



**Combining data from several sources to
produce a single set of contact spread
parameters for *NAADSM***

Aaron Reeves
Animal Population Health Institute
College of Veterinary Medicine & Biomedical Sciences
Colorado State University
Fort Collins, Colorado 80523
Aaron.Reeves@colostate.edu

NAADSM Technical Paper #2
Version 2010/09/23
<http://www.naadsm.org/techpapers/>

Copyright © 2010 Animal Population Health Institute at Colorado State University

Permission is granted to copy, distribute, and/or modify this document under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License: please see <http://creativecommons.org/licenses/by-nc-sa/3.0/us/> for complete terms of this license.

Document revision history

- 2010/09/23 – Document updated with reference to Patyk *et al.* (submitted for publication)
- 2010/08/15 – Initial public version released

Introduction

NAADSM allows users to model the spread of disease by direct and indirect contact. [The details of the contact spread components of NAADSM are described by Hill and Reeves (2006), Harvey *et al.* (2007), and Harvey and Reeves (2010).] Unlike some models, which allow the user to specify several levels or forms of contact spread (e.g., Bates *et al.* 2003), NAADSM uses only a single set of parameters to define all forms of disease spread by direct contact, and another single set of parameters to define disease spread by all forms of indirect contact.

In some cases, having only single parameter sets for these components may not be sufficient to construct a NAADSM-based model with the desired level of detail. For example, Patyk *et al.* (submitted for publication) had information that described up to 20 different sources of indirect contact among poultry flocks. Each source was associated with its own frequency of contact, as well as its own distribution of distances over which contacts occur.

This paper describes a approach that allows the combination of several sources of data to produce a single set of contact parameters suitable for use in NAADSM.

NAADSM parameters for contact spread

As described elsewhere (Hill and Reeves 2006, Harvey *et al.* 2007, Harvey and Reeves 2010), there are three critical parameters in NAADSM that describe contact spread between each combination of production types included in a model. These parameters are as follow:

- A frequency of contact (average daily contact rate)
- A probability that a contact will result in the transmission of disease
- A probability density function that represents the distance over which contacts occur

For purposes of demonstration, we will suppose that we wish to simulate three different sources of indirect contact, with parameters as shown in Table 1.

Combining multiple sources of contact into a single overall parameter set

The following sections will show how the three individual specific parameter values for each parameter type listed above can be combined to give a single value.

Calculating the overall average daily contact rate

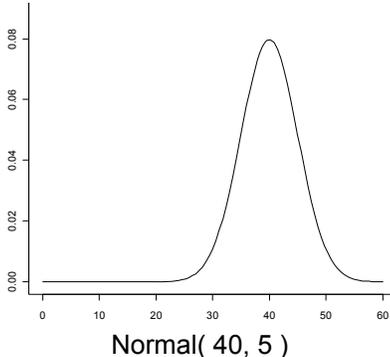
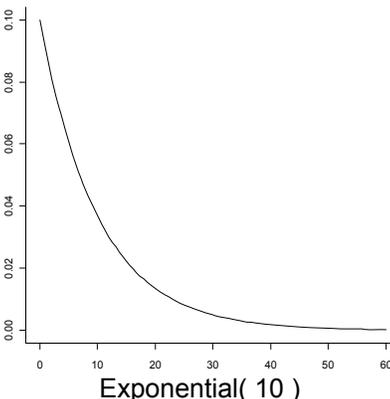
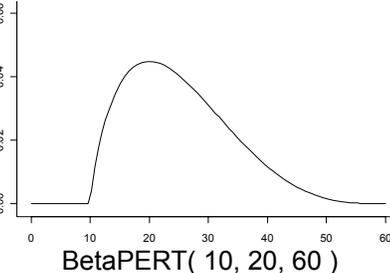
A simple sum of the average daily contact rates for the three individual parameters will give an overall daily frequency for all contacts:

$$\text{Total daily contact rate} = \sum_{n=1}^{\text{No. of contact types}} (\text{Daily contact rate})_n \quad (\text{eq. } 1)$$

For the example dataset:

$$1 + 0.14286 + 0.03333 = 1.17619 \text{ contacts per day} \quad (\text{eq. } 2)$$

Table 1. Sample parameters for three hypothetical sources of indirect contact spread.

