5.9 **Development of an Operational Northeast Snowfall Impact Scale**

Michael F. Squires * and Jay H. Lawrimore
NOAA National Climatic Data Center, Asheville, North Carolina

1. INTRODUCTION

While the Fujita and Saffir-Simpson Scales characterize tornadoes and hurricanes respectively, there is no widely used scale to classify snowstorms. The Northeast Snowfall Impact Scale (NESIS) developed by Paul Kocin of The Weather Channel and Louis Uccellini of the National Weather Service (Kocin and Uccellini, 2004) characterizes and ranks Northeast snowstorms. NESIS has five categories: Extreme, Crippling, Major, Significant, and Notable. The index differs from other meteorological indices in that it uses population information in addition to meteorological measurements. Thus NESIS gives an indication of a storm's societal impacts.

This paper describes a process being developed at the National Climatic Data Center (NCDC) to calculate NESIS operationally beginning with the 2005-06 winter season. Issues such as automation within a geographical information system (GIS), quality control, and estimating the uncertainty associated with a particular NESIS value are addressed.

2. NESIS ALGORITHM

The algorithm for computing the NESIS is:

$$NESIS = \sum_{n=4}^{n=30}\left[\frac{n}{10}\left(\frac{A_n}{A_{mean}} + \frac{P_n}{P_{mean}}\right)\right] \qquad (1)$$

where:

- n = snowfall category {4 for > 4", 10 for > 10", 20 for > 20", 30 > 30"}
- A_n = *area of snowfall greater than or equal to category n* (mi^2)
- P_n = *population affected by snowfall greater than category n (2000 census)*
- A_{mean} = *mean area of >10" snowfall within the 13-state Northeast region (91,000* mi^2*)*
- P_{mean} = *mean population affected by snowfall >10" within the 13-state Northeast region (35.4 million)*

The mean area and population constants are for 30 historical storms from 1956 to 2000 for the 13 northeastern states. These constants calibrate this index to northeast snowstorms.

* *Corresponding author address:* Michael F. Squires, NOAA/NCDC, 151 Patton Avenue, Asheville, NC 28801; e-mail: Mike.Squires@noaa.gov.

This algorithm typically results in values between 1 and 13 which are then transformed into one of five categories; Notable, Significant, Major, Crippling, and Extreme. Storms that cover a large area, with large snowfall amounts, and affect populated regions have the largest NESIS values. The highest severity case is the March 1993 super storm with a NESIS value of 12.52 placing it in the "Extreme" category. Kocin and Uccellini (K-U) used a combination of hand drawn maps and census data within a GIS to calculate their NESIS values.

2. NCDC CALCULATIONS

NESIS values produced at NCDC are calculated entirely within a GIS environment. Storm snowfall totals from point locations are brought into a GIS, quality controlled, converted to a snowfall grid, and then combined with population data in Equation 1 to estimate a NESIS score. This score is sensitive to the quality of the point snowfall data and the parameters used in the spatial interpolation scheme used to construct the snowfall grid.

2.1 *Quality Control*

NCDC computed historical NESIS scores to validate its process against the K-U NESIS scores and also to calculate new values of A_{mean} and P_{mean} so Equation 1 would be appropriate using the NCDC methods. In doing this, the calculation of a NESIS score for a particular storm using Equation 1 was found to be very sensitive to the snowfall grid created from the point snowfall values. Therefore it is imperative that the initial point values are accurate and any erroneous values are eliminated from the analysis. Although today's COOP data at NCDC is subject to rigorous quality control checks, all historical data has not been subjected to the same QC procedures which use multiple sensors (Angel et. al., 2003).

During development, two types of errors were noted in some of the storms;

1. questionable zero snowfall amounts, and
2. questionable non-zero snowfall amounts.

The first case refers to locations whose snowfall value is zero, but should most likely be set to missing since neighboring stations have rather high snowfall amounts. The second case refers to locations whose snowfall value is significantly different from its neighbors. These may be either errors or "interesting" cases. To help detect these errors a statistical tool known as the Local Moran's Index is used to identify locations whose values are suspect (Anselin 1995).

