ArcNLET 2.0: New ArcNLET Function of Monte Carlo Simulation for Uncertainty Quantification

User's Manual

Manuscript Completed: July, 2013

Prepared by
J. Fernando Rios, Ming Ye, and Huaiwei Sun

Department of Scientific Computing, Florida State University, Tallahassee, FL 32306

Rick Hicks, FDEP Contract Manager

Prepared for
Florida Department of Environmental Protection
Tallahassee, FL

Revision Sheet

Release No.	Date	Revision Description
Rev. 1.0.0	July 8, 2013	Initial release

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1   INTRODUCTION

This manual describes the new function of Monte Carlo (MC) simulation implemented in ArcNLET Version 2.0. ArcNLET, an ArcGIS-based Nitrate Load Estimation Toolkit, was developed to model fate and transport of nitrate in surficial groundwater aquifer from onsite wastewater treatment systems (OWTS) (a.k.a., septic systems) to surface water bodies. The previous versions of ArcNLET (1.0 and 1.1) produce an estimated value of nitrate load to specified surface water bodies. Descriptions of the development and application of ArcNLET 1.0 and 1.1 for load estimation are referred to Rios et al. (2013) and Wang et al. (2013), respectively. More information of ArcNLET can be found at the software website available at http://people.sc.fsu.edu/~mye/ArcNLET/index.html.

The ArcNLET-based nitrate load estimates are inherently uncertain, and the uncertainty needs to be quantified before the estimates (and/or their statistics) are used for environmental management and planning. The MC simulation of ArcNLET 2.0 is developed to facilitate the uncertainty quantification by addressing parametric uncertainty in the following seven model parameters:
(1)   Smoothing factor,
(2)   Longitudinal dispersivity,
(3)   Horizontal transverse dispersivity,
(4)   First-order denitrification coefficient,
(5)   Hydraulic conductivity,
(6)   Porosity, and
(7)   Source nitrate concentration.
Before running the MC simulation of ArcNLET 2.0, the users need to determine the random and deterministic parameters among the seven parameters above. ArcNLET 2.0 is designed to be flexible to consider a single or multiple random parameters. In addition, the user needs to characterize the random parameters by specifying their probabilistic distributions. ArcNLET 2.0 can handle four distributions: uniform, triangular, normal, and lognormal. In ArcNLET 2.0, the first four parameters are randomly homogeneous and the latter three are randomly heterogeneous. Details of characterizing the random variables are given in Chapter 2.

The outputs of the MC simulation consists of multiple values of nitrate concentrations at user specified locations (monitoring points) and nitrate load at all water bodies involved in the modeling. These values represent ArcNLET predictive uncertainty due to parametric uncertainty. The predictive uncertainty of the variables can be quantified by estimating their distributions and/or statistics (e.g., mean and variance) in post-processing within or outside of ArcNLET.

Like ArcNLET 1.0, ArcNLET 2.0 is implemented as an extension to ArcGIS from ESRI Inc. Integrating this model with ArcGIS makes it easy to incorporate the spatial nature of data, such as the locations of individual septic tanks and spatially variable hydraulic conductivity and porosity. This is particular useful when dealing with uncertainty of randomly heterogeneous fields. Finally, embedding the model within ArcGIS facilitates the pre- and post-processing of model data, as well as the visualization of model results.

The MC simulation is set up and executed via a graphical user interface (GUI), created as an extension to ArcGIS, and accessed as a tool on the toolbar of the main ArcMap window.   In order to facilitate the user interaction, a point and click approach is used, as it is more user friendly than the input file oriented interaction used in traditional groundwater modeling software such as MODFLOW and MT3DMS.

Readers of this manual should be familiar with the basics of uncertainty analysis and working with ArcNLET 1.0 and 1.1 by reading ArcNLET user's manual (Rios et al., 2011a) and technical manuals (Rios et al., 2011b). It is also recommended to be familiar with advanced uses of ArcNLET for model calibration by reading ArcNLET application manual (Wang et al., 2011). All the references can be downloaded from ArcNLET software website available at http://people.sc.fsu.edu/~mye/ArcNLET/index.html.

1.1   Organization of the Manual

The structure of the manual is as follows: the manual begins with an abbreviated description of the concepts and methods of uncertainty analysis employed in the software (Chapter 2), followed by the installation of the software and a general overview on its functioning (Chapter 3). Finally, in Chapter 4, an example problem is provided for preparing the input files and executing the software.

The measurement units used in this manual may vary between metric and imperial units. However required units will always be explicitly stated.

In order to make this manual easier to read, a certain typographic convention has been adopted:
- Model inputs and parameters always appear in **bold** font.
- Names of attributes in a shapefile's attribute table are always shown in a `typewriter` font.

1.2   Acronyms and Abbreviations

In this manual, acronyms or terms that are abbreviated are spelled out in full the first time they appear. Table 1-1 is a list of acronyms and abbreviations used in this manual:

Table 1-1. Abbreviations

DEM	Digital Elevation Model
GIS	Geographic Information System.
FDEP	Florida Department of Environmental Protection
GUI	Graphical User Interface
NED	National Elevation Dataset
NHD	National Hydrography dataset
OWTS	Onsite Wastewater Treatment System.
SA	Spatial Analyst (extension for ArcGIS)
MC	Monte Carlo
UQ	Uncertainty Quantification
LHS	Latin Hypercube Sampling

2   BACKGROUND

This chapter briefly describes the concepts and methods of uncertainty quantification used in the software. Generally speaking, uncertainty is lack of certainty for a quantity of interest (e.g., nitrate concentration or load estimate). For a process-based model like ArcNLET, there are four sources of uncertainty as follows:

(1)   Model parameters (e.g., hydraulic conductivity) vary in space and time, but the variability cannot be fully characterized by available data. This is always referred to as parametric uncertainty.

(2)   Model structure (e.g., flow and transport models) used for the modeling is inadequate to describe physical, chemical, and biological processes, and there may be multiple conceptual-mathematical models that are plausible with given data and information. This kind of uncertainty is always referred to as model uncertainty.

(3)   Model input data (e.g., recharge and water use per household in the future) are subject to spatial and temporal variability that cannot be fully characterized. There is no consensus on the name of this kind of uncertainty, and it is here referred to as scenario uncertainty.

(4)   Measurements of model parameters and observations of state variables (e.g., hydraulic head and nitrate concentration) are subject to measurement errors, spatial and temporal variability, scarcity, and/or scale discrepancy. These may give rise to data uncertainty.

The focus of this study is the **parametric uncertainty** in the seven ArcNLET parameters listed in Chapter 1. Model uncertainty is disregarded, since ArcNLET is the only model. Because the model is steady state, scenario uncertainty becomes irrelevant. It is further assumed that uncertainty caused by measurement error is small relative to that due to parametric uncertainty.

2.1   Basic Ideas for Quantifying Parametric Uncertainty

Below is the basic procedure of quantifying parametric uncertainty:

(1)   Identify uncertainty parameters whose values are not known with certainty. Selecting the random parameters is problem specific and depends on available data. For example, an uncertain parameter at one site may become more certain at another site where sufficient measurements are available.

(2)   Characterize uncertainty of random parameters using probabilistic and statistical methods. The common practice of characterizing parametric uncertainty is to specify probabilistic distribution of the parameters, e.g., uniform and normal distributions.

(3)   Conduct MC simulation to propagate the uncertainty in model parameters (e.g., hydraulic conductivity) to model outputs (e.g., ArcNLET load estimate). The uncertainty propagation involves two steps. The first is to generate random samples of the parameters according to their probability distributions, which is implemented using the Latin Hypercube Sampling (LHS) method in ArcNLET 2.0. The other step is to run the model (ArcNLET) for each set of the parameter samples, which is automated in ArcNLET 2.0 through the GUI.

(4)   Quantify the uncertainty in model outputs by calculating their statistical properties (e.g., mean and variance) and more comprehensively their probability distributions. This post-processing can be done within or outside of ArcNLET.

Instead of providing a single value of model outputs, the procedure above provides multiple output values. Different values occur with different probability, and the uncertainty is summarized in the statistics or probability distributions of the outputs. The distributions are more informative than the statistics and can be used to assess risk, i.e., the probability that the load exceeds a specific threshold or performance measure target value. This can provide more valuable information to decision/policy makers for making science-informed decisions.

When the number of random parameters is relatively large, it is useful to only work with the parameters most influential to model outputs so that the cost of running MC simulation and preparing input data for MC simulation can be reduced. Selecting the influential parameters can be done by conducting sensitivity analysis, which however is beyond the scope of ArcNLET 2.0. Discussions and examples of sensitivity analysis using ArcNLET are referred to Wang et al. (2012).

2.2   Deterministic and Random Parameters

A model parameter can be viewed as either deterministic or random, depending on the problem of interest and the amount of information and data used to measure the certainty. A deterministic parameter is the one known with high certainty (may not be 100% though), and a random parameter is the one known with low certainty (it is not completely unknown though). For example, topographic elevation obtained from digital elevation model (DEM) is an input (a model parameter in a broader sense) to ArcNLET. While the exact DEM values are unknown (even when the elevation is measured, the measurement is subject to measurement error), the accuracy of the DEM data used in ArcNLET is considered to be sufficient to load estimation, and the uncertainty associated with DEM is negligible in comparison with that of other parameters. As a result, DEM is treated as a deterministic variable in the ArcNLET MC simulation. For the same reason, the location of water bodies is also treated as a deterministic variable. On the contrary, the decay coefficient associated with denitrification needs to be considered as a random variable, because its exact value is entirely unknown for many ArcNLET modeling problems. The only available information is its range found in the literature that is of several orders of magnitude. In addition, the load estimate is highly sensitivity to the decay coefficient (Wang et al., 2012). As a result, uncertainty of the decay coefficient leads to large uncertainty in the load estimate. Treating the decay coefficient as a deterministic variable may lead to undesired decision-making.

In earth and environmental science and engineering, a spatially random parameter can be categorized into two classes: *randomly homogenous* and *randomly heterogeneous*. A randomly homogeneous parameter is homogeneous in space but its value is random. A randomly heterogeneous parameter is more complicated, because the parameter value varies in space and, at each location, its value is random. While all model parameters vary in space, certain parameters have significantly smaller variability than other parameters and can be viewed as spatially homogeneous. However, if a parameter is not homogeneous but there is no data available to characterize heterogeneity, one may treat the parameter as a randomly homogeneous parameter so that its heterogeneity is reflected in the probability distribution to some extent by virtue of the ergodicity assumption. This conceptualization is certainly less accurate than using

the concept of randomly heterogeneous parameters. In ArcNLET 2.0, the randomly homogeneous parameters include (1) smoothing factor, (2) longitudinal dispersivity, (3) horizontal transverse dispersivity, and (4) first-order denitrification coefficient. Following the common practice of groundwater solute transport modeling, the horizontal transverse dispersivity is assumed to be proportional to the longitudinal dispersivity. In other words, multiplying the longitudinal dispersivity to a multiplier (less than one) gives the horizontal transverse dispersivity. The multiplier is specified by ArcNLET users.

To fully describe a randomly heterogeneous parameter, in a strict sense, one needs to know probability distribution of the parameter at each location. Theoretically speaking, this is possible for the source nitrate concentration (C0) in ArcNLET, because one can specify the probability distribution for each septic system. Therefore, ArcNLET 2.0 is designed in this way that each C0 can have its own distribution. However, in practice, obtaining such distributions is difficult, if not impossible. In this case, one may use the same distribution of C0 for all the septic systems.

