

Agriculture

of Food and Agriculture

Visualization of Science Supported by NADP Measurements

Ray Knighton¹, Susan Braxton², David A. Gay², and Noor Syahira Nasarudin² ¹ US Department of Agriculture – National Institute of Food and Agriculture and ² University of Illinois at Urbana-Champaign

Introduction

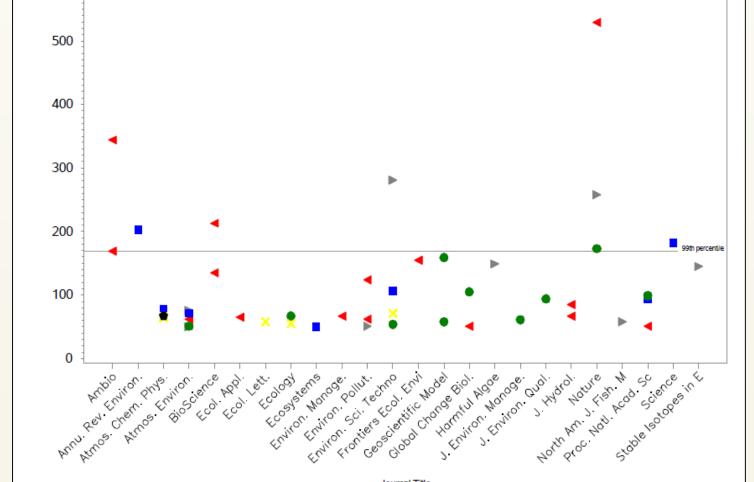
NADP measurements are widely used in research across a range of disciplines. NADP program staff identify publications using NADP measurements and maintain bibliographies on the NADP website as evidence of the impact of the program. Mapping connections among scientific publications using various attributes can increase understanding of research networks, the growth and diffusion of research topics, and relationships among authors and institutions (Börner et al., 2003). Bibliometric analyses were performed on the NADP-related publications to show disciplinary, author, and institutional connections among publications that use NADP measurements. New visualization tools are becoming available to display these connections/networks in 2dimensional space. The work described here is based on analyzing the words and meta data in a collection of documents to describe and categorize the documents and then visualize the documents in a spatial framework.

Methods

Data collection: Google Scholar was used to identify 1,277 unique publications that cited the use of NADP measurements from 2007-2014, by searching the full text for mentions of the National Atmospheric Deposition Program, then reviewing publications to determine whether or not NADP measurements were used. This approach was required because formal citation of data in scholarly publications is not practiced consistently. From this set of 1,277 works, full bibliographic metadata for analysis was retrieved from Scopus (http://www.scopus.com) for 963 publications.

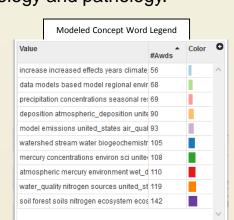
Semantic analyses: Semantic analyses were performed using VOSviewer (Van Eck and Waltman, 2015) and an instance of Pushgraph™ (Chalklabs, 2015) being developed for USDA-NIFA. Using natural language processing techniques, VOSviewer extracts terms from the corpus file (963 titles and abstracts), where a term is defined as a sequence of nouns and adjectives (ending with a noun). Based on the extracted terms, VOSviewer creates a term map. This is a map in which terms are located in such a way that the distance between two terms provides an indication of the number of co-occurrences of the terms. In general, the smaller the distance between two terms, the larger the number of co-occurrences of the terms. Two terms are said to co-occur if they both occur on the same line in the corpus file. Network maps of co-authorships and institutional affiliations were also created with VOSviewer. The National Institute of Food and Agriculture of USDA has semantically modeled approximately 130,000 (the grey dots in the adjacent map) project documents using PushgraphTM to create approximately 700 subject matter concepts that describe a wide range of scientific disciplines. Words from titles and abstracts of the 963 NADP-related publications were used to infer/model NIFA subject matter concepts and associated those concepts with the publications and then visualized the documents in 2-D space against the base NIFA science Pushgraph™ map.