The Moran's I_i is a statistic that quantifies the extent to which similar and dissimilar geographic features are clustered. A location that has nearby features with similar values will have a positive *Moran's* I_i. A feature with a value dissimilar to its neighbors will have a negative *Moran's* I_i. Because this statistic can help identify discontinuities in a field that should be continuous, it can be used in QC applications. The equation for Moran's I_i is;

$$I_i = \frac{\left(x_i - \bar{x}\right)}{s^2} * \sum_j w_{ij}\left(x_j - \bar{x}\right) \qquad (2)$$

Where x_i = value at target location
x_j = value at neighbor locations
s^2 = global variance for entire dataset
$\bar{x}$ = global mean for the entire dataset
w_{ij} = weights to be used with each target-neighbor pair.

Neighborhoods are defined by distance and some type of spatial relationship within the neighborhoods. Typical spatial relationships include inverse distance, inverse distance squared, and binary contiguity (fixed distance). An inverse distance relationship is used in this analysis. The weights (w_{ij}) are computed based on the spatial relationship that is chosen.

A negative value for I_i indicates that a feature is surrounded by features with dissimilar values. If the target feature value is above the mean, the ratio in Eq. (2) will be positive. If the neighboring feature values are below the mean, the sum will be negative. When the two are multiplied, the result (I_i) will be a negative number. The more the target value is above the mean and the more the neighboring values are below, the smaller (more negative) it will be. The inverse is also true.

Figure 1 shows how the Local Moran's I_i statistic is used to identify potential errors. The data is for the January 6-8, 1996 storm, which has the second to the largest NESIS score (11.54) of all the storms that have been analyzed. This particular storm had a number of questionable values in the raw data. The statistic is sensitive to the manner in which neighborhoods are defined and the mean value of the distribution being analyzed. To identify the two problems mentioned above, the statistic is applied twice with different strategies. In Figure 1(a), locations with questionable zero snowfall amounts are symbolized with a . These locations have zero snowfall amounts and a negative Moran's I_i. These cases identify features that have reported zero snowfall but are in the same neighborhood as features that have had considerable snowfall. In almost all of these situations, the observations should be eliminated from the analysis. It should be pointed out that there are some features that have zero snowfall amounts and are around other stations that have considerable snowfall, but do not have a negative Moran's I_i statistic. These situations occur around the gradient of the snowfall pattern where both the target and neighborhood locations have snowfall values that are less than the global mean, thus giving a positive result to Equation (2)

Locations with questionable non-zero snowfall amounts are symbolized with a . They were identified by only including stations with non-zero snowfall amounts and a Moran's I_i Z-score of -1.6 or less. This corresponds to stations that are significant at the 95% level for a one sided test. We are only interested in negative values of Z(I_i) since these are associated with dissimilar snowfall amounts. At this scale, one would expect snowfall amounts to be somewhat continuous. Therefore, values of Z(I_i) that are significant should be looked at critically. As mentioned before, these cases may be errors or simply locally heavy snowfall amounts caused by small scale process such as oragraphic lifting or convection. In 1(a), the in western North Carolina is associated with a large snowfall amount on Mt. Mitchell which is located at 6,240 ft. Nearby locations are situated between 1,500 and 3,000 ft, so have smaller snowfall amounts. Therefore in this case, a value that is statistically significant identifies a location that had high snowfall due to some local effects and is valid. Other locations were found to erroneous, and needed to be eliminated from the analysis.

In 1(b), a snowfall grid is generated using an inverse distance weighted spatial interpolation scheme without removing the suspect and erroneous snowfall values. It is apparent that this grid has problems and is not a realistic map of the January 1996 storm. There are numerous "holes" in the region of greater than 10" snowfall. The holes are caused by zero snowfall amounts mixed in with nearby values of 10" and greater. Most of these holes are centered on the markers, indicating that the Moran's I_i statistic does a good job at predicting this type of error.

Although the statistical quality control application identifies numerous possible errors, an analyst always has the final say on which values are eliminated. This is important because not all of the questionable zero snowfall amounts are identified by the Moran's I_i statistic. Also, some of the questionable non-zero snowfall amounts are valid values. Therefore, an "expert" is needed to evaluate the results of the statistical quality control algorithms and decide which features to eliminate from the analysis.

The erroneous and suspect snowfall values have been removed and a new snowfall grid is generated in Figure 2(c). The new map is much more realistic than 2(b). The artificial holes have been eliminated and the contours have been smoothed considerably. At the same time, large snowfall values that are valid (such as Mt. Mitchell) have been kept.

