Information Sciences

Algorithm for Identifying Erroneous Rain-Gauge Readings

What was previously a subjective manual analysis is automated and made objective.

Marshall Space Flight Center, Alabama

An algorithm analyzes rain-gauge data to identify statistical outliers that could be deemed to be erroneous readings. Heretofore, analyses of this type have been performed in burdensome manual procedures that have involved subjective judgements. Sometimes, the analyses have included computational assistance for detecting values falling outside of arbitrary limits. The analyses have been performed without statistically valid knowledge of the spatial and temporal variations of precipitation within rain events. In contrast, the present algorithm makes it possible to automate such an analysis, makes the analysis objective, takes account of the spatial distribution of rain gauges in conjunction with the statistical nature of spatial variations in rainfall readings, and minimizes the use of arbitrary criteria.

The algorithm implements an iterative process that involves nonparametric statistics. The steps of the algorithm are the following:

1. Raw rain-gauge data are subjected to qualitative tests of validity. The details of the tests are attuned to the details of the sources of data and data-entry procedures. For example, reports that include negative rain-gauge readings or incorrect dates are rejected. Data that pass these tests are accepted for processing in the next step.
2. Associated with each gauge is a neighborhood, defined as that gauge plus the five nearest gauges that (a) have reported, (b) are currently accepted, and (c) are more than 100 meters distant.

The 100-meter distance criterion is arbitrary, but not totally so: It has been chosen to ensure that each accepted gauge gives a reading independent of that of any other accepted gauge. Independence of readings is basic assumption of the statistical analysis performed in the subsequent steps.

The five-nearest-gauge criterion is also only partly arbitrary: It has been chosen as a compromise between (a) undesired sensitivity to numerical artifacts at fewer gauges per neighborhood and (b) undesired insensitivity to input errors (which the errors that one seeks to detect) at greater numbers of gauges per neighborhood.

3. The six readings from each neighborhood are ranked. If the reading of the gauge under consideration is a local minimum or maximum, then it is deemed erroneous if it is less than one-third or greater than three times the reading of the gauge of the adjacent rank.
4. After rejection of the gauges that have been thus deemed to give erroneous readings, a new set of neighborhoods is computed from the remaining accepted gauges, again following the logic of step 2.
5. The readings from gauges in the new neighborhoods are examined for errors, again following the logic of step 3.
6. The neighborhood of any gauge in step 3 or step 5 is examined to determine which, if any, other gauges in the neighborhood also were flagged as giving erroneous readings. If all of the gauges in the neighborhood have been flagged and if, in addition, their errors have all been found to be of the same sense (that is, all high or all low), then the readings from the neighborhood are assumed to be correct. The justification for this decision is that it is unlikely that two or more independent, spatially adjacent observations would both be extreme highs or extreme lows. In addition, when a gauge is flagged because of a low reading and the readings of at least three other gaugesare zero but are not local minima, then that gauge is not flagged.

The algorithm has been implemented as a series of subroutines in a computer program used to edit sets of rainfall data. The algorithm could also be implemented as a program in its own right or incorporated into other programs for the purpose of identifying erroneous input data pertaining to phenomena other than rainfall.

This work was done by Doug Rickman of **Marshall Space Flight Center.** *Further information is contained in a TSP (see page 1).*
MFS-31993-1

Condition Assessment and End-of-Life Prediction System for Electric Machines and Their Loads

System generates on-line, real-time condition assessment and end-of-life prediction.

Lyndon B. Johnson Space Center, Houston, Texas

An end-of-life prediction system developed for electric machines and their loads could be used in integrated vehicle health monitoring at NASA and in other government agencies. This system will provide on-line, real-time condition assessment and end-of-life prediction of electric machines (e.g., motors, generators) and/or their loads of mechanically coupled machinery (e.g., pumps, fans, compressors, turbines, conveyor belts, magnetic levitation trains, and others). In long-duration space flight, the ability to predict the lifetime of machinery could spell the difference between mission success or failure. Therefore, the system described here may be of inestimable value to the U.S. space program.

No known system (hardware, software, or hybrids) currently exists that