For randomly heterogeneous parameters, in a weaker sense, one can assume that the distribution is the same at all the locations and further assume that it is Gaussian with a constant mean and a covariance that is a function of distance between two locations. This function is always given by a variogram model, which can be estimated using spatial measurements. However, the measurements are severely limited in many cases of ArcNLET modeling. To resolve the problem of data scarcity and maintain certain level of accuracy for uncertainty quantification, ArcNLET 2.0 uses the concept of *randomly zonal heterogeneous parameters*, a compromise between the randomly homogeneous and heterogeneous parameters. Taking hydraulic conductivity as an example, it is delineated into soil zones as in the SSURGO database, and the delineation is deterministic. Within each soil zone, hydraulic conductivity is randomly homogeneous and its value follows a probability distribution. In ArcNLET 2.0, such parameters are hydraulic conductivity and porosity. For convenience of discussion, these parameters are still referred to as randomly heterogeneous parameter.

2.3   Probability Distributions of Random Parameters

As discussed above, probability distribution is needed to characterize uncertainty of both randomly homogeneous and heterogeneous variables. ArcNLET 2.0 considers four commonly used distributions: uniform, triangular, normal, and lognormal distributions. These distributions are described briefly in this section. The distribution here is the probability density function (PDF) of a random variable. Integrating a PDF from infinity to a value x give the cumulative distribution function (CDF). The concept of CDF is used for the Latin Hypercube Sampling in the next section.

If a random variable X follows uniform distribution, its PDF, f, is

$$f(x) = \frac{1}{B-A}, \quad A \leq x \leq B, \tag{1}$$

where A and B are the minimum and maximum values of X, respectively. The capital X denotes a random variable; the small x is a specific of X. Equation (1) means that the random variable X has the same probability to take any value between A and B. If one knows nothing about X but its minimum and maximum values, the uniform distribution should be used.

If one knows not only the minimum and maximum but also the mode of X (the mode corresponds to a value of X that has the highest probability), then the triangular distribution can be used. Its PDF is defined as follows:

$$
\begin{aligned}
f(x) &= \frac{2(x-A)}{(B-A)(C-A)} \quad A \leq x \leq C \\
&= \frac{2(B-x)}{(B-A)(B-C)} \quad C \leq x \leq B
\end{aligned}
\tag{2}
$$

where A, B, and C are the minimum, maximum, and mode values of X, respectively. Figure 2-1 shows an example PDF of the triangular distribution with $A=0$, $B=4.0$, and $C=2.5$. The triangular distribution suggests that the probability of X decreases linearly from the mode to the minimum and maximum values.
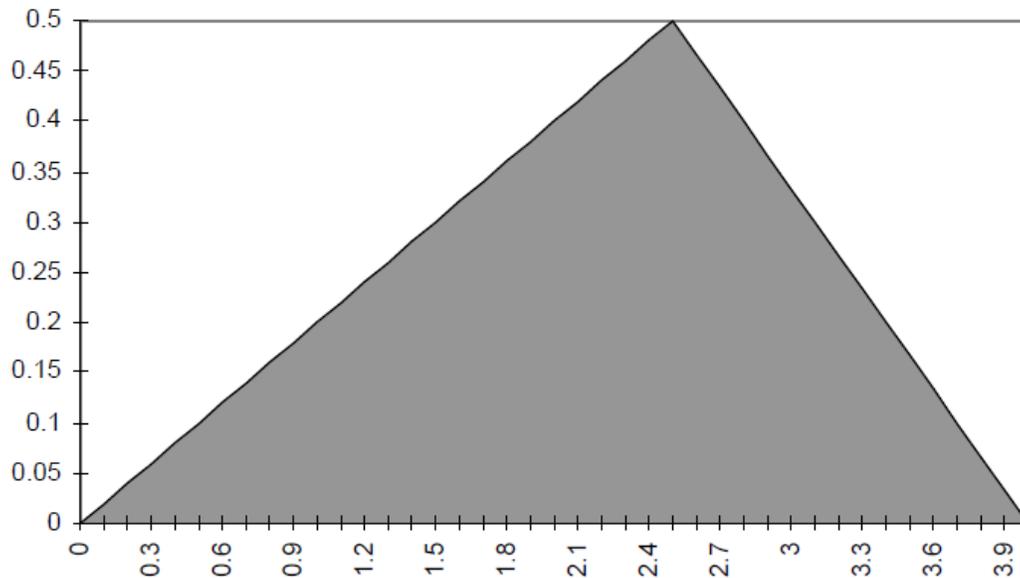
Figure 2-1. Triangular probability density function of a random variable with the minimum, mode, and maximum values are 0, 2.5, and 4.0, respectively.

The most commonly used probability distribution is the normal distribution, also known as Gaussian distribution, whose PDF is defined as

$$
f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad -\infty < x < \infty,
\tag{3}
$$

where μ and σ are mean and standard deviation, respectively. Since X of equation (3) can be negative, which may not be physically reasonable for certain parametrs, ArcNLET 2.0 uses the truncated normal distribution. Instead of providing to ArcNLET 2.0 the mean and standard deviation, one needs to provide its 1% and 99% percentiles, which include 98% of the random values of X. The percentiles are calculated with the mean and standard deviation, via

$$V_{0.01} = \mu - 2.33\sigma$$
$$V_{0.99} = \mu + 2.33\sigma \;.$$

(4)

For the truncated normal distribution, if the 1% percentile is nonnegative, all random samples generated from the distribution are nonnegative.

Another popular distribution is the lognormal distribution, and its PDF is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad -\infty < x < \infty,$$

(5)

where μ and σ are mean and standard deviation of ln$x$, respectively. If a random variable follows the lognormal distribution, its logarithm follows the normal distribution. Therefore, it is always convenient to work on $\ln(x)$, and then transform it back. For example, the 1% and 99% percentiles are calculated via

$$V_{0.01} = \exp\left(\mu - 2.33\sigma\right)$$
$$V_{0.99} = \exp\left(\mu + 2.33\sigma\right) \;.$$

(6)

Similar to the normal distribution, the 1% and 99% percentiles are needed in ArcNLET 2.0.

Table 2-1 lists the suggested parameter distributions and their defining statistics based on literature review and our experience of ArcNLET modeling. The specific reasons of selecting the distributions and suggested values of the defining statistics are as follows:

(1)   **Smoothing factor**. This parameter is specific to ArcNLET and it is used to smooth the topography to generate the shape of water table. Larger smoothing factor results in smoother water table shape and thus smaller hydraulic gradient. The parameter is site specific and strongly correlated to site topography. A fine resolution of topography requires large value of smoothing factor. Rios (2010) tested several smoothing factors for a groundwater model developed for the U.S Naval Air Station (NAS), Jacksonville. The model was calibrated by Davis et al. (1996) and the calibrated water table was used as the reference to evaluate the best smoothing factor. It was found that a value of 50 yielded a good approximation to the water table. However, this may not apply elsewhere. Therefore, the smoothing factor is assumed to follow uniform distribution and the range is set as 20 ~ 80 empirically. One may also use a triangular distribution and set the mode as 50.

(2)   **Longitudinal dispersivity** and horizontal transverse dispersivity. The longitudinal dispersivity is assumed to follow normal distribution. The 1% and 99% percentiles are taken as 0.21m and 21.34m, respectively. These values are determined based on literature data of Davis (2000) and Gelhar et al. (1992). Davis (2000) used a longitudinal dispersivity value of about 2.134m. Since longitudinal dispersivity may vary two orders of magnitude (Gelhar et al., 1992), the range of longitudinal dispersivity is determined as 0.21 m – 21.34 m. Following the tradition of contaminant transport modeling, the horizontal transverse dispersivity is assumed to be 10% of the longitudinal dispersivity.

(3)   **First-order decay coefficient of denitrification**. Based on a literature review, McCray et al. (2005) gave a range of 0.004 - 2.27 1/d for this parameter (Table 2-2). Heatwole and McCray (2007) plotted the cumulative frequency distribution of the literature data

(Figure 2-2) and showed that the parameter follows lognormal distribution. Figure 2-2 shows that although the possible maximum value is as large as 2.27/d, more than 80% of the data is below 0.5/day and 90% below 1.0/day, which indicates that the value is not likely to be larger than 1.0/day in most occasions. Heinen (2006) summarized the decay coefficient values used in several models (Table 2-3). Among the models, since models DAISY, DRAINMOD-N and SWATNIT have the same unit (mass per volume) of nitrate-nitrogen concentration as that used in ArcNLET, the parameter values used in these models are of special use to this study. The range used in DRAINMOD-N is 0.004-1.08/d, which supports the observation above that the coefficient values are unlikely to be larger than 1.0. In addition, the values used in model DAISY and SWATNIT (0.1 and 0.01/d, respectively) fall in the range of 0.004-1.08. Therefore, the range of the first-order denitrification coefficient in this study is determined to be 0.004-1.08/d.

(4)   **Hydraulic conductivity**. While there is a consensus that hydraulic conductivity follows lognormal distribution, obtaining the mean and standard deviation of the normal distribution is not easy due to lack of data. In many cases of ArcNLET modeling, the only data of hydraulic conductivity are those in the SSURGO database. It includes low, representative, and high values of hydraulic conductivity for each soil zone. These values do not support lognormal distribution. For example, the minimum and maximum values are more or less symmetric with the representative values. Since the SSURGO database is site specific, the distribution of hydraulic conductivity is assumed to be triangular, whose two ends are the low and high values and the mode is the representative value.

(5)   **Porosity**. Normal and lognormal distributions have been used in literature for porosity. Due to lack of data, porosity is handled in the same manner to that for hydraulic conductivity. In other words, triangular distribution is assumed for porosity, its minimum, mode, and maximum values are taken as the low, representative, and high values of porosity given in the SSURGO database. However, it has been found that these values are unavailable for some areas in SSURGO. In this case, one may have to treat porosity as a deterministic variable and find a representative value. This should not have adverse effect on uncertainty quantification, because the randomness of porosity is significantly smaller than that of other parameters, considering that porosity varies between 0.1 and 0.4 for most soils.

(6)   **Source nitrate concentration**. It is assumed that the source nitrate concentration follows normal distribution, and its range (the 1% and 99% percentile) is taken as 25 mg/L and 80 mg/L based on the review article of McCray et al. (2005).

Note that the distributions and defining statistics are site specific and problem specific. ArcNLET 2.0 users must be cautious when using them for their modeling.

Parameter	Distribution	Minimum	Mode	Maximum
Smoothing Factor	Uniform	Yes	N/A	Yes
Longitudinal Dispersivity	Normal	1% percentile	N/A	99% percentile
Decay Coefficient	Lognormal	1% percentile	N/A	99% percentile
Hydraulic Conductivity	Triangular	Yes	Yes	Yes
Porosity	Triangular	Yes	Yes	Yes
Source Plan Concentration	Normal	1% percentile	N/A	99% percentile

Table 2-1. Recommended parameter distributions and needed statistics for defining the distributions.

Process/Reaction Order	Median	Range	Number of Data
Nitrification—zero order, μ' (mg/L/d)	264	156 to 1464	7
Nitrification—first order, k_1 (per day)	2.9	0.0768 to 211.2	19
Denitrification—first order, k_2 (per day)	0.025	0.004 to 2.27	53

[1]Geng et al. (1996); Cho (1971); Ling and El-Kadi (1998); Yamaguchi et al. (1996); Starr et al. (1974); Starr and Gillham (1993); Misra et al. (1974); Ardakani et al. (1974a, 1974b); Lind (1983); Slater and Capone (1987); Anderson (1998); Trudell et al. (1986); Smith and Duff (1988); Bengtsson and Annadotter (1989); Francis et al. (1989); Obenhuber and Lowrance (1991); Smith et al. (1991); Ekpete and Cornfield (1965); Christensen et al. (1989); Bradley et al. (1992); Tesoriero et al. (2000); Smith et al. (1996); Lawrence and Foster (1986); Korom (1992); Hiscock (1991).