Source no.	Average daily contact rate	Probability of disease transmission	Distance distribution (km) ¹
1	1 per day	0.05	 <p>Normal(40, 5)</p>
2	1 per week 0.14286 per day	0.10	 <p>Exponential(10)</p>
3	1 per month 0.03333 per day	0.3	 <p>BetaPERT(10, 20, 60)</p>

¹ Probability density function parameters and calculations follow Hill and Reeves (2006) and Vose (1996)

Calculating the overall probability that a contact will result in the transmission of disease

Although we now have an expected 1.17619 contacts per day, not every contact will have the same probability of resulting in disease transmission. To account for this variation, we can use a weighted average, based on the frequencies of each individual source of contact and the probabilities that each type of contact will result in disease transmission:

$$\text{Overall } P(\text{Spread} | \text{Contact}) = \frac{\sum_{n=1}^{\text{No. of contact types}} [(\text{Daily contact rate})_n \times P(\text{Spread} | \text{Contact})_n]}{\sum_{n=1}^{\text{No. of contact types}} (\text{Daily contact rate})_n} \quad (\text{eq. 3})$$

For the example dataset:

$$[(1 \times 0.05) + (0.14286 \times 0.1) + (0.03333 \times 0.3)] / 1.17619 = 0.06316 \quad (\text{eq. 4})$$

Generating an overall contact distance distribution

A probability density function that represents the overall range of distances over which contacts can occur can be generated by creating a mixture distribution that considers the proportion of the overall number of contacts that is contributed by each form of contact.

To facilitate this process, and to allow the use of both theoretical and empirical probability density functions, we have created a simulation tool that can be used to create such mixture distributions. The application, called *DistMixer*, can be downloaded from <http://www.naadsm.org/distmixer>.

Using *DistMixer*

The *DistMixer* application accepts multiple individual probability density functions (PDFs) as inputs (Figure 1). Individual functions can be added or removed as needed. Clicking on the **Edit** button will allow users to select one of the supported probability density functions and enter the parameters for the selected function. Refer chapters 3 and 6 and appendix D of Hill and Reeves (2006) for details concerning the PDF editor used in *DistMixer*.

Each distribution is also associated with its relative frequency, or, in the case of producing an overall contact distribution, the proportion of all contacts that it will contribute to the overall distance distribution. The average daily contact rate associated with each distance distribution can be used directly as its relative frequency.

When *DistMixer* is run, a total of five million samples are drawn from the different distributions. The number of values drawn from each distribution will be proportional to the relative frequency of that distribution. For example, if one particular form of contact occurs twice as frequently as another, then twice as many values will be drawn from the distance distribution for the former form of contact than for the latter.

After *DistMixer* is run, the **Output** tab will be enabled (Figure 2). This tab displays a histogram produced from the samples taken during program run. The number of histogram bins can be altered dynamically, based either on one of several commonly used algorithms (Sturges 1926, Scott 1979, Freedman and Diaconis 1981), or based on a user-specified number of histogram bins. Larger numbers of histogram bins will produce a smoother mixture distribution, at the cost of having higher computational times when the mixture distribution is used elsewhere.

The resulting mixture distribution (in the form of an empirical histogram distribution) can be exported or copied from *DistMixer* for use in other application. In the case of our example, the mixture distribution can be copied from *DistMixer* and pasted into the appropriate NAADSM input screen as the overall contact distance distribution.

