MAPPING THE BACTERIAL CONTENT OF SURFACE WATERS WITH LANDSAT TM DATA: IMPORTANCE FOR MONITORING GLOBAL SURFACE SOURCES OF POTABLE WATER

Robert K. Vincent
Department of Geology, Bowling Green State University

R. Michael McKay
Mamoon M. D. Al-Rshaidat
Department of Biological Sciences, Bowling Green State University

Kevin Czajkowski
Thomas Bridgeman
Jeffrey Savino
Lake Erie Center, University of Toledo

ABSTRACT

An imaging algorithm for mapping early blooms of cyanobacteria, blue-green algae that are sometimes toxic to mammals, with LANDSAT TM data has recently been tested in Lake Erie and its tributaries. Testing is underway on similar LANDSAT TM algorithms that have been developed for mapping the surface freshwater contents of total coliform bacteria and E. coli, a subset of coliform bacteria that feeds on fecal matter. Cyanobacteria, coliform, and E. coli are all major bacterial threats to potable drinking water in surface freshwater sources; monitoring them is important to public health around the globe. Although the 16-day repeat cycle of LANDSAT TM is not optimum for this global public health application, the 30-meter spatial resolution of LANDSAT TM aids the mapping of bacteria in streams with widths $\geq 90$ m and drinking water reservoirs $\geq 5$ hectares in area, for water depths of $\geq 2$ m. Most other satellites either have insufficient spatial resolution to monitor streams and small drinking reservoirs or lack the appropriate bands (equivalent to TM bands 1-5 and 7) to do so. This is one of the biological applications of remote sensing that became possible only after the USGS/OhioView cooperation near the turn of the millennium that streamlined the data processing stream of LANDSAT TM data, such that freshly collected data could be delivered to a user from the EROS Data Center within 48 hours of collection. There is a global need for at least eight LANDSAT TM sensors in orbit continuously (2-day coverage) for these important biological applications.

INTRODUCTION

It may seem odd, on first impression, that papers about the mapping of bacterial contents of surface water in lakes, reservoirs, and streams were included in this remote sensing conference on Global Priorities in Land Remote Sensing. However, fresh water bodies are found on land masses, and in a world where the populations of humans doubled from 3 billion to 6 billion people during the years 1960-2000, the maintenance of potable sources of drinking water has become one of the most pressing needs of human survival. Groundwater is a non-renewable resource over the life-times of several generations of people, except when the water is from very shallow depths. For instance, most of the groundwater that has been produced from the Ogallala Aquifer in the states of Nebraska, Oklahoma, Texas, New Mexico, and Colorado originally came from the eastern slopes of the Rocky Mountains many millennia ago. Since then, uplifts and faulting have disconnected that source of recharge, and the groundwater mined from this once-prolific aquifer will not be replaced, except near the Sand Hills of Nebraska, where it is close enough to the surface to receive recharge sufficient for that local area.

Regarding human consumption, surface waters are at once the easiest to tap, most quickly replenished, and most readily contaminated. Though contamination of surface waters can come in many forms, no source of contamination is more immediate than that caused by increased bacterial content of those waters, regardless of whether the cause of bacterial proliferation is primarily anthropogenic or natural. Although engineers have done a remarkable job of learning to treat surface water for drinking purposes, at some point it becomes far less expensive to maintain a clean lake or reservoir as an ecological entity that to treat ever-degraded inputs of an ever-increasing
number of water treatment plants on that entity. An additional benefit of the former is that the lake or reservoir remains attractive for water recreation (including fishing) and tourism.

Satellite remote sensing has a lot to offer humanity in the maintenance of surface water sources of drinking water. First, it can give us the tools to determine the spatial and temporal patterns of bacterial contamination, thereby yielding information that will lead to an understanding of the ecology of a given body of water, which in turn leads to knowledge of what causes the contamination and how it can be corrected in the safest, most efficient manner. Second, it can give us the ability to monitor the body of water to detect significant changes in the status quo. The beauty of satellite remote sensing is that whatever we learn to do for one lake can be applied around the globe.

The purpose of this paper is to call attention to recently developed tools for mapping bacterial content of surface waters, to the unique abilities of the LANDSAT TM sensor for mapping these bacteria in lakes and reservoirs used for drinking water, and to the great need for orbiting new LANDSAT satellites in the time gap between now and the year 2010, when the next LANDSAT satellite is scheduled to be launched. The next section will briefly review two major types and one sub-type of bacteria that are important to water quality of lakes and reservoirs used for drinking water. The section after that will review three new algorithms that use LANDSAT TM data for mapping the content of these bacterial types in surface waters, and the conclusions and recommendations of this paper will be made in the final section.