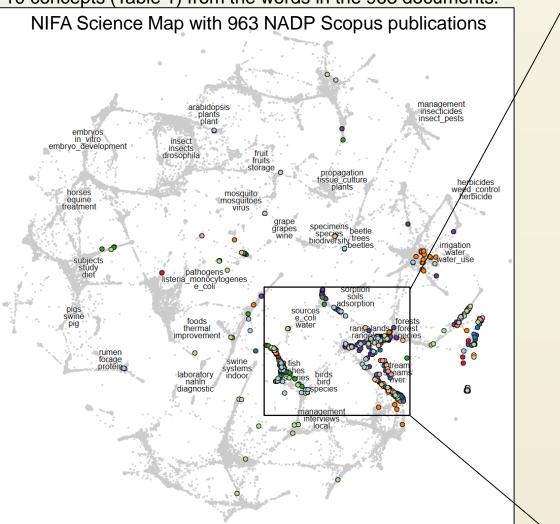
Citation of NADP-Data Journal Articles

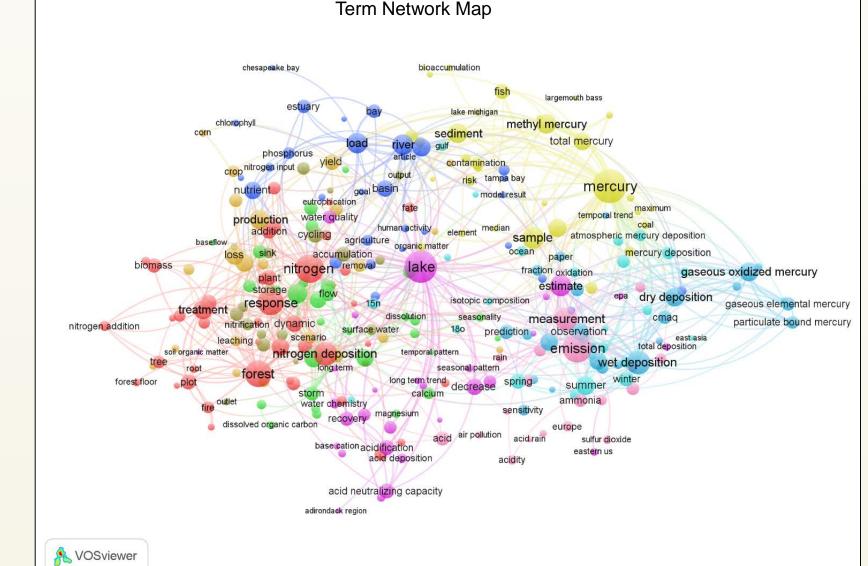


Journals which published the most cited articles using NADP's data, and the respective number of citations of each article. Articles with >50 citations are shown. These influential articles come from a wide variety of journals and general subject types. The most widely cited articles were published in Nature, Ambio, Environmental Science & Technology, and Bioscience. The most cited article was from Nature, "Dissolved organic carbon trends resulting from changes in atmospheric deposition chemistry" by Monteith et. al. in 2007.

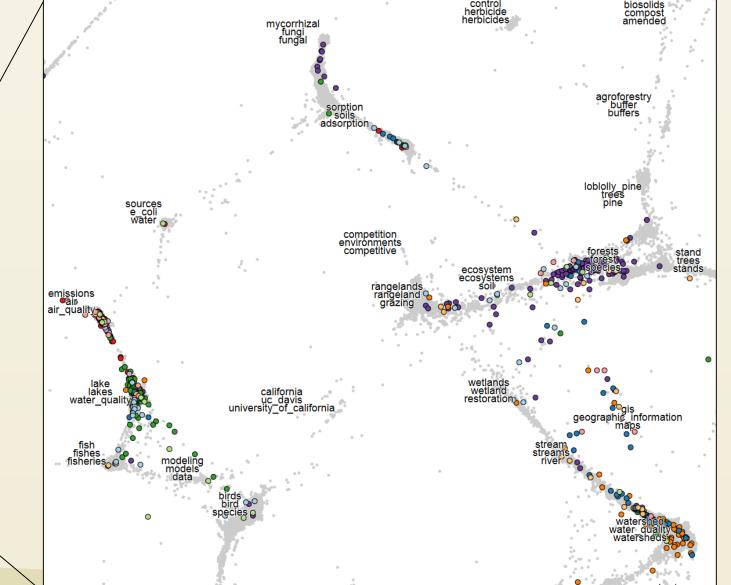
Pushgraph™ was used to model 10 concepts (Table 1) from the words in the 963 documents. Each NADP-related publication was inferenced against the USDA CRIS modeled concepts and mapped accordingly. The inset is zoomed to show the dominant science disciplines related to surface water, soils, air emissions, forestry, and disciplines have used NADP data, such as basic plant biology and pathology.



