Using Bibliometric Data and Analysis Tools to Identify Emerging Research Opportunities

Introduction
The Information Services Office (ISO) at the National Institute of Standards and Technology (NIST), a non-regulatory agency within the U.S. Department of Commerce, is working to identify emerging research areas where NIST can have the greatest impact on industry and society. ISO staff analyzed bibliometric data in order to reveal potential topics of interest in water quality metrology, a strategic priority area for NIST’s Material Measurement Laboratory (MML).

Methodology
ISO’s search strategy was created using measurement-related keywords identified from two MML water strategy documents and searching for those keywords within 31 water research journals. A Web of Science search identified 10,714 journal articles for the years 2012-2017. The full paper set was analyzed with the Sci2 tool. Analysis in R and CiteSpace focused on the 1,000 papers in this set that had received the most usage in Web of Science over the past 180 days. An analysis of Essential Science Indicators Research Fronts was used to evaluate the CiteSpace model.

Keyword Bursts
Using Kleinburg’s burst detection algorithm within the Sci2 tool and an approach identified by Guo et al. (2011), ISO identified keywords that have increased in frequency over time. An illustration of all bursts detected in 2017 focuses on the Ultratube and an activated carbon. The largest cluster of keywords to emerge in 2017 (army green) contained several cluster labels that may be of interest to NIST, including wireless sensor networks, optimal operations, neural networks, and hydrolysis.

Emerging Keyword Clusters
A 2016 cluster (brown) that emphasizes eutrophication also includes two references to SWAT (Soil and Water Assessment Tool) modelling, which appears as a bursting keyword in ISO’s Sci2 analysis.

The largest cluster of keywords to emerge in 2017 (army green) contained several cluster labels that may be of interest to NIST, including wireless sensor networks, optimal operations, neural networks, and hydrolysis.

A 2016 cluster (grass green) lists “microplastics” as its most prominent label. Other terms of interest include enhanced biological phosphorus removal and anaerobic digestion processes.

Model Comparison
Of the three approaches shown here, the emerging keywords model (pictured above) provided the most potential for further exploration; large clusters of keywords can be connected to pertinent papers and linked to source material in the CiteSpace interface. ISO compared the results of this model to 22 groups of papers found in Essential Science Indicators (ESI) Research Fronts. An ESI Research Front is a group of highly cited papers over a five-year period in a specialized topic defined by a cluster analysis.

Three of the most recent paper groups identified in the ESI search are all mirrored by a slightly broader CiteSpace cluster. For example, a Research Front for silver nanoparticles mirrored one of the larger CiteSpace clusters, Nanoparticles.

Conclusions
ISO’s analyses revealed several potential topics of interest to NIST researchers, including wireless sensor networks, neural networks, SWAT modeling, microplastics, biological phosphorus removal, and anaerobic digestion processes.

Future efforts will focus on smaller sets of papers identified with the help of subject matter experts in order to conduct a deep analysis of the topics that are most pertinent to the strategic planning process at NIST.

References