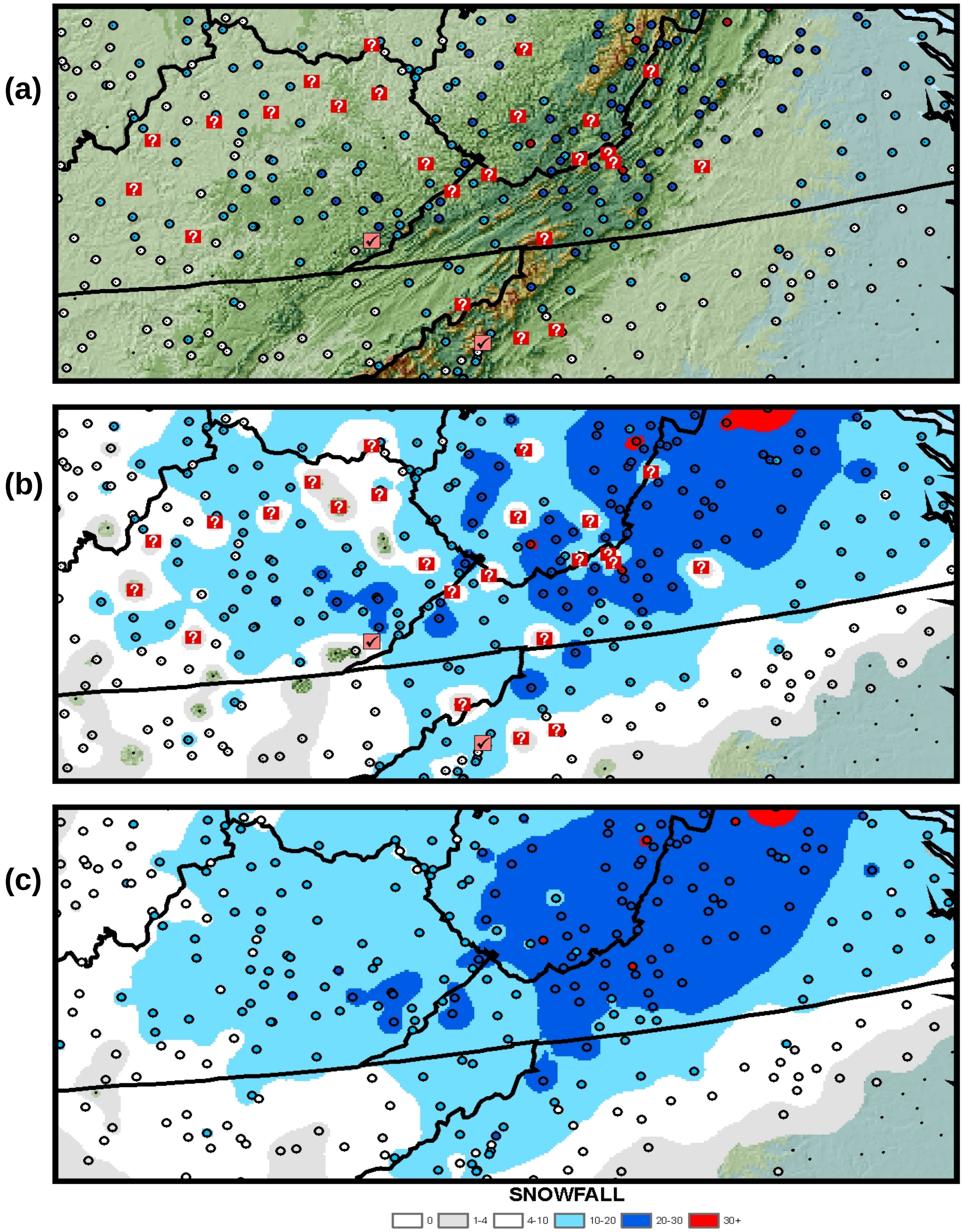

Figure 1. Maps showing how the Moran's I statistic is used to identify problem observations. (a) Locations with questionable observations are marked with a ? or a ✓. (b) These questionable observations diminish the quality of a grid produced from them. (c) Once the questionable values are removed, a realistic grid is produced.