Table 2-2: Range of decay coefficient of denitrification based on literature review. (source: Table 2 of McCray et al., 2005)
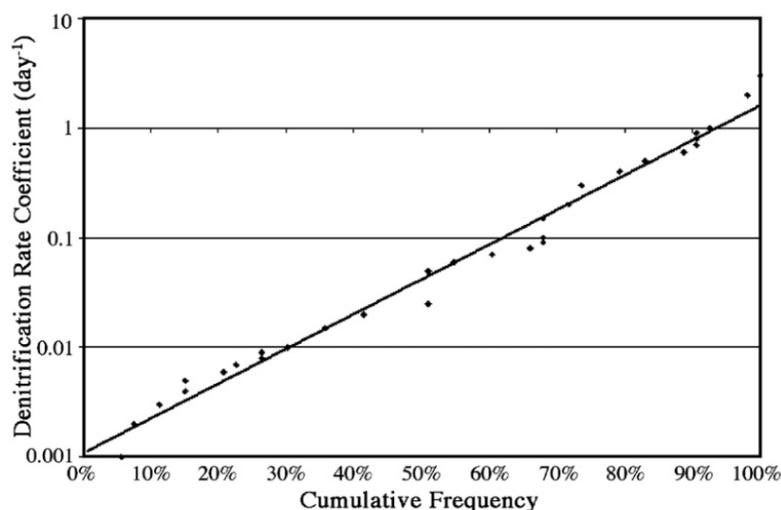
Figure 2-2. Cumulative frequency distribution for first-order denitrification coefficient. (source: Figure 4 of Heatwole and McCray, 2007)

k_d	Units	Units of N	Model
0.1	d^{-1}	kg N m^{-3}	DAISY
0.004–1.08	d^{-1}	g N m^{-3}	DRAINMOD-N
0.1	d^{-1}	kg N ha^{-1}	IMPACT
0.001–0.005	d^{-1}	g N m^{-2}	LEACHM
10	yr^{-1}	mol N ha^{-1}	NUCSAM
0.01	d^{-1}	mg N L^{-1}	SWATNIT

Table 2-3 Values of first-order denitrification coefficient (k_d) used in several models. (source: Table 3 of Heinen, 2006)

2.4   Latin Hypercube Sampling

After the random parameters are identified and their probability distributions are determined, MC simulation can be conducted to quantity predictive uncertainty due to parametric uncertainty. MC simulation consists of two steps, both of which have been automated in ArcNLET 2.0. The first step is to generate multiple random samples of the parameters, also called parameter realizations. The second step is to run the model for each parameter realization to yield simulated quantities of interest (e.g., nitrate concentration and load). Since the second step is straightforward, we only briefly describe the first step of random number generation.

ArcNLET 2.0 uses the Latin Hypercube Sampling (LHS) code of Iuzzolino (2003) to generate random parameter samples based on their probability distributions. When generating M samples for a single parameter, the sampling procedure is as follows (Swiler and Wyss, 2004):

(1)   The cumulative distribution function (CDF) is divided into M segments of equal width $1/M$, considering that the range of CDF is between 0 and 1.

(2)   For each segment, a probability value is randomly sampled using a uniform distribution.

(3)   Map the probability value to the parameter space according to the probability distribution function of the parameter.

In the LHS sampling, the random samples are generated from the entire parameter ranges, even with a relatively small sample number. If there are N random parameters, repeating this procedure for each parameter leads to N vectors of parameters, each of which has M numbers. If the parameters are correlated, the order of the M numbers in each vector is adjusted until the correlation is achieved. Note that LHS uses the Spearman rank correlation, instead of the Pearson linear correlation, because the former is applicable for nonlinear correlation. If the parameters are not correlated, the ordering is also performed to ensure that there is no correlation between the parameters. Parameter correlation is not considered in ArcNLET 2.0, because there is always lack of data to characterize the correlation. When characterizing parameter correlation becomes possible, it is straightforward to revise ArcNLET 2.0 to include the correlation into the random sampling process.

3   USAGE

3.1   Preliminaries

Before using this new version of ArcNLET to conduct Monte Carlo simulations for quantifying uncertainty in ArcNLET-estimated nitrate load estimates, it is necessary to become familiar with the previous versions of ArcNLET by reading ArcNLET user's manual (Rios et al., 2011a) and technical manuals (Rios et al., 2011b). It is also recommended to be familiar with advanced uses of ArcNLET for model calibration by reading ArcNLET application manual (Wang et al., 2011). This manual includes one example of Monte Carlo simulations; more examples of sensitivity analysis and uncertainty quantification using ArcNLET are referred to Wang et al. (2012), for which Monte Carlo simulations were conducted using a command line version of ArcNLET 1.1, not this new version. All the references can be downloaded from ArcNLET software website available at http://people.sc.fsu.edu/~mye/ArcNLET/index.html.

3.1.1 System Requirements

In order to use the software, the following is required:

- A Microsoft Windows computer (32 or 64 bit) that meets the minimum requirements for ArcGIS 10.0 service pack 5.

- ArcGIS 10.0 service pack 5

- SA (Spatial Analyst) extension for ArcGIS

- .NET Framework 3.5 or later versions (Available from www.microsoft.com/net)

A computer having a 3 GHz processor and at least 2 GB of system memory is recommended. Under normal program execution, a number of temporary files are created which may require a significant amount of disk space (depending on the current operation and model parameters). As a result, it is recommended to have several gigabytes of free disk space to store these and any other temporary files generated by ArcGIS.

3.1.2 Installation

Installation of the software is straightforward by running the appropriate setup package. The software must be installed by system administrator. To install an updated version, the previous version must be uninstalled by running Add/Remove Software under Control Panel. With ArcGIS installed, install the extension in the following procedure:

1. Double click the setup.exe installation file.

2. Follow the wizard's instructions.

After successful installation, it is necessary to add the extension to the ArcGIS toolbar as follows:

1. Open ArcMap.

2. Open the Customize → Customize Mode menu.

3. In the dialog that appears, select the entry ArcNLET (Figure 3-1).

In case the ArcNLET entry is not present, manual installation is required. To manually add the ArcNLET entry to the list in the Customize dialog,

1. Click the "Add from file…" button in the window of Figure 3-1.

2. Navigate to the installation directory (the default is C:\Program Files\ArcNLET).

3. Select the ArcMapCommand.tlb file.

The ArcNLET entry should now appear in the Customize dialog. Check the box as indicated in Figure 3-1 to enable the toolbar shown in Figure 3-2.
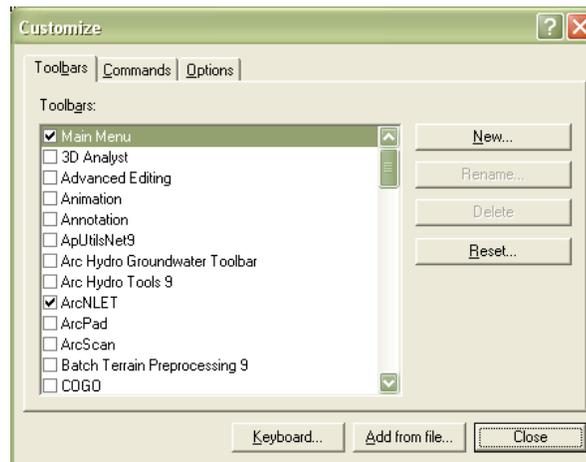

Figure 3-1. Adding the toolbar to ArcMap

Notes:

   The User's Manual (this document) and the data package for the example problem are not included as part of the installation file and must be downloaded separately from the website (http://people.sc.fsu.edu/~mye/ArcNLET/index.html).

   In some cases, the program cannot be uninstalled from the Windows Control Panel (only the option to repair the install appears). In this case, the program may be uninstalled by running the setup.exe file again and selecting the 'remove' option.

   If McAfee antivirus software is installed, it may block access to the system temporary folder, which is used by ArcNLET during model run. This may cause errors such as access to some files is denied. The users need to work with their system administrators to resolve this issue.

3.1.3 Starting the Program

   The software is executed from the buttons on the toolbar (Figure 3-2). The button on the left is the original ArcNLET program. The button in the middle corresponds to the interactive particle tracking functionality. The rightmost button is the uncertainty analysis module of ArcNLET.

Figure 3-2. Toolbar

3.2   Main Window

The main software window (Figure 3-3) is separated into tabs, much like in ArcNLET 1.0 and 1.1.   In the lower portion, there is a scrolling log window that shows the current software activity. The typical workflow progresses through each tab from left to right.
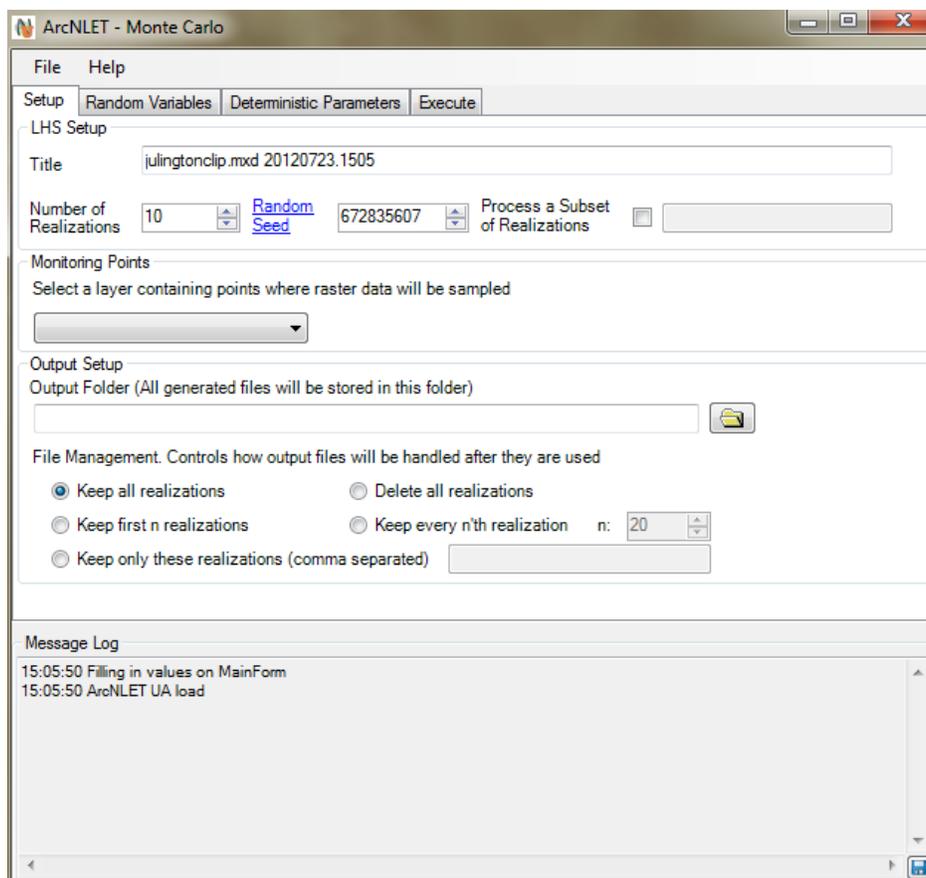

Figure 3-3. Main ArcNLET MC window

3.2.1 Setup Tab

This tab sets up the LHS parameters and output file handling options.

Title: The title of the project. The project file and associated realization files will have this string as their base name.

Number of Realizations: The number of samples to generate. Must be larger than 1 and less than 10,000 (LHS software limit).

Random Seed: The random seed to use for the LHS random number generator to generate random parameter samples. Must be an integer between 0 and 2147483647 (LHS software limit). A new seed is automatically generated each time the software is started. Whenever the random seed changes, the random parameter samples change. As a result, the results of Monte Carlo simulation change. <u>If one wants to duplicate the Monte Carlo simulation, the same random seed must be used.</u>

Process a Subset of Realizations: Allows for the processing of a subset of the total number of realizations given by **Number of Realizations.** This is useful for executing the Monte Carlo run in pieces, on separate computers. For example, if 200 Monte Carlo runs are desired, one computer could run realizations 1 to 100 and another one could run 101 to 200. For this to work, the **Random Seed** must be set to the same number on both computers.

