Statistical technique for automatically cleaning erroneous data from weather-balloon observations

February 28 2017

Weather balloons carrying disposable radiosondes are released twice a day at 700 locations around the world to make observations of the upper atmosphere. Credit: Alamy

Twice a day, weather balloons are released into the atmosphere from 700

locations around the world to observe conditions in the upper atmosphere. Since the 1920s, there have been tens of millions of these radiosonde launches, producing an enormous archive of data that is critical to weather forecasting and climate modeling. In such a large data set, inevitable errors can significantly affect modeling outcomes.

Ying Sun, Saudi Arabia's King Abdullah University of Science and Technology (KAUST) Assistant Professor of Applied Mathematics and Computational Science, collaborated with researchers from the Colorado School of Mines and Baylor University, US, to develop a method to remove these errors using on a robust statistical analysis of the data.

"A radiosonde is a small, expendable instrument package that is suspended below a two-meter-wide balloon filled with hydrogen or helium," explained Sun. "Sensors on the radiosonde measure height, pressure, temperature and dew point; they also calculate wind speed and direction by tracking the position of the radiosonde in flight. Radiosonde observations are the only direct measurements of the Earth's upper atmosphere, making them vital for satellite data, weather forecasting and climatology research.

The data's many errors are "far too many to correct by hand, so we need an automatic method for identifying such random errors," explained Sun.

There are automatic methods for removing systematic errors from the data, such as changes in location or measurement units. However, there has been no way to remove genuinely erroneous data, including data-entry mistakes, transmission errors or imprecise tracking of the balloon without also deleting extreme but real measurements—which are some of the most important data for forecasting. Looking specifically at wind data, Sun and her co-workers developed a statistical approach that achieves robust differentiation between extreme values and random

errors.

"Our approach considers a more realistic distribution of the wind vector that is skewed with a long tail of rare extreme values," said Sun. "This makes it possible to flag observations that are very likely to be errors as potential outliers without removing extreme values."

In addition to its application to new daily data, this error-detection scheme can also be used on the huge volumes of radiosonde observations held in archives around the world.

"We are developing an outlier-detection method that is fast and automatic. We will be able to use this method to quickly process the millions of records in the archive," said Sun. "We are also considering the possible effect of climatic change when developing the new method."

Provided by King Abdullah University of Science and Technology