

SCCWRP Annual Report 2005-06

A statistical appraisal of disproportional versus proportional microbial source tracking libraries

Brian J. Robinson¹, Kerry J. Ritter and Rudolph D. Ellender²

¹*National Oceanic and Atmospheric Administration, Charleston, NC*

²*The University of Southern Mississippi, Department of Biological Sciences, Hattiesburg, MS*

ABSTRACT

Library-based microbial source tracking (MST) can assist in efforts to reduce or eliminate fecal pollution in waters by predicting sources of fecal-associated bacteria. Library-based MST relies on an assembly of genetic or phenotypic "fingerprints" from pollution-indicative bacteria cultivated from known sources to compare with and identify fingerprints of unknown origin. The success of the librarybased approach depends on how well each source candidate is represented in the library and which statistical algorithm or matching criterion is used to match unknown sources. Because known source libraries are often built based on convenience or cost, some library sources may contain more representation than others. Depending on the statistical algorithm or matching criteria, predictions may become severely biased toward classifying unknown sources into the library's dominant source category. We examined prediction bias for three of the most commonly used statistical matching algorithms in library-based MST when applied to disproportionately represented known source libraries. These include maximum similarity (MS), average similarity (AS), and discriminant analyses (DA). We found that MS was particularly sensitive to disproportional source representation, while AS and DA were more robust. We discovered that nearest neighbor (NN) analyses provides a compromise between correct prediction and sensitivity to disproportional libraries among the three statistical procedures. This includes increased matching success and stability that should be considered when matching to disproportionately represented libraries.

Full Text

ftp://ftp.sccwrp.org/pub/download/DOCUMENTS/AnnualReports/2005_06AnnualReport/AR0506_327-332.pdf