

Improving Data Quality for a Dairy Pollutant Emissions Study

Ian K. Hahus, Purdue University and Albert J. Heber, Purdue University

The National Air Emissions Monitoring Study (NAEMS) was sanctioned by the EPA to determine the characteristics of airborne pollutant emissions from confined broiler, egg, pork, and dairy housing. Fifteen representative monitoring sites were selected around the U.S., at which influent and effluent pollutant concentrations were measured in conjunction with airflow and climatic data. Due to the monumental nature of this study and the potential ramifications of its findings, it is of vital importance that the data collected by the researchers and utilized by the EPA be as complete and accurate as possible. To improve the validity of the data collected at a dairy facility in New York, it was necessary to review the work of previous data analysts while studying the field notes that were logged by scientists onsite during data collection. This allowed for the correction of perceived errors in the handling of the data. When sensor data were deemed invalid or missing, redundant data were substituted. Any unnecessarily flagged out data were restored. The use of these strategies led to a significant improvement in data quality. For example, data completeness for ambient temperature and relative humidity were increased by over 6%, while atmospheric pressure data saw an improvement of more than 18% after substituting data from the nearest NWS weather station. These and other improvements to this data set will allow EPA to develop more accurate dairy facility emissions models that will have substantial, wide-ranging effects for both producers and consumers in the U.S. dairy industry.