Combining data from several sources to produce a single set of contact spread parameters
NAADSM Technical Paper #2

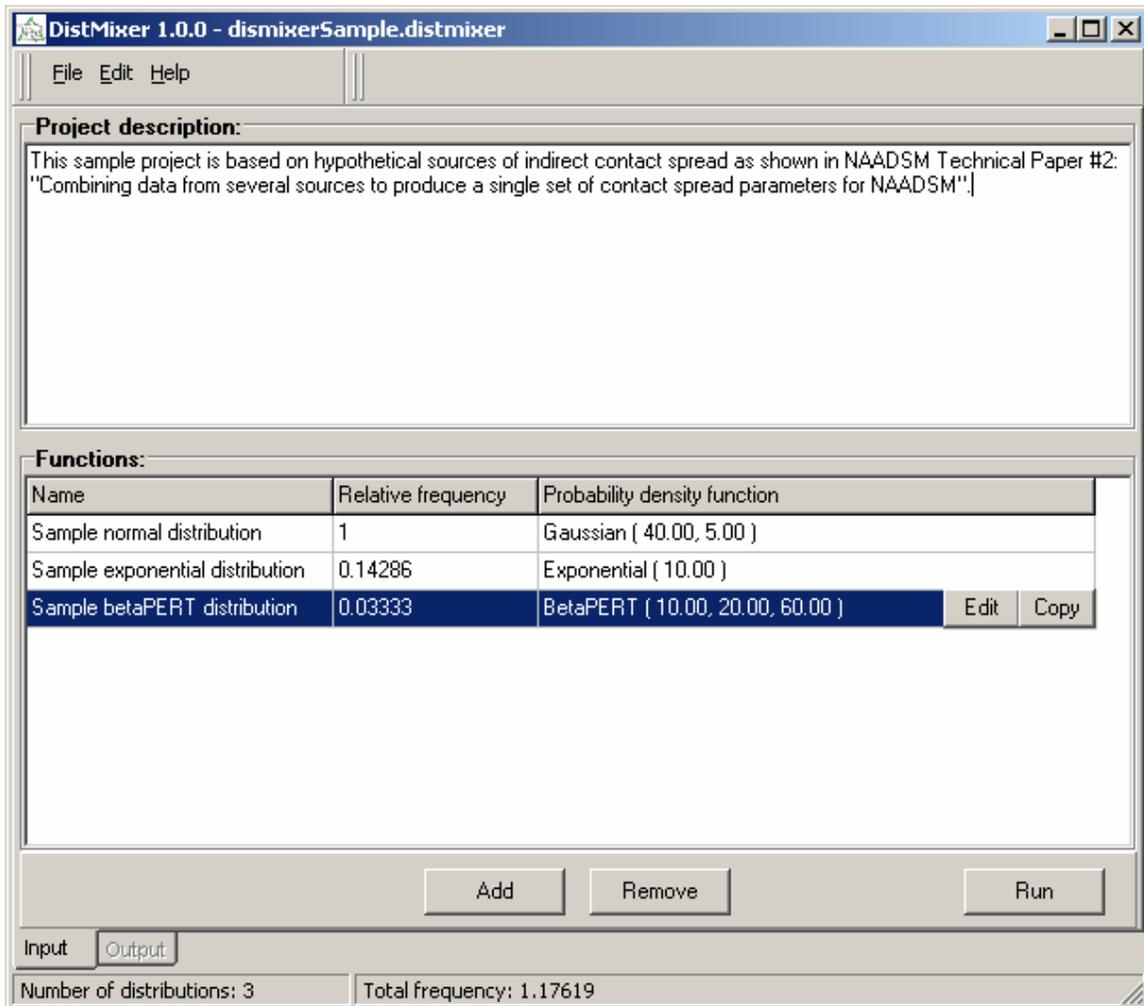


Figure 1. Input tab for *DistMixer*. Individual probability density functions are created, altered, or removed using the buttons in the lower panel. Each distribution is also characterized by its relative frequency, or the proportion that it will contribute to the overall mixture distribution when *DistMixer* is run.