Monitoring Points: A point layer specifying the locations at which heterogeneous random variables values and concentration values will be sampled. After the execution run completes, the file *out0MonitorPtsInputs.shp* contains the value of all random variables at the points specified for all realizations.

Output Folder: The folder to which the results will be saved. The selected folder must be empty before the run is initiated (except in the case where the run is executed using previously generated inputs, see the Execute tab).

File Management: Controls how generated output files are handled. There are several options:

- Keep all realizations: The output files of all realizations remain in the output folder on disk

- Delete all realizations: The output files of all realizations are deleted after each realization is complete

- Keep first n realizations: Does not delete the first n realizations, deletes the rest after each realization is complete. The value of n must be less than or equal to Number of Realizations.

- Keep every n'th realization: Does not delete every n'th realization. The value of n must be less than or equal to Number of Realizations:

- Keep only these realizations: A comma separated list of individual realization outputs to preserve.

Note that the **File Management** options only apply to the outputs generated from each run. The input files for each run are not deleted so that one can re-run the individual realizations if needed.

3.2.2 Random Variables Tab

This tab sets up the parameters for each random variable. It is separated into two sub-tabs: one for spatially homogeneous random variables and the other for spatially heterogeneous random variables.

3.2.2.1 Spatially Homogeneous Random Variables

Spatially homogeneous random variables are ones that are represented by a single probability distribution function everywhere in space. For ArcNLET modeling, these random variables are the smoothing factor (smthFac), longitudinal dispersivity (dispL), horizontal transverse dispersivity (dispTH), and decay coefficient (k) related to denitrification. Following the common

practice of transport modeling, it is assumed that dispTH is proportional to dispL, and the multiplier (should be less than one) is specified by the users of ArcNLET.
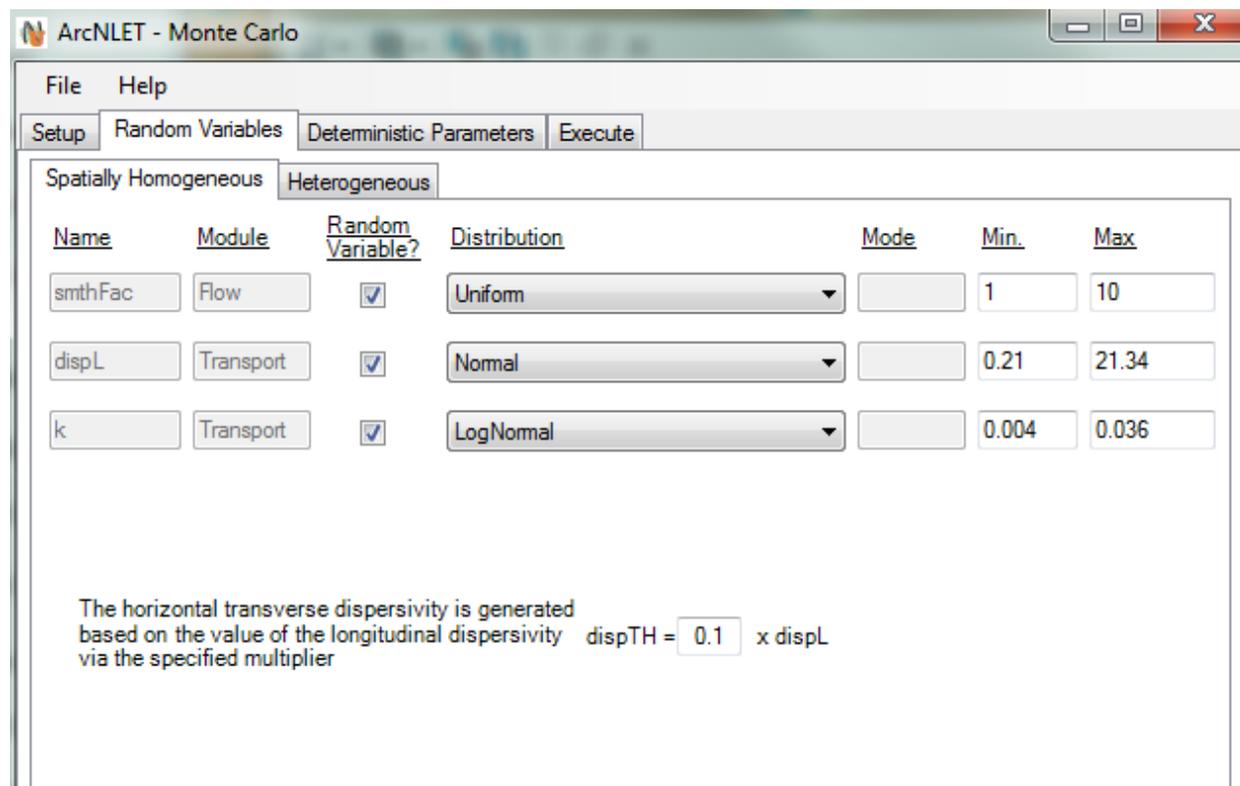

Figure 3-4. Spatially Homogeneous random variables tab

The parameters for the spatially homogenous random variables (Figure 3-4) are set in the same way for each variable. Hence, the settings have been separated into columns, which are described as follows.

Name: The name of the variable. Currently there are three smthFac (the smoothing factor), dispL (longitudinal dispersivity), k (decay coefficient). Refer to the ArcNLET manual for details on these parameters. The variable names cannot be changed.

Module: The module that a particular variable is associated with. This aids in recalling which module a particular variable is part of. These values cannot be changed.

Random Variable?: This check box, when checked, indicates that the particular variable is to be considered as a random variable. When checked, a **Distribution** must be specified along with the associated distribution parameters (mode, minimum, and maximum as required). When unchecked, the variable is treated as a deterministic parameter and does not vary during the Monte Carlo simulation. The deterministic value of the parameter is set by entering it in the **Mode** box.

Distribution: The statistical distribution that a particular variable follows. Details of the distributions are given in Section 2 of the manual. Depending on the selected distribution, the appropriate boxes for **Mode, Min,** and **Max** will be enabled. The supported distributions are:

- Uniform: Requires the **Min** and **Max** values.

- Triangular: Requires the **Mode**, **Min** and **Max** values.

- Normal: Requires the **Min** and **Max** values. In this case, the minimum and maximum values correspond to the 0.01 and 0.99 percentiles of a normal distribution.

- LogNormal: Requires the **Min** and **Max** values. In this case, the minimum and maximum values correspond to the 0.01 and 0.99 percentiles of a lognormal distribution.

Mode, Min, Max: The mode, minimum and maximum values of the selected distribution. Depending on the value of **Distribution,** not all of these values are required.

dispTH: The transverse dispersivity is calculated based on the longitudinal dispersivity, **dispL,** based on the provided multiplier. As a rule of thumb, **dispTH** is one tenth of **dispL**. The multiplier should be less than one.

3.2.2.2 Spatially Heterogeneous Random Variables

Spatially heterogeneous random variables are ones that are represented by different probability distribution functions at different locations in space. For ArcNLET modeling, these random variables are hydraulic conductivity (hyCon), porosity (porosity), and source plane concentration (C0/Source).

Similar to the case of spatially homogeneous random variables, spatially heterogeneous random variables are arranged in a column format (upper part of Figure 3-5). The difference is that each random variable is now composed of sub-entities consisting of a mix of random and deterministic parameters as shown in the lower part of Figure 3-5, which is the attribute tables of the layer files. These sub-entries are soil zones for hydraulic conductivity and porosity, and individual septic tanks for source plane concentration. The users can determine whether the parameters associated with the sub-entries are random or deterministic.
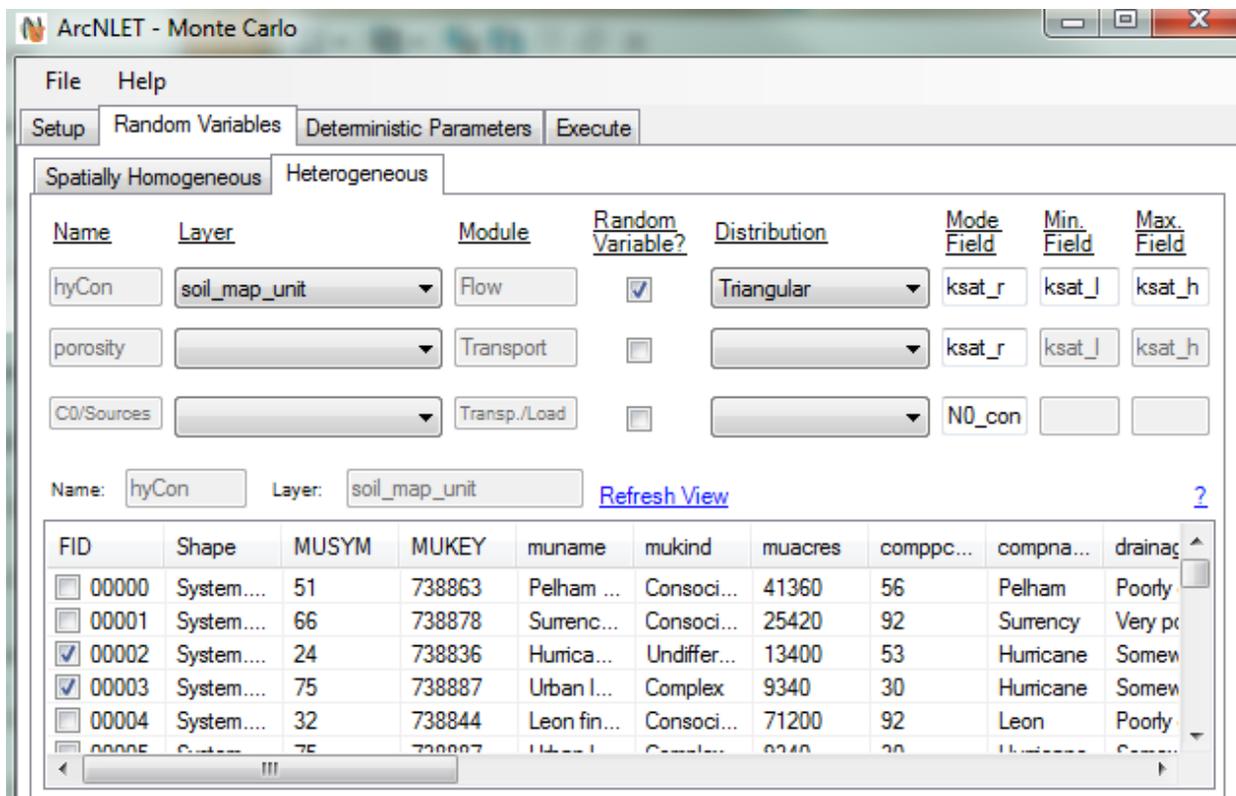
Figure 3-5. Spatially heterogeneous random variables

Name: The name of the variable. Currently, there are three: hyCon (hydraulic conductivity polygons), porosity (polygons), and C0/Sources (the point locations of the contaminant sources along with the corresponding value of the source plane concentration, C0, at each source). The variable names cannot be changed.

Layer: A polygon or point layer (depending on the variable) containing zones (in the case of a polygon layer) or points (in the case of a point layer) where each zone/point can be potentially considered an independent random variable.

Module: Same as the case of spatially homogeneous random variables. The module that a particular variable is associated with. This aids in recalling which module a particular variable is part of. These values cannot be changed.

