\mathrm{E(y)}^2} \tag{3.3.3a}$$

By neglecting third-order and higher terms, and second order and higher terms, the Taylor Series expansion can be simplified as:

$$E\left(\frac{\mathrm{x}}{\mathrm{y}}\right) \approx \frac{\mathrm{E(x)}}{\mathrm{E(y)}} \tag{3.3.3b}$$

This first-order approximation of equation 3.3.3b corresponds to small variances of $SSTO_{ST}$ and negligible covariance of $SSTO_{ST}$ and SSE_{ST}. It is used for all further derivations of adjusted GOF.

By using equation 3.3.3b, the expected value for NMSE can be approximated as:

$$E(NMSE_{st}) = E\left(\frac{SSE_{ST}}{SSTO_{ST}}\right) \approx \frac{E(SSE_{ST})}{E(SSTO_{ST})} \tag{3.3.4}$$

Expected value SSE

To simplify development, random errors of the observed data and the model parameter random errors are assumed to be homoscedastic and share a common mean, that is, $E(\varepsilon_{st}) = E(\varepsilon),\ E(\tilde{\varepsilon}_{st}) = E(\tilde{\varepsilon}), VAR(\varepsilon_{st}) = VAR(\varepsilon), and\ VAR(\tilde{\varepsilon}_{st}) = VAR(\tilde{\varepsilon})$. From equation 3.2.1, the expected value of SSE_{st} is then defined as:

$$E(SSE_{st}) = E\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\rho_{ij} + \varepsilon_{ij} - \tilde{\varepsilon}_{ij}\right)^2\right) \tag{3.4.1}$$
$$= M_1 + E(M_1') + B + E(F) + \tilde{X} + E(C)$$

where:

$$M_1 = \sum_{j=1}^{T}\sum_{i=1}^{S} E(\rho_{ij}^2) = \sum_{j=1}^{T}\sum_{i=1}^{S} \rho_{ij}^2 \tag{3.4.2a}$$

$$E(M_1') = E\left(2\sum_{j=1}^{T}\sum_{i=1}^{S} \rho_{ij}\varepsilon_{ij}\right) = 2E(\varepsilon)\sum_{j=1}^{T}\sum_{i=1}^{S} \rho_{ij} \tag{3.4.2b}$$

$$E(F) = \sum_{j=1}^{T}\sum_{i=1}^{S} E\left(\varepsilon_{ij}^2\right) = (TS)E(\varepsilon_{st}^2) = (TS)VAR(\varepsilon_{st}) + (TS)(E(\varepsilon_{st})^2) \tag{3.4.2c}$$

$$E(C) = 2\big(COV(\varepsilon, \tilde{\varepsilon}) + E(\varepsilon)E(\tilde{\varepsilon})\big) \tag{3.4.2d}$$

$$B = (TS)E(\tilde{\varepsilon}_{st})\left(E(\tilde{\varepsilon}_{st}) - \frac{2\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right)}{TS}\right) \tag{3.4.2e}$$

$$\tilde{X} = (TS)VAR(\tilde{\varepsilon}_{st}) \tag{3.4.2f}$$

All terms in equations 3.4.2 a-f must be positive values, with the exception of "B," which can be negative if the value of the model algorithm error is large enough relative to the mean of the parameter error.

Equation 3.4.1 can be further simplified for the conditions of mean of measurement error is zero (i.e. $E(\varepsilon) = 0$), and independent model parameter and observed data errors (i.e. $COV(\varepsilon, \tilde{\varepsilon}) = 0$):

$$E(SSE_{ST}) = E\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\rho_{ij} + \varepsilon_{ij} - \tilde{\varepsilon}_{ij}\right)^2\right) = M_{st} + B_{st} + X_{st} + \tilde{X}_{st} \tag{3.4.3}$$

Where M_{st} represents the model algorithm error; B_{st} is a combined term which represents the mean of the model parameter error; X_{st} represents the variance in the measurement error; and $\tilde{X}_{st}$ representative of the variance in the model parameter error. These terms are defined as:

$$M_{st} = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\rho_{ij}^2\right) \tag{3.4.4a}$$

$$B_{st} = (ST)\big(E(\tilde{\varepsilon}_{st})\big)\left(E(\tilde{\varepsilon}_{st}) - \frac{2\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right)}{ST}\right) \tag{3.4.4b}$$

$$X_{st} = (ST)VAR(\varepsilon_{st}) \tag{3.4.4c}$$

$$\tilde{X}_{st} = (ST)VAR(\tilde{\varepsilon}_{st}) \tag{3.4.4d}$$

The conditions of $E(\varepsilon) = 0$ and $COV(\varepsilon, \tilde{\varepsilon}) = 0$ are used for the remaining theoretical development. The expected value of SSE is then defined by equation 3.4.3.

Expected value SSTO

The sum of square total**,** $SSTO_{st,}$ can be written as:

$$\begin{aligned} E(SSTO_{ST}) &= E\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} + \varepsilon_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}}{ST}\right)^2\right) \\ &= Z_{st} + X_{st}\left(1 - \frac{1}{ST}\right) \end{aligned} \tag{3.5.1}$$

Where X_{st} is defined previously in equation 3.4.4c, and Z_{st} is the difference between the observed data and its mean with no error terms, equal to:

$$Z_{st} = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST}\right)^2 \qquad 3.5.2$$

Expected value GOF

From equations 3.4.3 and 3.5.1 for the expected values of SSE and SSTO, and the Taylor series expansion approximation from equation 3.3.3b, the expected value of the MSE and NMSE is equal to:

$$E\left(MSE_{ST}\right) = \frac{E(SSE_{ST})}{ST} = \frac{(M_{st} + B_{st} + X_{st} + \tilde{X}_{st})}{ST}$$
$$= \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\rho_{ij}^2\right)}{ST} + (\tilde{\varepsilon}_{st})\left(E(\tilde{\varepsilon}_{st}) - \frac{2\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right)}{TS}\right) \tag{3.6.1}$$
$$+ VAR(\varepsilon_{st}) + VAR(\widetilde{\varepsilon_{st}})$$

$$E(NMSE_{st}) = E\left(\frac{SSE_{ST}}{SSTO_{ST}}\right)$$
$$= \left(\frac{1}{1 + X_{st}^*\left(1 - \frac{1}{ST}\right)}\right)\left(M_{st}^* + B_{st}^* + X_{st}^* + \tilde{X}_{st}^*\right) \tag{3.6.2}$$

Where M_{st}^* is a normalized estimate of the algorithm error; B_{st}^* is a combined term that captures the mean of the model parameter error ; X_{st}^* represents variance in the measurement error; $\tilde{X}_{st}^*$ a normalized estimate of the variance in the model parameter error; and $\bar{Z}$ is an average of the difference between the observed data and its mean with no error terms. These terms are defined as:

$$X_{st}^* = \frac{VAR(\varepsilon_{st})}{\bar{Z}} \tag{3.6.3a}$$

$$M_{st}^* = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}E\left(\rho_{ij}^2\right)}{(TS)\bar{Z}} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\rho_{ij}^2\right)}{(TS)\bar{Z}} \tag{3.6.3b}$$

$$B_{st}^* = \frac{E(\tilde{\varepsilon}_{st})\left(E(\tilde{\varepsilon}_{st}) - \frac{2\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right)}{TS}\right)}{\bar{Z}} \tag{3.6.3c}$$

$$\tilde{X}_{st}^* = \frac{VAR(\tilde{\varepsilon}_{st})}{\bar{Z}} \tag{3.6.3d}$$

$$\bar{Z} = \frac{Z_{st}}{ST} \tag{3.6.3e}$$

As with the terms defined for E(SSE), all terms listed in equations 3.6.3 a through e must be positive values, with the exception of "B_{st}^{*}" which can be negative if the value of the model algorithm error is large enough relative to the mean of the model parameter error.

The full derivation of equations for the expected value of SSE and SSTO—including no assumptions used to simplify calculations—and the expected value of MSE and NMSE can be found in Appendix B.

Adjusted MSE and NMSE

To assess model accuracy, the modeler typically has a set of observed and predicted values that can be used to compute MSE and NMSE. These computations inherently include uncertainties in the observed and predicted data and therefore correspond to the GOF definitions given by equations 3.1.1 and 3.1.2. A better representation of model accuracy is to account for uncertainty in the observed data and, if the accuracy of the modeling algorithms themselves are of primary interest, errors introduced by parameter uncertainty. We wish to adjust the GOFs obtained by the modeler to account for uncertainty in both the observed data and for the model values obtained with uncertain parameters.

To distinguish among the different relationships for the GOFs, the subscript "obs" is used for the GOFs that are obtained for a given set of observed and predicted values and include random errors related to the observed data and parameter uncertainty. The subscripts "ε" and "$\varepsilon + \tilde{\varepsilon}$" are used for variables adjusted for errors in the observed values and for both errors in observed values and parameter uncertainty, respectively. By taking into account measurement error only, the adjusted MSE and NMSE is equal to:

$$MSE_{\varepsilon} = \frac{SSE_{\varepsilon}}{ST} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \tilde{P}_{ij} - \tilde{\varepsilon}_{ij}\right)^2}{ST} = \frac{\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2\right) - \Delta SSE_{\varepsilon}}{ST} \qquad 4.1.1a$$

$$NMSE_{\varepsilon} = \frac{SSE_{\varepsilon}}{SSTO_{\varepsilon}} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \tilde{P}_{ij} - \tilde{\varepsilon}_{ij}\right)^2}{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST}\right)^2} =$$

$$= \frac{\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2\right) \quad - \Delta SSE_{\varepsilon}}{\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}\right)^2\right) \quad - \Delta SSTO_{\varepsilon}} \qquad 4.1.1b$$

where

$$\Delta SSE_{\varepsilon} = \left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2\right) \quad - \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \tilde{P}_{ij} - \tilde{\varepsilon}_{ij}\right)^2 \qquad 4.1.2a$$

$$= SSE - SSE_{\varepsilon}$$

$$\Delta SSTO_{\varepsilon} = \left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}\right)^2\right) \quad - \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST}\right)^2 \qquad 4.1.2b$$

$$= SSTO - SSTO_{\varepsilon}$$

where SSE_{ε} is the difference between the observed and predicted data adjusted for uncertainty in the observed data, and $SSTO_{\varepsilon}$is the difference between the observed data and its mean, likewise adjusted for uncertainty in the datasets. The terms ΔSSE and $\Delta SSTO$ are equal to the difference between the SSE or SSTO with all error terms included and that adjusted for errors. Although SSE and SSTO in the theoretical development are random variables, they are related to the observed SSE and SSTO ($SSE_{obs}\ and\ SSTO_{obs}$) of specific set of conditions under which the model is used later in the paper.

By taking into account both measurement error and model parameter error, the adjusted actual MSE and NMSE equals:

$$MSE_{\varepsilon+\tilde{\varepsilon}} = \frac{SSE_{\varepsilon+\tilde{\varepsilon}}}{ST} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \tilde{P}_{ij}\right)^2}{ST}$$

$$= \frac{\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2\right) \quad - \Delta SSE_{\varepsilon+\tilde{\varepsilon}}}{ST} \qquad (4.1.3a)$$

$$NMSE_{\varepsilon+\tilde{\varepsilon}} = \frac{SSE_{\varepsilon+\tilde{\varepsilon}}}{SSTO_{\varepsilon+\tilde{\varepsilon}}} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \tilde{P}_{ij}\right)^2}{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST}\right)^2}$$

$$= \frac{\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2\right) \; - \Delta SSE_{\varepsilon+\tilde{\varepsilon}}}{\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}\right)^2\right) \; - \Delta SSTO_{\varepsilon}} \tag{4.1.3b}$$

where

$$\Delta SSE_{\varepsilon+\tilde{\varepsilon}} = \left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2\right) \; - \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \tilde{P}_{ij}\right)^2$$

$$= SSE \; - SSE_{\varepsilon+\tilde{\varepsilon}} \tag{4.3.4}$$

Approximation of adjusted MSE and NMSE

Equations 4.1.1a and b and 4.1.3 and b provide a framework to calculate random GOFs of the adjusted MSE and NMSE. In this paper, the focus is on the expected values of these random variables. Evaluation of the expected values only requires knowledge on the statistical characteristics of the random components (such as their mean and variance) that are more easily determined. By utilizing the Taylor series approximation from equation 3.3.3b, the expected value for the random variables of the adjusted terms are defined as:

$$E(MSE) = \widetilde{MSE} = E\left(\frac{SSE - \Delta SSE}{ST}\right) = \frac{E(SSE) - E(\Delta SSE)}{ST} \tag{4.4.1}$$

$$E(NMSE) = \widetilde{NMSE} = E\left(\frac{SSE - \Delta SSE}{SSTO - \Delta SSTO}\right) \approx \frac{E(SSE) - E(\Delta SSE)}{E(SSTO) - E(\Delta SSTO)} \tag{4.4.2}$$

where of generalized forms SSE-ΔSSE and SSTO-ΔSSTO are used to allow the above relationships for $\widetilde{MSE}$ and $\widetilde{NMSE}$ to be valid for consideration of both random observed components and parameter uncertainty, and the tilde over MSE or NMSE indicates that it is the expected value.

For sets of observed and predicted values, SSE=SSE_{obs} and SSTO=$SSTO_{obs}$. For the typical case where the exact value of the error is unknown, the values of ΔSSE and

$\Delta SSTO$ are unknown; however, they can be calculated based on their expected value. The values of SSE_{obs} and $SSTO_{obs}$ are considered non-random variables, because though they contain random numbers, they are the single values obtained by for a given set of observed and predicted values. By using SSE_{obs}=E(SSE) and $SSTO_{obs}$=E(SSTO) for a given set of observed and predicted values, the actual adjusted MSE and NMSE can be approximated by:

$$MSE \approx \widetilde{MSE} = \frac{SSE_{obs} - E(\Delta SSE)}{ST} \tag{4.4.3}$$

$$NMSE \approx \widetilde{NMSE} = \frac{SSE_{obs} - E(\Delta SSE)}{SSTO_{obs} - E(\Delta SSTO)} \tag{4.4.4}$$

The validity of using observed SSE and SSTO for the expected values of these variables is evaluated in the next chapter.

Adjustment for measurement error

From equations 4.3.1 and 4.3.2 the values of $E(\Delta SSE)$ and $E(\Delta SSTO)$ used in the adjusted GOF equations are defined as:

$$E(\Delta SSE_\varepsilon) = E(SSE_{st}) - E(SSE_{st,\varepsilon}) \tag{4.5.1}$$

$$E(\Delta SSTO_\varepsilon) = E(SSTO_{st}) - E(SSTO_{st,\varepsilon}) \tag{4.5.2}$$

Where the subscript ε is used to identify adjustment for measurement error in the observed data.

By the definitions of the expected values off SSE and SSTO from equations 3.4.3 and 3.5.1 (and the assumption that $E(\varepsilon_{st}) = 0$), these are defined as:

$$\begin{aligned} E(\Delta SSE_\varepsilon) &= \left(M_{st} + B_{st} + X_{st} + \tilde{X}_{st}\right) - \left(M_{st} + B_{st} + \tilde{X}_{st}\right) = X_{st} \\ &= (ST)VAR(\varepsilon_{st}) \end{aligned} \tag{4.5.3}$$

$$\begin{aligned} E(\Delta SSTO_\varepsilon) &= \left(Z_{st} + E(X_{st})\left(1 - \frac{1}{ST}\right)\right) - (Z_{st}) = X_{st}\left(1 - \frac{1}{ST}\right) \\ &= (ST)VAR(\varepsilon_{st})\left(1 - \frac{1}{ST}\right) \end{aligned} \tag{4.5.4}$$

Using the results of equations 4.5.3 and 4.5.3 above for the values of $E(\Delta SSE_\varepsilon)$ and $E(\Delta SSTO_\varepsilon)$, the adjusted MSE and NMSE for errors in the observed value are defined as:

$$\widetilde{MSE}_{\varepsilon} = \frac{SSE_{obs} - (ST)VAR(\varepsilon_{st})}{ST} \tag{4.5.6}$$

$$\widetilde{NMSE}_{\varepsilon} = \frac{SSE_{obs} - (ST)VAR(\varepsilon_{st})}{SSTO_{obs} - (ST)VAR(\varepsilon_{st})\left(1 - \frac{1}{ST}\right)} \tag{4.5.7}$$

The MSE and NMSE adjusted for measurement error can also be written in terms of their variance, equal to the term divided by the number of observations (ST):

$$\widetilde{MSE}_{\varepsilon} = \frac{SSE_{obs}}{ST} - VAR(\varepsilon_{st}) = VAR(SSE)_{obs} - VAR(\varepsilon_{st}) \tag{4.5.8}$$

$$\begin{aligned}\widetilde{NMSE}_{\varepsilon} &= \frac{SST_{obs}/ST - VAR(\varepsilon_{st})}{SSTO_{obs}/ST - VAR(\varepsilon_{st})\left(1 - \frac{1}{ST}\right)} \\ &= \frac{VAR(SSE)_{obs} - VAR(\varepsilon_{st})}{VAR(SSTO)_{obs} - VAR(\varepsilon_{st})\left(1 - \frac{1}{ST}\right)}\end{aligned} \tag{4.5.9}$$

Adjustment for measurement error and parameter error

The $E(\Delta SSE)$ and $E(\Delta SSTO)$ used to adjusted MSE and NMSE to account for the effect of both measurement error in the observed data and uncertainty in model predictions due to parameter values are defined as:

$$E(\Delta SSE_{\varepsilon+\tilde{\varepsilon}}) = E(SSE_{st}) - E(SSE_{st,\varepsilon+\tilde{\varepsilon}}) \tag{4.6.1}$$

$$E(\Delta SSTO_{\varepsilon}) = E(SSTO_{st}) - E(SSTO_{st,\varepsilon}) \tag{4.6.2}$$

Where the subscript ε is used to identify adjustment for measurement error in the observed data, and $\tilde{\varepsilon}$ denotes the adjustment for model parameter error. For NMSE, the solution to $E(\Delta SSTO_{\varepsilon})$ is the same as that given in equation 4.5.4 for $E(\Delta SSTO_{\varepsilon})$ since SSTO only contains values associated with the observed data. Equation 3.4.3 is used to solve for the value of $E(\Delta SSE_{\varepsilon+\tilde{\varepsilon}})$ (using the assumption that the mean of the observed error is equal to zero) such that $E(\Delta SSE_{\varepsilon+\tilde{\varepsilon}})$ is defined as:

$$E(\Delta SSE_{\varepsilon+\tilde{\varepsilon}}) = \left(M_{st} + B_{st} + X_{st} + \tilde{X}_{st}\right) - (M_{st}) = B_{st} + X_{st} + \tilde{X}_{st}$$
$$= (ST)\left(\left(E(\tilde{\varepsilon}_{st})\right)\left(E(\tilde{\varepsilon}_{st}) - \frac{2\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right)}{ST}\right) + VAR(\varepsilon_{st}) + VAR(\tilde{\varepsilon}_{st})\right) \tag{4.6.3}$$

Solving for $E(\Delta SSE_{\varepsilon+\tilde{\varepsilon}})$ in equation 4.6.3 above requires knowledge of the value of the model algorithm error (ρ). For the case where the mean of the parameter uncertainty is equal to zero ($E(\tilde{\varepsilon}) = 0$), the value of $E(\Delta SSE_{\varepsilon+\tilde{\varepsilon}})$ can be solved for without knowledge of the value of the model algorithm error. The expected value of $\Delta SSE_{\varepsilon+\tilde{\varepsilon}}$ can be approximated in this case by:

$$E\left(\Delta SSE_{\varepsilon,\tilde{\varepsilon}}\right) = \left(M_{st} + B_{st} + X_{st} + \tilde{X}_{st}\right) - (M_{st} + B_{st}) = X_{st} - \tilde{X}_{st}$$
$$= (ST)\left(VAR(\varepsilon_{st}) + VAR(\tilde{\varepsilon}_{st})\right) \tag{4.6.4}$$

The assumption that the mean of the model parameter error is equal to zero has been shown to be an appropriate approximation in some applications (as shown in Haan and Skaggs 2003), but is not always appropriate (SWAT case study presented in Shirmohammadi et al 2006).