2.2 *Calculating NESIS Within a GIS Environment*

GIS applications are becoming more numerous in weather and climate applications (Yuan 2005, Shipley 2005, Wilhelmi and Betancourt 2005, and Habermann 2005). NCDC uses a GIS with customized scripts to compute NESIS scores. After the storm total snowfall data has been quality controlled, the NESIS value can be calculated. Figure 2 shows the process by which this is done within a GIS. The point snowfall is converted to a 5 km by 5 km grid. A 5 km population density grid that shows the number of people per grid cell is represented by the earth tone shades in Figure 2. The population data is based on the 2000 census. Since the population and snowfall grids are aligned, the GIS can calculate the number of people living in each snowfall category. The area is calculated as from total number of grid cells in each snowfall category multiplied times 25 km^2. This is shown conceptually in Figure 2. The area and population values represent the totals for each category (4, 10, 20, 30). So the area of snowfall between 10" and 20" is 107,786 mi^2 which is represented on the map by the turquoise band. However, Equation (1) defines the area and population as "greater than" 4", greater than 10", etc. For example, the area for category 10 would include the areas for categories 10, 20, and 30. So the population and area values must summed with the values from the categories above. This information is then fed into Equation (1) and a NESIS score is estimated. In this case, a NESIS value of 11.78 is produced which translates to Category 5 (Extreme).

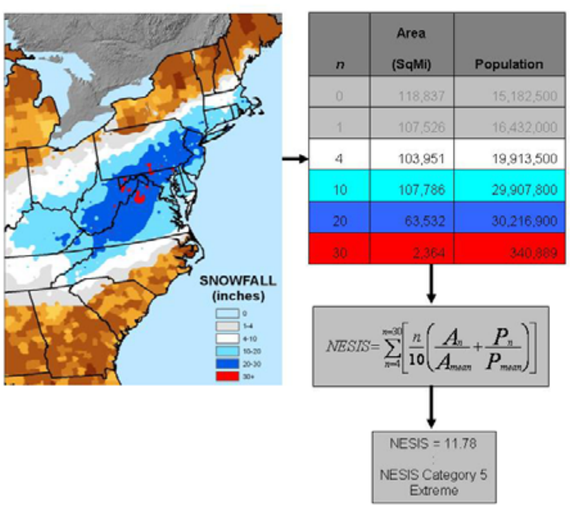

Figure 2. Diagram of the process by which NESIS values are estimated within a GIS. The earth tone background map is a population density grid. Both the population and snowfall grids are 5x5 km and are aligned. The table represents the values needed for the NESIS algorithm calculated from the GIS map layers.

2.3 *Estimating Uncertainty and Confidence Intervals*

The NESIS score estimated using the GIS method in the previous section for the January 1996 snowstorm was 11.78. The value estimated by K-U was 11.54. The difference is small and is a result of the different ways in which the snowfall map was drawn since differences in area and population for the various snowfall categories will result in different NESIS values. The differences between NESIS scores for the 30 historical storms as calculated by K-U and NCDC are small. There was a mean absolute difference of 0.51, a bias of 0.15, and a Spearmen correlation 0f 0.95 (Squires 2005). The Spearmen correlation was used because it compares the ranks of the data. See Figure 3. When computing NESIS values using two different methods, it is highly desirable that there final ranks be the same or similar.

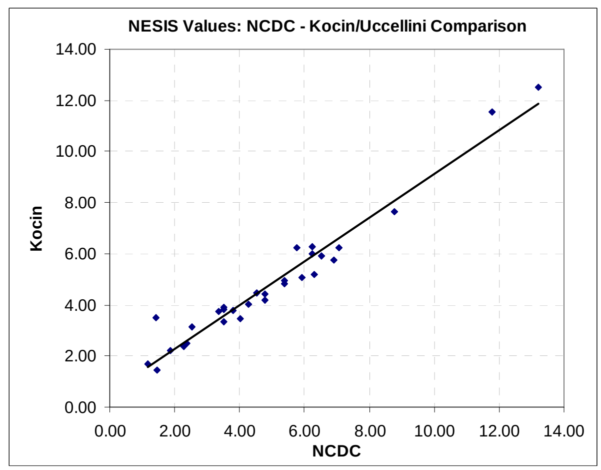

Figure 3. Comparison of the original NESIS values for the 30 historical storms as computed by K-U and NCDC.

Although the differences in NESIS scores between the two methods are small, there are still differences. Even using the NCDC method can yield different values if slightly different spatial interpolation schemes are used. Also, as Figure 3 shows, many of the storms have NESIS values that are quite close to each other. Differences less than 0.1 are not uncommon. Since one of the primary purposes of the NESIS is to rank storms, these small differences are critical. Therefore, it is desirable to have some indication of uncertainty or confidence for the final NESIS score.

