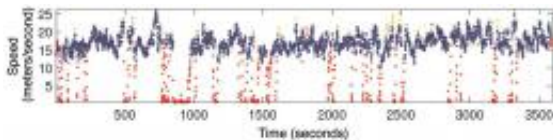# Innovative technique can spot errors in key technological systems

January 27 2010

IODA separated good and bad data from an anemometer: blue=high-quality; red=low-quality. Credit: NCAR

An innovative computational technique that draws on statistics, imaging and other disciplines has the capability to detect errors in sensitive technological systems ranging from satellites to weather instruments.

The patented technique, known as the Intelligent Outlier Detection Algorithm, or IODA, is described this month in the *Journal of Atmospheric and Oceanic Technology*.

IODA offers the potential to alert operators to faulty readings or other problems associated with failing sensors. If sensors malfunction and begin transmitting bad data, computers programmed with the algorithm could identify the problem and isolate that bad data.

IODA was developed by researchers at the National Center for Atmospheric Research (NCAR) and the University of Colorado at Boulder (CU).

The National Science Foundation (NSF), NCAR's sponsor, funded the research. "This technology will have broad applicability in many new areas," says Steve Nelson, NSF program director for NCAR.

The developers of the algorithm say its principles can eventually be used in a vast range of technological settings, including cars and other transportation systems, power plants, satellites and space exploration, and data from radars and other observing instruments.

"This could, at least in theory, enable operators to keep a system performing even while it's failing," says Andrew Weekley, a software engineer at NCAR who led the algorithm development effort. "When a system starts to fail, it's absolutely critical to be able to control it as long as possible. That can make the difference between disaster or not."

IODA is designed to perform quality control on time series data--that is, data collected over time, such as wind speeds over the course of a month.

The algorithm, an expert system that draws on statistics, graph theory, image processing and decision trees, can be applied in cases where the correct assessment of data is critical, the incoming data are too numerous for a human to easily review, or the consequences of a sensor failure would be significant.

At present the algorithm consists of several thousand lines of a technical computing language known as MATLAB. The researchers may expand and translate it into a computer programming language such as C so it can be used for commercial purposes.

Ensuring the quality of incoming time series data is a priority for virtually any organization involved in complex operations. If sensors begin relaying inaccurate information, it can be highly challenging for

personnel or automated systems to separate good data from bad, especially in cases involving enormous amounts of information.

Typically, to identify bad data, complex operations may rely on multiple sensors, as well as algorithms that characterize specific relationships among the data being collected, and identify failures when the data unexpectedly change.

A drawback in most of these algorithms, however, is they are designed for a particular type of time series and can fail catastrophically when applied to different types of data, especially in situations where there are numerous and sometimes subtle errors.

IODA, however, compares incoming data to common patterns of failure--an approach that can be applied broadly because it is independent of a specific sensor or measurement.

Weekley and co-authors took a new approach to the problem when they began developing IODA 10 years ago. Whereas existing methods treat the data as a function of time, Weekley conceived of an algorithm that treats the data as an image.

This approach mimics the way a person might look at a plot of data points to spot an inconsistency.

For example, if a person looked at a line drawn between points on a graph that represented morning temperatures rising from 50 to 70 degrees, and then spotted a place where that smooth line was broken, dipping precipitously because of numerous data points down at 10 degrees, the person would immediately suspect there was a bad sensor reading.

In cases where there are thousands or even millions of data points about

temperature or other variables, pinpointing the bad ones can be more difficult.

But Weekley thought that a computer could be programmed to recognize common patterns of failure through image processing techniques.

Then, like a person eyeing data, the computer could identify problems with data points such as jumps and intermittency; view patterns in the data; and determine not only whether a particular datum is bad but also characterize how it is inaccurate.

"Our thought was to organize a sequence of data as an image and apply image processing techniques to identify a failure unambiguously," Weekley says. "We thought that, by using image processing, we could teach the system to detect inconsistencies, somewhat like a person would."

The research team came up with ways of arranging data points in a time series into clusters, both in a domain that represents the data points over time and in another domain known as delay space.

Delay space, which offers another way to detect differences in the data, is a technique that pairs a data point in the time series with the previous value.

Using the clusters from both the time domain and delay space, bad data are separated into their own cluster, clearly distinct from the cluster of accurate data. At the same time, IODA can calculate quality scores indicating if each individual data point is good or bad.

"I would say the approach we report in the paper is a radical departure from the usual techniques found in the time series literature," says Kent Goodrich, a CU mathematician and a co-author of the paper.

"The image processing and other techniques are not new, but the use of these images and techniques together in a time series application is new. IODA is able to characterize good and bad points very well in some commonly encountered situations."

When the research team tested IODA, they found it accurately isolated incorrect data in several cases.

For example, they applied the algorithm to wind readings from anemometers in Alaska that contained faulty errors due to a loose nut, which left the anemometers unable to consistently measure gusts in high-wind situations. The algorithm identified the bad readings, separating them into a series of clusters away from the good data.

"This technique has very broad implications," Weekley says. "Virtually all control systems rely on time series data at some level, and the ability to identify suspect data along with the possible failure is very useful in creating systems that are more robust.

"We think it is a powerful methodology that could be applied to almost all sequences of measurements that vary over time."

Provided by National Science Foundation