Random Variable?: Similar to the case of spatially homogeneous random variables. This check box, when checked, indicates that the particular variable is to be considered as a random variable. When checked, a **Distribution** must be specified along with the associated distribution parameters (mode, minimum, and maximum as required). This distribution applies to all the sub-entries shown in the lower part of Figure 3-5. When unchecked, the variable is treated as a deterministic parameter and does not vary during the Monte Carlo simulation. The deterministic value of the parameter is set by entering it in the **Mode Field** box.

Distribution: Similar to the case of spatially homogeneous random variables. The statistical distribution that a particular variable follows. Details of the distributions are given in Section 2 of the manual.

Mode Field, Min Field, Max Field: The field within the attribute table of the selected **Layer** containing the value of the mode/minimum/maximum for each random variable. Depending on the selected **Distribution**, not all of these fields are required.

The lower part of Figure 3-5 is the feature selector. When a layer is selected using the **Layer** drop down, the Name and Layer boxes of the feature selector are set to the name of the variable currently being edited and the layer associated with that variable. The table shows all of the features in the layer indicated by the value of the Layer box. In Figure 3-5, the attribute table of the soil_map_unit layer is being shown. This layer is, in turn, associated with the hyCon variable.

Each feature shown in the feature selector can potentially be a random variable. To set a particular feature as a random variable, check the box in the FID column of the feature to consider as a random variable. To select all the FIDs, right click the list and Select All. Note that the selection boxes disappear if the **Random Variable?** box is unchecked. In Figure 3-5, the features with FID 2 and 3 will be considered as random variables. The unchecked features are considered as deterministic parameters with value equal to the value contained in the **Mode Field.** When selecting which features should be considered as random variables, attention must be given to the Name and Layer boxes to ensure that the correct layer and variable are being edited.

3.2.3 Deterministic Parameters Tab

This tab of **Deterministic Parameters** (Figure 3-6) sets the remaining inputs and parameters required by ArcNLET that are not considered to be random. Parameters that are spatial in nature are arranged in a column layout on the left side of the tab.

Name: The name of the spatial parameter. There are currently two: DEM (the input digital elevation model raster), and waterBodies (the polygon layer indicating the locations of water bodies).

Module: The module that a particular variable is associated with. This aids in recalling which module a particular variable is part of. These values cannot be changed.

Layer: A raster layer for the DEM and a polygon layer for the water bodies.
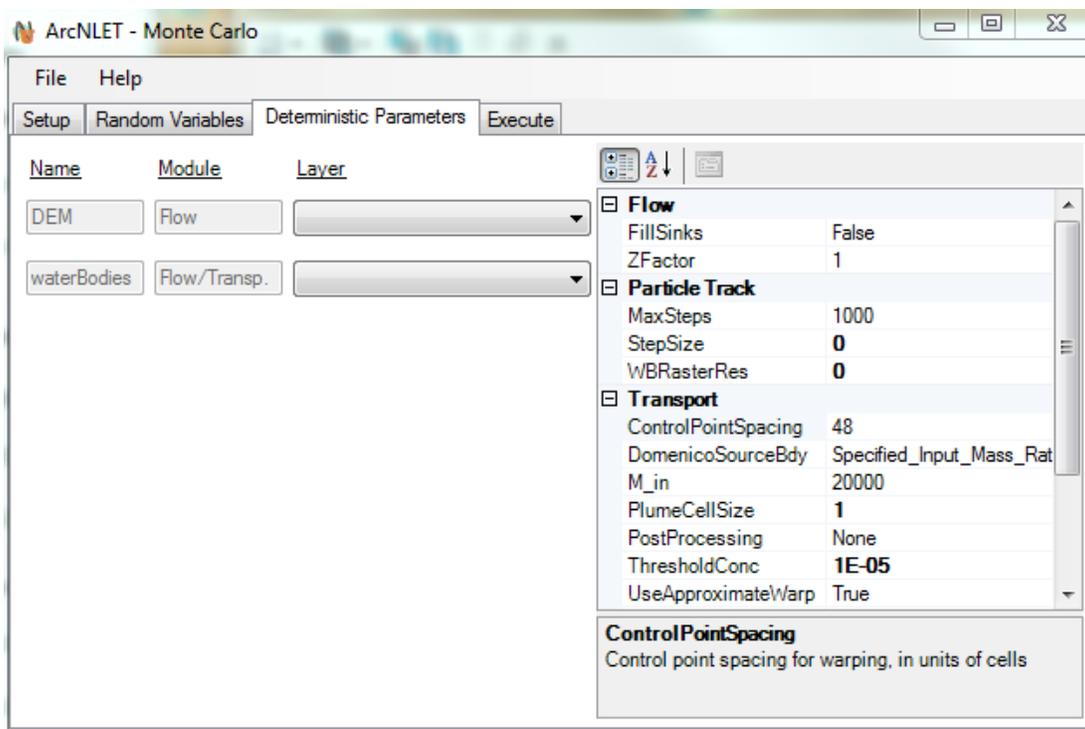
Figure 3-6. Deterministic Parameters tab.

The non-spatial parameters are presented in a compact list on the right hand side of Figure 3-6. The parameters are grouped according to the module to which they belong. See the ArcNLET manual for an explanation of each parameter. A brief description of each parameter is given in the box at the bottom-right corner of the tab.

3.2.4   Execute Tab

The execute tab (Figure 3-7) handles the execution of the Monte Carlo run.

Run Type: Sets the following three ways of conducting the Monte Carlo simulation:

- Generate Input Files Only: Generates the input files for all realizations but does not begin the Monte Carlo run. This run type is useful to check that the realizations indeed have the statistical properties desired before conducting the Monte Carlo run.

- Run Using Existing Input Files. Runs Monte Carlo using a set of previously generated input files. Upon clicking this option, an ArcNLET UQ project file (extension .proj) generated previously by the Generate Input Files Only option must be selected. When this run type is selected, all inputs from all other tabs are ignored except the Monitoring Points and Process a Subset of Realizations inputs of the Setup tab.

- Generate Input Files and Run: Generates the input files for each realization and immediately begins the Monte Carlo run.
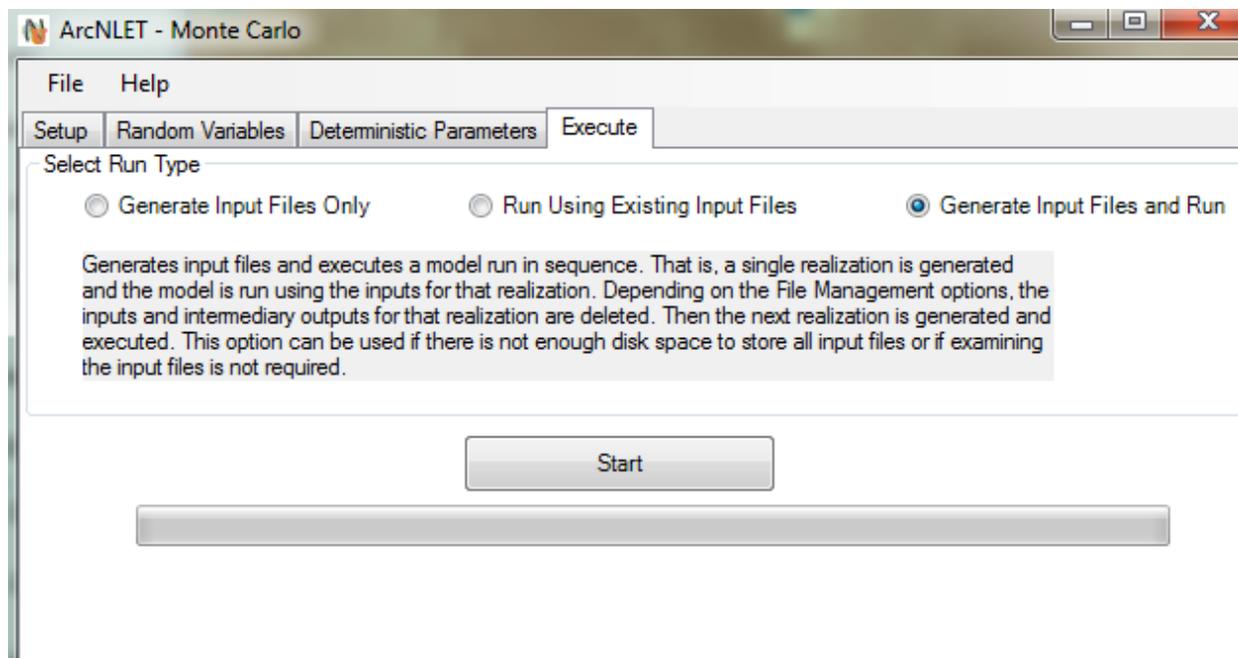
Figure 3-7. Execute tab.

3.3   Generated Input and Output Files

After the run is initiated in the Execute tab, a series of files are generated, which are now described. For simplicity, it is assumed that the **Run Type** of Generate Input Files and Run was selected, as this run type includes the other two run types.

As discussed above, the first part of a Monte Carlo run consists of generating the input files required by ArcNLET. This is done in two steps. The first step is to generate multiple realizations of the random parameters using the LHS code. The next step is to generate ArcNLET input files for each parameter realization. The output files of the LHS run are as follows:

- A text file with name *Title_in.txt*, where **Title** corresponds to the project title specified in the Setup Tab (Section 3.2.1). This file is the input file passed from the GUI to the LHS code for the generation of random numbers. The LHS execution is invisible to ArcNLET users.

- Two text files with names *Title_out1.txt* and *Title_out2.txt*. These files are generated by the LHS software. The first file contains a formatted list containing the generated sample values as well as detailed information on the generated samples; the second file contains just the generated samples.

In general, the user does not need to understand these files. The generated ArcNLET input files based on the LHS samples are as follows:

- An input file with name *Title_<realization>.input*, where <realization> corresponds to some realization number. This file contains all the information required by ArcNLET to execute the realization in batch mode. For example, this file contains the names of the raster files of DEM, hydraulic conductivity, porosity as well as the shape files of septic tanks and water bodies.

- For each heterogeneous random variable specified in the randomly heterogeneous variables tab, a shapefile with the name *<variable name>_<realization>.shp* is generated. This shapefile contains the value(s) of the heterogeneous random variable(s) generated by LHS in the column with name *<variable name>*.

  For polygon random variables (i.e., hydraulic conductivity and porosity), the shape file is converted to a raster format with the same name but extension .img. For example, hyCon_0001.shp is converted to hyCon_0001.img.

- A shapefile with name *out0MonitorPtsInputs.shp*. This file contains the values of all the randomly heterogeneous variables at each location given by **Monitoring Points** for each realization. Table 2 details the meaning of each field.

All of these files can be found in the **Output Folder** specified by the user in the Setup tab.

The second part of the Monte Carlo run consists of running ArcNLET for each realization, using the values specified in each file with the .input extension. Each realization generates a series of output files, one set per ArcNLET module. The naming convention is as follows:

- The Flow module generates *outFlowDir_<realization>.img* and *outFlowMag_<realization>.img* corresponding to the raster files of flow velocity direction and magnitude.

- The Particle Tracking module generates a file *outPtrack_<realization>.shp*.

- The Transport module generates *outPlumes_<realization>.img*, which is a raster containing the calculated plumes and *outPlumes_<realization>_info.img* which is the associated "info" file needed for the load calculation.

- The Load Estimation module generates outLoad_<realization>.csv. This is a comma separated ASCII file containing the load for each water body (identified by the water body's FID).

Additionally, two files are generated that record the results of each realization:

- The file *out0MonitorPtsOutputs.shp* contains the value of concentration at each monitoring point for all realizations. The linkID and realizNum fields have the same meaning as the same fields in *out0MonitorPtsInputs.shp* (see Table 3-1).