Using the simplified version of the expected value of $\Delta SSE_{\varepsilon+\tilde{\varepsilon}}$ from equation 4.6.4 and $\Delta SSTO_{\varepsilon+\tilde{\varepsilon}}$ from equations 4.5.4, the adjusted MSE and NMSE for errors in the observed values and model parameters is defined as:

$$\widetilde{MSE}_{\varepsilon+\tilde{\varepsilon}} = \frac{SSE_{obs} - (ST)\left(VAR(\varepsilon_{st}) + VAR(\tilde{\varepsilon}_{st})\right)}{ST} \tag{4.6.5}$$

$$\widetilde{NMSE}_{\varepsilon+\tilde{\varepsilon}} = \frac{SSE_{obs} - (ST)\left(VAR(\varepsilon_{st}) + VAR(\tilde{\varepsilon}_{st})\right)}{SSTO_{obs} - (ST)VAR(\varepsilon_{st})\left(1 - \frac{1}{ST}\right)} \tag{4.6.6}$$

These equations can also be written in terms of variance, as:

$$\widetilde{MSE}_{\varepsilon+\tilde{\varepsilon}} = \frac{SSE_{obs}}{ST} - VAR(\varepsilon_{st}) - VAR(\tilde{\varepsilon}_{st})$$
$$= VAR(SSE)_{obs} - VAR(\varepsilon_{st}) - VAR(\tilde{\varepsilon}_{st}) \tag{4.6.7}$$

$$\widetilde{NMSE}_{\varepsilon+\tilde{\varepsilon}} = \frac{SSE_{obs}/ST - VAR(\varepsilon_{st}) - VAR(\tilde{\varepsilon}_{st})}{\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}\right)^2\right)_{obs}/ST - VAR(\varepsilon_{st})\left(1 - \frac{1}{ST}\right)}$$
$$= \frac{VAR(SSE)_{obs} - VAR(\varepsilon_{st}) - VAR(\tilde{\varepsilon}_{st})}{VAR(SSTO)_{obs} - VAR(\varepsilon_{st})\left(1 - \frac{1}{ST}\right)} \quad (4.6.8)$$

Influence of component terms on MSE and NMSE

Expected value of MSE and NMSE not adjusted for errors

The application of the theoretical framework to modeling conditions is explored in greater detail in the next chapter. Insight into the framework is given here by considering the influence of the individual components of MSE and NMSE. The influence of the components of the expected value of MSE and NMSE not adjusted for errors (equations 3.6.1 and 3.6.2) are summarized in figure 15. The value of the components used in this figure were chosen arbitrarily, but maintain the relationships established in Table 4. The "sum" in the legend for the expected value of MSE (fig. 15a) corresponds to the sum of the model algorithm error, the mean of the model parameter error, and variance of the parameter error, $\left(M_{st} + B_{st} + \tilde{X}_{st}\right)/ST$, where each term was defined in equations 3.4.4a-d. For the expected value of the NMSE (fig. 2b), the "sum" corresponds to the sum of the normalized model algorithm error, the normalized mean model parameter error, and the normalized variance of the parameter error, $M_{st}^{*} + B_{st}^{*} + \widetilde{X^{*}}_{st}$, where terms were defined equations 3.6.3a-e. In figure 2a, the x-axis is defined as the variance of the measurement error divided by the number of sites and time (eq. 3.4.4c), and in figure 15b the x-axis is equal to the normalized estimate of variance in the measurement error (eq. 3.6.3a)

A perfect fit for the expected value of MSE and NMSE not adjusted for errors is obtained when the observed and predicted values are the same ($O_{st} = P_{st}$), resulting in a GOF equal to zero. The mean and variance of model parameter error ($E(\tilde{\varepsilon}_{st}), VAR(\tilde{\varepsilon}_{st})$), model algorithm error (ρ), and the mean and variance of the measurement error

$(E(\varepsilon_{st}), VAR(\varepsilon_{st}))$, all must equal zero to achieve perfect fit. This is shown in figure 15, when the values of $M_{st}, B_{st}, X_{st}, and\ \tilde{X}_{st}$ are all equal to zero.

For the expected value of MSE, a poor fit occurs when it is greater than the variance of the observed data (where $VAR(O) = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}\right)^2/ST$). Under these circumstances, the mean of the observed data is a better predictor of the observed value, than the model predicted value. This can occur when the sum of model algorithm error, mean and variance of the model parameter error, and variance of the measurement error (terms $M_{st}, B_{st}, X_{st}, and\ \tilde{X}_{st}$) are greater than the sum of the differences between each observed value, O_{ij}, and the overall observed mean, $\bar{O}$. The value corresponding to a poor MSE fit is dependent on the values of the model algorithm error, mean and variance of model parameter error, and variance of the measurement error from equation 3.6.1. Generally, the higher the value of the error terms, the worse the fit between the observed and predicted data is (fig. 15a).

Poor fit for the expected value of NMSE occurs when it is equal to or greater than one. A NMSE equal to one results when the values of SSE_{st} and $SSTO_{st}$ are equal, which indicates that the mean of the observed data is as good a predictor of the observed data as is the model-predicted value. A NMSE greater than one signifies that the mean of the observed data is better than P_{ij}. As with the MSE, this poor fit is dependent on the sum of the component terms—though generally, the higher the values of the error terms, the worse the fit between the observed and predicted data are (fig. 15b).

The expected value of MSE has a linear relationship with the model error and measurement error so that any change in these will result in an equal change in the expected value of MSE (Fig. 15a). For the expected value of NMSE, when the sum of terms that represent normalized model error $(M_{st}^*, B_{st}^*, and\ \tilde{X}_{st}^*)$ is large, the fit between the observed and predicted data improves as the variance of the measurement error increases (fig. 15b). However, for low model error, the opposite is true, and an increase in measurement error results in a worse fit. This trend occurs because of the variance in the observed data present in the NMSE denominator. If there is a high amount of uncertainty in the model predictions, but not in the observed data, the numerator is large relative to the denominator, resulting in a large NMSE and poor fit. However, for higher amounts of uncertainty in both the model predictions and the observed data (relatively similar numerator and denominator), the NMSE is smaller, resulting in better fit (fig. 15b).

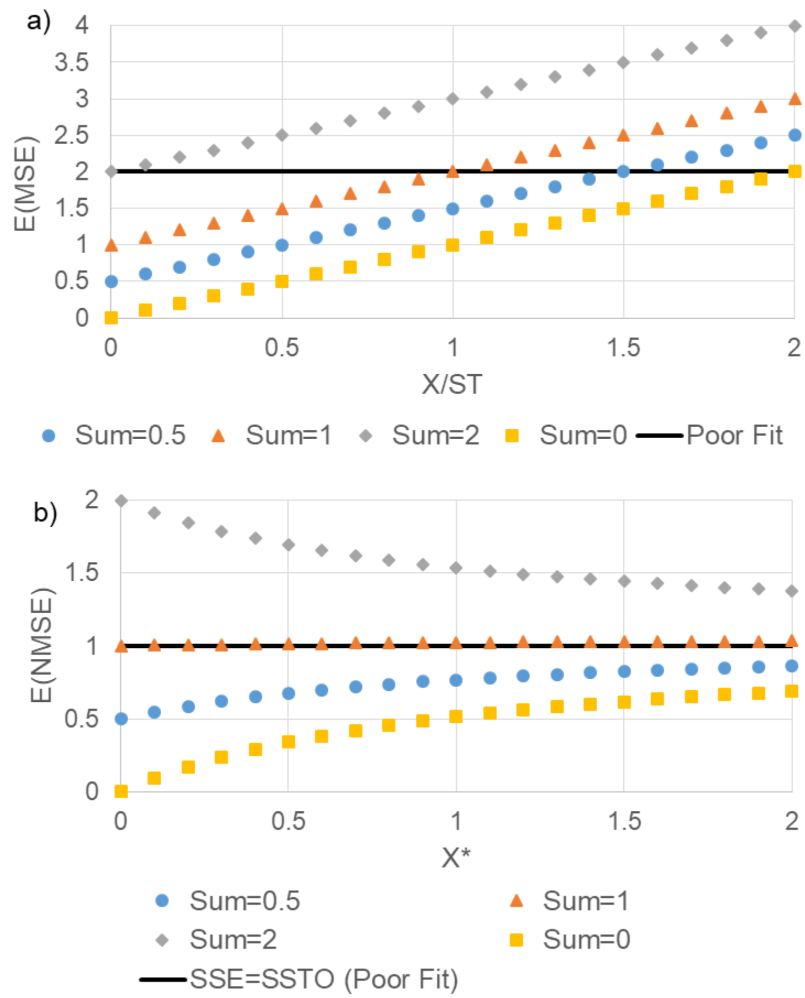

Figure 15: Relationship between the component terms and the resulting expected value of a) MSE and b) NMSE. A perfect fit occurs for both MSE and NMSE of zero; a very poor fit corresponds to a MSE equal to or greater than the variance of the observed data (shown here to be equal to 2) and corresponds to a NMSE equal to or greater than one.

Expected value of MSE and NMSE adjusted for errors

The influence of the different components on adjusted MSE and NMSE ($\widetilde{MSE}_{\varepsilon+\tilde{\varepsilon}}$ and $\widetilde{NMSE}_{\varepsilon+\tilde{\varepsilon}}$) is shown in figure 16. This figure shows the adjusted MSE or NMSE plotted for one set of possible components and illustrates how each term can influence the value of the adjusted MSE and NMSE. Here, the value of ST and $VAR(O)_{obs} = 40$ remained constant. The adjusted MSE and NMSE are plotted as a function of two values for the observed SSE (low = 200; high =360) and two values of $VAR(\hat{\varepsilon})$.

A perfect fit for the NMSE and MSE adjusted for measurement error, or adjusted for measurement and parameter error, is obtained if the difference in the observed and predicted values (SSE_{obs}), and that with error terms removed (ΔSSE) is equal to zero. For NMSE and MSE adjusted for measurement error (eq. 4.5.6 and 4.5.7), and adjusted for measurement and parameter error (eq. 4.6.5 and 4.6.6), this requires that in addition to the difference between the actual observed and predicted values equaling zero, the variance of the measurement error and variance of model parameter error also must equal zero. Perfect fit is also theoretically possible even if SSE_{obs} is not equal to zero, so long as the sum of the variances of the model parameter error and the measurement error is equal to the difference between observed and predicted data divided by the number of observations, ST. This is illustrated in figure 3, where the value of MSE or NMSE can be equal to zero even when the actual SSE is not. Though the fit is not mathematically perfect (since $O_{st} \neq P_{st}$), it must be considered a perfect fit because the measurement and parameter errors are large relative to the actual observed and predicted values.

The value for the adjusted MSE and NMSE can become negative when the variance of the measurement error plus the variance of the model parameter error is greater than the variance in the SSE term (fig. 16). Though mathematically possible, the implication of this result is that that the measurement error plus model parameter error is a better indicator of the observed and predicted data than the observed and predicted data alone. Similar to the perfect fit that can be obtained even when the difference between the observed and predicted does not equal zero, a negative adjusted NMSE or MSE should be interpreted as being equal to zero even though it is not actually a perfect fit.

For both the adjusted MSE and NMSE, fit between the predicted and observed data is worse for larger values of SSE_{obs}. However, in those situations, there is improvement in the adjusted MSE or NMSE when the values of variance of the measurement and model parameter error are larger (fig. 16). A clearly poor fit for the NMSE adjusted for measurement or measurement and parameter error occurs when the adjusted NMSE is greater than one. The adjusted NMSE is greater than one when the numerator, $SSE_{obs} - E(\Delta SSE)$, is greater than the denominator, $SSTO_{obs} - E(\Delta SSTO)$. The value of $E(\Delta SSTO)$ is always less than $E(\Delta SSE)$ from equations 4.5.4 and 4.6.4. Therefore, in order for the adjusted numerator to be greater than the denominator, $SSTO_{obs}$ must be less than the value of SSE_{obs}. This will only occur for situations where the observed data were homogenous, resulting in little difference between the observed values and their mean. For the adjusted MSE, the worst fit occurs when it is equal to or greater than the estimate of the variance of the actual data with error terms removed, which is approximated as the variance of the actual observed data minus the variance of ΔSSTO ($\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}\right)^2/ST$)-$\Delta SSTO/ST$).

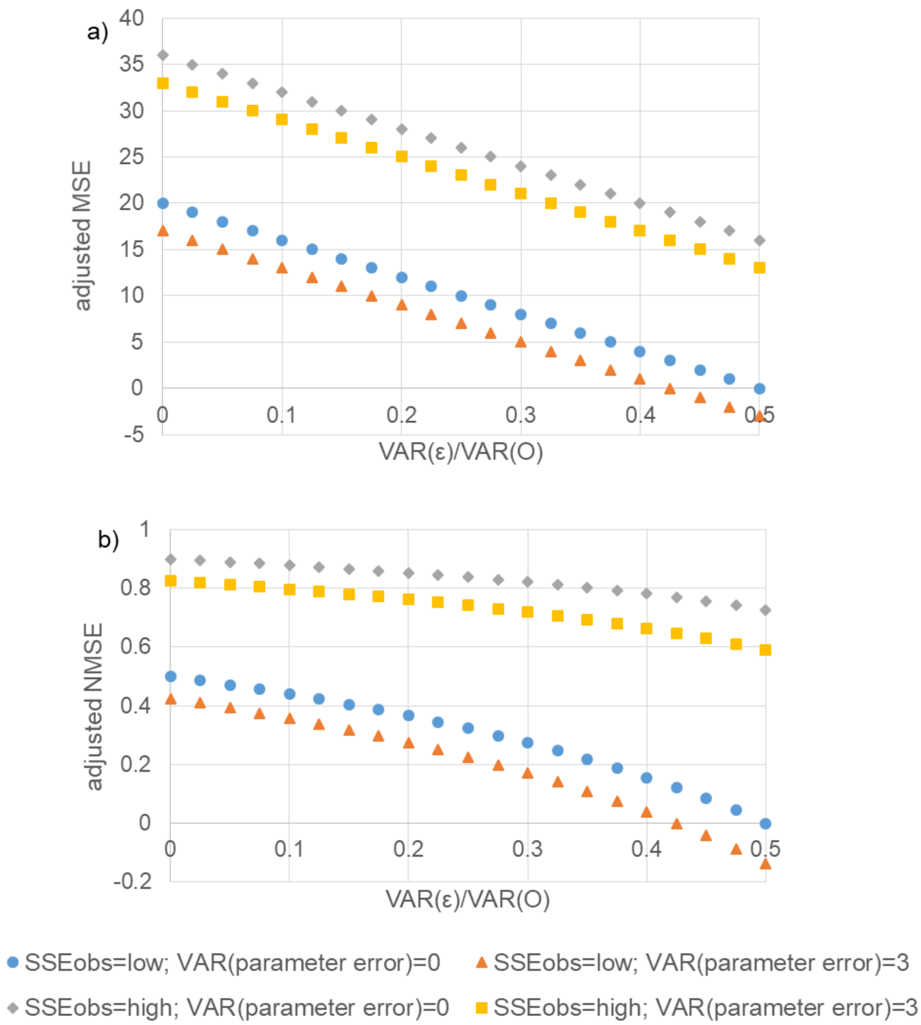

Figure 16: The adjusted a) MSE and b) NMSE plotted against the ratio of the variance of the measurement error to the variance of the observed data ($VAR(\varepsilon_{st})\ /\ VAR(O)_{obs}$).

Summary

A fundamental component of hydrologic and water quality modeling is assessing the accuracy of the simulation using goodness-of-fit (GOF) relationships. Two useful GOF measures are the mean square error (MSE) and the normalized mean square error (NMSE) that are defined using the squared difference between observed and predicted values. Although often reported as a single value, they are inherently random variables. The observed values have a random component from measurement errors and possibly with the sampling from a population of sites or watersheds. The predicted values have a random component with parameter uncertainty.

A theoretical framework was developed to improve the use of MSE and NMSE that takes into account random errors in the observed data due to measurement uncertainty, and in the model predictions due to incorrect model algorithms and model parameter uncertainty. These relationships were developed for conditions of multiple sites at different times. These key relationships were incorporated into estimates of MSE and NMSE for conditions where the errors show homoscedasticity, share a common mean, that the mean of the measurement error is equal to zero and the model parameter uncertainty and observed errors are independent of each other. MSE and NMSE defined from this approach are random variables. The expected values of these variables were then derived. This derivation was done so that the modeler's single value obtained from a set of observed and predicted values could be modified for the uncertainties in the observed and predicted data. These adjusted values are defined using the mean and variance of the error terms as well as the non-random error associated with using incorrect modeling algorithms.

Insight into the framework was obtained by considering the influence of (1) variance of the observed data and its random component, (2) mean and variance of the random component caused by parameter uncertainties, (3) non-random errors obtained from incorrect modeling algorithms on the values of unadjusted and adjusted MSE and NMSE statistics. A clearly poor fit to observed data corresponds to a NMSE= 1, and a perfect fit corresponds to NMSE = 0. The adjusted value for NMSE decreases as the ratio of the variances of the measurement errors to the observed value increases. For an initial value of NMSE=0.5, the NMSE is approximate equal to zero when the variance ratio is approximately equal to 0.5. For an initial NMSE = 0.8, the NMSE decreases to a value of approximately 0.6 for the same variance ratio.

Chapter 5. A new framework for assessing the goodness-of-fit of hydrologic and water quality models. Part II: Evaluation and Application

Synopsis

Hydrologic and water quality model performance is evaluated using statistical goodness-of-fit (GOF) measures that compare the observed data measured in the field to model predictions. Though both the observed and predicted data contain uncertainty due to measurement error and model input parameter error, current methods of calculating GOF do not take these uncertainties into account. A more robust approach of evaluating model GOF which does take these uncertainties into account was developed in Part I. In this paper, the accuracy of the expected value approach defined in Part I to adjust the GOF value for measurement and parameter errors was evaluated by comparing the expected value theoretical approach to the value adjusted using known values of random error. Additionally, the expected value theoretical approach to adjusting the GOF statistics, mean square error (MSE) and normalized mean square error (NMSE), was applied to two datasets for observed and predicted subsurface drainage flow in southern Minnesota, to see how the adjusted GOF compared to the standard approach where errors are not taken into account. The results show that the expected value approach results in adjusted MSE and NMSE that are very similar to those obtained using known values of the random numbers. Applying the MSE and NMSE adjusted for errors resulted in better fit between the observed and predicted values than using GOF indicators that do not take into account random measurement and model parameter errors.

Introduction

An important component of model development and use is the assessment of model accuracy. Many different goodness-of-fit (GOF) statistics are available (Moriasi et al. 2015). Common statistical measures to evaluate this model goodness-of-fit (GOF) include mean square error (MSE) and normalized mean square error (NMSE) (Legates 1999, Gershenfeld and Weigend 1993):

$$MSE = \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N} \tag{1.1}$$

$$NMSE = \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \tag{1.2}$$

where O_i is the observed data for a specified time step, i, P_i is the corresponding predicted data, N is the number of observations, and $\bar{O}$ is the mean of the observed data.

A good GOF statistic allows the accuracy of different models under different conditions to be fairly compared. Since it is more difficult to predict accurately a larger range in observed values, most modelers prefer to use GOF statistics that consider the error between predicted and observed values relative to an indicator of a range in the observed values. Good examples of this adjustment are the Nash-Sutcliffe statistic (Nash and Sutcliffe 1970) and normalized mean square error of equation 1.2. Since the measurement errors in the observed values and the uncertainty in parameter estimates vary with different conditions, GOF that account for these uncertainties allow for a more fair assessment of the model accuracy. Although it is well documented that both the predicted and observed data have associated uncertainties (i.e. Shirmohammadi et al 2006, Guzman et al 2015), additional work is needed to easily incorporate them into the calculation of the statistical GOF. In part I, a new framework was developed which allowed for the GOF measures MSE and NMSE to be adjusted for measurement error in the observed data and model input parameter errors in the predicted data. In this framework, the observed and predicted data were defined to allow a spatial and temporal component as:

$$O_{st} = \mu_{**} + \delta_{s*} + \eta_{*t} + \varepsilon_{st} = \mu_{st} + \varepsilon_{st} \tag{1.3}$$

$$P_{st} = \tilde{P}_{st} + \tilde{\varepsilon}_{st} = \mu_{st} - \rho_{st} + \tilde{\varepsilon}_{st} \tag{1.4}$$

For calculation of the observed data (O_{st}) for a given site (s) and time (t) in equation 1.3: μ_{**} is the overall true mean, averaged for all spatial locations and time in the observed data set; δ_{s*} is the difference between the mean for a given site for all times and the overall mean; η_{*t} is the difference between the mean for a given time for all sites and the overall mean; μ_{st} is the population value; and ε_{st} is the value of the measurement error. For the predicted data in equation 1.4, $\tilde{P}_{st}$ is the value obtained using the correct model inputs; and $\tilde{\varepsilon}_{st}$ is the value of the error introduced by using the wrong model input parameters. The value of $\tilde{P}_{st}$ is equal to the true observed value (μ_{st}) minus the error introduced in modelling by using the incorrect algorithms (ρ_{st}).

The MSE and NMSE obtained by the modeler for a specific set of measured and predicted data, can be defined as:

$$MSE_{obs} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2}{ST} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} + \varepsilon_{ij} - \tilde{P}_{ij} - \tilde{\varepsilon}_{ij}\right)^2}{ST} = \frac{SSE_{obs}}{ST} \tag{2.1a}$$

$$\begin{aligned} NMSE_{obs} &= \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2}{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}_{**}\right)^2} \\ &= \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} + \varepsilon_{ij} - \tilde{P}_{ij} - \tilde{\varepsilon}_{ij}\right)^2 =}{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} + \varepsilon_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}}{ST}\right)^2} \\ &= \frac{SSE_{obs}}{SSTO_{obs}} \end{aligned} \tag{2.1b}$$

Where the subscript "obs" is used to denote the value is obtained using the set of observed and predicted data used by the modeler. Though MSE_{obs} and $NMSE_{obs}$ contain random components, they are considered a non-random term for a particular set of values.

