A statistical approach to find the origin of outliers in a multibeam data set

Furey Tony[1], Church Ian[1]

[1] Department of Geodesy and Geomatics Engineering, University of New Brunswick, Canada
afurey1@unb.ca

A typical multibeam echosounder (MBES) survey platform is equipped with an inertial measurement unit (IMU), global navigation satellite system (GNSS) receivers, sound speed profiler, surface sound speed sensor, MBES transducer, and a nearby tide gauge or ocean model data. These sensors combine to produce a high dimensional data set with native features of time, heave, pitch, roll, heading, beam number, beam angle, latitude, longitude, elevation, across-track distance, along-track distance, and depth. Additional data processing steps and cleaning after acquisition add a status flag feature to the data set, differentiating between valid and invalid samples.

MBES data processing is time-intensive because of large data volumes, high data dimensionality, and the required involvement of an experienced hydrographer. The current automated outlier detection methods fail to explain the origin of an outlier; thus, an experienced hydrographer is needed to make final decisions on the validation or rejection of depth outliers. Incorporating raw, processed, and derivative data from the input sensors of an MBES system during outlier detection could explain a given outlier and help facilitate or automate the decision-making process. The objective is to determine what relationships exist in MBES data sets between the raw sensor data and the status flag feature. Statistical methods and data clustering tools are used to identify connections. A sample data set is extracted from processed MBES data acquired by the CCGS Amundsen.

Results of the analysis will provide a better understanding of the complex relationships between features in a high dimensional MBES data set and will identify characteristics of importance with respect to detected outliers or invalid samples. This work supports the automation of validating MBES data and the application of online data processing and machine learning in the MBES processing workflow.