BACKGROUND ON CYANOBACTERIA AND COLIFORM BACTERIA

Bacteria range in linear dimension from 1-5 µm (J. A. and C.A. Ingraham, 2000), which makes bacteria larger than most the wavelengths of light in LANDSAT TM bands 1-5 and 7 that cover parts of the wavelength range from 0.45-2.35 µm. Therefore, it is theoretically possible to map bacteria in the water directly. However, mapping surrogates of the bacteria, such as what they feed on, offers another possible way to map where the bacteria are located.

Of the many types of bacteria that can affect water quality adversely, the two most frequently cited are cyanobacteria, often called blue-green algae, and coliform bacteria. There are three types of cyanobacteria: anabaena, aphanizomenon, and microcystis. All three can produce toxins that are harmful to humans and other mammals. Some Lake Erie strains of *Microcystis* spp. produce the peptide hepatotoxin microcystin, which is harmful to waterfowl or other animals that might drink the untreated water (Brittain et al, 2000). Microcystin has also been identified as a tumor promoter, making long-term ingestion of even low levels of the toxin a topic of concern (Falconer and Humpage, 1996; Carmichael, 2001).

Coliform bacteria feed on decaying animal and plant tissue. A particularly dangerous subset of coliforms is *E. coli*, which feed on fecal matter. *E. coli* is one of the most important contaminants of water quality, and it can cause diarrhea and gastroenteritis in humans (Hoeprich et al, 1994; Xu et al, 1994). There is at least one type of *E. coli* that is usually fatal to humans. Although chlorine kills all sorts of coliform bacteria, knowledge of the level of coliform content that is expected to enter a water treatment plant helps assure the plant operators that enough chlorine is on hand for killing larger amounts of such bacteria.

The first step in mapping a particular type of bacterium from a satellite sensor is to find an algorithm (or recipe) for combining spectral bands of that sensor such that an image is produced that gets brighter (or redder, if in color) as the water’s content of that bacterium increases. Algorithms for estimating the content of all three types of bacteria from LANDSAT TM data have recently been developed. Whereas all three of the sometimes toxic cyanobacteria listed above contain two pigments, chlorophyll *a* and phycocyanin, almost all forms of algae contain the former and very few contain phycocyanin. To map cyanobacteria, Vincent et al (2004) created an algorithm that maps phycocyanin content (PC) in surface waters. The algorithm was created by performing multiple regression between dark-object-corrected (for atmospheric haze and additive signal offset of the sensor) spectral ratios of LANDSAT TM data (from bands 1-5 and 7) and phycocyanin content measured for 30 water samples collected in Lake Erie within ±2 hours of LANDSAT 7 overpass. The best model was taken to be the one with the highest $R^2$ (adjusted) that also passed the Durbin-Watson test, which assured minimal correlation among input variables. The best spectral ratio model, which had an $R^2$ (Adjusted) = 77.6% and $S = 0.5921$ micrograms/liter (about 14.8% of the total PC range for the July 1, 2000 overpass), is given by

\[
PC (\mu g/L) = 47.7 – 9.21(R31) + 29.7(R41) – 118(R43) – 6.81(R53) + 41.9(R73)-14.7(R74) \quad \text{Eqn. 1}
\]
where $RIJ$ (formerly called $R_{ij}$) stands for the dark-object-subtracted spectral ratio of the $i$th band over the $j$th band. The algorithm was tested for robustness on data from a September 27, 2000 LANDSAT 7 overpass, and the root mean square errors in PC found on the later date, from the algorithm derived on the earlier date) was 2 micrograms/liter ($\mu$g/L), or about 18% of the total range of PC on the September 27 date.

Vincent et al (2005) have more recently applied the same methodology to develop algorithms that estimate total coliform content (TC) and $E. coli$ content (EC) from LANDSAT 7 data collected in Lake Erie and the mouth of the Maumee River on August 21, 2001. There was evidence to limit the EC algorithm to waters of $\geq 2$ m depth. The algorithms were tested on a second date, October 6, 2001, for only 8 water samples collected in the Maumee and Portage Rivers, which are tributaries of the western basin of Lake Erie. The root mean square error for TC ($\Delta TC_{rms}$) was 91 colonies/ml, or 27% of the range of actual TC measurements on October 8, 2001, and for EC ($\Delta EC_{rms}$) was 250 colonies/100 ml (2.5 colonies/ml), or 11% of the range of actual EC measurements on October 8, 2001. Within the goal of mapping large (50% or more) changes in these two parameters from the August 21, 2001 models, the rms errors of both the TC and EC algorithms are acceptable. Better testing is needed, especially with lower bacterial content for the $E. coli$ algorithm, but such testing has become much more difficult because of a scan line converter (SLC) failure on LANDSAT 7 on May 31, 2003, and development of a striping problem over water in LANDSAT 5 data.