The vast majority of variability between NESIS scores is due to different ways in which gridded snowfall maps are constructed from the point snowfall observations. Even if maps are drawn by hand, there is no doubt that 10 different meteorologists would draw 10 different maps even though they were using the same data. If the maps were drawn by "experts", the maps would be similar but would yield different NESIS values because the areas and population counts for each of the snowfall categories would be slightly different. The different maps would be reflections of the differing styles and biases of the experts drawing the maps. In reality, the correct map is not known. That is because the snowfall observations at discrete points do not provide all the information necessary to construct a continuous snowfall grid that is 100% accurate. This is especially true in complex terrain where topography or land water-boundaries can modify patterns of snowfall. Therefore, there will always be some uncertainty associated with the snowfall grids and the subsequent NESIS values.

One way to estimate the uncertainty associated with a particular NESIS value would be to construct multiple snowfall grids by varying the spatial interpolation parameters. A NESIS score would be estimated from each of the grids by using Equation (1). If enough NESIS scores are estimated for a particular storm, the collection of these values could be viewed as a sampling distribution from which confidence intervals could be estimated.

The snowfall grids for this study are generated using an inverse distance weighted (IDW) spatial interpolation scheme;

$$\hat{s} = \frac{\sum_{i=1}^{n} \frac{1}{d_i^p} s_i}{\sum_{i=1}^{n} \frac{1}{d_i^p}} \qquad (3)$$

Where:

- s = snowfall
- d = distance
- p = **power parameter** to which d is raised
- n = number of stations in a particular search neighborhood as defined by a **radius parameter (*r*)**

The snowfall grids generated from this interpolation technique can vary by varying the power and radius parameters. Typical values of p range from 1 to 4. A value of zero for p would make all the weights equal to one and result in averaging all the values within a search neighborhood to estimate the snowfall at a grid point. This would result in a very smooth map. Large values of p (> 4) result in maps that appear less smooth and often contain "bull's eyes" around individual station values. Larger values of p (> 9) result in maps that appear "blocky" and come close to emulating Thiessen polygons. Appropriate values for the neighborhood radius (r) are less straightforward to determine and are based on the average station density across a study area. Small values of r yield maps that are rather noisy and also contain "bull's eyes". Large values of r result in much smoother maps, but important details could be lost. Therefore the choice of the spatial interpolation parameters p and r must be decided on a case by case basis.

An interesting analogy can be drawn between different spatial parameters and different experts. The spatial parameters can be thought of as the different styles and biases possessed by a collection of experts. Just as some combinations of parameters work well in some situations, others do well in different situations. Some experts may analyze maps well in some situations, but have difficulty with others. Some experts routinely draw maps that are smooth; others draw maps that contain more detail. Whether the detail is real or not is often subject to debate. A consensus among analysts typically produces the most useful maps.

In order to produce a "consensus" NESIS value that arises from different maps produced with an IDW interpolation scheme, the parameters p and r can be varied. The resulting NESIS values represent a sampling distribution from which confidence values can be estimated.

One must take care in choosing the spatial parameters. It is possible to choose parameters that yield maps which are not realistic. Therefore the spatial parameters are chosen to;

1. produce maps that are realistic, and
2. produce a sampling distribution of NESIS values that approximate a Gaussian distribution.

The first criterion is obvious. The second criterion is needed for the construction of valid confidence intervals.

A series of 48 snowfall grids were generated for the January 1996 storm. Different map simulations were created by allowing the power parameter to vary from 1.75 to 3.00 and the radius parameter to vary from 60 to 100 km. A NESIS score was calculated for each map. The collection of these scores represents a sampling distribution. The maps associated with the minimum, mean, and maximum NESIS scores are shown in Figure 4.

Figure 5 is a histogram of NESIS sampling distribution for 48 simulations. The distribution is approximately Gaussian and should provide an adequate basis for constructing confidence intervals.