- The file *out0WBLoadOutputs.shp* contains the load value for each water body for all realizations. Table 3-2 details the meaning of each field.

Table 3-1. out0MonitorPtsInputs.shp

Field Name	Description
linkID	Links a feature to the FID of the **Monitoring Points** shapefile
realizNum	The realization to which this measurement belongs to
<various random variables>	The remaining columns are variable depending on which variables were set to be heterogeneous random variables in the setup. Out of those, only ones that are area-based will be sampled. In the current version of the software, the remaining columns can represent hydraulic conductivity or porosity

Table 3-2. out0WBLoadOutputs.shp

Field Name	Description
linkID	Links a feature to the FID of the water bodies input shapefile (the waterBodies parameter in the Deterministic Parameters tab)
realizNum	The realization to which this measurement belongs to
outLoad	The load to the water body in units of M/T
outLoadRF	The outLoad multiplied by the risk factor
outDecay	Out of all sources whose flow path terminates at the water body indicated by linkID, this value indicates the total rate of mass lost due to first-order decay in units of M/T.
inLoad	Out of all sources whose flow path terminates at the water body indicated by linkID, this value indicates the total input load to groundwater from the constant concentration plane due to those sources in units of M/T.

4   EXAMPLE

4.1   Introduction and Objective

The final goal of this user's manual is to show an example workflow process for using the new feature of Monte Carlo (MC) simulation. In this chapter, the basic steps required to set up MC simulation, run MC simulation, and post-processing of MC results are discussed. Program features related to this specific example will also be explained. The desired output of the model in this example is the probability distribution and statistics of nitrate load due to septic tanks on the target water body in the Julington Creek neighborhood of Jacksonville, Florida. This neighborhood is west to the St. Johns River, and ArcNLET modeling for this neighborhood was conducted by Wang et al. (2011, 2013). In particular, Wang et al. (2012) conducted sensitivity analysis and uncertainty for this neighborhood.

Since MC simulation is time consuming, the tutorial is not to conduct the MC simulation for the entire neighborhood, but for a clipped area with a relatively small number of septic systems, so that the reader can quickly practice the MC simulation. The procedure of conducting MC simulation for the clipped area is the same as that for the entire neighborhood and any area of interest. The data used in this tutorial and output files generated by the MC simulation are provided as part of the "julingtonclip_MC_example.zip" package which can be downloaded from the software website available at http://www.people.sc.fsu.edu/~mye/ArcNLET under the Downloads section. Extract the julingtonclip_MC_example.zip file to a folder on your computer. Preferably, the path should not contain folders with spaces in their names as this may cause some ArcGIS tools to fail to execute.

Among the seven random parameters considered in ArcNLET 2.0, the five below are considered to be random:

(1)   Smoothing factor,

(2)   Longitudinal and transverse dispersivity,

(3)   Horizontal transverse dispersivity,

(4)   First-order denitrification coefficient, and

(5)   Hydraulic conductivity.

Porosity and source nitrate concentration are treated as deterministic parameters. This is to demonstrate the flexibility of the MC simulation for handling the mixture of random and deterministic parameters.

It should be noted that this example serves only for the purpose of training, not for realistic estimation of nitrate load in the Julington Creek neighborhood.

4.2   Description of Sample Data

In this example, the files needed for ArcNLET run are prepared by following the instruction of the user manual of ArcNLET 1.0 (Rios et al., 2011). These files are:

(1)   *lidardem10m.img*: Lidar DEM with 10m resolution. This domain is a small portion of the Julington Creek neighborhood.

(2)   *waterbodies.shp*: Water bodies in the neighborhood.

(3)   *septic_tanks_all.shp*: Septic tank involved in the modeling

(4)   *soil_map_unit.shp*: SSURGO database of hydraulic conductivity and porosity

Note that the raster files of hydraulic conductivity and porosity needed for ArcNLET modeling are not prepared at this stage. The two files are generated for each MC realization during the software execution. Other variables needed for the ArcNLET modeling are specified in the **Deterministic Parameters** Tab as described in the next section.

In addition to the files above, an optional point layer is prepared as file "*monitorpts_soil_map_unit.shp*". It specifies the locations at which heterogeneous random variables values and concentration values will be sampled. The locations can be the points of significance to environmental management. In the example MC run, these points are arbitrarily selected.

4.3   Running the MC Simulation

4.3.1 Setup Tab

After add the four shape files described in Section 4.2 into ArcMap, save the ArcMap file as "julingtonclip.mxd". This file name will be used by ArcNLET 2.0 to generate the default **Title** name. We here go through the steps of setting the MC simulation in the **Setup** Tab:

Title: After clicking the uncertainty analysis button shown in Figure 3-2, the default title name is "julingtonclip.mxd DDDDDDDD.TTTT", where DDDDDDDD (e.g., 20130703) and TTTT (e.g., 0930) are the date and time, respectively. The user can give any name to the title. We use "*julingtonclip*" in this manual.

Number of Realizations: The number of random samples to generate. ArcNLET 2.0 can run up to 10,000 realizations (LHS software limit). A MC simulation with larger number of realizations is better as long as computational time and resource allow. While one can conduct a statistical analysis to determine how many realizations are needed to reach MC convergence, a simulation with more than 100 realizations should be sufficient for many cases of ArcNLET modeling. In this manual, we only run *10* realizations so that the download package julingtonclip_MC_example.zip is not too large to be downloaded.

Random Seed: The default random seed of *1609248789* is used for the LHS sampling. The user can change it to any integer between 0 and 2147483647 (LHS software limit). A new seed is automatically generated each time the software is started. Whenever the random seed changes, the random parameter samples change. As a result, the results of MC simulation change. If one wants to duplicate the MC simulation, the same random seed must be used. After the MC simulation run, the random seed can be found in the LHS output file "*julingtonclip_in.txt*" in the output folder specified below.

Process a Subset of Realizations: Since the number of MC realization is small, this feature is not used.

Monitoring Points: The layer file is set as "*monitorpts_soil_map_unit.shp*" described in Section 4.2. After the execution run completes, the file *out0MonitorPtsInputs.shp* contains the value of all random variables at the points specified for all realizations.

Output Folder: The folder to which the results will be saved. The folder name is specified as "*Results*". The full file path should be specified by locating the desired folder through clicking the folder icon.

File Management: The option "*Keep all realizations*" is selected so that the output files of all realizations remain in the output folder on disk and in the download file.

Figure 4-1 shows the snapshot of the **Setup** tab after entering the data and options above.
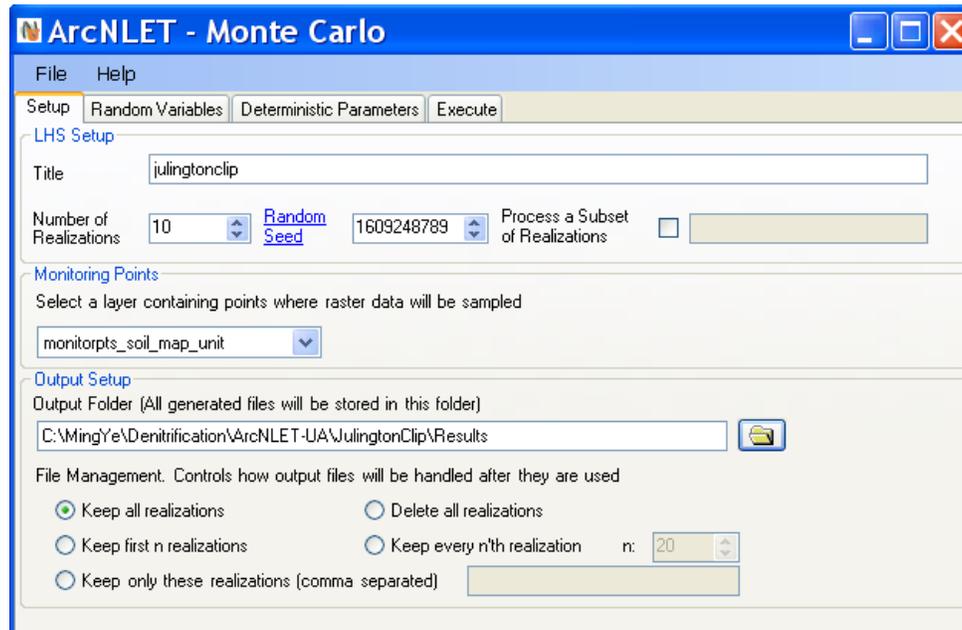

Figure 4-1. Snapshot of the **Setup** tab for the example MC simulation.

4.3.2 Random Variables Tab

The next step is to select the random parameters and their probability distributions and defining statistics through the **Random Variables** Tab.

In the tab "**Spatially Homogeneous**", we choose all the three parameters as random variables. Below are their probability distributions and corresponding defining statistics:

- **smthFac**: the smoothing factor. We choose the uniform distribution and assign 20 and 80 as its minimum and maximum values.

- **disL**: the longitudinal dispersivity. We choose the normal distribution as assign 0.21m and 21.34m as its minimum and maximum values, which are actually the 1% and 99% percentiles of the underlying normal distribution.

- **k**: the first-order decay coefficient of denitrification. We choose the lognormal distribution and assign 0.004/d and 0.036/d as its minimum and maximum values, which are actually the 1% and 99% percentiles of the underlying log normal distribution.

The multiplier between the longitudinal dispersivity and horizontal transverse dispersivity is set as 0.1. Figure 4-2 shows the snapshot of the **Spatially Homogeneous** tab after entering the data and options above.
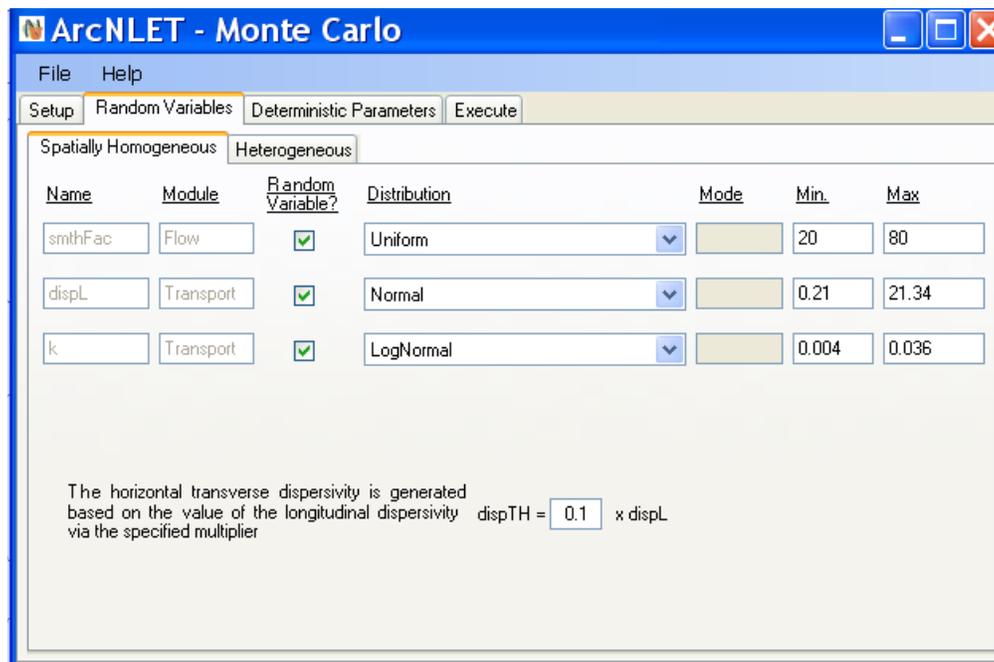

Figure 4-2. Snapshot of the **Spatially Homogeneous** tab for the example MC simulation.