This framework allows for the MSE and NMSE to be adjusted for uncertainty in the observed and predicted data. If the random error is known for each of the observed and predicted values the MSE and NMSE adjusted for measurement errors ($MSE_{\varepsilon},, NMSE_{\varepsilon}$) or measurement errors and parameter errors ($MSE_{\varepsilon+\tilde{\varepsilon}}$, $NMSE_{\varepsilon+\tilde{\varepsilon}}$) is derived in the previous chapter as:

$$MSE_{\varepsilon} = \frac{SSE_{\varepsilon}}{ST} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \varepsilon_{ij} - P_{st}\right)^2}{ST} \tag{3.1a}$$

$$MSE_{\varepsilon+\tilde{\varepsilon}} = \frac{SSE_{\varepsilon+\tilde{\varepsilon}}}{ST} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \varepsilon_{ij} - P_{ij} - \tilde{\varepsilon}_{ij}\right)^2}{ST} \tag{3.1b}$$

$$NMSE_{\varepsilon} = \frac{SSE_{\varepsilon}}{SSTO_{\varepsilon}} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \varepsilon_{ij} - P_{st}\right)^2}{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST}\right)^2} \tag{3.2a}$$

$$NMSE_{\varepsilon+\tilde{\varepsilon}} = \frac{SSE_{\varepsilon+\tilde{\varepsilon}}}{SSTO_{\varepsilon+\tilde{\varepsilon}}} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \varepsilon_{ij} - P_{ij} - \tilde{\varepsilon}_{ij}\right)^2}{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST}\right)^2} \tag{3.2b}$$

Where SSE and SSTO are the sums of squares of error, and sums of square total adjusted for random error terms based on the subscripts "ε" and "$\varepsilon + \tilde{\varepsilon}$" for measurement error or measurement error and parameter error, respectively.

In practice, adjusted GOF statistics of equations 3.1-3.2 cannot be computed because random error for individual observed and predicted values is unknown. However, the mean and variance of the errors can often be estimated. As shown in Part I (chapter 4), the expected value of adjusted MSE and NMSE of equations 3.1-3.2 are functions of the means and variances of the random error terms. In addition, the modeler often has a set of observed and predicted values that is embedded into their computational framework. For conditions of the mean of measurement error of zero, the expected value of the adjusted MSE and NMSE of Part I are defined as

$$\widetilde{MSE}_{\varepsilon} = \frac{SSE_{obs} - (ST)VAR(\varepsilon_{st})}{ST} \tag{4.1a}$$

$$\widetilde{NMSE}_{\varepsilon} = \frac{SSE_{obs} - (ST)VAR(\varepsilon_{st})}{SSTO_{obs} - (ST)VAR(\varepsilon_{st})\left(1 - \frac{1}{ST}\right)} \tag{4.1b}$$

$$\widetilde{MSE}_{\varepsilon+\tilde{\varepsilon}} = \frac{SSE_{obs} - (ST)\left(\left(E(\tilde{\varepsilon}_{st})^2 - \frac{2E(\tilde{\varepsilon}_{st})\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right)}{ST}\right) + VAR(\varepsilon_{st}) + VAR(\tilde{\varepsilon}_{st})\right)}{ST} \quad (4.2a)$$

$$\widetilde{NMSE}_{\varepsilon+\tilde{\varepsilon}} = \frac{SSE_{obs} - (ST)\left(\left(E(\tilde{\varepsilon}_{st})^2 - \frac{2E(\tilde{\varepsilon}_{st})\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right)}{ST}\right) + VAR(\varepsilon_{st}) + VAR(\tilde{\varepsilon}_{st})\right)}{SSTO_{obs} - (ST)VAR(\varepsilon_{st})\left(1 - \frac{1}{ST}\right)} \quad (4.2b)$$

where $\widetilde{MSE}_{\varepsilon}$ *and* $\widetilde{NMSE}_{\varepsilon}$ are the adjusted MSE and NMSE accounting for measurement error only, $\widetilde{MSE}_{\varepsilon+\tilde{\varepsilon}}$ *and* $\widetilde{NMSE}_{\varepsilon+\tilde{\varepsilon}}$ are adjusted MSE and NMSE accounting for measurement and model parameter errors), $E(\varepsilon_{st})$ *and* $VAR(\varepsilon_{st})$ are the mean and variance of the measured data; $E(\tilde{\varepsilon}_{st})$ *and* $VAR(\tilde{\varepsilon}_{st})$ are the mean variance of the model parameter error; $SSE_{obs} = \left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2\right)_{obs}$ and $SSTO_{obs} = \left(\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}\right)^2\right)_{obs}$ are values obtained using the observed and predicted data used in equations 2.1 and 2.2 and inherently include all error terms. The tilde symbol is used to indicate MSE and NMSE was derived using expected values theoretical framework of Wilson.

The key issues related to using equations 4.1 and 4.2 to adjust MSE and NMSE for random errors examined in this paper are: (1) how well the adjusted expected values represent the random variables of MSE and NMSE, and (2) how do the MSE and NMSE adjusted for random errors in the observed and predicted data using equations 4.1 and 4.2 compare to the values of these GOF measures if no adjustments for errors are made. These issues are examined using Monte Carlo simulations to generate values of MSE and NMSE of known statistical characteristics. The paper also illustrates the application of the methods to datasets of observed and predicted subsurface drainage flow for two sites in southern Minnesota. The adjusted GOF was compared to that obtained using the typical approach which does not take into account errors in the observed and predicted data.

Methods

Accuracy of Adjusted MSE and NMSE using Expected Value Approach

To determine the accuracy of the theoretical approach (eq. 4.1 and 4.2) in adjusting MSE and NMSE for measurement, or measurement and model parameter errors, Monte Carlo (MC) simulations were used. The adjusted MSE and NSME were computed using the expected value of equations 4.1 and 4.2 and the true, random values obtained directly using equations 3.1 and 3.2. The Monte Carlo simulations were done using code written in Visual BASIC of the Microsoft Excel software.

Values were assigned to each of the component terms in equations 1.3 and 1.4 to generate values of the predicted (P_{st}) and observed (O_{st}) data. The values of the component variables used in generating observed and predicted data were chosen to reflect the magnitude of their values for a subsurface drainage system. The random measurement and parameter error terms, ε_{ij} and $\tilde{\varepsilon}_{ij}$, were assigned by a random number generator. This generator required the specification of the mean and variance of the error terms, and used a normal distribution. Though the values of these errors may not be normally distributed in the real world, this assumption was acceptable in our application because our primary goal is to compare different GOFs under the same set of conditions. The values of SSE_{obs} and $SSTO_{obs}$ (no adjustments for error) were calculated using equations 1.3 and 1.4 using the unique error terms, as were the MSE and NMSE adjusted for random errors in equations 3.1 and 3.2.

Selection of Non-random Parameters

Important non-random parameters that remained constant for all MC simulations include the overall population mean (μ_{**},), the overall spatial factor (δ_{s*}), and the overall temporal factor (η_{*t}). The dependency of the observed values on these parameters is given by equations 1.3 and 1.4. Values for these parameters were chosen based on measured field data from 2003 to 2008 for subsurface drainage from three plots located in Waseca in southern Minnesota (data from chapter 2). They are summarized in table 5. The values for δ_{s*} and η_{*t} were calculated as $\delta_{s*} = \mu_{s*} - \mu_{**}$ and $\eta_{*t} = \mu_{*t} - \mu_{**}$, respectively, and rounded to the nearest tenth. The average value for a given year ($\boldsymbol{\mu_{*t}}$), for a given plot over all years ($\boldsymbol{\mu_{s*}}$), and the overall mean, were all calculated based on the measured yearly data for all three plots.

Table 5: Field measured values of subsurface drainage from Waseca, MN (units of cm/year). The values for the replicate were used to calculate the values of $\mu_{**}, \delta_{s*}, and\ \eta_{*t}$ used in generating the observed data.

Time		**Plot s=1**	**Plot s=2**	**Plot s= 3**	$\boldsymbol{\mu_{*t}}$	**η_{*t}**
Year	**t**	μ_{1t}	μ_{2t}	μ_{3t}		
2003	**1**	6.43	6.32	4.95	5.90	-6.7
2004	**2**	17.81	18.22	15.60	17.21	4.5
2005	**3**	14.47	17.40	13.19	15.02	2.5
2006	**4**	19.02	19.42	11.79	16.74	4.0
2007	**5**	9.01	19.85	9.82	12.89	0.2
2008	**6**	6.10	13.65	5.00	8.25	-4.5
$\boldsymbol{\mu_{s*}}$		12.14	15.81	10.06	$\boldsymbol{\mu_{**} = 12.67}$	
δ_{s*}		-0.5	3.0	-2.5		

Selection of Parameters for Measurement Errors and Parameter Uncertainty

To generate observed and predicted values, parameters needed to be selected for: 1) the mean and variance of the random error in the observed data ($E(\varepsilon), VAR(\varepsilon)$), 2) the mean and variance of the random error due to model parameter uncertainty ($E(\tilde{\varepsilon}), VAR(\tilde{\varepsilon})$), and 3) the value for the uncertainty introduced by the non-random model algorithm error (ρ) . Parameters for the random errors were estimated for a range of values corresponding to high, medium, and low uncertainties. The calculations for the values corresponding to medium uncertainty are described in the following sections. The value of high was taken to be double the medium value, while low was equal to half the medium value. The values of the error terms used to calculate O_{st} and P_{st} in the MC simulations are summarized in table 6.

Table 6: Values of error terms used in Monte Carlo simulations.

Error Term	Low (0.5x medium)	Medium	High (2x medium)
$E(\varepsilon)$	0	0	0
$VAR(\varepsilon)$	0.25	0.5	1
$E(\tilde{\varepsilon})$	0.15	0.3	0.6
$VAR(\tilde{\varepsilon})$	1.2	2.4	4.8
$\rho_{relative}$		0.064	

Measurement Error: In generating observed data for MC simulations, the mean of measurement error was assumed to equal zero ($E(\varepsilon) = 0$). Its variance was calculated assuming a uniform distribution of the measured data:

$$VAR(\varepsilon) = \frac{(b-a)^2}{12} \tag{5.1}$$

The range of this distribution (values of b and a) was defined by a relative percent error around the overall mean (μ_{**}). The relative percent error was taken as plus or minus 10%, based on the measurement error reported for weirs and flumes reported in Herschy (1971). The value of μ_{**} was defined by the overall mean of the subsurface drainage data reported in table 1 (12.67 cm/d). Using these values, “b” and “a” were equal to 10.8 and 13.2, respectively, resulting in a variance of the measurement error ($VAR(\varepsilon)$) equal to 0.5 (table 6). Though the uniform distribution was used to calculate the value for the variance of the measurement data, in the MC simulations, the random numbers representing the measurement error were generated with a normal distribution.

Model Parameter Error: The mean of the model parameter error ($E(\tilde{\varepsilon})$) was estimated based on the mean model output for predicted subsurface drainage reported by Haan and Skaggs (2003). In their work, estimates of the mean of model output were developed using both First-Order Analysis (FOA) and MC methods (Haan and Skaggs 2003). If the model response is approximately linear, then the mean of predicted values is well approximated by the predicted value obtained using the mean of the parameter values. For this condition, $E(\tilde{\varepsilon}) = 0$. Model response is, in general, nonlinear. The following relationship was used between the mean model output calculated using FOA and MC simulations to estimate $E(\tilde{\varepsilon})$:

$$relative\ error\ of\ E(\tilde{\varepsilon}) = \frac{E(model\ output)_{FOA} - E(model\ output)_{MC}}{E(model\ output)_{FOA}} \tag{5.2}$$

For our analysis, we estimated the expected values in equation 4.2 using the analysis of Haan and Skaggs (2003). They reported a value of 35.3 for mean of the output for FOA and 34.4 for the mean of the predicted value using MC simulations. The relative error is then 0.026.

Based on the relative error, the mean of the parameter error is given by equation 5.3.

$$E(\tilde{\varepsilon}) = \mu_{**} - \left(\frac{\mu_{**}}{1 + RE}\right) \tag{5.3}$$

For the simulation of subsurface drainage in southern Minnesota, the data given in Table 1 are used. By using a value of $\mu_{**} = 12.67$ cm/yr, we estimate $E(\tilde{\varepsilon}) = 0.3$ for the simulation of subsurface drainage in southern Minnesota.

The variance of the model parameter error ($VAR(\tilde{\varepsilon})$) was estimated using the coefficient of variation (C_v) obtained by Haan and Skaggs (2003) for the MC method. The coefficient of variation was converted into a variance and scaled to the magnitude of the values in table 5, by multiplying the reported value of 0.14 by a mean model output for plot 2 in table 5. DRAINMOD (Skaggs et al. 2012) was used to simulate plot 2 from table 5; details on the DRAINMOD simulation are described later in the paper. The estimated mean model output corresponding to plot 2 from table 5 was equal to 10.8 cm/year, resulting in the variance of the model parameter error equal to 2.4.

Model algorithm error: One of the most difficult parameters to estimate is the error in the model algorithms. This error is usually confounded by measurement errors and uncertainties in model parameters. Future expansion of the theoretical framework has the potential to be useful in understanding the algorithm errors from observed data. This application is beyond the scope of this paper.

Some researchers have evaluated the limitations of a simple model by comparing its results to those predicted by a more complex model (Mein and Larson 1978, Wilson and Ruffini 1988, Rice and Larson 1974). If the predictions from the complex model are taken as "observed" values, and if the parameters of the two models have no uncertainties because they are directly related to each other, then the predicted differences can be used to assess algorithm errors of the simpler model. In this paper, the algorithm error was evaluated using the simulated data of Mein and Larson (1971). Mein and Larson compared the infiltration obtained by using the simple Green-Ampt-Mein-Larson model to that obtained by the complex infiltration model obtained by solving Richards equation (Haan et al., 1982). Algorithm error was estimated from this data set by calculating the average percent difference between the cumulative infiltration depths using Richards equation to those obtained using Green-Ampt-Mein-Larson model.

Monte Carlo Simulations

Monte Carlo simulations were used to determine: 1) the values of MSE and NMSE corresponding to the "observed" MSE and NMSE from equation 2.1 using the random and non-random parameters discussed in the previous section, 2) the adjusted MSE and NMSE values that accounts for the errors in measurements and parameter uncertainty ($MSE_{\varepsilon}, MSE_{\varepsilon+\tilde{\varepsilon}}, NMSE_{\varepsilon}$ and $NMSE_{\varepsilon+\tilde{\varepsilon}}$ from equations 3.1 and 3.2), and 3) adjusted values of MSE and NMSE using the theory of Wilson ($\widetilde{MSE}_{\varepsilon}, \widetilde{MSE}_{\varepsilon+\tilde{\varepsilon}}, \widetilde{NMSE}_{\varepsilon}$, and $\widetilde{NMSE}_{\varepsilon+\tilde{\varepsilon}}$ from equations 4.1 and 4.2) Simulations were run for each combination of the error statistics shown in table 6.

Before running each of these combinations, the number of MC simulations needed to obtain a good estimate of the statistical properties of the generated adjusted MSE and NMSE terms was determined. For a single combination of the error terms, 1K, 10K, 20K and 30K simulations were evaluated 10 times. The minimum number of simulations needed to obtain correct statistical properties of the GOF measures was determined based on convergence of the mean of the difference between the true and theoretical GOF measures towards a single value.

Once the minimum number of MC simulations was determined, that number of simulations was run for each of the 36 combinations of random terms from table 6. Each

MC simulation generated the following terms, where the subscript "i" denotes an individual simulation:

1. The unadjusted MSE and NMSE with no error terms removed (MSE_{MC} and $NMSE_{MC}$). This value is analogous to MSE_{obs} and $NMSE_{obs}$ from equations 2.1 a and b; however those terms are non-random numbers for a given set of observed and predicted data, whereas $MSE_{i,MC}$ and $NMSE_{i,MC}$ are random numbers generated by MC simulation:

$$MSE_{i,MC} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2}{ST} = \frac{SSE_{i,MC}}{ST} \tag{6.1a}$$

$$NMSE_{i,MC} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2}{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}_{**}\right)^2} = \frac{SSE_{i,MC}}{SSTO_{i,MC}} \tag{6.1b}$$

2. The MSE and NMSE adjusted for measurement error ($MSE_{i,\varepsilon}$ and $NMSE_{i,\varepsilon}$), and measurement and parameter error ($MSE_{i,\varepsilon+\tilde{\varepsilon}}$ and $NMSE_{i,\varepsilon+\tilde{\varepsilon}}$) by removing the value of the generated random number (from equations 3.1 and 3.2)
3. The MSE and NMSE adjusted using the expected-value theory for measurement error ($\widetilde{MSE}_{i,\varepsilon}$ and $\widetilde{NMSE}_{i,\varepsilon}$), and measurement and parameter error ($\widetilde{MSE}_{i,\varepsilon+\tilde{\varepsilon}}$ and $\widetilde{NMSE}_{i,\varepsilon+\tilde{\varepsilon}}$) (from equations 4.1 and 4.2)
4. The difference between the MSE and NMSE adjusted using the known random numbers (MSE and NMSE) and the theory ($\widetilde{MSE}$ and $\widetilde{NMSE}$); and the difference between the adjusted using the known random numbers and the unadjusted (no errors taken into consideration) MSE and NMSE (MSE_{MC} and $NMSE_{MC}$). These difference terms are defined as:

$$\Delta MSE_i = MSE_i - \widetilde{MSE_\iota} \tag{7.1a}$$

$$\Delta NMSE_i = NMSE_i - \widetilde{NMSE_\iota} \tag{7.1b}$$

$$\Delta MSE_{i,MC} = MSE_i - MSE_{i,MC} \tag{7.2a}$$

$$\Delta NMSE_{i,MC} = NMSE_i - NMSE_{i,MC} \tag{7.2b}$$

Where the generalized notation of $MSE, \widetilde{MSE}, NMSE,$ and $\widetilde{NMSE}$ are used to allow the above relationships for ΔMSE and $\Delta NMSE$ to be valid for consideration of both adjustments in measurement error and measurement + parameter error.

The overall mean and standard deviation for values generated for each MC simulation for the terms listed above were also calculated. The expected value notation is used to denote the mean of the MC simulations; for example, the overall mean for all MC simulations for the difference between the MSE and $\widetilde{MSE}$ is $E(\Delta MSE_i)$.

Comparisons among MSE and NMSE Values

The similarity between the adjusted MSE and NMSE using known random parameters (MSE and $NMSE$) and those obtained by the theoretical approach of Part I ($\widetilde{MSE}$ and $\widetilde{NMSE}$) was determined by the difference between the two terms (ΔMSE and $\Delta NMSE$ from equations 7.1 a and b). A similar comparison was done between the same MSE and NMSE and those obtained without adjusting for errors (MSE_{MC} and $NMSE_{MC}$), where this difference is defined in equations 7.2 a and b as ΔMSE_{MC} and $\Delta NMSE_{MC}$. A value of zero indicated a perfect representation of the method; whereas larger values indicate problems with the approach. The overall mean and the standard deviation of the difference term from all MC simulations were used to calculate the probability that the adjusted GOF using the theoretical method was similar to that obtained with known values for the random errors.

The accuracy of the proposed method was evaluated by considering the probability that the difference between $E(NMSE_i)$ and $E(\widetilde{NMSE_i})$ was less than or equal to 0.1. The value of 0.1 is rather arbitrary but it does represent a relatively close value for hydrologic and water quality models (Moriasi et al. 2007). For the adjusted MSE, the threshold for accuracy assessment was based on the coefficient of variation around the predicted data. The probability was determined for the differences between $E(MSE_i)$ and $E(\widetilde{MSE_i})$ that were less than or equal to the variance of the non-adjusted predicted value for a specified coefficient of variation. The coefficient of variation used was equal to 0.15, while the mean predicted value was chosen to equal the mean value predicted by DRAINMOD for plot 2, equal to 13.11 cm/year (table 5).

***Assumptions of Normality*:** The probability calculations assumed normally distributed values for $\Delta MSE_i, \Delta NMSE_i, \Delta MSE_{i,MC}$ and $\Delta NMSE_{i,MC}$. These difference terms were calculated for values of MSE and NMSE generated in the MC simulations using random errors were which were generated using a normal distribution. To check for the impact that the probability distribution function of the random errors had on the distribution for the difference terms, a MC simulation was also run for a single combination of variables where in addition to generating the random error terms were assuming a normal distribution, they were also generated assuming a non-normal, uniform distribution.

Results

Accuracy of Adjusted MSE and NMSE using Expected Value Approach

Number of Monte Carlo simulations

The first step in the Monte Carlo simulations was to determine the number of runs necessary to obtain robust statistics of the results. The mean of the difference between the actual and theoretical adjusted MSE and NMSE showed convergence with increasing number of simulations (fig. 17). The graph shows the error between the mean differences for each run relative to the mean value obtained over all 10 of the 30K. For the MSE measurement error, the mean value from the 30K simulations was equal to -0.0028; for the NMSE measurement and parameter error the mean difference over all simulations of 30K was equal to 0.000457. As shown in figure 1, 20K simulations provided robust estimates of the statistical characteristics of the adjusted MSE and NMSE measures and allowed MC simulations to be completed within a reasonable computational time.