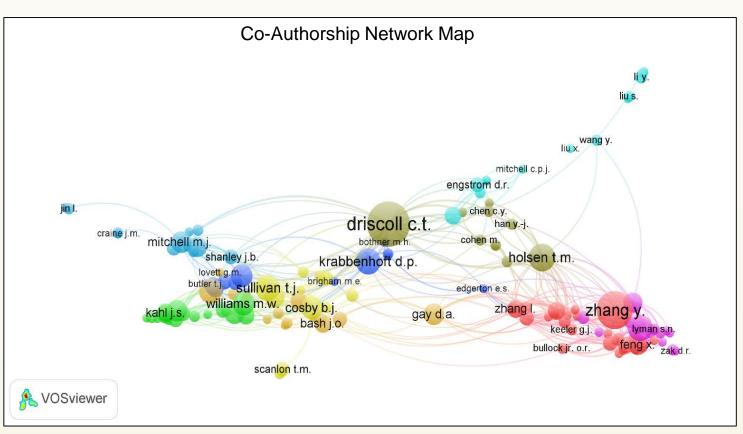




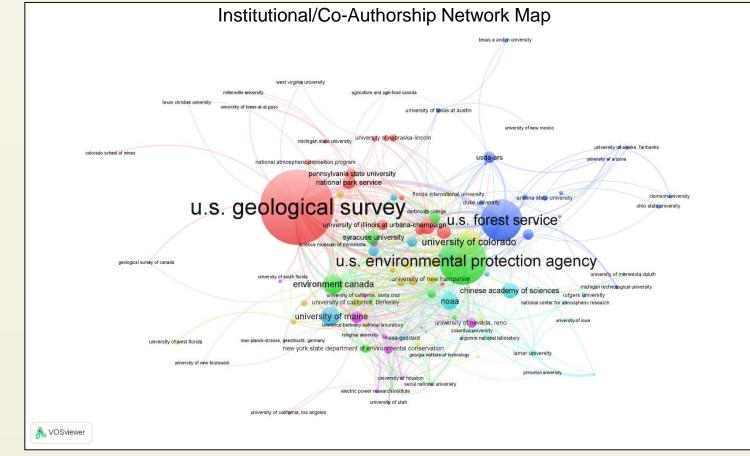
VOSviewer software was used to find and display the top 273 most relevant terms from the 963 titles and abstracts that had a minimum of 15 occurrences. Colored clusters illustrate the dominant word topics. The size of the circle and label is proportional to the number of occurrences of the word. The closer the words the higher the association of the words to each other. The thicker the lines the stronger the link between two items.



Results



VOSviewer software was used to find and display the top 140 most relevant authors from the 963 titles and abstracts that had a minimum of 5 publications. Colored clusters illustrate the dominant co-author affiliations. The size of the circle and label is proportional to the number of publications. The lines show the connections between co authors. For example, author C.T. Driscoll had 27 publications and 56 coauthors.



VOSviewer software was used to find and display 128 of 818 of the most relevant institutions determined from the author affiliations in the 963 Scopus publications that had a minimum of 5 publications. The size of the circle and label is proportional to the number of publications for each institution. The lines show the connections to institutions as determined by co-authorships. The thickness of the line is the strength of the connection which is based on the number of co-authorships. For example, the U.S. Geological Survey had the largest number of publications (239) and the most coauthorships (345). The thick line between The U.S. Geological Survey and the U.S. Forest Service indicates a relative link strength of 26 compared to the link strength of 4 for the link between the U.S. Geological Survey and the Geological Survey of Canada (thin red line to the left of the U.S. Geological Survey circle).

