

A Citation Content Analysis

of "Data Publications" in
Earth Systems Science

Nic Weber
nmweber@illinois.edu
University of Illinois
Center for Informatics Research
in Science & Scholarship

Matthew Mayernik
mayernik@ucar.edu
NCAR Library
Research Data Services

Steven Worley
worley@ucar.edu
NCAR - Computational &
Information Systems Laboratory
Research Data Archive

Motivation

- > Better understand the impact of the **International Comprehensive Ocean and Atmosphere DataSet (ICOADS)**
- > Innovate with traditional bibliometric techniques to study the context of citations made to data products
- > Explore potential for text-mining "sentiment" of a citation
- > Inform future data citation analyses

Publications

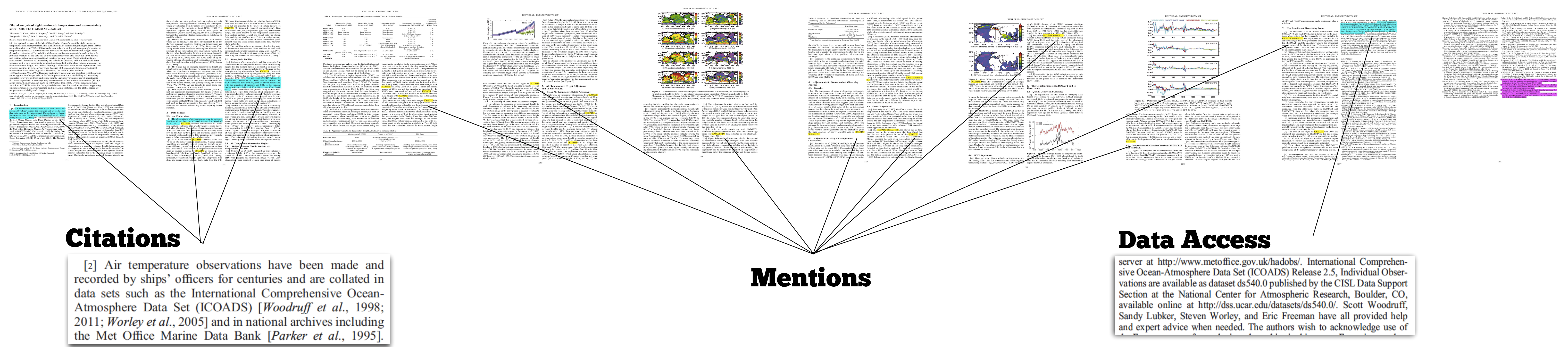
We first gathered all citations made to the following ICOADS data release papers:

- 1987 Woodruff et al. COADS 1.0 (n = 786)
- 1998 Woodruff et al. COADS 2.0 (n= 162)
- 2005 Worley et al. ICOADS 2.1 (n =190)
- 2011 Woodruff et al. ICOADS 2.5 (n= 57)

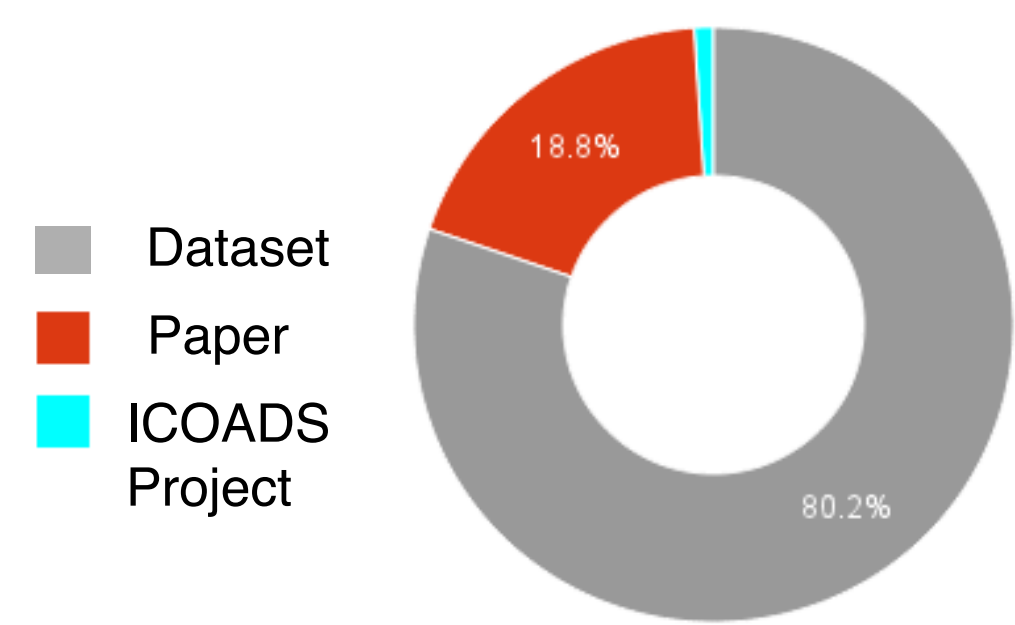
We then sampled a subset of the citing documents (n = 112) and coded the **content** and **context** in which ICOADS was **cited** (formally in the reference list) or **mentioned** (informally by name).