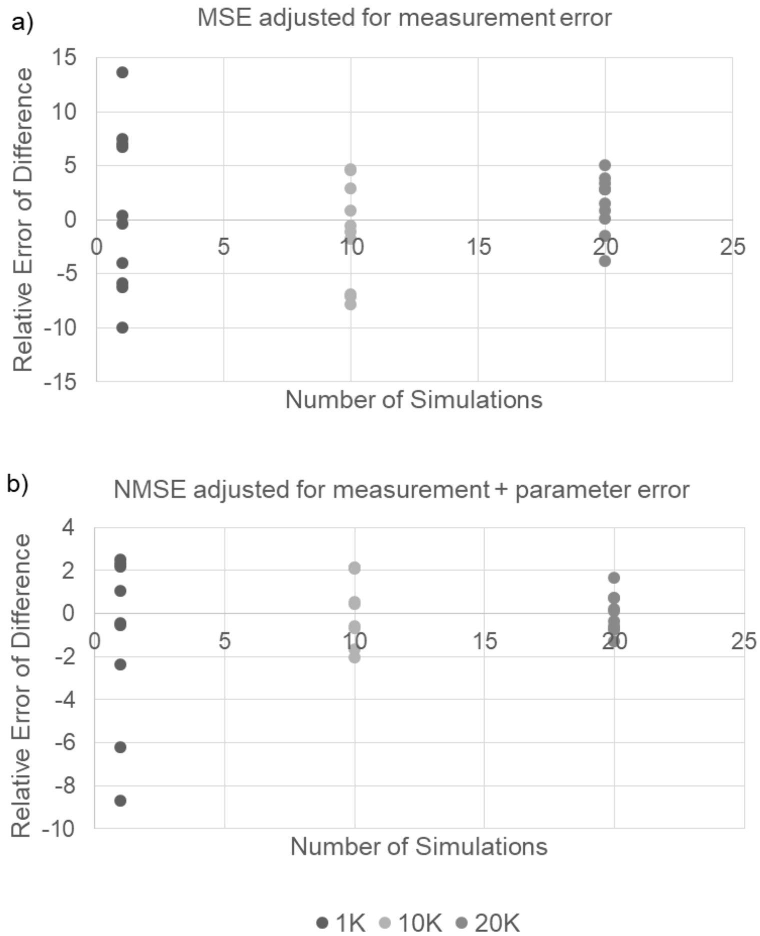

Figure 17: Convergence of the mean difference between the GOF calculated using the expected value-theory approach vs the actual values of the random numbers for: a) the MSE adjusted for measurement error; and b) the NMSE adjusted for measurement and parameter error.

Mean and coefficient of variation of the true and theory adjusted MSE and NMSE

A comparison of the mean of the adjusted GOF using the theoretical approach of Wilson (i.e.: $E(\widetilde{MSE}_{i,\varepsilon}), E(\widetilde{MSE}_{i,\varepsilon+\tilde{\varepsilon}}), E(\widetilde{NMSE}_{i,\varepsilon})$, and $E(\widetilde{NMSE}_{i,\varepsilon+\tilde{\varepsilon}})$) and that obtained using known random parameters (i.e.: $E(MSE_{i,\varepsilon}), E(MSE_{i,\varepsilon+\tilde{\varepsilon}}), E(NMSE_{i,\varepsilon})$ and $E(NMSE_{i,\varepsilon+\tilde{\varepsilon}})$) is shown in figure 18. As an average, the theory does a good job of approximating the GOF adjusted using known values of the error terms. As shown in figure 18, whether adjusted for measurement error only or measurement and parameter error, the values of $E(MSE_i)$ and $E(\widetilde{MSE}_i)$ are nearly identical, as are the values of $E(NMSE_i)$ and $E(\widetilde{NMSE}_i)$, with nearly all points falling on the 1:1 line.

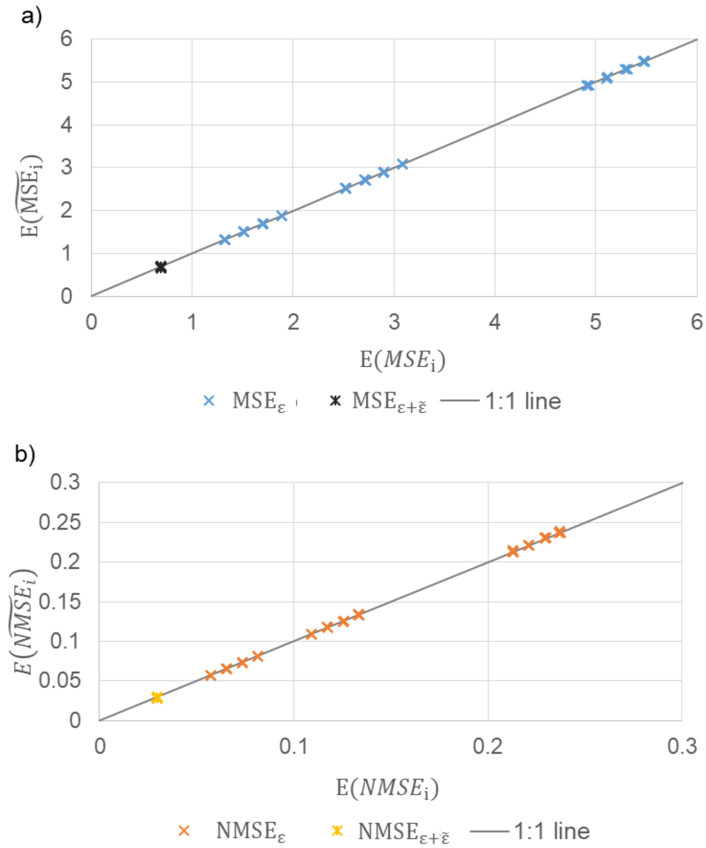

Figure 18: Mean value from MC simulations of the GOF obtained using known random numbers and using the theory of Wilson for: a) MSE adjusted for measurement error (MSE_ε), and measurement and parameter error ($MSE_{\varepsilon+\tilde{\varepsilon}}$), and b) NMSE adjusted for measurement error ($NMSE_\varepsilon$), and measurement and parameter error ($NMSE_{\varepsilon+\tilde{\varepsilon}}$).

Although there is close agreement in the overall mean of the MC simulation, the variability in adjusted GOF is also useful in evaluating the accuracy of the theoretical approach of Part I. The coefficient of variation for the results of MC simulations for GOF adjusted using the known random values (C_v for GOF of $MSE_\varepsilon, MSE_{\varepsilon+\tilde{\varepsilon}}, NMSE_\varepsilon$, and $NMSE_{\varepsilon+\tilde{\varepsilon}}$) and using the theoretical approach (C_v for GOF of $\widetilde{MSE}_\varepsilon, \widetilde{MSE}_{\varepsilon+\tilde{\varepsilon}}, \widetilde{NMSE}_\varepsilon, and\ \widetilde{NMSE}_{\varepsilon+\tilde{\varepsilon}}$) are shown in figures 19a and b. Figure 19a shows the variability in GOF when uncertainties in measurement errors are removed, and

Figure 19b shows the variability when both uncertainties in measurement errors and parameter uncertainties are removed. Since the variability in MSE_i and $NMSE_i$is directly related to random terms in the relationships, the removal of uncertainties in both measurement and parameter uncertainty results in very small variabilities in adjusted GOF, that is, the only term in the GOF is the non-random algorithm error. However for the theory of Part I, the variability includes the random values of $SSE_{i,MC}$ and $SSTO_{i,MC}$ as defined in equation 6.1. The coefficient of variation for the $\widetilde{MSE}$ and $\widetilde{NMSE}$ statistics are therefore quite high. This is especially pronounced for the adjustment which takes in to account both measurement and parameter error (fig. 19b). The high C_v indicates that the user may have less confidence in the accuracy of the theoretical approach for their specific simulation (representative of 1 out of the 20K MC simulations) in matching the corresponding actual value.

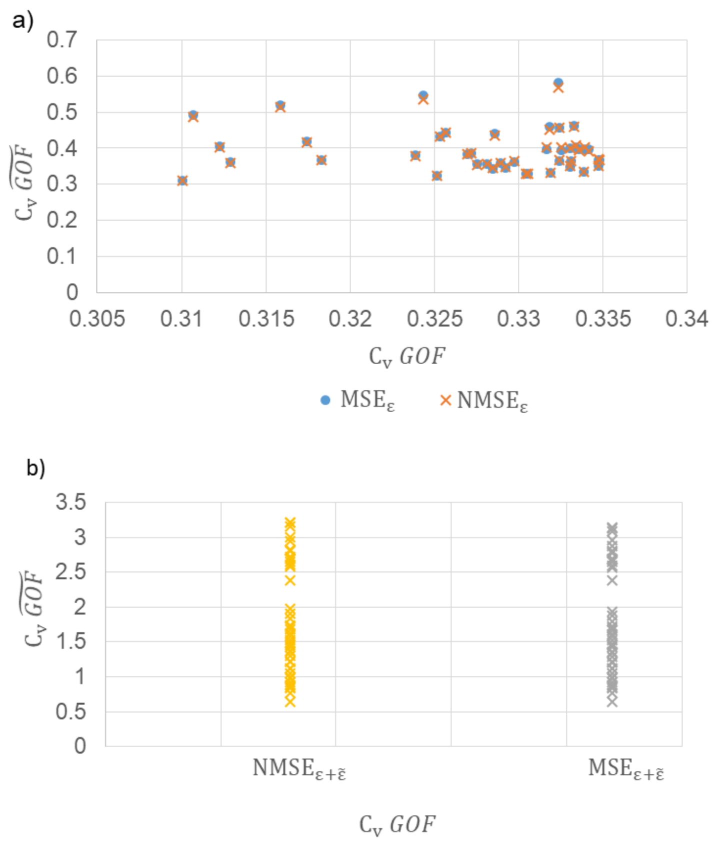

Figure 19: Coefficient of variation calculated from MC simulations for the known random numbers (MSE and $NMSE$) vs theoretical MSE and NMSE ($\widetilde{MSE}$, $\widetilde{NMSE}$) for: a) measurement error only (MSE_ε and $NMSE_\varepsilon$), and b) measurement and parameter error ($MSE_{\varepsilon+\tilde{\varepsilon}}$ and $NMSE_{\varepsilon+\tilde{\varepsilon}}$). In figure 19a, the C_v for MSE and NMSE adjusted for measurement error is similar for both simulations using known random errors (C_v GOF) and those using the theoretical approach ($C_v\ \widetilde{GOF}$). However, in figure 19b, there is only one value of C_v for MSE and NMSE adjusted for measurement error and parameter error for MC simulations using known random errors, while the C_v varies for the theoretical approach. This large difference in the variance of the GOF using known random errors vs. the theoretical approach indicates that the user may have less confidence in the accuracy of the theoretical approach.

Probability Assessment for the Accuracy of Theory and Unadjusted GOF

Insight into the accuracy of the adjusted GOF using the theory of Part I (i.e. $\widetilde{MSE}_\varepsilon, \widetilde{MSE}_{\varepsilon+\tilde{\varepsilon}}, \widetilde{NMSE}_\varepsilon$, and $\widetilde{NMSE}_{\varepsilon+\tilde{\varepsilon}}$) can be obtained by computing the probability that the GOF adjusted using the theoretical approach is similar to that using known random numbers. This probability is computed using the difference measure, ΔMSE_i and

$\Delta NMSE_i$ from equation 7.1. Since the difference measure is computed using the sum of random variables, it is expected that this measure is normally distributed due to the Central Limit Theorem (Beck and Arnold, 1977). The relative frequency plot as well the normal distribution of the difference measure is shown in figure 20. The difference measure between the theory and random number adjusted GOF is well approximated by the normal distribution when the random errors terms were generated as normally distributed values and when generated by a uniform distribution.

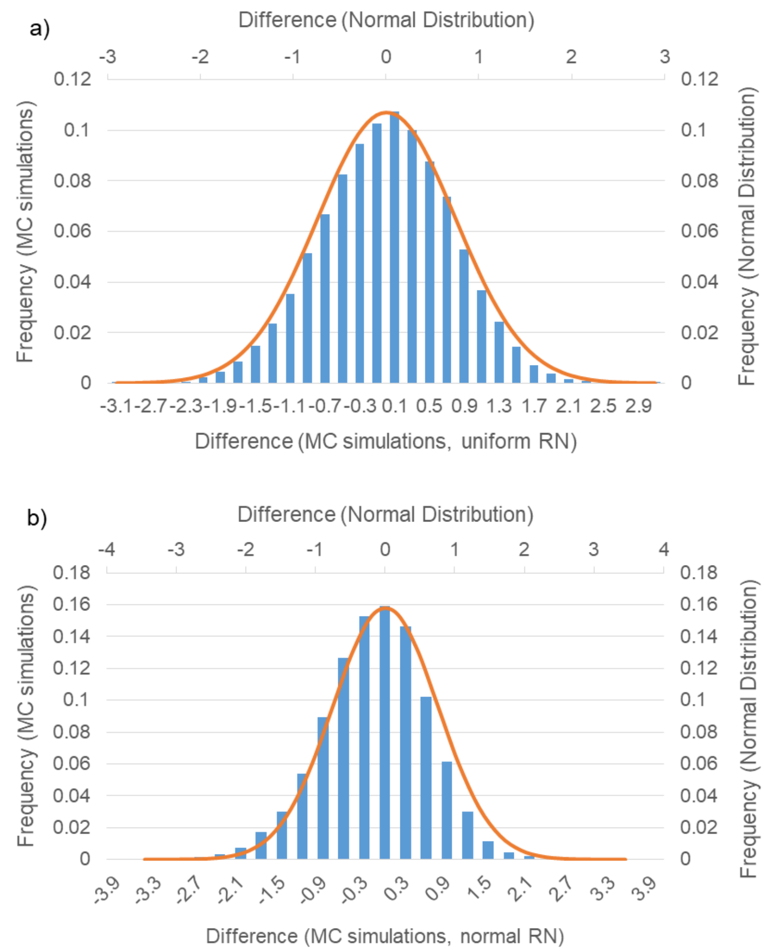

Figure 20: Distribution of $\Delta NMSE_{i,\varepsilon+\tilde{\varepsilon}}$ for a) random numbers (RN) generated with a normal distribution, and b) random numbers generated with a uniform distribution. Bars show the frequency of the difference terms generated via MC simulation, while the normal curve shows the corresponding normal distribution. Distributions are shown for the MC simulations which used the values of: $E(\tilde{\varepsilon}) = 0.6, VAR(\tilde{\varepsilon}) = 4.8, and\ VAR(\varepsilon) = 0.5$.

For a difference measure that is normally distributed, the probability of difference being greater than or equal to any specified value can be computed for a known mean and variance. These computations were performed for the NMSE and MSE. The results are discussed below.

NMSE: Table 7 and figure 21 show the probability that the theoretically-adjusted NMSE ($\widetilde{NMSE}_{i,\varepsilon}$ and $\widetilde{NMSE}_{i,\varepsilon+\tilde{\varepsilon}}$), and the NMSE without any adjustment made ($NMSE_{i,MC}$), were less than or equal to 0.1 of the actual adjusted NMSE for a subset of the 36 combinations of error terms used in the MC simulations. A value close to one suggests that all of the simulations are well approximated by theory of Part I. A value close to zero suggest that none of the simulations are well approximated by this theory.

Table 7: Probability that $E(\Delta NMSE_{i,\varepsilon})$, $E(\Delta NMSE_{i,\varepsilon+\tilde{\varepsilon}})$, $E(\Delta NMSE_{i,MC,\varepsilon})$ and $E(\Delta NMSE_{i,MC,\varepsilon+\tilde{\varepsilon}})$ are less than or equal to 0.1.

			Probability for NMSE: measurement error		Probability for NMSE: measurement + parameter error	
$E(\tilde{\varepsilon})$	$VAR(\tilde{\varepsilon})$	$VAR(\varepsilon)$	$E(\Delta NMSE_{i,\varepsilon})$	$E(\Delta NMSE_{i,MC,\varepsilon}$	$E(\Delta NMSE_{i,\varepsilon+\tilde{\varepsilon}})$	$E(\Delta NMSE_{i,MC,\varepsilon+\tilde{\varepsilon}})$
0	1.2	0.25	1.00	1.00	1.00	0.91
		0.5	1.00	1.00	1.00	0.81
		1	1.00	0.98	0.99	0.60
	2.4	0.25	1.00	1.00	0.97	0.39
		0.5	1.00	1.00	0.95	0.32
		1	0.99	0.95	0.92	0.23
	4.8	0.25	1.00	1.00	0.78	0.08
		0.5	0.99	0.99	0.75	0.07
		1	0.94	0.89	0.71	0.06
0.6	1.2	0.25	1.00	1.00	1.00	1.00
		0.5	1.00	1.00	1.00	0.98
		1	1.00	0.99	1.00	0.85
	2.4	0.25	1.00	1.00	0.99	0.61
		0.5	1.00	1.00	0.98	0.51
		1	1.00	0.97	0.95	0.36
	4.8	0.25	1.00	1.00	0.82	0.11
		0.5	1.00	0.99	0.80	0.10
		1	0.95	0.90	0.75	0.08

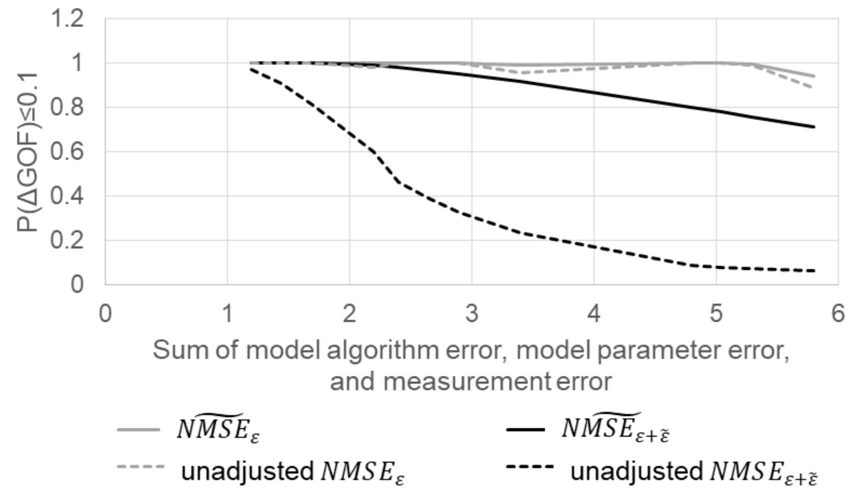

Figure 21: Probability that the difference between the NMSE (P(ΔGOF)) calculated using conventional methods with no errors taken into consideration (unadjusted $NMSE_\varepsilon$ or $NMSE_{\varepsilon+\tilde{\varepsilon}}$) and that adjusted using the expected value approach ($\widetilde{NMSE}_\varepsilon$ or $\widetilde{NMSE}_{\varepsilon+\tilde{\varepsilon}}$), is less than or equal to 0.1 of the true NMSE solution.

The NMSE adjusted for measurement error using the approach given in Part I has a very high probability of being within 0.1 of the NMSE adjusted using known values of the random error, as shown in the table for a high probability for the difference term $E(\Delta NMSE_{i,\varepsilon})$ being less than or equal to 0.1. The probability decreases with increases in the error terms, but remains high (94% or greater) for all combinations of variables run here. The NMSE without any adjustments also shows high probability of being within 0.1 of the actual NMSE adjusted for measurement error. This is also shown in the table with the high probability that the difference term, $E(\Delta NMSE_{i,MC,\varepsilon})$ is less than or equal to 0.1. Though slightly lower than the probability of the theoretically adjusted values, it is also has a quite high (89% or greater) probability of being within 0.1 of the actual adjusted NMSE for measurement error.

In contrast to the NMSE adjusted for measurement error only, the theoretically adjusted NMSE is much more likely than the unadjusted to be within 0.1 of the actual adjusted term (high probability of the difference term $E(\Delta NMSE_{i,\varepsilon+\tilde{\varepsilon}})$ being less than or equal to 0.1). With the highest values of error ($VAR(\varepsilon) = 1, VAR(\tilde{\varepsilon}) = high, and\ E(\tilde{\varepsilon}) = high$), the theoretically adjusted value has a 71% probability of being within 0.1 of the actual adjusted NMSE. The unadjusted value has a very low probability of being close to the actual NMSE adjusted for measurement and parameter error; for the same high variance in parameter error, the unadjusted value only has a 6% probability of being within 0.1 of the actual NMSE.

In figures 22 and 23, we can see that the lower probability for the unadjusted NMSE accurately representing the value adjusted using known random variables, is larger due to the difference in their means. Figure 22 shows the standard deviations of the difference terms plotted against each other ($\Delta NMSE_i$ vs. $\Delta NMSE_{i,MC}$). From this figure it is apparent that the range in $\Delta NMSE_i$ and $\Delta NMSE_{i,MC}$ is similar. The lower probabilities for $\Delta NMSE_{i,MC}$ compared to $\Delta NMSE_i$ shown in table 7 can be explained by figure 23, which shows the mean value of the difference terms for measurement error ($\Delta NMSE_{i,\varepsilon}$ and $\Delta NMSE_{i,MC,\varepsilon}$) (fig. 23a), and the mean value of the difference terms for measurement error and parameter error ($\Delta NMSE_{i,\varepsilon+\tilde{\varepsilon}}$ and $\Delta NMSE_{i,MC,\varepsilon+\tilde{\varepsilon}}$) (fig. 23b). The differences obtained using the theoretical approach ($\Delta NMSE_{i,\varepsilon}$ and $\Delta NMSE_{i,\varepsilon+\tilde{\varepsilon}}$) are shown as circles; those obtained using NMSE not adjusted for random errors ($\Delta NMSE_{i,MC,\varepsilon}$ and $\Delta NMSE_{i,MC,\varepsilon+\tilde{\varepsilon}}$) as dashes. For measurement error, the difference NMSE and MSE term is still affected by the value of the model parameter error, even though it is not accounted for in the adjustment. It is clear from the figure 23 that for the combination of measurement error and parameter error in particular, the unadjusted NMSE does a poor job of representing the NMSE adjusted with the known random error terms, with values substantially larger than zero.