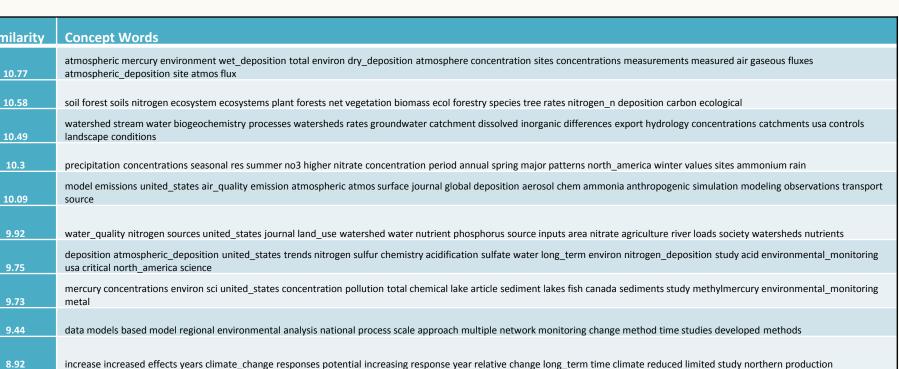


Table 1. The concept words modeled from the 963 Scopus publications using Pushgraph™. Also shown is the relative percentage of each publication that is similar to the concept defined by the concept words

		# of	# of Co-			# of	# of Co-
ank	Institution	# 01 Publications	authorships	Rank	Institution	# 01 Publications	authorships
1	U.S. Geological Survey	239	345	11	Syracuse University	35	80
2	U.S. Environmental Protection			12			
	Agency	143	367		National Park Service	32	80
3	U.S. Forest Service	122	221	13	North Carolina State University	30	43
4	University of Colorado	66	159	14	Harvard University	29	178
5	Environment Canada	53	232	15	USDA-ARS	28	54
6	University of Maine	50	86	16	University of Minnesota	27	34
7	NOOA	44	161	17	Pennsylvania State University	26	49
8	Colorado State University	40	59	18	University of Michigan	26	41
9				19	University of Illinois at Urbana-		
	Chinese Academy of Sciences	40	21		Champaign	26	29
LO	Cornell University	36	82	20	University of Virginia	25	63

Table 2. The top 20 institutions based on the number of publications found in the 963 Scopus publications as well as the number of co-authorships for each institution.

Conclusions

VOSviewer identified subject matter term clusters around **nitrogen** and **mercury**, and indicated the research topics connections around nitrogen are more numerous and more integrated than around mercury.

VOSviewer also revealed major clusters around processes and specific environments.

On the PushgraphTM concept map, NADP-related documents mapped most frequently to concepts around **air** quality, water quality, forests, and watersheds.

Most commonly appearing author was **Driscoll** with 56 co-authors.

Key author affiliations consisted of authors at federal agencies that support NADP monitoring. The top five affiliations in rank order were : U.S. Geological Survey, U.S. Environmental Protection Agency, U.S. Forest Service, University of Colorado, and Environment Canada. The Chinese Academy of Sciences was the most prolific foreign entity using NADP data.

Of the 963 total records retrieved from Scopus, 441 (46%) of them cited the NADP product used in the bibliography. There is a widespread practice of referencing data within the text but omitting formal citations from the references cited section. NADP needs to promote the proper citation of their data to the scientific community to make it easier to find and identify the impact of NADP data on science.

References

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. Ann. Rev. Info. Sci. Tech., 37(1), 179–255. doi:10.1002/aris.1440370106

PushgraphTM [software]. Bloomington, IN: Chalklabs. Retrieved from http://chalklabs.com

Van Eck, N. J., & Waltman, L. (2015). VOSviewer Version 1.6.2. [software] Leiden: Center for Science and Technology Studies. Retrieved from http://www.vosviewer.com/Home