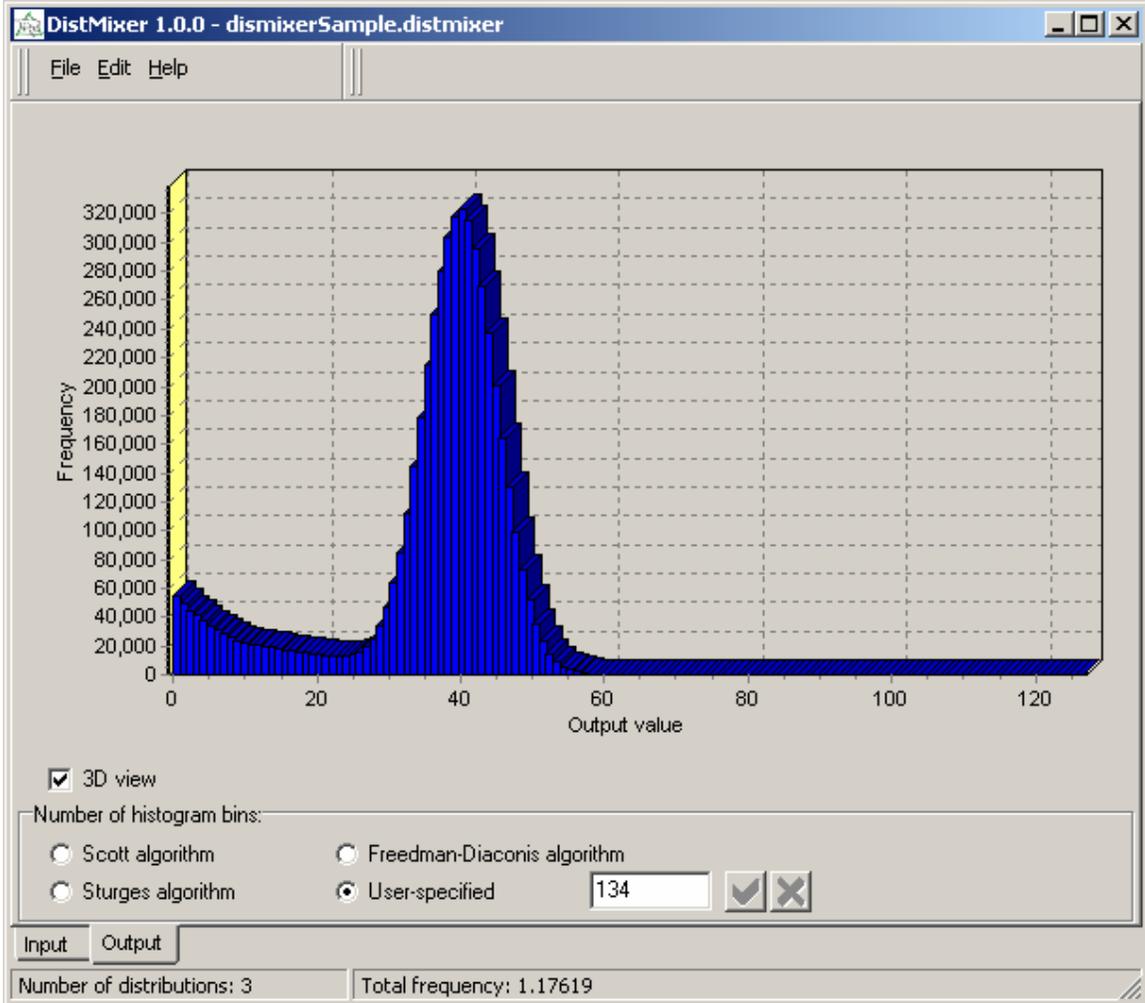


Figure 2. Output tab for *DistMixer*. An empirical histogram distribution is generated after the program is executed. The user may specify the number of histogram bins as described above. The resulting distribution can be exported or copied for use in other applications, such as *NAADSM*.

Limitations of this approach to modeling different sources of contact spread

DistMixer is a stochastic model: every time that it is run, even for the same set of input distributions, the output will differ slightly from previous runs. Because a large number of samples are generated, however, these differences should be minor. For most, if not all, of the probability density functions currently supported by *DistMixer*, it should be possible to arrive at an analytical (but still discretized) solution for the mixture distribution, rather than relying on simulation. A future version of *DistMixer* may support this alternative approach. Users are urged to contact the developer if such a feature would be desirable.