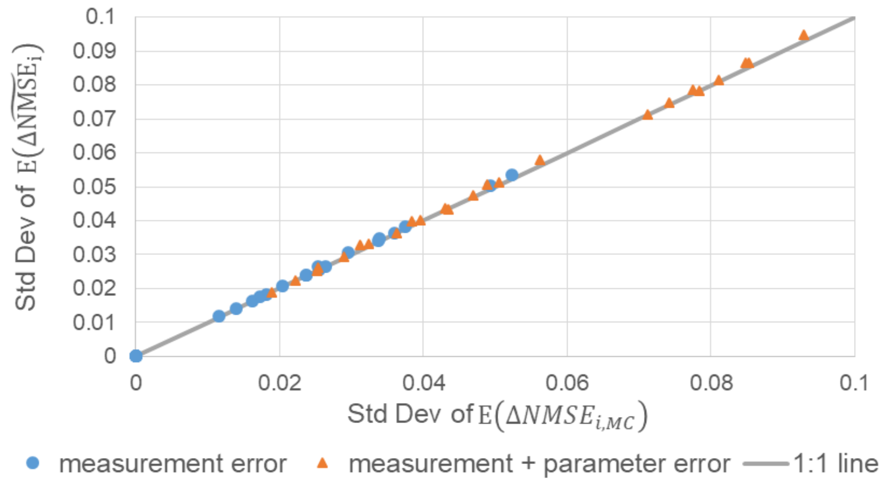

Figure 22: Standard deviation from MC simulations of the difference between the NMSE adjusted using known random numbers and unadjusted vs. the difference for NMSE adjusted using known random numbers and adjusted using the expected value approach.

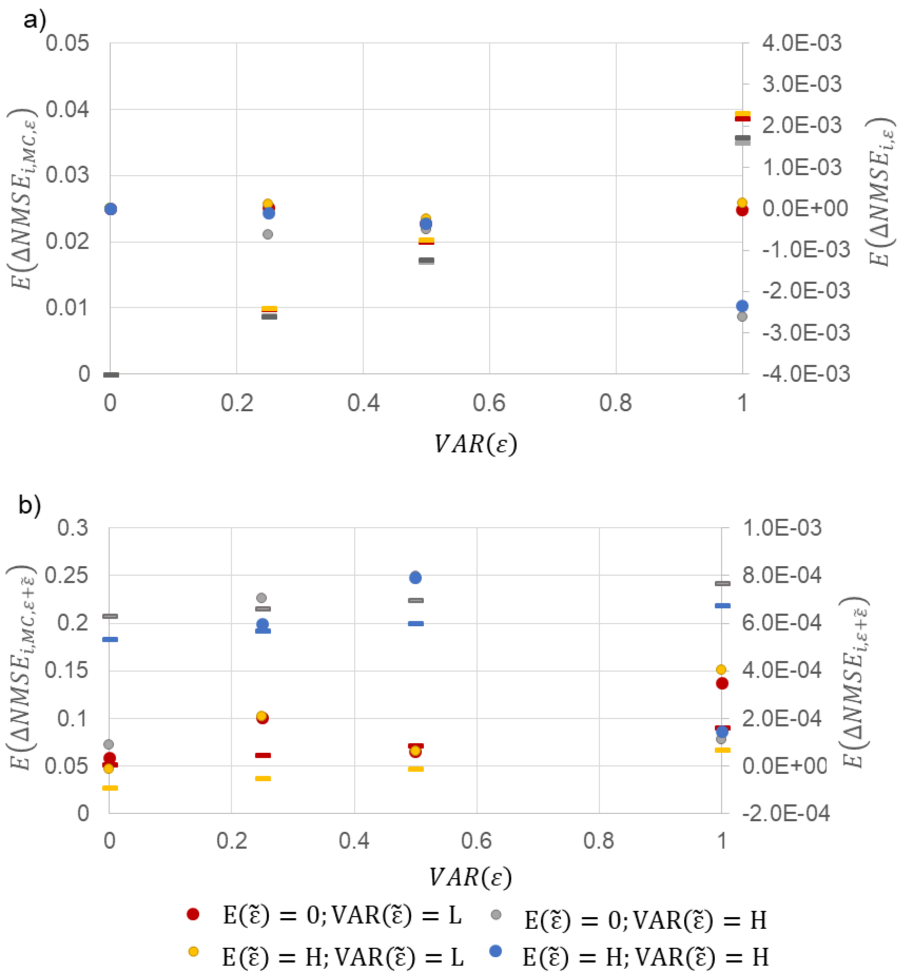

Figure 23: The mean difference between the NMSE adjusted using known random numbers and the expected value theory ($E(\Delta NMSE_i)$ represented by circles), and the difference in the NMSE adjusted using known random numbers and the NMSE not adjusted for random errors ($E\left(\Delta NMSE_{i,MC}\right)$ represented by dashes). The difference when adjusting for measurement error only is shown in figure 23a, while the difference when adjusting for both measurement and parameter errors is shown in figure 23b. The different colors correspond to different combinations of the mean and variance of the model parameter error (table 6).

MSE: Table 8 shows the probability that the theoretically-adjusted MSE ($\widetilde{MSE}_{i,\varepsilon}$ and $\widetilde{MSE}_{i,\varepsilon+\tilde{\varepsilon}}$), and the MSE without any adjustment made ($MSE_{i,MC}$), were within 2.83 of the actual adjusted MSE for a subset of the variable combinations run in the MC simulations. The value of 2.83 is the variance for the predicted mean equal to 13.11 for a coefficient of variation of 0.15.

Table 8: Probability that $E\left(\Delta MSE_{i,\varepsilon}\right)$, $E\left(\Delta MSE_{i,\varepsilon+\tilde{\varepsilon}}\right)$, $E\left(\Delta MSE_{i,MC,\varepsilon}\right)$ and $E\left(\Delta MSE_{i,MC,\varepsilon+\tilde{\varepsilon}}\right)$ are less than or equal to 2.83.

			Probability for MSE: measurement error		Probability for MSE: measurement + parameter error	
$E(\tilde{\varepsilon})$	$VAR(\tilde{\varepsilon})$	$VAR(\varepsilon)$	$E\left(\Delta MSE_{i,\varepsilon}\right)$	$E\left(\Delta MSE_{i,MC,\varepsilon}\right)$	$E\left(\Delta MSE_{i,\varepsilon+\tilde{\varepsilon}}\right)$	$E\left(\Delta MSE_{i,MC,\varepsilon+\tilde{\varepsilon}}\right)$
0	1.2	0.25	1.00	1.00	1.00	0.77
		0.5	1.00	1.00	0.99	0.62
		1	1.00	1.00	0.96	0.39
	2.4	0.25	1.00	1.00	0.92	0.26
		0.5	1.00	1.00	0.90	0.21
		1	1.00	1.00	0.85	0.14
	4.8	0.25	1.00	1.00	0.70	0.05
		0.5	1.00	1.00	0.67	0.05
		1	1.00	1.00	0.63	0.04
0.6	1.2	0.25	1.00	1.00	1.00	0.98
		0.5	1.00	1.00	1.00	0.91
		1	1.00	1.00	0.99	0.65
	2.4	0.25	1.00	1.00	0.96	0.44
		0.5	1.00	1.00	0.95	0.35
		1	1.00	1.00	0.90	0.22
	4.8	0.25	1.00	1.00	0.74	0.07
		0.5	1.00	1.00	0.72	0.06
		1	1.00	1.00	0.67	0.05

For the MSE adjusted for measurement error only, both the theory-adjusted and unadjusted values show good fit to the actual value for the specified range. Similar to the results of NMSE, the theory-adjusted values show much closer fit to the actual values than the unadjusted when both measurement error and parameter uncertainties are included. For the highest level of uncertainty in the observed and predicted data ($VAR(\varepsilon) = 1, VAR(\tilde{\varepsilon}) = 4.8, E(\tilde{\varepsilon}) = 0.6$), the theoretically adjusted MSE shows a 67% probability of being close to the true value, while the unadjusted value shows only a 5% probability.

In figures 24 and 25, we can see that the lower probability for the unadjusted MSE accurately representing the value adjusted using known random variables, is larger due to the difference in their means. Figure 24 shows the standard deviations of the difference terms (ΔMSE_i vs. $\Delta MSE_{i,MC}$), where the standard deviations have been normalized by the corresponding mean value obtained using known random numbers ($E\left(MSE_{i,\varepsilon}\right)$ or $E\left(MSE_{i,\varepsilon+\tilde{\varepsilon}}\right)$). From this figure it is apparent the standard deviations among the two approaches are very similar.

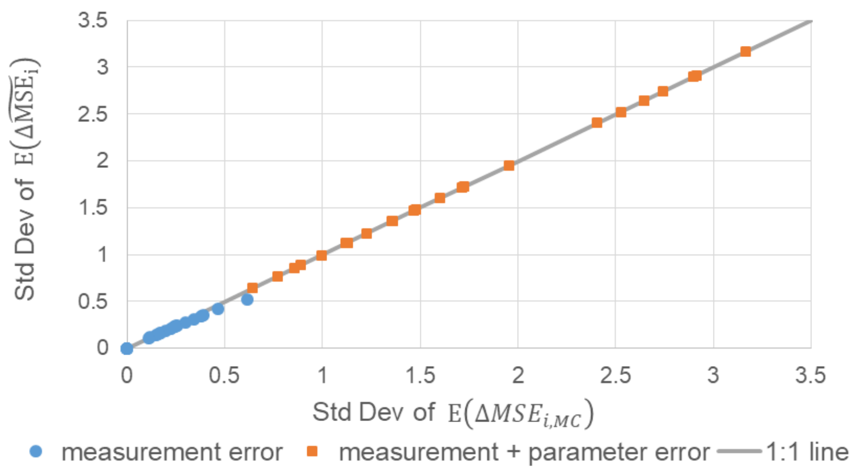

Figure 24: Standard deviation from MC simulations of the difference between the MSE adjusted using known random numbers and unadjusted vs. the difference for MSE adjusted using known random numbers and adjusted using the expected value approach.

The lower probabilities for $\Delta MSE_{i,MC}$ compared to ΔMSE_i shown in table 8 can be explained by figure 25, which shows the mean value of the difference terms for measurement error ($\Delta MSE_{i,\varepsilon}$ and $\Delta MSE_{i,MC,\varepsilon}$) (fig. 25a), and the mean value of the difference terms for measurement error and parameter error ($\Delta MSE_{i,\varepsilon+\tilde{\varepsilon}}$ and $\Delta MSE_{i,MC,\varepsilon+\tilde{\varepsilon}}$) (fig. 25b). All of the difference terms in figure 25 have been normalized by the corresponding mean value obtained using the known random numbers. The differences obtained using the theoretical approach ($\Delta MSE_{i,\varepsilon}$ and $\Delta MSE_{i,\varepsilon+\tilde{\varepsilon}}$) are shown as circles; those obtained using MSE not adjusted for random errors ($\Delta MSE_{i,MC,\varepsilon}$ and $\Delta MSE_{i,MC,\varepsilon+\tilde{\varepsilon}}$) as dashes. As with the NMSE, it is clear from the figure that for the combination of measurement error and parameter error in particular, the unadjusted MSE does a poor job of representing the MSE adjusted with the known random error terms. Interestingly, the values of $\Delta MSE_{i,MC,\varepsilon}$ are highest for the low values of the variance in the model parameter error, but the values of $\Delta MSE_{i,\varepsilon+\tilde{\varepsilon}}$ are highest for high values for the variance in the model parameter error.

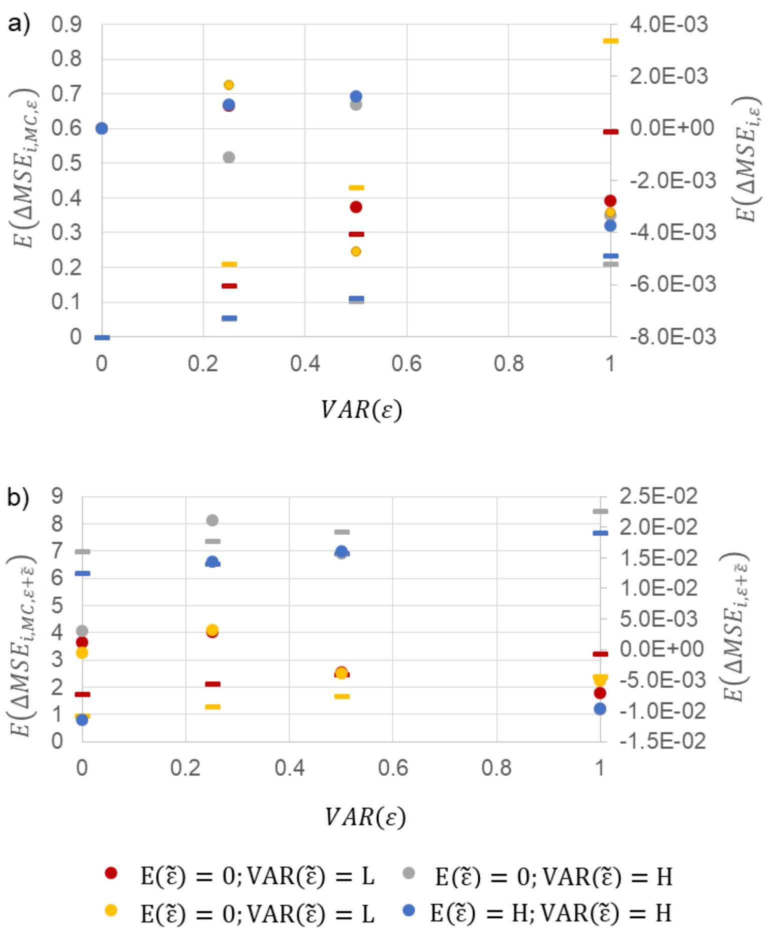

Figure 25: The mean difference between the MSE adjusted using known random numbers and the expected value theory ($E(\Delta MSE_i)$ represented by circles), and the difference in the MSE adjusted using known random numbers and the NMSE not

adjusted for random errors ($E(\Delta NMSE_{i,MC})$ represented by dashes). The difference when adjusting for measurement error only is shown in figure 25a, while the difference when adjusting for both measurement and parameter errors is shown in figure 25b.

Illustration of Method Using a Case Study

The adjusted MSE and NMSE proposed in Part I was calculated for model simulations and field observations were calculated for two locations in southern Minnesota: Waseca and Lamberton. These GOF statistics were then compared to the traditional definitions of MSE and NMSE given by equations 2.1 and 2.2. The hydrologic simulation model DRAINMOD was used to simulate the yearly subsurface drainage at these locations. DRAINMOD is a process-based, distributed, field-scale model, developed to describe the hydrology of poorly or artificially drained lands (Skaggs et al. 2012). It conducts water balances on hourly and daily time scales, and predicts hydrologic parameters including infiltration, runoff, evapotranspiration, seepage, water table depth, and subsurface drainage on a daily, monthly, and annual time step.

Experimental Sites

Waseca

The drainage plot simulated using DRAINMOD was located in southern Minnesota at the University of Minnesota Southern Research and Outreach Center, at Waseca, Minnesota. The drainage plot was designed with subsurface drainage installed at a depth and spacing of 1.2 m, resulting in a drainage intensity of 1.3 cm/day (Sands et al. 2008). Soils and topography found at the site are typical for south-central Minnesota. The major soil types at the site include Webster silty clay loam (fine-loamy, mixed, superactive, mesic Typic Endoaquolls) and Nicollet clay loam (fine-loamy, mixed, superactive, mesic Aquic Hapludolls). Average annual precipitation is approximately 889 mm; precipitation during the growing season (May through September) is approximately 533 mm (averages are the 30 year normal, 1982-2010). Plots were outfitted with tipping buckets to provide a continuous measurement of subsurface drainage flow rates for the years 2003-2008.

The plots were planted in a 5-year soy-corn-soy-corn-corn rotation, starting with soybean in 2003.

Initial input parameters describing the soil water and chemical properties of the Waseca plots were based on values determined by Luo et al. (2010). Soil samples taken from the site were used to determine model inputs including: soil particle distributions, soil water characteristic curves, and saturated hydraulic conductivity. Hydraulic conductivity showed great variability in both the lab measured samples and county soil survey data (Luo et al. 2010) and an average value was chosen to represent the site. Climatic data measured at the Waseca experiment station site for the years 2003-2008 was used as model input, including: daily maximum and minimum temperatures, and hourly precipitation.

Lamberton:

Measured drainage and nitrate loads for the years 1990-1998 were obtained from plot studies at the University of Minnesota Southwest Research and Outreach Center (SROC) in Lamberton, MN and used for DRAINMOD calibration. Plots at this site were planted in a corn-soybean rotation, and had dimensions of 13.7 m by 15.24 m, with subsurface tile drainage placed at a depth of 1.2 m, and spacing of approximately 27 m (personal communication, Jeff Strock, University of MN Extension). Soils at SROC in Lamberton are predominately poorly draining Webster silty clay loam (fine-loamy, mixed, superactive, mesic Typic Endoaquolls), moderately well-drained Normania loam (fine-loamy, mixed, superactive, mesic Aquic Hapludolls), and well drained Ves loam (fine-loamy, mixed, superactive, mesic Calcic Hapludolls) (Oquist et al., 2006). Average annual precipitation is approximately 709 mm; precipitation during the growing season (May through September) is approximately 497 mm. Climate input data, including hourly precipitation, daily maximum and minimum temperatures, and daily PET, were obtained from Professor Gary Sands' group at the University of Minnesota.

Model Performance using adjusted and un-adjusted MSE and NMSE

Model performance was evaluated using the expected value theoretical approach in equations 4.1 and 4.2, and using the unadjusted MSE and NMSE from equation 2.1. It was assumed that the expected value of the measurement error was equal to zero, and that the mean of the model parameter error was also equal to zero ($E(\varepsilon) = 0\ and\ E(\tilde{\varepsilon}) = 0$). Two sets of measurement error and measurement + parameter error were used. For the first, the values for the variance in the measurement error and model parameter error were assumed to be the same as the "medium" values used in generating data for the MC simulations reported in table 2, equal to 0.5 and 2.4, respectively. For the second set of probable errors, the values for the variance were assumed to be equal to the "high" values in table 2, equal to 1 for the variance of the measurement error and 4.8 for the variance of model parameter error. The value of the measurement error for subsurface flow is a good approximation of what would be likely for the measured flow data from both locations. The variance of the model parameter error is not specific to the DRAINMOD predicted dataset for Waseca and Lamberton, but is a reasonable approximation of likely model parameter error.

Comparison of GOF for DRAINMOD results using adjusted and un-adjusted MSE and NMSE

The results of the observed and DRAINMOD-predicted subsurface flow for Waseca and Lamberton are summarized in table 9. The predicted data for Waseca for the year 2006 shows significant deviation from the observed data, so for comparison of the adjusted and unadjusted GOF statistics, 2006 was not used in the GOF for Waseca.

Table 9: Results of the annual observed and DRAINMOD predicted subsurface drainage flow for two sites located in southern Minnesota: Waseca and Lamberton.

Subsurface Drainage (cm/year)					
Waseca			Lamberton		
Year	Observed	Predicted	Year	Observed	Predicted
2003	6.32	4.42	1990	2.07	6.96
2004	18.22	18.69	1991	27.44	23.43
2005	17.40	14.01	1992	17.72	7.75
2006	19.42	5.80	1993	48.90	35.33
2007	19.85	23.01	1994	13.34	13.1
2008	13.65	12.71	1995	23.14	16.03
AVG	**15.81**	**13.11**	1996	16.69	10.78
			1997	16.73	18.14
			1998	15.24	6.84
			AVG	**20.14**	**15.37**

For the low error, taking into account measurement error had only a small effect on the resulting adjusted MSE or NMSE for both locations, reducing the NMSE by only 0.01, and the MSE by less than one (table 10). Taking into account the measurement error and model input parameter error resulted in greater improvements to the MSE and NMSE. For the Waseca datasets, the adjustment resulted in a change of 0.1 to the NMSE, though for Lamberton, it only resulted in a 0.02 change. The MSE was improved by about a value of 3 for both locations.

Taking into account higher estimates of the error still had little effect on the adjusted NMSE and MSE when accounting for measurement error only. Once again, the representation of the measurement error was based on general statistics given in table 6. A rigorous assessment for this data set was not performed and the actual measurement error maybe larger. Accounting for both higher measurement error and model parameter error, resulted in more pronounced improvements in the adjusted GOF, particularly for Waseca. For Waseca, taking into account the higher measurement error only, reduced the

NMSE to less than zero. This negative result should be interpreted as being a perfect fit. The less pronounced improvements to the GOF for the Lamberton data compared to Waseca, is likely due to the differences in the variation in the given observed and predicted datasets. For Waseca, the SSE_{obs} was equal to 26.21, and $SSTO_{obs}$ equal to 117. For Lamberton, the SSE_{obs} was equal to 481, while the $SSTO_{obs}$ was equal to 1520. These higher values of SSE_{obs} and $SSTO_{obs}$ for Lamberton indicate that there is much greater variability in the observed and predicted data for Lamberton than for Waseca. The magnitude of error chosen for the variance in the measurement error and model parameter error were likely not a good representation for the circumstances at Lamberton. For Lamberton, the adjusted GOF values calculated here are likely not a good approximations as they do not take into account the actual model parameter uncertainty for the specific datasets, and they use the assumption that the mean of the model parameter error is equal to zero.

Table 10: GOF statistics calculated for Waseca (with the year 2006 removed), and for Lamberton. GOF statistics included the MSE and NMSE 1) not taking into account error terms (unadjusted), 2) adjusted for measurement error, and 3) adjusted for measurement and model input parameter error.