Table 1 shows the relationship between the power and radius parameters for the 48 NESIS simulations of the January 1996 storm. The yellow cells contain the mean NESIS score (11.78) from the sampling distribution. The green cells represent scores between the 25th percentile and the mean. The red cells represent scores between the mean and the 75th percentile. The grey cells represent the marginal mean NESIS scores for different values of power and radius. As the power parameter increases the NESIS scores also increase. The marginal means for power (bottom row) show a clear trend from 11.75 to 11.80 as the power increases from 1.75 to 3.00. The red cells on the right-hand side of the table also highlight this trend. The relationship between the radius parameter and NESIS scores is not as strong. There is no discernable trend in the marginal means for radius (right-most column). Overall, parameters that produce smoother grids (small power and large search radius) produce smaller NESIS scores. Parameters that produce less smooth grids (large power and small radii) produce

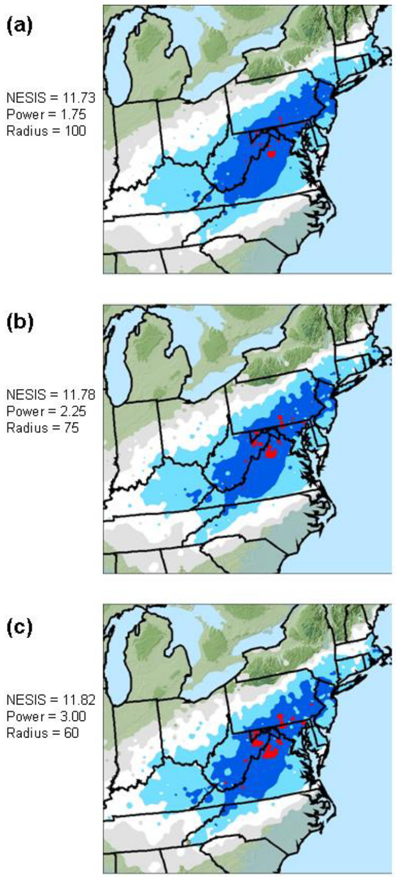

FIG 4. Snowfall maps associated with the (a) minimum, (b) mean, and (c) maximum NESIS scores from the Jan 1996 sampling distribution.

higher NESIS values. This choice of parameters appears to produce a reasonable sampling distribution from which confidence intervals may be inferred. However, a wider sampling distribution would probably be more realistic. The K-U value for this storm is 11.54, which is less than the minimum value of this distribution.

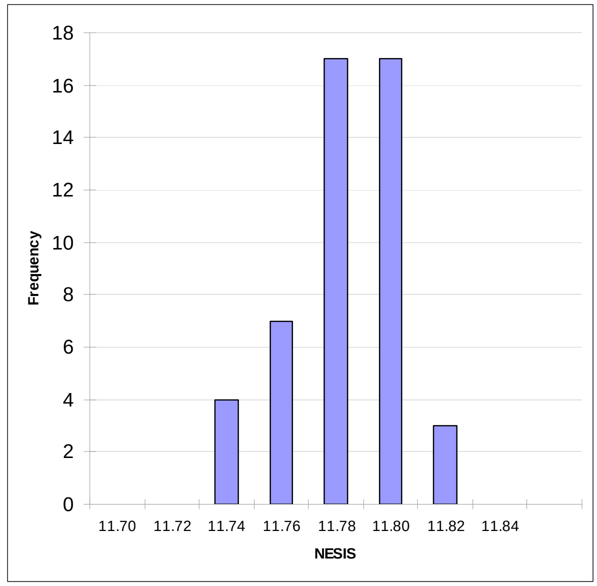

Figure 5. Histogram of NESIS sampling distribution.

Table 2 contains descriptive statistics for the NESIS sampling distribution. With a mean of 11.78 and a standard deviation of 0.02, the NESIS score for the January 1996 storm is taken to be 11.78 +/ 0.04.

NESIS	
Mean	11.78
Median	11.78
Standard Deviation	0.02
Skewness	-0.36
Range	0.09
Minimum	11.73
Maximum	11.82
Count	48.00

Table 2. Descriptive statistics from the NESIS sampling distribution.

	POWER						
RADIUS	**1.75**	**2.00**	**2.25**	**2.50**	**2.75**	**3.00**	**Mean**
60	11.78	11.77	11.78	11.80	11.81	11.82	*11.79*
70	11.76	11.77	11.77	11.78	11.80	11.81	*11.78*
75	11.76	11.77	11.77	11.78	11.79	11.80	*11.78*
77	11.76	11.78	11.78	11.79	11.80	11.80	*11.79*
80	11.75	11.77	11.78	11.79	11.80	11.80	*11.78*
85	11.74	11.76	11.77	11.79	11.80	11.80	*11.78*
90	11.74	11.76	11.77	11.77	11.80	11.80	*11.77*
100	11.73	11.74	11.76	11.77	11.79	11.80	*11.77*
Mean	*11.75*	*11.77*	*11.77*	*11.78*	*11.80*	*11.80*	*11.78*

Table 1. Table showing the relationship between the spatial parameters (power and radius) and NESIS scores for 48 simulations of the January 1996 storm.