EXAMPLES OF MAPPING BACTERIA IN A SMALL RESERVOIR FOR DRINKING WATER

An important application of LANDSAT TM data has been overlooked in the past: monitoring of small fresh water reservoirs designed to store drinking water. An example of how the above three algorithms can view small drinking reservoirs is shown in Figures 1-2 below. The two reservoirs are owned and operated by the village of Archbold, Ohio, which is located in the same LANDSAT TM frame center as the Ohio cities of Bowling Green and Toledo and the Michigan cities of Ann Arbor and Detroit, namely, Path 20, Row 31. Figures 1-2 show a small area on the east side of Archbold in false color (TM bands 2, 3, and 4 displayed as blue, green, and red, showing vigorous vegetation as red), as a phycocyanin content (PC) image, as a total coliform image (TC), and as an $E. coli$ content (EC) image. The PC image shows pretty low values, especially for the non-border pixels (likely mixtures of land and water), but this is the time of year when cyanobacterial blooms are very low. However, in this same frame (and same date), an incipient bloom of cyanobacteria was imaged at the mouth of the Maumee River in the southwest corner of Lake Erie, with values up to 5 $\mu$g/l of PC. In some years, the manager of the Archbold water treatment plant has said that cyanobacterial blooms have occurred in these reservoirs, which were treated with chemicals to kill the blooms because of the possible toxins they might produce. However, he wants to know where incipient blooms are occurring, so that the small, early blooms can be killed with fewer chemicals, which themselves can create problems for drinking water quality.

The TC image shows values of total coliform content that are not especially high, compared to parts of Lake Erie, where values more than ten times higher have been found at different times in the summer. The red in the TC image outside the reservoirs are places where the land mask did not recognize the land as such, and the TC algorithm produced erroneous values $>20,000$ CFU/100 ml. The EC image shows the red, orange, and most of the yellow pixels to be above reportable values to USEPA (160 CFU/100 ml), but the EC algorithm has an rms error of 250 CFU/100ml, so these values are suspect. However, on August 21, 2001, there were places on the Portage River (about 50 miles east of Archbold and in the same frame) where EC values were estimated by this EC algorithm to be as high as 1,083 CFU/100 ml (Vincent et al, 2005). Further work is needed to confirm or improve the EC algorithm for values below 300 CFU/100 ml.
Figure 1. (A.) False color image of Archbold, Ohio water reservoirs on July 1, 2000, with LANDSAT 7 ETM+ bands 1, 2, and 4 displayed as blue, green, and red, respectively. Deep water is black, green vegetation is red, and commercial building roofs are white in this image. Each small square is 1 pixel (30m x 30m) of LANDSAT 7 ETM+ data; the small reservoir is approximately 237m x 226 m (about 64 pixels) and the larger reservoir is approximately 533m x 290m (about 160 pixels). (B.) Same area as (A.), but for the PC algorithm (Vincent et al, 2004) with PC ≥ 4.9 µg/l displayed as red, PC = 1.5-4.8 µg/l displayed as yellow-orange, PC = 0.2-1.4 µg/l as green and PC = 0-0.2 µg/l as blue. Land is shown as yellow-green background.

Figure 2. (A.) Total coliform content (TC) from Vincent et al (2005) on July 1, 2000 (from model of Aug. 21, 2001) of Archbold, Ohio, water reservoirs. Red in the water (but not on a pixel bordering land) is TC = 7,700-16,100 CFU/100ml, yellow and orange is 7,000-7,699 CFU/100ml, light blue and green is 5,600-6,999, and blue is 0-1,100 CFU/100ml, with land (TC=0) shown as dark blue background. (B.) E. Coli content (EC) from Vincent et al (2005) on July 1, 2000 (from model of Aug. 21, 2001) of Archbold, Ohio water reservoirs. Red in the water is EC = 214-245 CFU/100ml, orange is 183-213 CFU/100ml, yellow is 148-182 CFU/100ml, green is 111-147 CFU/100ml, and blue is 0-110 CFU/100ml. Land is also shown as red.