	Waseca		**Lamberton**	
GOF indicator	**Low errors**	**High errors**	**Low errors**	**High errors**
NMSE-unadjusted	0.22	0.22	0.32	0.32
NMSE adjusted measurement error	0.21	0.19	0.31	0.31
NMSE adjusted measurement + parameter error	0.10	-0.02	0.30	0.28
MSE-unadjusted	5.24	5.24	53.51	53.51
MSE adjusted measurement error	4.74	4.24	53.01	52.51
MSE adjusted measurement +parameter error	2.34	-0.56	50.61	47.71

Summary and Conclusions

In the previous chapter, we proposed a theoretical framework to adjust the MSE and NMSE statistics to account for measurement errors and parameter uncertainties. This framework used expected values of random variables and incorporated a single set of observed and predicted values. In this chapter, the limitations of using these conditions was explored. In addition, the application of the method was explored using a case study of a model simulation of drainage in southern Minnesota.

Monte Carlo simulations were used to generate values of MSE and NMSE for the following conditions: 1) adjusted using known values of random error, 2) adjusted using the expected value framework proposed in Part I (chapter 4), and 3) not adjusted for random errors. Parameters used to generate these terms in MC simulation were based on literature review, and were chosen to reflect the magnitude of their values for a subsurface drainage system.

On average, the adjustment derived using the expected-value theory from Part I for the mean square error and normalized mean square error matches the value obtained using known values of the random errors. However, the spread of the theory-adjusted MSE and NMSE is larger than the values obtained using the true random errors, as evidenced by their large coefficient of variation. Even with this variability, the probability that the theory-adjusted value is close to the actual is quite high, and the theory-adjusted MSE and NMSE are much more likely to match the actual values than the MSE and NMSE that are not adjusted for errors.

When applied to an actual dataset, the theory-adjusted MSE and NMSE improve the fit compared to the traditional approach which does not take into account random errors in the observed and predicted data. The adjusted GOF was much better for the data from Waseca than the data from Lamberton. This result indicates that the assumptions on the mean and variance of the model parameter error and variance of the measurement error that were used to assign values to the errors were acceptable for the Waseca data, but not for the Lamberton data. Despite this, these results show the potential that this new framework has for reporting model GOF values that are more accurate representations of true model performance.

Bibliography

Baker J.L., K.L. Campbell, H.P. Johnson and J.J. Hanway. (1975). Nitrate, phosphorus, and sulfate in subsurface drainage water. Journal of Environmental Quality 4:406-412.

Bakhsh, A., R.S. Kanwar, T.B. Bailey, C.A. Cambardella, D.L. Karlen, and T.S. Colvin. (2002). Cropping system effects on NO_3-N loss with subsurface drainage water. Transactions of the ASAE, 45(6), 1789-1797.

Beck, J.V., and K.J. Arnold. (1977). Parameter Estimation in Engineering and Science. John Wiley and Sons, New York, NY.

Beven, K., and A. Binley. (1992). The future of distributed models: model calibration and uncertainty prediction. Hydrological processes, 6(3), 279-298.

Bierman, P. M., C. J. Rosen, R. T Venterea, and J. A. Lamb. (2012). Survey of nitrogen fertilizer use on corn in Minnesota. Agricultural Systems, 109, 43-52.

Busman L.M. and G.R. Sands. (2002). Agricultural drainage: Issues and answers. University of Minnesota Extension Service.

Casella, G., and R. L. Berger. (2002). Statistical inference (Vol. 2). Pacific Grove, CA: Duxbury.

CENR (Committee on Environment and Natural Resources) (2010). Scientific assessment of hypoxia in U.S. coastal waters. Interagency working group on harmful algal blooms, hypoxia, and human health of the Joint Subcommittee on Ocean Science and Technology, Washington, DC.

David, M. B., L.E. Drinkwater, and G.F. McIsaac. (2010). Sources of nitrate yields in the Mississippi River Basin. Journal of Environmental Quality, 39(5), 1657-1667.

Des Moines Water Works. (2009). Fact sheet: Nitrate removal facility. Accessed 2013.

Dimitri C., A. Effland and N.C. Conklin. (2005). The 20th century transformation of US agriculture and farm policy. Rep. Economic Information Bulletin Number 3. US Department of Agriculture, Economic Research Service

Dinnes, D. L., D.L. Karlen, D. Jaynes, T.C. Kaspar, J. Hatfield, T. Colvin, and C. Cambardella. (2002). Nitrogen management strategies to reduce nitrate leaching in tile-drained Midwestern soils. Agronomy Journal, 94(1), 153-171.

Edwards, D. R. and C.T. Haan. (1989). Incorporating Uncertainty into Peak Flow Estimates. Transactions of the ASAE. 32(1): 113-119

Engel, B., D. Storm, M. White, J. Arnold, and M. Arabi. (2007). A Hydrologic/Water Quality Model Application. Journal of the American Water Resources Association, 43(5), 1223-1236.

Erickson, B., and D.A. Widmar. (2015). 2015 Precision agricultural services dealership survey results. Purdue University, West Lafayette, Indiana.

Evans R.O. and N.R. Fausey. (1999). Effects of inadequate drainage on crop growth and yield. p. 13-54. In R.W. Skaggs and J. Van Schilfgaarde (eds.) Agricultural drainage. American Society of Agronomy, Madison.

Follett R.F. (2008). Transformation and transport processes of nitrogen in agricultural systems. p. 19-50. In J.L. Hatfield and R.F. Follett (eds.) Nitrogen in the environment: Sources, problems, and management. 2nd ed. Elsevier, Inc.

Foster, G. R., C.T. Haan, H.P. Johnson, and D.L. Brakensiek, D. L. (1982). Hydrologic Modeling of Small Watersheds. ASAE Monograph #5, St. Joseph, MI.

Freer, J., K. Beven, and B. Ambroise. (1996). Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach. Water Resources Research 32(7): 2161-2173.

Garen, D. C. and S.J. Burges. (1981). Approximate error bounds for simulated hydrograph. Journal of Hydraulics Division, Proc. ASCE. 107(HY11): 1519-1534.

Gast, R. G., W. W. Nelson, and G.W. Randall. (1978). Nitrate accumulation in soils and loss in tile drainage following nitrogen applications to continuous corn. Journal of Environmental Quality, 7(2), 258-261.

Ghane, E., A. Z. Ranaivoson, G. W. Feyereisen, C. J. Rosen, and J. F. Moncrief. (2016). Comparison of contaminant transport in agricultural drainage water and urban storm water runoff. PLoS ONE 11(12), e0167834

Gershenfeld, N.A., and A.S. Weigend. (1993). The future of time series: Learning and understanding. In Time Series Prediction: Forecasting the Future and Understanding the Past. SFI Studies in the Sciences of Complexity, Proc Vol XV, Addison-Wesley

Guzman, J. A., A. Shirmohammadi, A.M. Sadeghi, X. Wang, M.L. Chu, M.K. Jha, P.B. Parajuli, R. D. Harmel, Y. P Khare, and J.E. Hernandez. (2015). Uncertainty Considerations in Calibration and Validation of Hydrologic and Water Quality Models. Transactions of the ASABE, 58(6), 1745-1762.

Haan, C. T. (1977). Statistical methods in hydrology. The Iowa State University Press.

Haan, C. T. (1989). Parametric uncertainty in hydrologic modeling. Transactions of the ASAE, 32(1), 137-0146.

Haan, C.T., B. Allred, D. E. Storm, G. J. Sabbagh, and S. Prabhu. (1995). Statistical Procedure for Evaluating Hydrologic/Water Quality Models. Transactions of the ASAE, 38(3), 725-733.

Haan, P. K., and R.W. Skaggs. (2003). Effect of parameter uncertainty on DRAINMOD predictions: I. Hydrology and yield. Transactions of the ASAE, 46(4), 1061.

Harmel, R. D., R. J. Cooper, R. M. Slade, R. L. Haney, and J. G. Arnold. (2006). Cumulative uncertainty in measured streamflow and water quality data for small watersheds. Transactions of the ASABE, 49(3), 689-701.

Harmel, R. D. and P. K. Smith. (2007). Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. Journal of Hydrology, 337(3), 326-336.

Harmel, R. D., P. K. Smith, and K. W. Migliaccio. (2010). Modifying Goodness-of-Fit Indicators to Incorporate Both Measurement and Model Uncertainty in Model Calibration and Validation. Transactions of the ASABE, 53(1), 55-63.

Harmel, R. D., P. K. Smith, K. W. Migliaccio, I. Chaubey, K. R. Douglas-Mankin, B. Benham, S. Shukla, R. Muñoz-Carpena, and B. J. Robson. (2014). Evaluating, interpreting, and communicating performance of hydrologic/water quality models considering intended use: A review and recommendations. Environmental modelling & software. 57, 40-51.

Harr, M.E. (1987). Reliability-Based Design in Civil Engineering. McGraw-Hill, New York

Herschy, R. W. (1971). The magnitude of errors at flow measurement stations. Water Resources Board.

Jaynes, D. B., T.S. Colvin, D. L. Karlen, C. A. Cambardella, and D. W. Meek. (2001). Nitrate loss in subsurface drainage as affected by nitrogen fertilizer rate. Journal of Environmental Quality, 30(4), 1305-1314.

Jaynes, D. B. (2015). Corn yield and nitrate loss in subsurface drainage affected by timing of anhydrous ammonia application. Soil Science Society of America Journal, 79(4), 1131-1141

Kladivko, E. J., J. R. Frankenberger, D. B. Jaynes, D. W. Meek, B. J. Jenkinson, and N. R. Fausey. (2004). Nitrate leaching to subsurface drains as affected by drain spacing and changes in crop production system. Journal of Environmental Quality, 33(5), 1803-1813.

Legates, D. R., and G.J. McCabe. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water resources research, 35(1), 233-241.

Luo, W., G. R. Sands, M. Youssef, J. S. Strock, I. Song, and D. Canelon. (2010). Modeling the impact of alternative drainage practices in the northern Corn-belt with DRAINMOD-NII. Agricultural Water Management, 97(3), 389-398.

Mamo, M., G. L. Malzer, D. J. Mulla, D. R. Huggins, and J. S. Strock. (2003). Spatial and temporal variation in economically optimum nitrogen rate for corn. Agronomy Journal, 95(4), 958-964.

Mantovan, P. and E. Todini. (2006). Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. Journal of hydrology, 330(1-2), 368-381

Mein, R. G., and C.L. Larson. (1971). Modeling the infiltration component of the rainfall-runoff process. Water Resources Research Center, Bulletin 43. Minneapolis, MN.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3), 885-900.

Moriasi, D.N., R. W. Zeckoski, J. G. Arnold, C. B. Baffaut, R. W. Malone, P. Daggupati, J. A. Guzman, D. Saraswat, Y. Yuan, B. N. Wilson, A. Shirmohammadi, and K. R. Douglas-Mankin. (2015). Hydrologic and water quality models: Key calibration and validation topics. Transactions of the ASABE, 58(6), 1609-1618

MPCA (Minnesota Pollution Control Agency). (2013). Nitrogen in Minnesota surface waters: Conditions, trends, sources, and reductions. Minnesota Pollution Control Agency, Saint Paul, Minnesota.

Mulla, D. J., and T.M. Addiscott. (1999). Validation Approaches for Field-, Basin-, and Regional-Scale Water Quality Models. Assessment of non-point source pollution in the vadose zone. Geophysical Monograph 108. 63-78.

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems Engineering, 114(4), 358-371.

Nangia, V. P.H. Gowda, D. J. Mulla, and G. R. Sands. (2008). Water quality modeling of fertilizer management impacts on nitrate losses in tile drains at the field scale. Journal of Environmental Quality, 37(2), 296-307.

Nelson, D. (1985) Minimizing nitrogen losses in non-irrigated eastern areas. Proceedings of Plant Nutrient Use and the Environment Symposium Proceedings. The Fertilizer Institute, Washington, DC, 173-209.

NOAA (National Centers for Environmental Information, NCDC) (2016). https://www.ncdc.noaa.gov/cdo-web/datatools/normals. Accessed 11 December 2017.

Oquist, K. A., J. S. Strock, and D. J. Mulla. (2006). Influence of alternative and conventional management practices on soil physical and hydraulic properties. Vadose Zone Journal, 5:356-364.

Power, J. F., R. Wiese, and D. Flowerday. (2000). Managing nitrogen for water quality-lessons from Management Systems Evaluation Area. Journal of Environmental Quality, 29(2), 355-366.

Randall G.W. (1998). Implications of dry and wet cycles on nitrate loss to subsurface tile drainage. p. 53-60. In Drainage in the 21st century: Food production and the environment. Proceedings of the seventh international drainage symposium, Orlando, Florida, USA. 8-10 March 1998 1998.

Randall G.W. and D. J. Mulla. (2001). Nitrate nitrogen is surface waters as influenced by climatic conditions and agricultural practices. Journal of Environmental Quality 30,337-344.

Randall, G. W., J.A. Vetsch, and J. R. Huffman. (2003). Nitrate losses in subsurface drainage from a corn-soybean rotation as affected by time of nitrogen application and use of nitrapyrin. Journal of Environmental Quality, 32(6), 1764-1772.

Randall, G. W. (2004). Subsurface drain flow characteristics during a 15-year period in Minnesota. Proceedings of the 8th International Drainage Symposium Drainage VIII, ASAE, Sacramento, CA, 21-24.

Randall, G. W., and M. J. Goss. (2008). Nitrate losses to surface water through subsurface, tile drainage. In J.L Hatfield and R. F. Follett (Eds), Nitrogen in the environment: Sources, problems, and management (pp. 145-175). New York: Elsevier.

Refsgaard, J. C., J. P. Van der Sluijs, J. Brown, and P. Van der Keur. (2006). A framework for dealing with uncertainty due to model structure error. Advances in Water Resources, 29(11), 1586-159

Rice, C. E., and C. L. Larson. (1974). Storage Routing Methods without Fitted Coefficients. Transactions of the ASAE, 17(1), 152-0156.

Ritzema H.P. (1994). Subsurface flow to drains. p. 263. In H.P. Ritzema (ed.) Drainage principles and applications. 2nd ed. International Institute for land Reclamation and Improvement, Netherlands.

Robertson G.P. and P.M. Vitousek. (2009). Nitrogen in agriculture: Balancing the cost of an essential resource. Annual Review of Environment and Resources 34, 97-125.

Russelle M.P., M. H. Entz, and A. J. Franzluebbers. (2007). Reconsidering integrated crop–livestock systems in North America. Agron. J. 99:325-334.

Sands G.R., I. Song, L. M. Busman and B. Hansen. (2006). Water quality benefits of shallow subsurface drainage systems. 2006 ASABE annual international meeting, Portland, OR. 2006.

Sands, G. R., I. Song, L. M. Busman, and B. J. Hansen. (2008). The effects of subsurface drainage depth and intensity on nitrate loads in the Northern Cornbelt. Transactions of the ASABE, 51(3), 937-946.

Scharf, P. C., V. C. Hubbard, J. A. Lory, N. R. Kitchen, K. A. Sudduth, and J. G. Davis. (2005). Field-scale variability in optimal nitrogen fertilizer rate for corn. Agronomy Journal, 97(2), 452-461.

Schulte, L. A., M. Liebman, H. Asbjornsen, and T. R. Crow. (2006). Agroecosystem restoration through strategic integration of perennials. Journal of Soil and Water Conservation, 61(6), 164A-169A.

Sela, S., H. M. van Es, B. N. Moebius-Clune, R. Marjerison, J. Melkonian, D. Moebius-Clune, R. Schindelbeck, and S. Gomes. (2016). Adapt-N outperforms grower-selected nitrogen rates in northeast and Midwestern United States strip trials. Agronomy Journal, 108(4), 1726-1734

Shirmohammadi, A., I. Chaubey, R. D. Harmel, D. D. Bosch, R. Muñoz-Carpena, C. Dharmasri, .A. Sexton, M. Arabi, M. L. Wolfe, J. Frankenberger, C. Graff, and T. M. Sohrabi. (2006). Uncertainty in TMDL Models. Transactions of the ASABE, 49(4), 1033-1049.

Skaggs, W. R., N. R. Fausey, and R. O. Evans. (2012). Drainage water management. Journal of Soil and Water Conservation, 67(6), 167A-172A.

Skaggs, R. W., M. A. Youssef, and G. M. Chescheir (2012). DRAINMOD: Model use, calibration, and validation. Transactions of the ASABE, 55(4): 1509-1522

Sorooshian, S., V.K. Gupta, and J. L. Fulton. (1983). Evaluation of Maximum Likelihood Parameter Estimation Techniques for Conceptual Rainfall-Runoff Models: Influence of Calibration Data Variability and Length on Model Credibility. Water Resources Research. 19(1): 251-259.

Stow, C. A., K. H. Reckhow, S. S Qian, E. C. Lamon, G. B. Arhonditsis, M. E. Borsuk, and D. Seo. (2007). Approaches to evaluate water quality model parameter uncertainty for adaptive TMDL implementation. Journal of the American Water Resources Association, 43(6), 1499-1507.

US EPA. (2007). Hypoxia in the northern gulf of Mexico: An update by the EPA science advisory board. Rep. EPA-SAB-08-003. U.S. Environmental Protection Agency, Washington, D.C., U.S.

Van Der Ploeg, R. R., R. Horton and D. Kirkham. (1999). Steady flow to drains and wells. p. 213-263. In R.W. Skagsg and J. Van Schilfgaarde (eds.) Agricultural drainage. Agronomy Monograph 38 ed. American Society of Agronomy.

Walker, W.W. 1996. Simplified procedures for eutrophication assessment and prediction: User manual, instruction report W-96-2. Vicksburg, MS: Army Engineer Waterways Experiment Station.

Wilson, B. N., and J. R. Ruffini. (1988). Comparison of physically-based Muskingum methods. Transactions of the ASAE, 31(1), 91-0097.

Wilson, B. N. and C. T. Haan. 1991. Bayesian estimation of erosion parameters. Part I: Theoretical development. Transactions of the ASAE. 34(3): 809-820

Yang, J., P. Reichert, K. C. Abbaspour, J. Xia, and H. Yang. (2008). Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. Journal of Hydrology, 358(1), 1-23.

Youssef, M. A., R. W. Skaggs, G. M. Chescheir, and J. W. Gilliam. (2005). The Nitrogen Simulation Model, DRAINMOD-N II. Transactions of the ASAE, 48(2), 611-626.

Appendix

A. General Relationships for Expected Value and Variance

General Relationships

We have the following definitions for variance, covariance, and expected value of a product:

$$VAR(x) = E(x^2) - \left(E(x)\right)^2 \quad \text{A.1.1a}$$
$$COV(x) = E(xy) - E(x)E(y) \quad \text{A.1.1b}$$
$$E(x,y) = COV(x,y) + E(x)E(y) \quad \text{A.1.1c}$$

The expected value and variance of a variable that represents a sum of terms is described below. We begin by generalizing the form of a random variable z that is a sum of random variables x_1 through x_n:

$$z = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n \quad \text{A.1.2a}$$

Where a_0 through a_n are non-random terms. The expected value for z is defined as:

$$E(z) = a_0 + a_1E(x_1) + a_2E(x_2) + \cdots + a_nE(x_n) \quad \text{A.1.2b}$$

The variance of z can be defined as:

$$\begin{aligned} VAR(z) &= \sum_{j=1}^{n}\sum_{i=1}^{n} a_ia_jCOV(x_i,x_j) \\ &= a_1^2VAR(x_1) + \cdots + a_n^2VAR(x_n) \\ &\quad + \left(\sum_{j=1}^{n}\sum_{i=1}^{n} a_ia_jCOV(x_i,x_j)\right)_{i\neq j} \end{aligned} \quad \text{A.1.2c}$$

B. Full derivation of E(SSE) and E(SSTO) for separate parameter for each time and space

Used in the derivation of: 1) the expected value of the non-adjusted MSE and NMSE, $E(MSE_{st})$ *and* $E(NMSE_{st})$, and 2) the value for the adjusted MSE and NMSE ($\widetilde{MSE}$ *and* $\widetilde{NMSE}$) adjusted for both measurement errors and measurement errors + model parameter errors.