2.4 *Operational GIS Application to Produce NESIS*

NESIS values are calculated within ArcMap by using a customized set of tools. There are three general tasks that are done to estimate a NESIS value; get the snow data into GIS, quality control the snow data, and calculate the actual value along with estimates of the uncertainty. In the first step, a text file containing total storm snowfall values is converted to a shapefile and put in an Albers Equal Area projection. The second step uses the Local Moran's I statistic to identify problem data as discussed in section 2.1. The final step calculates the actual NESIS value along with 95% confidence intervals as discussed in sections 2.2 and 2.3. These three steps have been automated with Python scripts and are implemented within the ArcMap interface. See Figure 5.

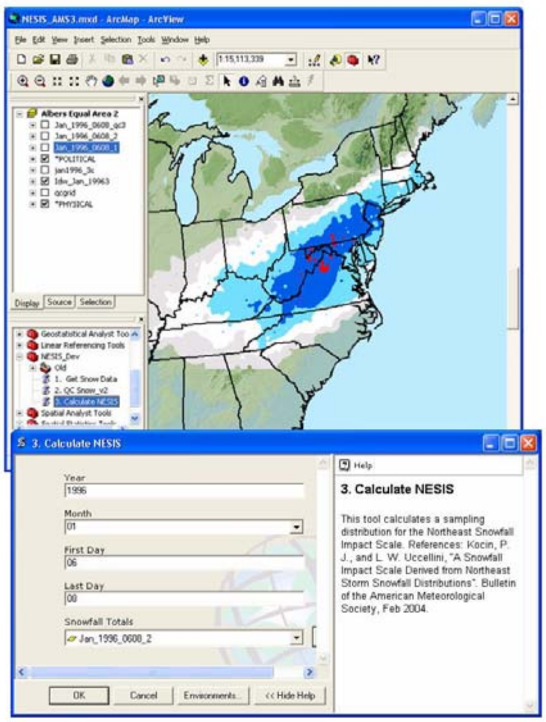

Figure 5. ArcMap interface with customized tools to estimate NESIS scores operationally.

3. SUMMARY

This paper has described a preliminary process that NCDC will use to compute NESIS scores operationally beginning with the 2005-06 winter season. The new process has incorporated a statistical quality control procedure, a method to estimate the uncertainty associated with NESIS values, and specialized tools embedded within the ArcMap interface to carry out these tasks.

Although the technique to produce a sampling distribution for each storm appears sound, more work needs to be done to identify appropriate parameters for the spatial interpolation scheme. The current parameters produce reasonable maps and a sampling distribution that is approximately Gaussian, however there is a need to ensure that the distribution has enough variance to include all likely simulations.

4. REFERENCES

Angel, W.E., M.L. Urzen., S.A. Del Greco, and M.W. Bodosky, 2003: Automated validation for summary of the day temperature data (TempVal). *83rd AMS Annual Meeting, combined preprints CD-ROM, 9-13 February 2003, Long Beach CA, 19th Conference IIPS [International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology]*, American Meteorological Society, Boston, Mass., File 15.3, 4 pp. (February 2003).

Anselin, L., 1995: Local Indicators of Spatial Association - LISA. *Geographical Analysis*, **27**, no. 2, 93-115.

Habermann, T., 2005: What Is GIS (for UNIDATA)?. *Bull. Amer. Meteor. Soc*., **86**, 174-175.

Kocin, P. J. and L. W. Uccellini, 2004: A Snowfall Impact Scale Derived From Northeast Storm Snowfall Distributions. *Bull. Amer. Meteor. Soc*., **85**, 177-194

Squires, M. F., 2005: The Northeast Snowfall Impact Scale. *25th Annual ESRI International User's Conference*, July 25-29, San Diego, CA.

Shipley, S. T., 2005: GIS Applications in Meteorology, or Adventures in a Parallel Universe. *Bull. Amer. Meteor. Soc*., **86**, 171-173.

Wilhelmi, O. V., and T. L. Betancourt, 2005: Evolution of NCAR's GIS Initiative. *Bull. Amer. Meteor. Soc*., **86**, 176-178.

Yuan, M., 2005: Beyond Mapping in GIS Applications to Environmental Analysis. *Bull. Amer. Meteor. Soc*., **86**, 169-170.