Context refers to the structural information of a citation: location, number of times cited, number of authors, type of publication, etc.

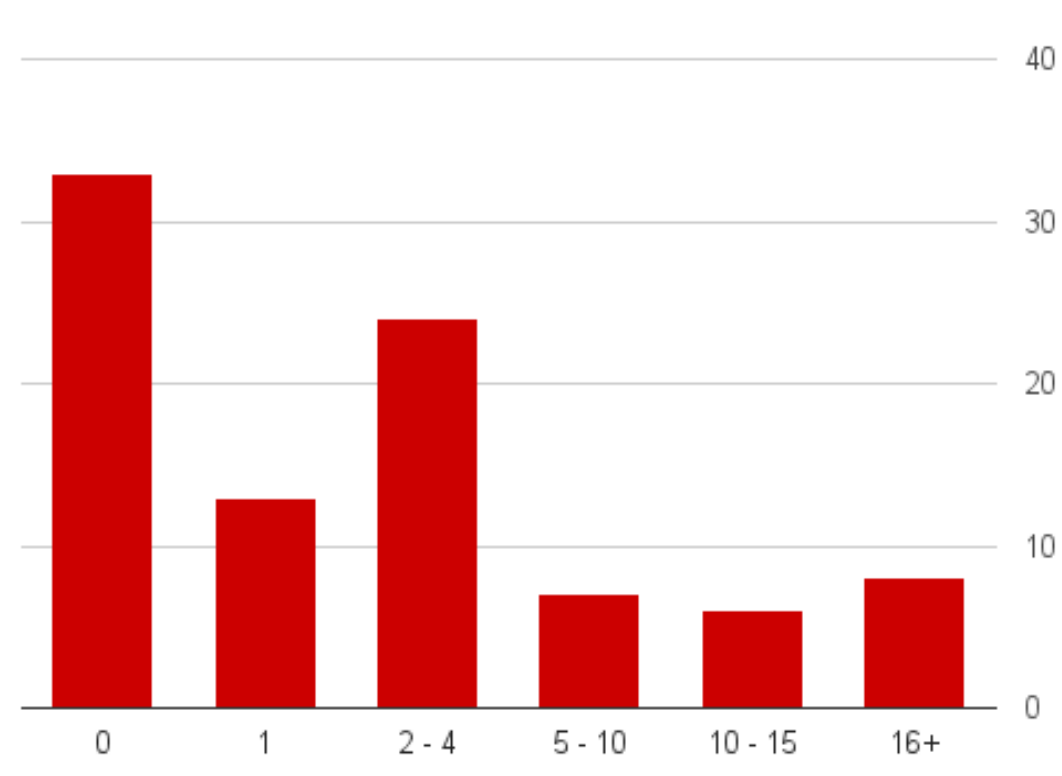
Content refers to the functional information of a citation: sentiment (positive, negative or neutral), the purpose, the meaning, etc.



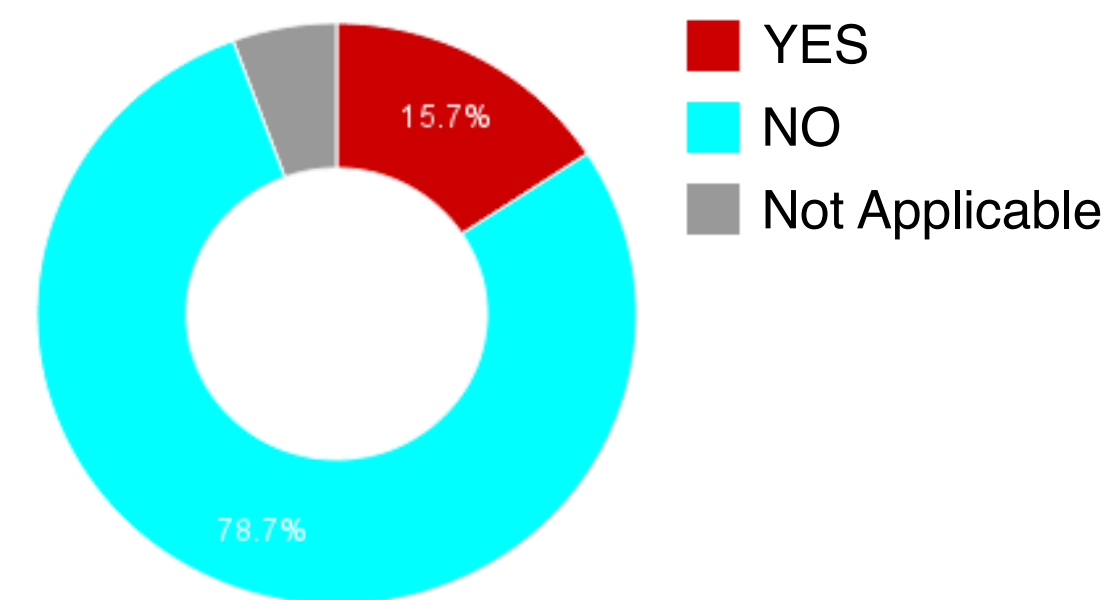
What were authors citing?