Sums of squares error: SSE_{st}

Expanded SSE: The expansion of SSE_{ST} is:

$$SSE_{ST} = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - P_{ij}\right)^2 = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\rho_{ij} + \varepsilon_{ij} - \tilde{\varepsilon}_{ij}\right)^2$$

$$= \sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}^2 + 2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}(\varepsilon_{ij} - \tilde{\varepsilon}_{ij}) \qquad \text{B.1.1a}$$

$$+ \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\varepsilon_{ij} - \tilde{\varepsilon}_{ij}\right)^2$$

Which we alternatively write as:

$$SSE_{ST} = M_1 + M_1' - M_2 + F - C + \tilde{F}_T \qquad \text{B.1.1b}$$

Where

$$M_1 = \sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}^2, \quad M_1' = 2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\varepsilon_{ij}, \quad M_2 = 2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\tilde{\varepsilon}_{ij} \qquad \text{B.1.2a}$$

$$F = \sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}^2 \qquad and\ \widetilde{F_T} = \sum_{j=1}^{T}\sum_{i=1}^{S}\tilde{\varepsilon}_{ij}^2 \qquad \text{B.1.2b}$$

$$C = 2\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}\tilde{\varepsilon}_{ij} \qquad \text{B.1.2c}$$

Expected value SSE: The expected value of SSE_{ST} can be evaluated as:

$$E(SSE_{st}) = M_1 + E(M_1') - E(M_2) + E(F) - E(C) + E(\widetilde{F_T}) \qquad \text{B.1.6}$$

Where:

$$E(M_1) = \sum_{j=1}^{T}\sum_{i=1}^{S}E(\rho_{ij}^2) = \sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}^2 = M_1 \qquad \text{B.1.7a}$$

$$E(M_1') = E\left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\varepsilon_{ij}\right) = 2E(\varepsilon)\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij} \qquad \text{B.1.7b}$$

$$E(M_2) = 2\sum_{j=1}^{T}\sum_{i=1}^{S}E\left(\rho_{ij}\tilde{\varepsilon}_{ij}\right) = 2E(\tilde{\varepsilon})\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij} \qquad \text{B.1.7c}$$

$$E(F) = \sum_{j=1}^{T}\sum_{i=1}^{S}E\left(\varepsilon_{ij}^2\right) = (TS)E(\varepsilon_{st}^2) = (TS)VAR(\varepsilon_{st}) + (TS)(E(\varepsilon_{st})^2) \qquad \text{B.1.7d}$$

$$E(C) = 2\big(COV(\varepsilon, \tilde{\varepsilon}) + E(\varepsilon)E(\tilde{\varepsilon})\big) \quad \text{B.1.7e}$$

$$E(\widetilde{F_T}) = \sum_{j=1}^{T}\sum_{i=1}^{S} E(\tilde{\varepsilon}_{ij}^2) = (TS)E(\tilde{\varepsilon}_{st}^2)$$
$$= (TS)VAR(\tilde{\varepsilon}_{st}) + (TS)(E(\tilde{\varepsilon}_{st})^2) \quad \text{B.1.7f}$$

To further simplify the way we write the equation, we divide $E(\tilde{F}_T)$ into its mean and variance components and combine the mean component with $E(M_2)$, creating the new term B:

$$B = (TS)E(\tilde{\varepsilon}_{st})\left(E(\tilde{\varepsilon}_{st}) - \frac{2\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right)}{TS}\right) \quad \text{B.1.7g}$$

And $E(\widetilde{F_T})$ becomes:

$$E(\tilde{F}) = \tilde{X} = (TS)VAR(\tilde{\varepsilon}_{st}) \quad \text{B.1.7f}$$

Such that the expected value of the error sums of squares is:

$$E(SSE_{st}) = M_1 + E(M_1') + B + E(F) + \tilde{X} + E(C) \quad \text{B.1.8}$$

***Simplify E(SSE)**:* If we simplify using the assumption that the mean of the measurement error is equal to zero, $E(\varepsilon) = 0$, and that the model parameter uncertainty and observed errors are independent of each other (i.e. $COV(\varepsilon, \tilde{\varepsilon}) = 0$), then the expected value of $M_1', F, and\ C$ become:

$$E(M_1') = E\left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\varepsilon_{ij}\right) = 0 \quad \text{B.1.9a}$$

$$E(F) = X = \sum_{j=1}^{T}\sum_{i=1}^{S} E(\varepsilon_{ij}^2) = (TS)E(\varepsilon_{st}^2)$$
$$= (TS)VAR(\varepsilon_{st}) + (TS)(E(\varepsilon_{st})^2) = (TS)VAR(\varepsilon_{st}) \quad \text{B.1.9b}$$

$$E(C) = 2\big(COV(\varepsilon, \tilde{\varepsilon}) + E(\varepsilon)E(\tilde{\varepsilon})\big) = 0 \quad \text{B.1.9c}$$

And thus the expected value of the error sums of squares is:

$$E(SSE_{st}) = M_1 + B + X + \tilde{X} \quad \text{B.1.10}$$

Sums of squares total: $SSTO_{st}$

The sums of square of total (SSTO), is defined as:

$$SSTO_{ST} = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}\right)^2 = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}}{ST}\right)^2$$
$$= \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} + \varepsilon_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}}{ST}\right)^2 \qquad \text{B.2.1}$$

For later derivation of the E(NMSE), we separate $SSTO_{ST}$ into its random and non-random components by defining κ as:

$$\kappa = \frac{SSTO_{ST}}{Z_{ST}} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} + \varepsilon_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}}{ST}\right)^2}{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST}\right)^2} \qquad \text{B.2.2a}$$

Where Z_{ST} equals the non-random component:

$$Z_{ST} = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST}\right)^2 \qquad \text{B.2.2b}$$

Combining our definition of κ from equation B.2.2a with SSTO in equation B.2.1, we can define SSTO as:

$$SSTO_{ST} = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}\right)^2 = \kappa\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST}\right)^2 = \kappa Z_{st} \qquad \text{B.2.3}$$

Expanded SSTO: The SSTO expanded is equal to:

$$SSTO_{ST} = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij} - \bar{O}\right)^2 = \sum_{j=1}^{T}\sum_{i=1}^{S}\left(O_{ij}^2 - 2\bar{O}\left(O_{ij}\right) + \bar{O}^2\right)$$
$$= \sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}^2 - 2\bar{O}\sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij} + (ST)\bar{O}^2$$
$$= \sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}^2 - 2(ST)\bar{O}^2 + (ST)\bar{O}^2 \qquad \text{B.2.4}$$
$$= \sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}^2 - (ST)\bar{O}^2$$

Expected value SSTO: The expected value of $SSTO_{st}$ is written as:

$$\begin{aligned}
E(SSTO_{ST}) &= \sum_{j=1}^{T}\sum_{i=1}^{S} E\left(O_{ij}^2\right) - (ST)E(\bar{O}^2) \\
&= \sum_{j=1}^{T}\sum_{i=1}^{S}\left(VAR\left(O_{ij}\right) + \left(E\left(O_{ij}\right)\right)^2\right) \\
&- ST(VAR(\bar{O}) + E(\bar{O})^2) \\
&= \sum_{j=1}^{T}\sum_{i=1}^{S}\left(VAR\left(O_{ij}\right) + \left(E\left(O_{ij}\right)\right)^2\right) \\
&- ST\left(VAR\left(\frac{\sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}}{ST}\right) + E\left(\frac{\sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}}{ST}\right)^2\right) \\
&= \sum_{j=1}^{T}\sum_{i=1}^{S} VAR\left(O_{ij}\right) + \sum_{j=1}^{T}\sum_{i=1}^{S} \mu_{ij}^2 \\
&- \frac{1}{ST}\left(VAR\left(\sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}\right) + \left(\sum_{j=1}^{T}\sum_{i=1}^{S} \mu_{ij}\right)^2\right)
\end{aligned} \qquad \text{B.2.5a}$$

From equation A.1.2c we can write, the variance of the summation of O_{ij} as:

$$\begin{aligned}
VAR\left(\sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}\right) &= VAR(O_{11} + \cdots + O_{1T} + O_{21} + \cdots + O_{TS}) \\
&= \sum_{j=1}^{T}\sum_{i=1}^{S} COV\left(O_{11}, O_{ij}\right) + \cdots + \sum_{j=1}^{T}\sum_{i=1}^{S} COV\left(O_{1T}, O_{ij}\right) \\
&+ \sum_{j=1}^{T}\sum_{i=1}^{S} COV\left(O_{21}, O_{ij}\right) + \cdots + \sum_{j=1}^{T}\sum_{i=1}^{S} COV\left(O_{ST}, O_{ij}\right) \\
&= \sum_{l=1}^{T}\sum_{k=1}^{S}\sum_{j=1}^{T}\sum_{i=1}^{S} COV\left(O_{kl}, O_{ij}\right)
\end{aligned} \qquad \text{B.2.6a}$$

When k=l=i=j the covariance is defined by the variance. Thus we can also write the variance of O_{ij} as:

$$VAR\left(\sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}\right)$$

$$= VAR\left(\sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}\right) + \sum_{l=1}^{T}\sum_{k=1}^{S}\sum_{j=1}^{T}\sum_{i=1}^{S} COV\left(O_{kl}, O_{ij}\right)_{i\neq j\neq k\neq l} \qquad \text{B.2.6b}$$

Using this form of the variance of O_{ij}, the expected value of SSTO can be evaluated as:

$$E(SSTO_{ST}) = \sum_{j=1}^{T}\sum_{i=1}^{S} VAR\left(O_{ij}\right) + \sum_{j=1}^{T}\sum_{i=1}^{S} \mu_{ij}^2 - \frac{1}{ST}\left(VAR\left(\sum_{j=1}^{T}\sum_{i=1}^{S} O_{ij}\right) + \sum_{l=1}^{T}\sum_{k=1}^{S}\sum_{j=1}^{T}\sum_{i=1}^{S} COV\left(O_{kl}, O_{ij}\right)_{i\neq j\neq k\neq l} + \left(\sum_{j=1}^{T}\sum_{i=1}^{S} \mu_{ij}\right)^2\right) \qquad \text{B.2.5b}$$

To simplify the analysis, we assume that the **observed values are uncorrelated**, so that $\sum_{l=1}^{T}\sum_{k=1}^{S}\sum_{j=1}^{T}\sum_{i=1}^{S} COV\left(O_{kl}, O_{ij}\right)_{i\neq j\neq k\neq l} = 0$. For the condition of homoscedasticity, and assuming that that the **variance in O_{ij} is due to variance in the measurement error, $\boldsymbol{\varepsilon_{ij}}$,** we then obtain:

$$E(SSTO_{ST}) = (ST)VAR(\varepsilon_{st}) + \sum_{j=1}^{T}\sum_{i=1}^{S} \mu_{ij}^2 - \frac{1}{ST}\left((ST)VAR(\varepsilon_{st}) + \left(\sum_{j=1}^{T}\sum_{i=1}^{S} \mu_{ij}\right)^2\right) \qquad \text{B.2.5c}$$

$$= (ST-1)VAR(\varepsilon_{st}) + \sum_{j=1}^{T}\sum_{i=1}^{S}\left(\mu_{ij} - \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\mu_{ij}}{ST}\right)^2$$

Or using our definition of Z in equation B.2.2b:

$$E(SSTO_{ST}) = Z_{st} + (ST)VAR(\varepsilon_{st})\left(1 - \frac{1}{ST}\right) \qquad \text{B.2.5d}$$

The expected value for κ is defined as:

$$E(\kappa) = \frac{E(SSTO_{st})}{E(Z_{st})} = \frac{1}{Z_{st}} E(SSTO_{st}) \qquad \text{B.2.5d}$$

Mean Square Error and Normalized Mean Square Error

The mean square error is equal to:

$$MSE_{st} = \frac{SSE}{ST} \qquad \text{B.3.1a}$$

Using our definition of SSE from equation C.1.1b, this is equal to:

$$MSE_{st} = \frac{1}{ST}\left(M_1 + 2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\varepsilon_{ij} - M_2 + F - 2\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}\tilde{\varepsilon}_{ij} + \widetilde{F_T}\right) \qquad \text{B.3.2b}$$

The normalized mean square error is equal to:

$$NMSE_{st} = \frac{SSE}{SSTO} \qquad \text{B.3.3a}$$

Using our definitions of SSE from equation B.1.1b, and the simplification of SSTO from equation B.2.3, where $SSTO = \kappa Z_{st}$, we have:

$$NMSE_{st} = \frac{1}{\kappa Z_{st}}\left(M_1 + 2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\varepsilon_{ij} - M_2 + F - 2\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}\tilde{\varepsilon}_{ij} + \widetilde{F_T}\right) \qquad \text{B.3.3b}$$

Expected value of (non-adjusted) MSE and NMSE

We will simplify by using the conditions that random errors for the observed data share a common variance ($VAR(\varepsilon_{et})$) and mean equal to zero ($E(\varepsilon_{et}) = 0$); that random errors from parameters uncertainty share a common mean and variance ($E(\tilde{\varepsilon}_{st}) and\ VAR(\tilde{\varepsilon}_{et})$); and that the parameter uncertainty and observed errors are independent of each other and therefore $E(\varepsilon_{ij}, \tilde{\varepsilon}_{ij}) = COV(\varepsilon, \tilde{\varepsilon}) + E(\varepsilon)E(\tilde{\varepsilon}) = 0$.

The expected value of the MSE is equal to the expected value of SSE divided by the total number of sites (S) and times (T).

$$E(MSE) = \frac{E(SSE_{st})}{ST} = \frac{M_1 + B + X + \tilde{X}}{ST} \tag{B.4.1}$$

The expected value of the NMSE is the expected value of the ratio of SSE to SSTO. Taking the expected value of equation B.3.3.b, we have the expected value of the NMSE is:

$$\begin{aligned} E(NMSE_{st}) &= E\left(\frac{SSE}{SSTO}\right) \\ &= E\left(\frac{1}{\kappa Z_{st}}\left(M_1 + 2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\varepsilon_{ij} - M_2 + F \right.\right. \\ &\left.\left. - 2\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}\tilde{\varepsilon}_{ij} + \widetilde{F_T}\right)\right) \end{aligned} \tag{B.4.2a}$$

The normalized mean square error is the ratio of two random numbers, so we must use the Taylor series approximation such that:

$$E(NMSE_{st}) = E\left(\frac{SSE_{ST}}{SSTO_{ST}}\right) \approx \frac{E(SSE_{ST})}{E(SSTO_{ST})} \tag{B.4.2b}$$

Using the definition of SSTO from equation B.2.3, this can also be written as:

$$E(NMSE_{st}) = \frac{E(SSE_{ST})}{Z_{st}E(\kappa)} \tag{B.4.2c}$$

Using the approximation from the Taylor Series, equation B.1.8 for the expected value of SSE, and equation B.2.5d for κ, the NMSE can be written as:

$$E(NMSE_{st}) = \frac{M_1 + B + X + \tilde{X}}{Z_{st}\frac{1}{Z_{st}}\left(Z_{st} + (ST)VAR(\varepsilon_{st})\left(1 - \frac{1}{ST}\right)\right)} \tag{B.4.2d}$$

Which can be simplified to:

$$E(NMSE_{st}) = \left(\frac{1}{1 + X_{st}^{*}\left(1 - \frac{1}{ST}\right)}\right)\left(M_{st}^{*} + B_{st}^{*} + X_{st}^{*} + \tilde{X}_{st}^{*}\right) \tag{B.4.2e}$$

where terms are defined as follows:

$$X_{st}^* = \frac{VAR(\varepsilon_{st})}{\bar{Z}}$$ B.4.3a

$$\bar{Z} = \frac{Z_{st}}{ST}$$ B.4.3b

$$M_{st}^* = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S} E\left(\rho_{ij}^2\right)}{(TS)\bar{Z}} = \frac{\sum_{j=1}^{T}\sum_{i=1}^{S}\left(\rho_{ij}^2\right)}{(TS)\bar{Z}}$$ B.4.3c

$$B_{st}^* = \frac{E(\tilde{\varepsilon}_{st})\left(E(\tilde{\varepsilon}_{st}) - \frac{2\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right)}{TS}\right)}{\bar{Z}}$$ B.4.3d

$$\tilde{X}_{st}^* = \frac{VAR(\tilde{\varepsilon}_{st})}{\bar{Z}}$$ B.4.3e

C. Variance of the actual adjusted and theoretically derived GOF

Numerator, SSE:

The SSE is that obtained by not taking the expected value of any term. Thus, for SSE (all error terms included), this is equal to equations B.1.1b-B1.2c in Appendix Section B:

$$SSE = M_1 + M_1' - M_2 + F - C + \tilde{F}_T$$ B.1.1b

Where:

$$M_1 = \sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}^2, \quad M_1' = 2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\varepsilon_{ij}, \quad M_2 = 2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\tilde{\varepsilon}_{ij}$$ B.1.2a

$$F = \sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}^2 \qquad and\ \widetilde{F_T} = \sum_{j=1}^{T}\sum_{i=1}^{S}\tilde{\varepsilon}_{ij}^2$$ B.1.2b

$$C = 2\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}\tilde{\varepsilon}_{ij}$$ B.1.2c

Meanwhile, the theoretically-derived SSE is equal to the expected value of these terms. Using the assumption that the mean of the measurement error is equal to 0, this is:

$$E(SSE_{st}) = \widetilde{SSE}_{st} = M_1 - E(M_2) + E(F) + E\left(\widetilde{F_T}\right)$$ B.1.10

Where:

$$E(M_1) = \sum_{j=1}^{T}\sum_{i=1}^{S}E(\rho_{ij}^2) = \sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}^2 = M_1$$ B.1.7a

$$E(M_1') = E\left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\varepsilon_{ij}\right) = 2E(\varepsilon)\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij} = 0$$ B.1.7b

$$E(M_2) = 2\sum_{j=1}^{T}\sum_{i=1}^{S} E\left(\rho_{ij}\tilde{\varepsilon}_{ij}\right) = 2E(\tilde{\varepsilon})\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij} \qquad \text{B.1.7c}$$

$$\begin{aligned} E(F) &= \sum_{j=1}^{T}\sum_{i=1}^{S} E\left(\varepsilon_{ij}^2\right) = (TS)E(\varepsilon_{st}^2) = (TS)VAR(\varepsilon_{st}) + (TS)(E(\varepsilon_{st})^2) \\ &= (TS)VAR(\varepsilon_{st}) \end{aligned} \qquad \text{B.1.7d}$$

$$E(C) = 2\left(COV(\varepsilon,\tilde{\varepsilon}) + E(\varepsilon)E(\tilde{\varepsilon})\right) = 0 \qquad \text{B.1.7e}$$

$$\begin{aligned} E\left(\widetilde{F_T}\right) &= \sum_{j=1}^{T}\sum_{i=1}^{S} E\left(\tilde{\varepsilon}_{ij}^2\right) = (TS)E(\tilde{\varepsilon}_{st}^2) \\ &= (TS)VAR(\tilde{\varepsilon}_{st}) + (TS)(E(\tilde{\varepsilon}_{st})^2) \end{aligned} \qquad \text{B.1.7f}$$

Taking the variance of both definitions of SSE, we obtain:

$$\begin{aligned} VAR(SSE) &= VAR\left(M_1 + M_1' - M_2 + F - C + \tilde{F}_T\right) \\ &= VAR(M_1) + VAR\left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\varepsilon_{ij}\right) \\ &- VAR\left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\tilde{\varepsilon}_{ij}\right) + VAR\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}^2\right) \\ &- VAR\left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}\tilde{\varepsilon}_{ij}\right) + VAR\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\tilde{\varepsilon}_{ij}^2\right) \end{aligned} \qquad \text{C.1.1a}$$

$$\begin{aligned} VAR\left(\widetilde{SSE}\right) &= VAR\left(M_1 - E(M_2) + E(F) + E\left(\widetilde{F_T}\right)\right) \\ &= VAR(M_1) - VAR\left(2E(\tilde{\varepsilon})\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right) \\ &+ VAR\left((TS)VAR(\varepsilon_{st})\right) \\ &+ VAR\left((ST)\left(VAR(\tilde{\varepsilon}_{st}) + \left(E(\tilde{\varepsilon}_{st})\right)^2\right)\right) \end{aligned} \qquad \text{C.1.1b}$$

Adjusted SSE, measurement error:

The variance of the actual SSE adjusted for measurement error is the variance of equation C.1.1a above, with measurement error terms removed:

$$VAR(SSE_\varepsilon) = VAR\big(M_1 - M_2 - C + \tilde{F}_T\big)$$
$$= VAR(M_1) - VAR\left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\tilde{\varepsilon}_{ij}\right) - VAR\left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}\tilde{\varepsilon}_{ij}\right) + VAR\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\tilde{\varepsilon}_{ij}^{2}\right) \qquad \text{C.1.2a}$$

The variance for the adjusted SSE uses the definition that it is the difference between the actual SSE, no error terms removed, and the difference between the expected value of SSE no error terms removed and that with error terms removed ($SSE - E(\Delta SSE_\varepsilon)$). For measurement error only, $E(\Delta SSE_\varepsilon)$ is defined as:

$$E(\Delta SSE_\varepsilon) = \big(M_{st} + B_{st} + X_{st} + \tilde{X}_{st}\big) - \big(M_{st} + B_{st} + \tilde{X}_{st}\big) = X_{st} = (ST)VAR(\varepsilon_{st}) \qquad \text{C.2.1}$$

Thus the variance of the adjusted theoretical SSE is:

$$VAR\big(\widetilde{SSE}\big) = VAR\big(SSE - E(\Delta SSE_\varepsilon)\big)$$
$$= VAR\Big(\big(M_1 + M_1' - M_2 + F - C + \tilde{F}_T\big) - X_{st}\Big)$$
$$= VAR\left((M_1) + \left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\varepsilon_{ij}\right) - \left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\tilde{\varepsilon}_{ij}\right) + \left(\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}^{2}\right) - \left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}\tilde{\varepsilon}_{ij}\right) + \left(\sum_{j=1}^{T}\sum_{i=1}^{S}\tilde{\varepsilon}_{ij}^{2}\right) - (ST)VAR(\varepsilon_{st})\right) \qquad \text{C.2.2}$$

Adjusted SSE, measurement + parameter error:

The variance of the actual adjusted SSE which accounts for both measurement and parameter error is:

$$VAR(SSE_{\varepsilon+\tilde{\varepsilon}}) = VAR(M_1) \qquad \text{C.2.3a}$$

For the theoretical value, the variance is equal to $SSE - E(\Delta SSE_{\varepsilon+\tilde{\varepsilon}})$, where $E(\Delta SSE_{\varepsilon+\tilde{\varepsilon}})$ is defined as:

$$E(\Delta SSE_{\varepsilon+\tilde{\varepsilon}}) = \left(M_{st} + B_{st} + X_{st} + \tilde{X}_{st}\right) - (M_{st}) = B_{st} + X_{st} + \tilde{X}_{st}$$
$$= (ST)\left(\left(E(\tilde{\varepsilon}_{st})\right)\left(E(\tilde{\varepsilon}_{st}) - \frac{2\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right)}{ST}\right) + VAR(\varepsilon_{st}) + VAR(\tilde{\varepsilon}_{st})\right) \quad \text{C.2.3b}$$

Thus the overall variance of the adjusted theoretically-derived SSE is:

$$VAR\left(\widetilde{SSE}_{\varepsilon+\tilde{\varepsilon}}\right) = VAR\left(SSE - E(\Delta SSE_{\varepsilon+\tilde{\varepsilon}})\right)$$
$$= VAR\left(M_1 + M_1' - M_2 + F - C + \tilde{F}_T - \left(B_{st} + X_{st} + \tilde{X}_{st}\right)\right)$$
$$= VAR\left((M_1) + \left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\varepsilon_{ij}\right) - \left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\tilde{\varepsilon}_{ij}\right) + \left(\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}^2\right) - \left(2\sum_{j=1}^{T}\sum_{i=1}^{S}\varepsilon_{ij}\tilde{\varepsilon}_{ij}\right) + \left(\sum_{j=1}^{T}\sum_{i=1}^{S}\tilde{\varepsilon}_{ij}^2\right) - (ST)\left(\left(E(\tilde{\varepsilon}_{st})\right)\left(E(\tilde{\varepsilon}_{st}) - \frac{2\left(\sum_{j=1}^{T}\sum_{i=1}^{S}\rho_{ij}\right)}{ST}\right) + VAR(\varepsilon_{st}) + VAR(\tilde{\varepsilon}_{st})\right)\right) \quad \text{C.2.3c}$$

Variance of the difference between actual and theoretical

Difference for variance of actual and theoretical adjusted SSE, measurement error only:

$$VAR\left(SSE_{\varepsilon} - \widetilde{SSE}_{\varepsilon}\right)$$
$$= VAR\left(M_1 - M_2 + \tilde{F}_T - \left(M_1 + M_1' - M_2 + F - C + \tilde{F}_T - X_{st}\right)\right) \quad \text{C.3.1a}$$
$$= VAR(M_1' + F - C - X_{st})$$

Difference for variance of actual and theoretical adjusted SSE, measurement + parameter error:

$$
\begin{aligned}
VAR\left(SSE_{\varepsilon+\tilde{\varepsilon}} - \widetilde{SSE}_{\varepsilon+\tilde{\varepsilon}}\right) \\
&= VAR\Bigg(M_1 \\
&- \left(M_1 + M_1' - M_2 + F - C + \tilde{F}_T - \left(B_{st} + X_{st} + \tilde{X}_{st}\right)\right)\Bigg) \qquad \text{C.3.1b} \\
&= VAR\left(M_1' - M_2 + F - C + \tilde{F}_T - \left(B_{st} + X_{st} + \tilde{X}_{st}\right)\right)
\end{aligned}
$$

D. Visual Basic Code for generating Monte Carlo Simulations

The following is the VB Code used in Excel for generating the MC simulations of Part II.

```
Sub nonlumped()
'Use new syntax/theory development (SSE, SSTO)
'For taking into account measurement error (b) and model parameter error (c)
'Uses ij for all variables -- i.e. no "lumping" of time or space components
   'However, it is currently set up so that eta_ij=eta_*j and delta_ij=delta_i*
'calculates adjusted NMSE and MSE
Dim output, output2, output3, output4, int_var As Worksheet
Dim rowout, NumRow, row As Double
Dim i, j, x, y, pt, p, t As Double
Dim mu_pt(1 To 3, 1 To 6), mu, O_pt(1 To 3, 1 To 6), z(1 To 3, 1 To 6) As Double
Dim delta(1 To 3, 1 To 6), eta(1 To 3, 1 To 6), P_pt(1 To 3, 1 To 6) As Double
Dim p_tilde(1 To 3, 1 To 6) As Double
Dim y1_c1, y1_c2, mean, mean_t, std_t, var_t As Double
Dim var, std As Double
Dim RandomNum(1 To 3, 1 To 6), RN_t(1 To 3, 1 To 6) As Double
Dim C, O_sum, e_sum, O_bar, e_bar As Double
Dim Ptilde_pt(1 To 3, 1 To 6), rho_pt(1 To 3, 1 To 6) As Double
Dim Res(1 To 3, 1 To 6), Res_b(1 To 3, 1 To 6) As Double
Dim SSTO_obs_a, SSTO_obs_b, SSE_obs_b, SSE_obs_a As Double
Dim v(1 To 3, 1 To 6), v_b(1 To 3, 1 To 6) As Double
Dim SSE_E_a, SSTO_E_a, SSE_E_b, SSTO_E_b, d_SSE_E_c, SSE_E_c As Double
Dim d_SSE_E_b, d_SSTO_E, d_NMSE_E, a_NMSE_b, a_NMSE_c, NMSE_obs As Double
Dim rho, rho_s, rho_bar As Double
Dim NMSE_a_E, NMSE_b_E, a_NMSE_check, NMSE_obs_a, NMSE_obs_b As Double
Dim SSE_obs_c, SSTO_E_c, Res_c(1 To 3, 1 To 6) As Double
Dim eta_s(1 To 3, 1 To 6), eta_st(1 To 6), eta_star(1 To 6) As Double
Dim C_alt, SSTO_E_a2, z_alt(1 To 3, 1 To 6) As Double
Dim rho_t(1 To 6), d_p(1 To 3), n_t(1 To 6) As Double
Dim mu_t(1 To 6), Ptilde_t(1 To 6), P_t(1 To 6) As Double
Dim RN_tl(1 To 6), Res_alump(1 To 3, 1 To 6), SSE_obs_alump, Res_d1(1 To 3, 1 To 6), SSE_obs_d1 As
Double
Dim v_d1(1 To 3, 1 To 6), SSTO_obs_d1 As Double
```

```
Dim RN_d(1 To 3, 1 To 6), O_rt(1 To 3, 1 To 6), O_rt_sum As Double
Dim Res_a2lump(1 To 3, 1 To 6), SSE_obs_a2lump As Double
Dim v_a2(1 To 3, 1 To 6), SSTO_E_a_alt, SSTO_obs_a2, Ort_bar As Double
Dim Res_d2(1 To 3, 1 To 6), SSE_obs_d2, d_bar, d_sum As Double
Dim v_d2(1 To 3, 1 To 6), SSTO_obs_d2, mean_d, var_d, std_d As Double
Dim rho_lump, rho_slump, rho_bar_lump, delta_lump, delta_bar_lump, delta_lump_sum As Double
Dim SSE_E_alump, SSE_E_a2lump, SSE_E_d1, SSTO_E_d1, delta_sum, delta_ssum, delta_bar As
Double
Dim a_NMSE_d1, d_SSE_E_d1, d_SSE_E_d2, a_NMSE_d2 As Double
Dim SSE_E_d2, SSTO_E_d2, d_SSTO_E_2 As Double
Dim z2(1 To 6), C2 As Double
Dim Res_1(1 To 3, 1 To 6), Res_2(1 To 3, 1 To 6), Res_3(1 To 3, 1 To 6), SSE_obs_1, SSE_obs_2,
SSE_obs_3 As Double
Dim NMSE_obs_1, NMSE_obs_2, NMSE_obs_3, SSTO_obs_1, SSTO_obs_2, SSTO_obs_3, O_sum1,
O_sum2, O_sum3 As Double
Dim v_1(1 To 3, 1 To 6), v_2(1 To 3, 1 To 6), v_3(1 To 3, 1 To 6) As Double
Dim O_bar1, O_bar2, O_bar3, rho_relative As Double
Dim MSE_obs, MSE_obs_b, a_MSE_b, MSE_obs_c, a_MSE_c, NMSE_obs_c, a_NMSE_c2,
d_SSE_E_c2 As Double
Dim adj_denom_b, adj_num_b, adj_num_c As Double
Dim SSE_obs_a_alt, Res_alt(1 To 3, 1 To 6) As Double
Dim m1_final, m2_final, c_final, f_sum, ft_sum, m_p(1 To 3, 1 To 6), m_p_sum, m2(1 To 3, 1 To 6),
m2_sum, c1(1 To 3, 1 To 6), c_sum As Double

Set output = ActiveWorkbook.Sheets("output")
rowout = 1
output.Cells(1, 1) = "run #"
output.Cells(1, 2) = "SSE_obs_a (all terms)"
output.Cells(1, 3) = "E(SSEa)"
output.Cells(1, 4) = "SSTO_obs_a"
output.Cells(1, 5) = "E(SSTOa)"
output.Cells(1, 6) = "SSE_obs_b (no measure error)"
output.Cells(1, 7) = "E(SSEb)"
output.Cells(1, 8) = "SSTO_obs_b"
output.Cells(1, 9) = "E(SSTOb)"
output.Cells(1, 10) = "SSE_obs_c (no measure or parameter error)"
output.Cells(1, 11) = "E(SSEc)"
output.Cells(1, 12) = "adjMSE_true_b"
output.Cells(1, 13) = "E(adjMSEb)"
output.Cells(1, 14) = "adjNMSE_true_b"
output.Cells(1, 15) = "E(adjNMSEb)"
output.Cells(1, 16) = "adjMSE true_c"
output.Cells(1, 17) = "E(adjMSEc)"
output.Cells(1, 18) = "adjNMSE true_c"
output.Cells(1, 19) = "E(adjNMSEc)"
output.Cells(1, 20) = "E(adjNMSEc2)"
output.Cells(1, 21) = "adjDenominator"
output.Cells(1, 22) = "adjNumerator_b"
output.Cells(1, 23) = "adjNumerator_c"
```

```
output.Cells(1, 24) = "RN measurement"
output.Cells(1, 25) = "RN parameter"
output.Cells(1, 26) = "Diff MSEb"
output.Cells(1, 27) = "Diff NMSEb"
output.Cells(1, 28) = "Diff MSEc"
output.Cells(1, 29) = "Diff NMSEc"

'set delta and eta arrays
delta(1, 1) = -0.5
delta(1, 2) = -0.5
delta(1, 3) = -0.5
delta(1, 4) = -0.5
delta(1, 5) = -0.5
delta(1, 6) = -0.5
delta(2, 1) = 3
delta(2, 2) = 3
delta(2, 3) = 3
delta(2, 4) = 3
delta(2, 5) = 3
delta(2, 6) = 3
delta(3, 1) = -2.5
delta(3, 2) = -2.5
delta(3, 3) = -2.5
delta(3, 4) = -2.5
delta(3, 5) = -2.5
delta(3, 6) = -2.5

eta(1, 1) = -6.7
eta(2, 1) = -6.7
eta(3, 1) = -6.7
eta(1, 2) = 4.5
eta(2, 2) = 4.5
eta(3, 2) = 4.5
eta(1, 3) = 2.5
eta(2, 3) = 2.5
eta(3, 3) = 2.5
eta(1, 4) = 4
eta(2, 4) = 4
eta(3, 4) = 4
eta(1, 5) = 0.2
eta(2, 5) = 0.2
eta(3, 5) = 0.2
eta(1, 6) = -4.5
eta(2, 6) = -4.5
eta(3, 6) = -4.5

'Set "errors": (baseline for all is 0).
'for model parameter error and measuremnt error, baseline of 0 is set below in loop
   'Set model algorithm error (bias):
```

```
    rho_relative = 0.064

    'Set mean and stdev for measurement error:
    '**the way eqs written, requires assumptions than mean of measurement error=0
    mean = 0
    var = 0.5
    std = (var) ^ (0.5)

    'Set mean and stdev for model parameter error:
    mean_t = 0.6 '0.6 '1 '3
    var_t = 4.8 '1 '4
    std_t = (var_t) ^ 0.5

'set number of sites and years:
p = 3
t = 6
pt = p * t '6*3
rowout = 1
row = 1

'overall mean for observed data:
mu = 12

'Loop for each row:
For x = 1 To 20000
'"reset" summed variables:
C = 0
C_alt = 0
C2 = 0
e_sum = 0
O_sum = 0
SSE_obs_a = 0
SSE_obs_b = 0
SSE_obs_c = 0
SSTO_obs_b = 0
SSTO_obs_a = 0
SSTO_obs_d1 = 0
SSE_obs_alump = 0
SSE_obs_d1 = 0
rho = 0
rho_s = 0
rho_lump = 0
rho_slump = 0
O_rt_sum = 0
d_sum = 0
delta_lump = 0
delta_lump_sum = 0
delta_sum = 0
delta_ssum = 0
```

```
SSE_obs_a2lump = 0
SSE_obs_d2 = 0
SSTO_obs_a2 = 0
SSTO_obs_d2 = 0
SSE_obs_1 = 0
SSE_obs_2 = 0
SSE_obs_3 = 0
O_sum1 = 0
O_sum2 = 0
O_sum3 = 0
SSTO_obs_1 = 0
SSTO_obs_2 = 0
SSTO_obs_3 = 0
f_sum = 0
ft_sum = 0
m_p_sum = 0
m2_sum = 0
c_sum = 0
'row = 1


'switched i and j, so run j loop 1st, i loop 2nd
'Loop for number of years within a row:
For j = 1 To t

    'loop for number of sites within a row:
    For i = 1 To p
        'calculate observed terms, non-lumped:
        mu_pt(i, j) = mu + delta(i, j) + eta(i, j)
        'RandomNum(i, j) = 0
        RandomNum(i, j) = (1.224745 - (-1.224745)) * Rnd() + (-1.224745)

'WorksheetFunction.RandBetween(-1.22474, 1.22474) 'WorksheetFunction.Norm_Inv(Rnd(), mean, std)
        O_pt(i, j) = mu_pt(i, j) + RandomNum(i, j)
        z(i, j) = (mu_pt(i, j) - mu) ^ 2
        'z_alt(i, j) = (mu_pt(i, j)) ^ 2 - (mu) ^ 2
        C = C + z(i, j)
        'C_alt = C_alt + z_alt(i, j)
        O_sum = O_pt(i, j) + O_sum
        e_sum = e_sum + RandomNum(i, j)
        delta_sum = delta_sum + delta(i, j)
        delta_ssum = delta_ssum + (delta(i, j)) ^ 2

        'calculate predicted terms
        Ptilde_pt(i, j) = mu_pt(i, j) * (1 - rho_relative)
        RN_t(i, j) = (4.394733 - (-3.19473)) * Rnd() + (-3.19473) 'WorksheetFunction.RandBetween(-
3.19473, 4.394733) 'WorksheetFunction.Norm_Inv(Rnd(), mean_t, std_t)
        P_pt(i, j) = Ptilde_pt(i, j) + RN_t(i, j)
        rho_pt(i, j) = mu_pt(i, j) - Ptilde_pt(i, j)
```

```
        rho = rho + rho_pt(i, j)
        rho_s = rho_s + (rho_pt(i, j)) ^ 2

        'calculate residual--"true" numerator SSE
        Res(i, j) = (O_pt(i, j) - P_pt(i, j)) ^ 2
        SSE_obs_a = SSE_obs_a + Res(i, j)
        Res_b(i, j) = (mu_pt(i, j) - P_pt(i, j)) ^ 2
        SSE_obs_b = SSE_obs_b + Res_b(i, j)
        Res_c(i, j) = (mu_pt(i, j) - Ptilde_pt(i, j)) ^ 2
        SSE_obs_c = SSE_obs_c + Res_c(i, j)


        'cacluate residual for individual NMSE?
           'Res_1(1, j) = (O_pt(1, j) - P_pt(1, j)) ^ 2
           'Res_2(2, j) = (O_pt(2, j) - P_pt(2, j)) ^ 2
           'Res_3(3, j) = (O_pt(3, j) - P_pt(3, j)) ^ 2

        'output to view random numbers generated:
        row = row + 1
        output.Cells(row, 24) = RandomNum(i, j)
        output.Cells(row, 25) = RN_t(i, j)
        m_p(i, j) = rho_pt(i, j) * RandomNum(i, j)
        m_p_sum = m_p_sum + m_p(i, j)
        m2(i, j) = rho_pt(i, j) * RN_t(i, j)
        m2_sum = m2_sum + m2(i, j)
        f_sum = f_sum + (RandomNum(i, j)) ^ 2
        ft_sum = ft_sum + (RN_t(i, j)) ^ 2
        c1(i, j) = RandomNum(i, j) * RN_t(i, j)
        c_sum = c1(i, j) + c_sum


Next i
Next j

'average terms:
O_bar = O_sum / pt
e_bar = e_sum / pt
rho_bar = rho / pt
rho_bar_lump = rho_lump / t
m1_final = 2 * m_p_sum
m2_final = 2 * m2_sum
c_final = 2 * c_sum

SSE_obs_a_alt = rho_s + m1_final - m2_final + f_sum - c_final + ft_sum

'calculate "true" denominator SSTO
For i = 1 To p
   For j = 1 To t
        v(i, j) = (O_pt(i, j) - O_bar) ^ 2
```

```
        SSTO_obs_a = SSTO_obs_a + v(i, j)
        v_b(i, j) = ((O_pt(i, j) - RandomNum(i, j)) - (O_bar - e_bar)) ^ 2
        SSTO_obs_b = SSTO_obs_b + v_b(i, j)


    Next j

Next i

'calculate "true" MSE and NMSE
NMSE_obs = SSE_obs_a / SSTO_obs_a
MSE_obs = SSE_obs_a / pt
NMSE_obs_b = SSE_obs_b / SSTO_obs_b
MSE_obs_b = SSE_obs_b / pt
NMSE_obs_c = SSE_obs_c / SSTO_obs_b
MSE_obs_c = SSE_obs_c / pt

'calculate theoretical derived terms:

    'single ij:
SSE_E_a = rho_s + pt * mean_t * (mean_t - 2 * rho_bar) + pt * var + pt * var_t
SSTO_E_a = C + pt * var * (1 - (1 / pt))
SSTO_E_a_alt = (pt - 1) * var + C_alt
SSE_E_b = rho_s + pt * mean_t * (mean_t - 2 * rho_bar) + pt * var_t
SSTO_E_b = C
SSE_E_c = rho_s '+ pt * mean_t * (mean_t - 2 * rho_bar)
SSTO_E_c = SSTO_E_b

    'difference terms:
d_SSE_E_b = pt * var
d_SSTO_E = pt * var * (1 - (1 / pt))

    'difference of SSE_c assuming B=0:
    d_SSE_E_c = pt * (var + var_t)

    'difference of SSE_c not using that assumption:
    d_SSE_E_c2 = SSE_E_a - SSE_E_c
NMSE_a_E = (1 / (1 + (pt * var - var) / C)) * (rho_s / C + pt * mean_t * (mean_t - 2 * rho_bar) / C + pt *
var_t / C + pt * var / C) 'SSE_E_a / SSTO_E_a
adj_denom_b = SSTO_obs_a - d_SSTO_E
adj_num_b = SSE_obs_a - d_SSE_E_b
adj_num_c = SSE_obs_a - d_SSE_E_c2

'adjusted MSE and MSE calculations from theory:
a_NMSE_b = (SSE_obs_a - d_SSE_E_b) / (SSTO_obs_a - d_SSTO_E)
a_MSE_b = (SSE_obs_a - d_SSE_E_b) / pt
a_NMSE_c = (SSE_obs_a - d_SSE_E_c) / (SSTO_obs_a - d_SSTO_E)
a_NMSE_c2 = (SSE_obs_a - d_SSE_E_c2) / (SSTO_obs_a - d_SSTO_E)
a_MSE_c = (SSE_obs_a - d_SSE_E_c2) / pt
```

```
'output results:
rowout = rowout + 1
output.Cells(rowout, 1) = x
output.Cells(rowout, 2) = SSE_obs_a
output.Cells(rowout, 3) = SSE_E_a
output.Cells(rowout, 4) = SSTO_obs_a
output.Cells(rowout, 5) = SSTO_E_a
output.Cells(rowout, 6) = SSE_obs_b
output.Cells(rowout, 7) = SSE_E_b
output.Cells(rowout, 8) = SSTO_obs_b
output.Cells(rowout, 9) = SSTO_E_b
output.Cells(rowout, 10) = SSE_obs_c
output.Cells(rowout, 11) = SSE_E_c
output.Cells(rowout, 12) = MSE_obs_b
output.Cells(rowout, 13) = a_MSE_b
output.Cells(rowout, 14) = NMSE_obs_b
output.Cells(rowout, 15) = a_NMSE_b
output.Cells(rowout, 16) = MSE_obs_c
output.Cells(rowout, 17) = a_MSE_c
output.Cells(rowout, 18) = NMSE_obs_c
output.Cells(rowout, 20) = a_NMSE_c2
output.Cells(rowout, 21) = adj_denom_b
output.Cells(rowout, 22) = adj_num_b
output.Cells(rowout, 23) = adj_num_c
output.Cells(rowout, 26) = MSE_obs_b - a_MSE_b
output.Cells(rowout, 27) = NMSE_obs_b - a_NMSE_b
output.Cells(rowout, 28) = MSE_obs_c - a_MSE_c
output.Cells(rowout, 29) = NMSE_obs_c - a_NMSE_c2

Next x

End Sub
```