In the tab "**Heterogeneous**", we only choose hydraulic conductivity (hyCon) as a random variable. As discussed in Section 2.3, the triangular distribution is selected. For the modeling area, there are 43 soil zones with FID from 00000 to 00042. We consider that hydraulic conductivity values of all the soil zones are random, and check the FID boxes of all the soil zones by right clicking the list and selecting All. This is an extreme case for demonstration only. The real-world modeling, one should only consider the soil zone influential to the load estimate. Having too many random parameters makes MC simulation computationally expensive and result analysis difficult. The **Mode Field, Min Field, Max Field** are set as the fields of "hyCon_r", "hyCon_l", and "hyCon_h" of the attribute table. The fields names can be anything specified by users and allowed by ArcGIS. For the two deterministic variables, porosity and source plane concentration (C0/Source), we assign the feature of "porosity" and "N0_conc" as in the **Mode Field**, which is actually the deterministic values of the parameters. Figure 4-3 shows the snapshot of the **Heterogeneous** tab after entering the data and options above.
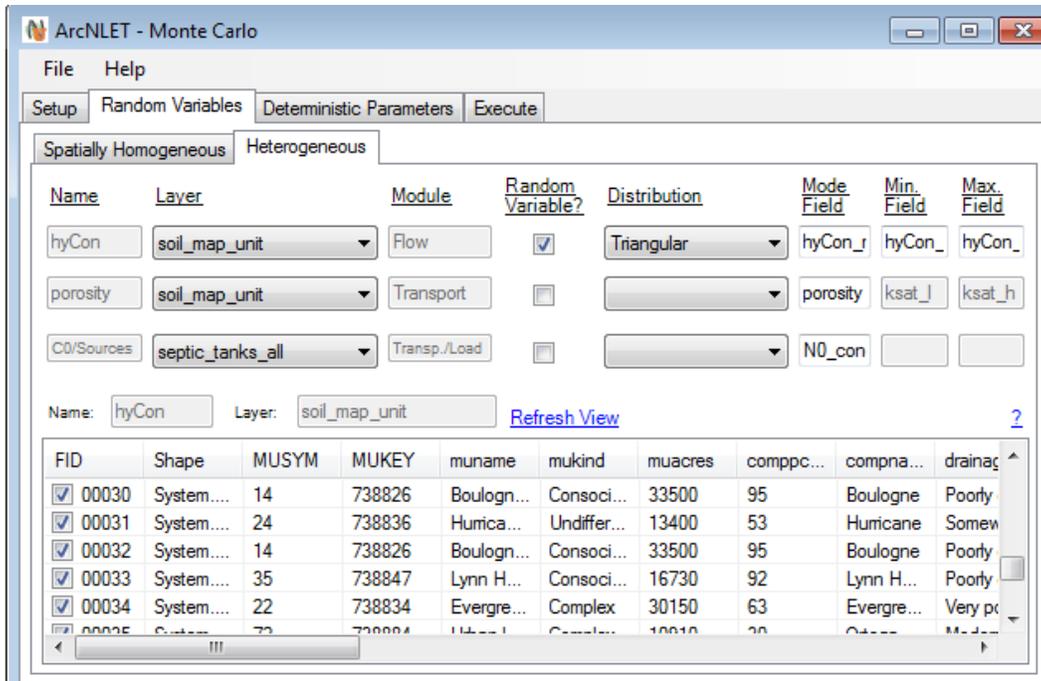
Figure 4-3. Snapshot of the **Heterogeneous** tab for the example MC simulation.

4.3.3   Deterministic Parameters Tab

The DEM and waterBodies files used in this tab are "*lidardem10m.img*" and "*waterbodies.shp*", respectively. The default values are used for other parameters, as shown in Figure 4-4, except that the option of "*Specified_Z*" instead of "*Specified_Input_Mass_Rate*" is used to calculate the load. The default values can be adjusted directly in this tab.
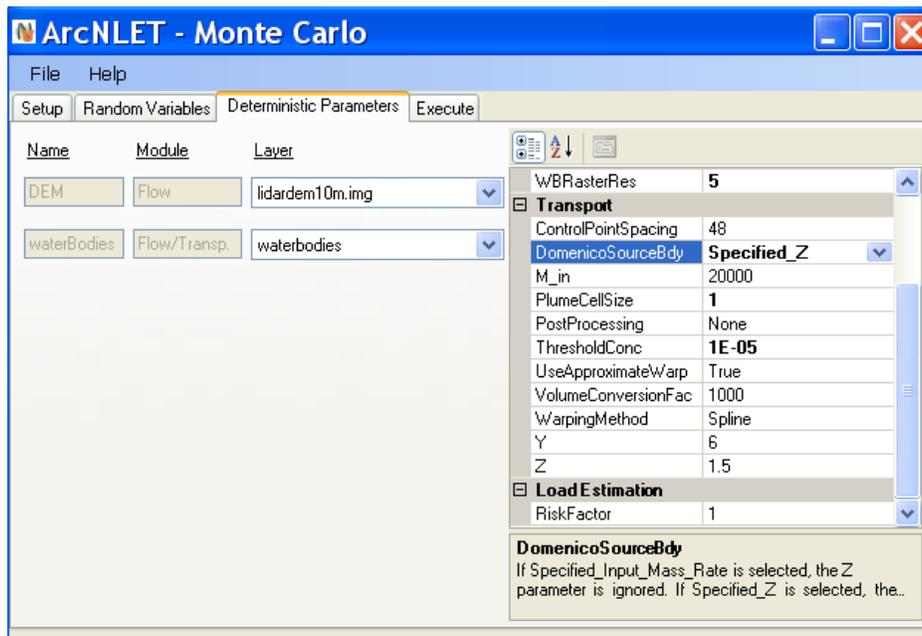

Figure 4-4. Snapshot of the **Deterministic Parameters** tab for the example MC simulation.

4.3.4 Execute Tab

The option of "Generate Input files and Run" is selected in this tab. Clicking the "Start" button triggers the MC simulation.

4.4 Output Files and Post-Processing for Uncertainty Quantification

All the files generated during the MC simulation are saved in the "**Output Folder**" of *results* specified in Section 4.3.1. These files are generated below. The MC simulation generates three auxiliary folders (i.e., g_g42, g_g82, and info); they are only used to save temporary files during the model run, the users do not need to pay any attention to them.

4.4.1 Files for LHS Sampling

There are three files associated with the LHS sampling, one input and two output files, as described below.

- *julingtonclip_in.txt*. This is the LHS input file generated by ArcNLET 2.0 using the information of parameter distributions passed from the GUI. The random seed value can be found at the fourth line of the file. It lists the name of the parameters, their distributions, and defining statistics of the distributions. Note that the information of the decay-coefficient (*k*) is given at the bottom of the file after the information of hydraulic conductivity.

- *julingtonclip_out1.txt.* This is a LHS output files, and it is informative to include all necessary information of the sampling. For those who are familiar with the LHS method, they can evaluate success of the sampling by investigate the information contained in this file.

- *julingtonclip_out2.txt*. This is the other LHS output file, but only includes the generated random samples. These samples are organized by realizations, from realization 1 to realization 10 in this example. The realization index is the first number of each data block. The second value in each block is the number of the random parameters (46 in this example), and the rest values are the random parameter values, in the same order listed in *julingtonclip_in.txt*. For example, the first two values are for the smoothing factor and longitudinal dispersivity.

  This file is important for users of ArcNLET 2.0 to understand different results of ArcNLET modeling between different realizations due to the difference in model parameters. Take the first two realizations as an example. The values of smoothing factor, longitudinal dispersivity, and decay coefficient are 70, 14.9, and 0.015, respectively in realization 1; their corresponding values in realization 2 are 48, 4.8, and 0.025. These different parameter values indicate that the load of the second realization should be smaller than that of the first realization, as shown below.

4.4.2 Input Files of ArcNLET MC Simulation

ArcNLET modeling of each realization requires specific input files of the realization. This is handled by using two input files of ArcNLET described below.

- *julingtonclip_<realization>.input*, where <realization> corresponds to realization numbers. This text file contains all the information required to run ArcNLET in the batch mode; details

---

of the batch run is skipped in the manual so that readers are not distracted from the MC simulation.

- o Values of the **Deterministic Parameters**. Note that porosity and source plane concentration are also treated as deterministic variables in this example. The raster file of porosity, *porosity.img*, is generated during the MC simulation and located in the output folder.

- o Values of the **Random Variables**. For the randomly homogeneous variables, their values vary in each input file. For randomly heterogeneous variables, their file names vary in each input file.

- *hyCon_<realization>.shp* and *hyCon_<realization>.img*, where <realization> corresponds to realization numbers. These are the shape and raster files for randomly heterogeneous hydraulic conductivity. Figure 4-5 plots the hydraulic conductivity of the first two realizations, and shows the difference in hydraulic conductivity between the two realizations.

- *out0MonitorPtsInputs.shp*. This file contains the values of all the randomly heterogeneous variables at each location given by **Monitoring Points** for each realization. This file can be used to directly find the random parameter values at the monitoring points, which can help understand simulated nitrate plume at the points in each realization. Figure 4-6 shows the summary statistics and the histogram of the hydraulic conductivity values at the first monitoring point specified in file *monitorpts_soil_map_unit.shp*.
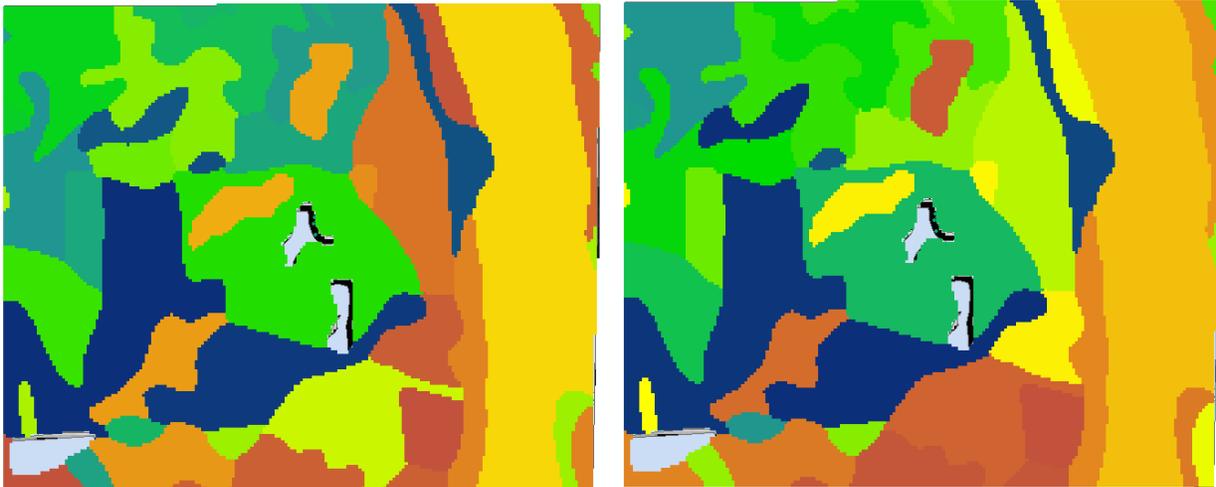

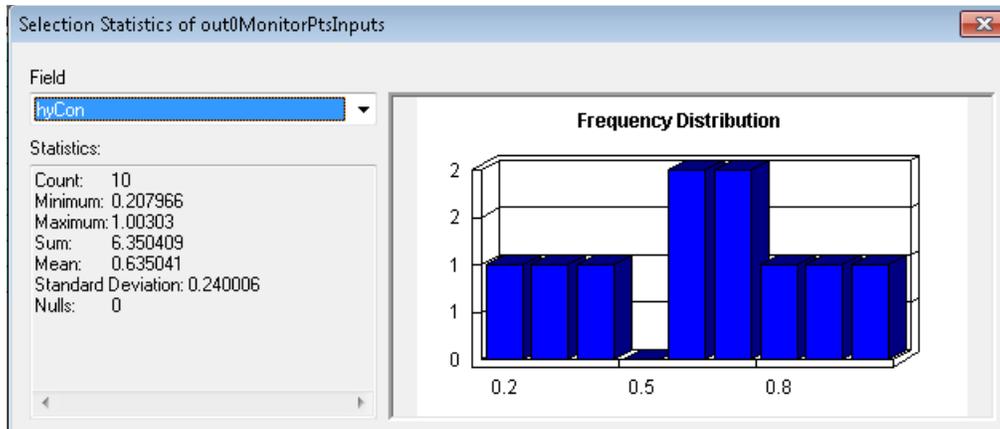Figure 4-5. Plots of hydraulic conductivity for the first (left) and second (right) realization.