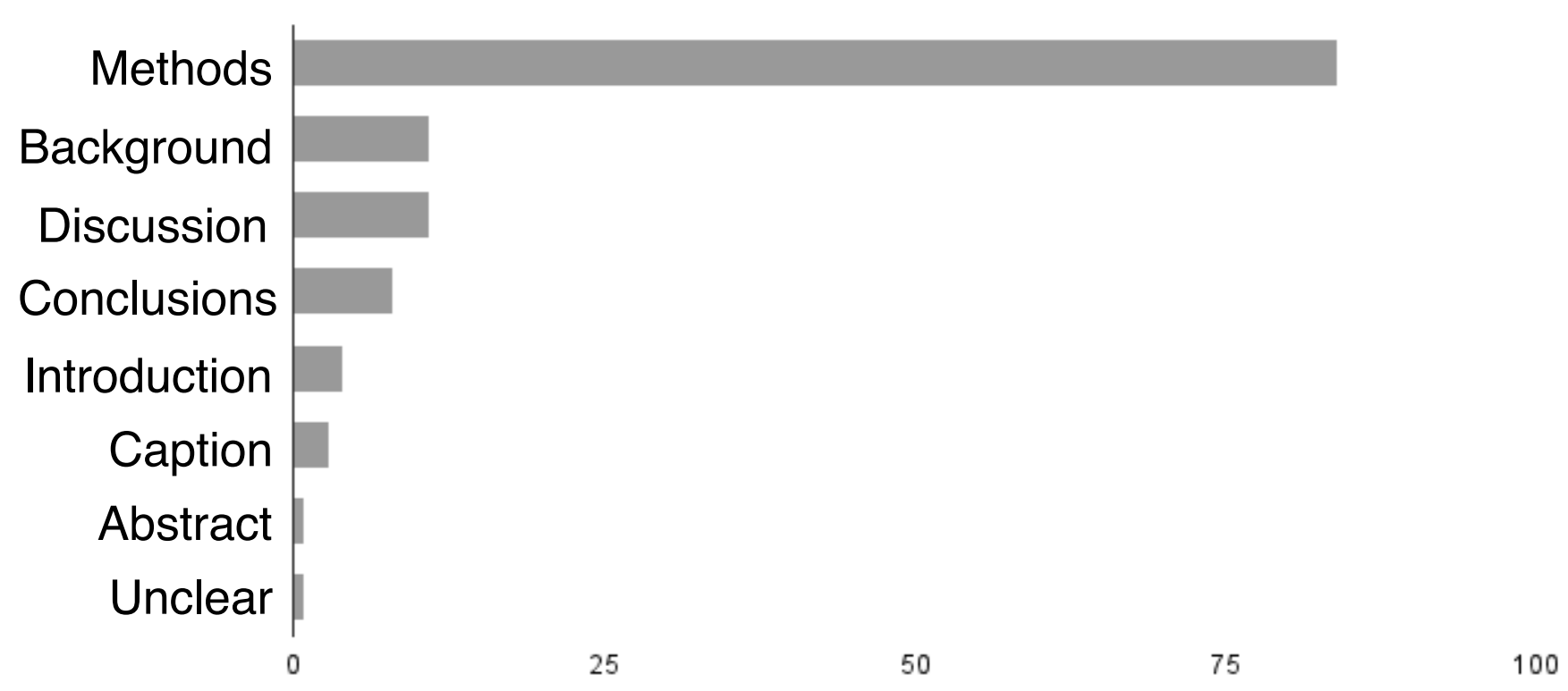
Times mentioned per publication



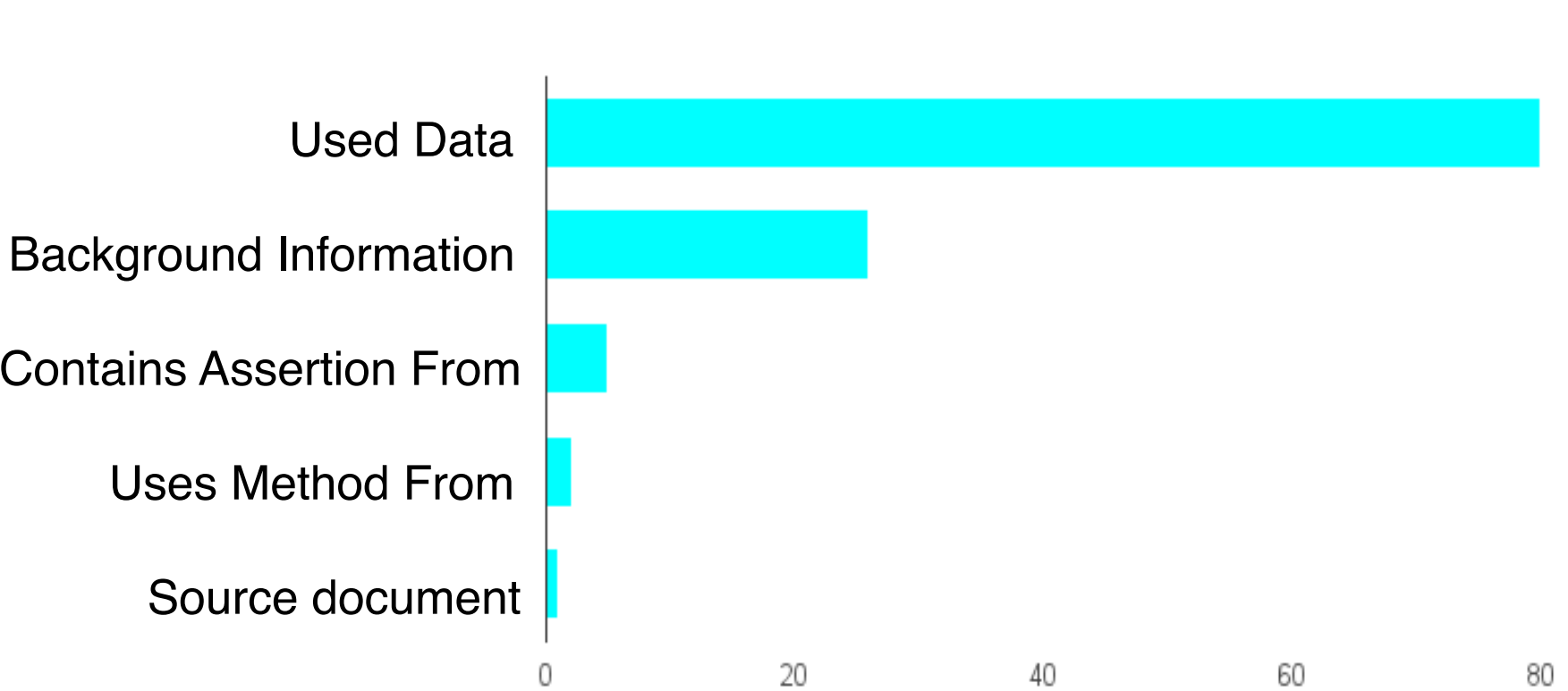
Did authors provide direct access?



Where did citation appear?



Why was ICOADS cited / mentioned?



Future Work

Unfortunately, our biggest dilemma is access to full text of ICOADS publications.

Many of our tasks might be automated in future work. This includes:

- Refining our classifications in order to code additional data publications
- Training language models to recognize sentiment & motivation of a citation
- Improving recognition and extraction of "mentions"

Works Cited

1. Woodruff, S.D., R.J. Slutz, R.L. Jenne, and P.M. Steurer, 1987: A comprehensive ocean-atmosphere data set. Bulletin of the American Meteorology Society 68, p. 1239-1250.
2. Woodruff, S.D., H.F. Diaz, J.D. Elms, and S.J. Worley, 1998: COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields. Physical Chemistry of the Earth, 23, p. 517-526.
3. Worley, S.J., S.D. Woodruff, R.W. Reynolds, S.J. Lubker, and N. Lott, 2005: ICOADS Release 2.1 data and products. International Journal of Climatology (CLIMAR-II Special Issue), 25, 823-842. doi:10.1002/joc.1166
4. Woodruff, S.D., S.J. Worley, S.J. Lubker, Z. Ji, J.E. Freeman, D.I. Berry, P. Brohan, E.C. Kent, R.W. Reynolds, S.R. Smith, and C. Wilkinson, (2011) ICOADS Release 2.5: Extensions and enhancements to the surface marine meteorological archive. International Journal Climatology (CLIMAR-III Special Issue), 31, 951-967. doi: 10.1002/joc.2106