The approach described here ultimately uses only single values for the frequency of contact and the probability of effective contact in *NAADSM*. This will cause the overall variability in outcomes produced by a stochastic model (like *NAADSM*) to be somewhat reduced, compared to an alternative scheme in which each of the several forms of contact would be simulated separately. The practical impact of this reduced variability, although likely to be minor, is not known. It would

be helpful to compare the results generated by this approach to a more fully stochastic implementation. The method described here, however, has the advantages that it is computationally simpler (and likely considerably faster) than using multiple parameter sets for different types of contact spread, and that it can be used with existing software without substantial modification.

One additional potential limitation is related to the way in which NAADSM simulates the effect of restrictions on movements (either of animals, or of people, equipment, and other indirect sources of disease spread) in outbreak situations (Harvey *et al.* 2007, Harvey and Reeves 2010). When this approach is used in NAADSM, the modeler is making the assumption that the effect of movement restrictions applies equally to all forms of contact that have been combined into a single parameter set. This assumption may not be realistic, depending on the type and nature of the individual forms of contact.

Summary

Using the approach outlined here, it is possible to use the existing NAADSM model framework and still take advantage of information on several sources of contact spread. This approach is likely to be more beneficial for combining sources of indirect contact (which represent the movement of people, equipment, animal products, or other fomites among premises) than for direct contact (which is typically used in NAADSM to represent the introduction of animals from one herd or flock into a new herd or flock). There is likely to be much more variation among different forms of indirect contact.

The NAADSM Development Team relies on feedback from users for the refinement and improvement of NAADSM as framework for the creation of epidemiologic simulation models. We encourage users to contact us to discuss the potential need for a model that simulates many forms of contact simultaneously, or any other topic related to the use of NAADSM. Contact information for members of the Development Team can be found at our website, <http://www.naadsm.org>.

References

- Bates, T.W., Thurmond, M.C., and Carpenter, T.E., 2003. Description of an epidemic simulation model for use in evaluating strategies to control an outbreak of foot-and-mouth disease. *Am J Vet Res* 64: 195-204.
- Freedman, D. and Diaconis, P. 1981. On the histogram as a density estimator: L_2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 57: 453-476.
- Harvey, N., and Reeves, A. (eds.), 2010. Model description: *North American Animal Disease Spread Model* 3.1, version 1.1.2. Available online at <http://www.naadsm.org/documentation/specification>. Last accessed August 14, 2010.
- Harvey, N., Reeves, A., Schoenbaum, M.A., Zagmutt-Vergara, F.J., Dubé, C., Hill, A.E., Corso, B.A., McNab, W.B., Cartwright, C.I., and Salman, M.D., 2007. The *North American Animal Disease Spread Model*: A simulation model to assist decision making in evaluating animal disease incursions. *Prev Vet Med* 82: 176-197.
- Hill, A., and Reeves, A. 2006. User's Guide for the *North American Animal Disease Spread Model*, 2nd ed. Fort Collins, Colorado: Animal Population Health Institute, Colorado State University. Available online at <http://www.naadsm.org/documentation/guide>. Last accessed August 14, 2010.
- Patyk, K.A., Helm, J., Martin, M.K., Forde-Folle, K.N., Olea-Popelka, F.J., Hokanson, J., Fingerlin, T., and Reeves, A., submitted for publication. Establishing input parameters for an epidemiologic simulation model scenario of the spread and control of highly pathogenic avian influenza (H5N1) among commercial and backyard poultry flocks in South Carolina.
- Scott, D. W. 1979. On optimal and data-based histograms. *Biometrika* 66: 605-610.

Sturges, H. A. 1926. The choice of a class interval. *J. American Statistical Association*. 21: 65–66.

Vose, D., 1996. *Quantitative Risk Analysis: A Guide to Monte Carlo Simulation Modelling*. John Wiley and Sons, Inc., New York.