Figure 4-6. Summary statistics and histogram of random samples of hydraulic conductivity generated at the first monitoring point.

4.4.3 Output Files of ArcNLET MC Simulation

Among the output files described in Section 3.3, the most important files are *outLoad_<realization>.csv*. These are comma separated ASCII file containing the load for each water body (identified by the water body's FID). The load estimates are calculated based on the output files *outPlumes_<realization>_info.img*, which is the associated "info" file needed for the load calculation. One can process the *.csv* files to quantify uncertainty in the ArcNLET load estimate, such as calculating the summary statistics and plotting the histogram. To facilitate the quantification, ArcNLET 2.0 produces a shape file, *out0WBLoadOutputs.shp*, which contains the load value for each water body for all realizations (Table 3-2). Figure 4-7 shows the summary statistics and histogram of the load to the water body with FID of 15, which is close to the septic tanks. The load estimate varies significantly, ranging between 7000 mg/d and 60,000 mg/d, but with low probability being large. Note that these results are for demonstration only. More MC simulations should be conducted to yield more meaningful statistics.
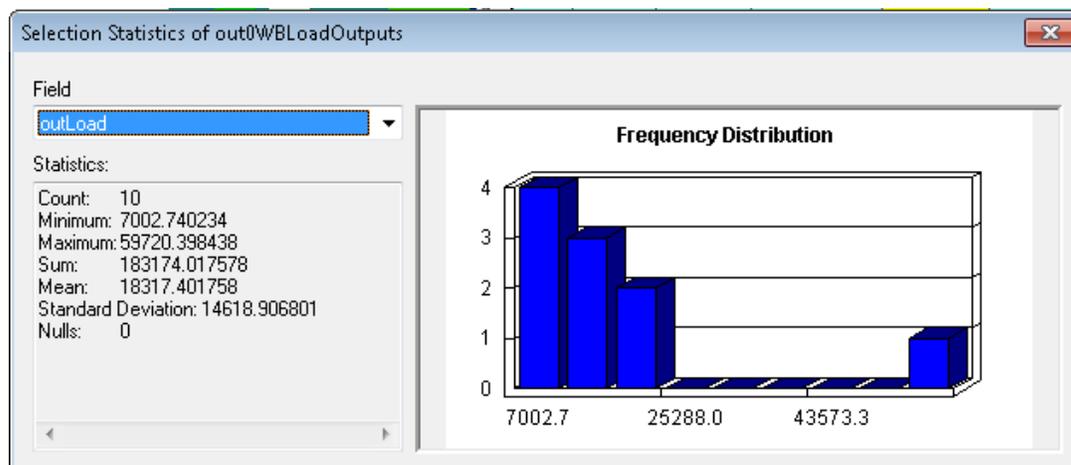

Figure 4-7. Summary statistics and histogram of random load (mg/day) to the water body with FID of 15.

Uncertainty in the load estimate can be better understood by examining the simulated plumes of each realization. We here present an example for the first two realizations, whose nonzero load

estimates to the water bodies are listed in Table 4-1. The table shows that load estimates are significantly different in the two realizations. Figure 4-8 plots the simulated plumes for the two realizations using the output files *outPlumes_<realization>.img*, which is a raster containing the calculated plumes. The plume of realization 1 is wider spread than that of realization 1. As a result, the load to water body with FID of 14 (the one in the north) is zero in realization 2, and the load to water body with FID 15 (the one highlighted) is significantly smaller in realization 2 than in realization 1. The difference in the plumes is caused by the different parameters of the two realizations. For example, the first-order decay coefficients of realizations 1 and 2 are 0.015 and 0.025, respectively. The larger value of decay coefficient in realization 2 is one of the major reasons for the small plume spread. More discussion on effect of model parameters on simulated nitrate plume and load can be found in Wang et al. (2012).

WB FID	Realization 1	Realization 2
14	2833.943	0
15	10859.79	8127.129
19	3930.785	4403.746
21	2944.117	4293.879
26	57136.66	24140.02

Table 4-1. Load estimates (mg/day) to five water bodies in the first two realizations.
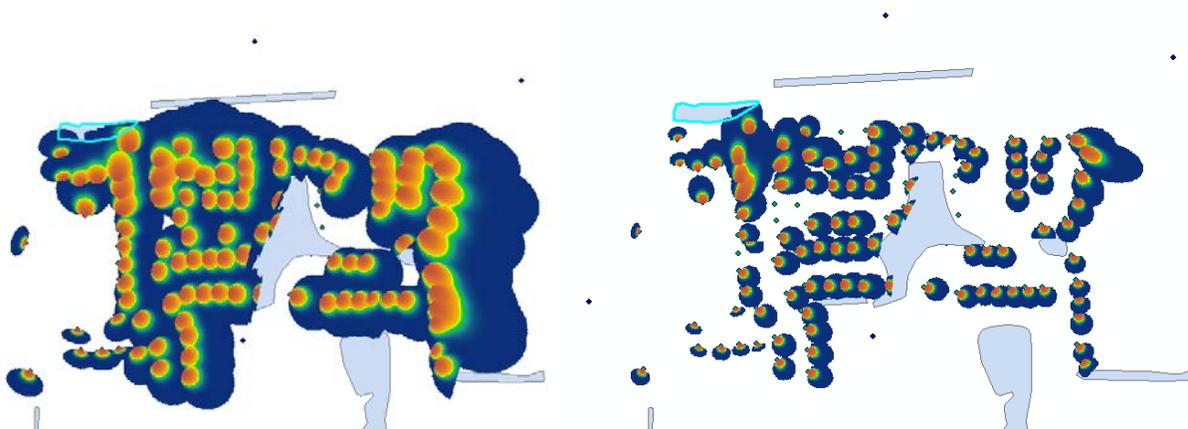

Figure 4-8. Simulated nitrate plumes for the first (left) and second (right) realization.

If one is interested in the nitrate concentration at the monitoring points, the output file *out0MonitorPtsOutputs.shp* can be used. It contains the values of concentration at each monitoring point for all realizations. Figure 4-9 shows the summary statistics of histogram of nitrate concentration at the monitoring point with FID of 8. It shows again the large variability in the simulated concentration.
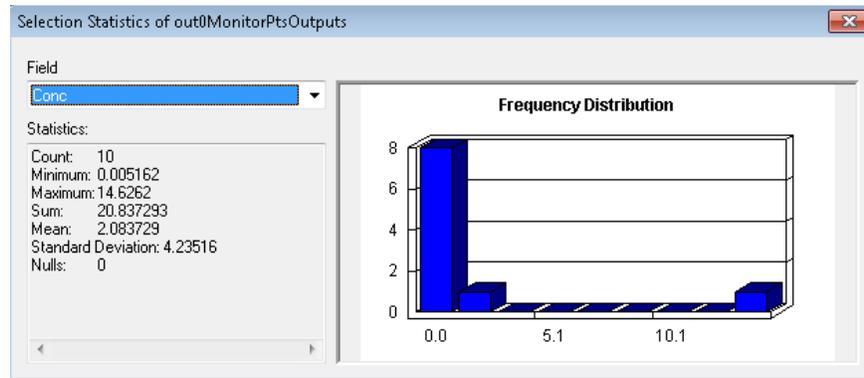
Figure 4-9. Summary statistics and histogram of simulated nitrate concentration at the monitoring point with FID of 8.

5   REFERENCES

Davis, J.H., Planert, M., Andrews, W.J. (1996), *Simulation of Ground-Water Flow at the U.S. Naval Air Station, Jacksonville, Florida, with an Evaluation of Changes to Ground-Water movement Caused by Proposed Remedial Designs at Operable Unit 1*. U.S. Geological Survey, Tallahassee FL. Open-file report 96-597.

Davis, J.H. (2000), *Fate and Transport Modeling of Selected Chlorinated Organic Compounds at Operable Unit 3, U.S. Naval Air Station*, Jacksonville, Florida. U.S. Geological Survey, Tallahassee FL. Open-file report 00-255.

Heatwole, K.K., McCray, J.E. (2007), Modeling potential vadose-zone transport of nitrogen from onsite wastewater systems at the development scale. *Journal of Contaminant 91*, 184-201.

Heinen, M. (2006), Simplified denitrification models: overview and properties. *Geoderma 133*:444–463.

Iuzzolino, H. (2003), Users Manual (Rev 00) for LHS, Version 2.51. Civilian Radioactive Waste Managemenet System Management and Operating Contractor, MOL.20040210.0391, Albuquerque, New Mexico.

Gelhar, L.W., Welty, C., Rehfeldt, K.R. (1992), A critical review of data on field-scale dispersion in aquifers. *Water Resources Research 28 (7)*:1955–1974.

McCray, J.E., Kirkland, S.L., Siegrist, R.L., Thyne, G.D. (2005), Model parameters for simulating fate and transport of on-site wastewater nutrients. *Ground Water 43*, 4, 628–639.

Rios, J.F. (2010), *A GIS-based model for estimating nitrate fate and transport from septic systems in surficial aquifers*. Master's thesis, Florida State University.

Rios, J.F., M. Ye, L. Wang, and P.Z. Lee (2011a) ArcNLET User's Manual. Department of Scientific Computing, Florida State University, Tallahassee, FL. Available online at http://people.sc.fsu.edu/mye/ArcNLET.

Rios, J.F., M. Ye, L. Wang, and P.Z. Lee (2011b) ArcNLET Technical Manual. Department of Scientific Computing, Florida State University, Tallahassee, FL. Available online at http://people.sc.fsu.edu/mye/ArcNLET.

Rios, J.F., M. Ye, L. Wang, P.Z. Lee, H. Davis, and R.W. Hicks (2013), ArcNLET: A GIS-based software to simulate groundwater nitrate load from septic systems to surface water bodies, *Computers and Geosciences*, 52, 108-116, 10.1016/j.cageo.2012.10.003.

Swiler L.P., Wyss G.D. (2004), *A user's guide to Sandia's Latin Hypercube Sampling software: LHS UNIX library/standalone version.* Technical Report SAND2004-2439, Sandia National Laboratories, Albuquerque, NM.

Wang, L., M. Ye, J.F. Rios, and P.Z. Lee Lee (2011) ArcNLET Application Manual. Department of Scientific Computing, Florida State University, Tallahassee, FL. Available online at http://people.sc.fsu.edu/mye/ArcNLET.

Wang, L., M. Ye, J.F. Rios, and P.Z. Lee (2012) Sensitivity Analysis and Uncertainty Assessment for ArcNLET-Estimated Nitrate Load from Septic Systems to Surface Water Bodies.

Department of Scientific Computing, Florida State University, Tallahassee, FL. Available online at http://people.sc.fsu.edu/mye/ArcNLET.

Wang, L., M. Ye, J.F. Rios, R. Fernandes, P.Z. Lee, and R.W. Hicks (2013), Estimation of nitrate load from septic systems to surface water bodies using an ArcGIS-based software, *Environmental Earth Sciences*, doi:10.1007/s12665-013-2283-5.