Finding similar algorithms for satellites with more frequent repeat cycles than LANDSAT TM will certainly improve the monitoring of large lakes, like Lake Erie, for water quality. An example of that will be given in the next paper of this conference session (Dash et al, 2005b) for a PC algorithm that was developed for SeaWiFS data, by application of the same methodology as used by Vincent et al (2004) for the LANDSAT TM PC algorithm shown.
above. However, such coarse spatial resolution satellite sensors cannot monitor small drinking water reservoirs like the Archbold reservoirs. For example, one of the SeaWiFS pixels (1,000 m x 1,000 m) would approximately cover the entire area shown in each of the individual images in Figure 1 and 2 above, and neither of the reservoirs could be resolved. Thus, LANDSAT TM has a unique role to play in water quality monitoring of small drinking reservoirs of cities, towns, and villages throughout the world.

CONCLUSIONS AND RECOMMENDATIONS

Algorithms for producing image of PC, TC, and EC have been developed with LANDSAT TM data to map cyanobacterial (blue-green algae) blooms in their nascency and coliform bacterial content, including fecal coliforms, in surface waters. All three of these types of bacteria are important for water quality, especially in drinking water reservoir, such as the twin reservoirs of the village of Archbold, OH, where examples were shown. It is possible that similar algorithms may be found to image one or all of these bacterial parameters with satellites that have a daily repeat cycle, such as SeaWiFS and MODIS, but the coarser spatial resolution of those satellites render them useless for mapping water quality changes in small drinking reservoirs, such as those of Archbold. However, the coarser spatial resolution sensors are much more useful than LANDSAT TM data for showing the sequence of water quality spatial patterns in Lake Erie over long periods (months) of time.

An example of a PC algorithm for SeaWiFS data was recently developed (Dash et al, 2005a) and was used to map PC in the whole of Lake Erie for all the non-cloudy days from July 1 – October 15, 2003, a year of large cyanobacterial blooms in Lake Erie. That algorithm was used to create an animated video of changes in PC throughout that time period in Lake Erie, which will be included in the paper following this one.

These recently developed water quality algorithms need to be tested further in the vicinity of Lake Erie and applied to fresh water reservoirs at different latitudes before they are fully operational. It is particularly important that many more drinking water reservoirs be monitored with the LANDSAT TM algorithms as soon as possible. Yet, both LANDSAT 7 and LANDSAT 5 are having technical difficulties that greatly hinder or preclude such improvements, and the use of archived data from earlier times (before May 13, 2003, when LANDSAT 7 lost the function of its scan line converter) is only possible when these three water quality parameters were measured on one of these past days of overpass. Because LANDSAT 7 has a 16-day repeat cycle and cloudy days are recorded about half the time in Ohio, there are only about 12 times a year where such serendipitous archived in situ data have a chance to be coincident with a day when water samples were collected and measured for one or all of the three water quality parameters PC, TC, and EC. Yet there are no plans announced by any country or company to orbit a new LANDSAT TM sensor before the year 2010.

Of course, there are now many other examples of how the LANDSAT TM sensor is needed for monitoring events of importance, such as the monitoring of tree diseases (Hurley et al, 2004), evapo-transpiration of plants (Boegh et al, 2002), and other biological remote sensing applications. However, with the world population doubling between 1960 and 2000, and inexpensive drinking water becoming scarcer almost every decade, the importance of monitoring surface water reservoirs for drinking water quality elevates the need for “filling the LANDSAT gap” to emergency status all on its own merits. The world greatly needs another one or more LANDSAT TM sensors with at least bands 1-5 and 7 orbited within the next three years. If the gap-fillers last past the next scheduled LANDSAT TM launch in 2010, all the better, because with two LANDSAT TM sensors we would have 8-day repeat coverage (assuming the sensors are identical, though LANDSAT 7 and LANDSAT 5 now clearly are not). With eight LANDSAT TM sensors in the same orbit, but delayed within that orbit, we would have a 2-day repeat cycle, and still have the spatial resolution (30 m) to monitor small drinking water reservoirs. The human species needs LANDSAT TM on a two-day monitoring cycle basis for this and other biological remote sensing applications. It is urgent that we get a start on that number with one or two LANDSAT TM gap-fillers before the end of year 2008.

ACKNOWLEDGEMENTS

This project was partially supported by OhioView NASA Glenn grants from 2000-2004. The authors wish to thank Dennis Howell, Village Administrator of Archbold, OH for his information about the Archbold drinking water reservoirs.
REFERENCES


