Journal of Exposure Science and Environmental Epidemiology **(2010) 20, 176–185**

www.nature.com/jes

Use of land surface remotely sensed satellite and airborne data for environmental exposure assessment in cancer research

SUSAN K. MAXWELL[a], JAYMIE R. MELIKER[b] AND PIERRE GOOVAERTS[c]

[a]*U.S. Geological Survey Earth Resource Observation and Science Center, Sioux Falls, South Dakota, USA*
[b]*Department of Preventive Medicine, Stony Brook University, Stony Brook, New York, USA*
[c]*BioMedware, Ann Arbor, Michigan, USA*

In recent years, geographic information systems (GIS) have increasingly been used for reconstructing individual-level exposures to environmental contaminants in epidemiological research. Remotely sensed data can be useful in creating space-time models of environmental measures. The primary advantage of using remotely sensed data is that it allows for study at the local scale (e.g., residential level) without requiring expensive, time-consuming monitoring campaigns. The purpose of our study was to identify how land surface remotely sensed data are currently being used to study the relationship between cancer and environmental contaminants, focusing primarily on agricultural chemical exposure assessment applications. We present the results of a comprehensive literature review of epidemiological research where remotely sensed imagery or land cover maps derived from remotely sensed imagery were applied. We also discuss the strengths and limitations of the most commonly used imagery data (aerial photographs and Landsat satellite imagery) and land cover maps.
Journal of Exposure Science and Environmental Epidemiology (2010) **20,** 176–185; doi:10.1038/jes.2009.7; published online 25 February 2009

Keywords: *GIS, remote sensing, satellite imagery, exposure assessment, cancer.*

Introduction

In recent years, geographic information systems (GIS) have increasingly been used for assessing individual-level exposures to environmental contaminants in epidemiological research (Beyea and Hatch, 1999; Ward et al., 2000; Nuckols et al., 2004). The growing body of literature on this application most commonly includes studies that assess geographic proximity of individuals to sources of environmental contaminants such as pesticide application on farms (Reynolds et al., 2005), landfill sites (O'Leary et al., 2004), and hazardous waste sites (Elliott et al., 2001; McNamee and Dolk, 2001). On occasion, investigators implement environmental fate and transport models to improve the exposure assessment (Reif et al., 2003; Yu et al., 2006; Riggs, 2007). These spatial applications are valuable for prospective data collection when monitoring campaigns are uneconomical, and for retrospective data collection in which collecting individual biomarkers or environmental measures is unattainable. Advancing exposure modeling, therefore, is beneficial for many epidemiological studies that require individual level exposure reconstruction. A specific focus on historical exposure reconstruction is essential for assisting the epidemiological study of diseases with long latency, such as many types of cancer (Nieuwenhuijsen et al., 2006). Adopting GIS in historical exposure reconstruction, however, has been limited by (1) the technical obstacle of linking individual mobility histories with space-time environmental models (Meliker et al., 2007) and (2) the difficulty in constructing compelling historical models of environmental measures.

Addressing the technical obstacle of linking space-time data sets was identified as the second most pressing issue during a recent National Cancer Institute sponsored workshop on *The Crossroads of GIS and Health Information* (Pickle et al., 2006). The continual expansion of space-time databases, coupled with the recognized need to incorporate residential history in environmental epidemiology, has highlighted the deficiencies of current software to visualize and process space-time information (Mather et al., 2004; Pickle et al., 2005). Recent development of space-time technology (aka Time-GIS) overcomes this technical obstacle by enabling visualization and analysis of space-time data sets (e.g., see STIS, TerraSeer Inc., Ann Arbor, Michigan, USA and STNexus, GeoVista, University Park, Pennsylvania, USA) (AvRuskin et al., 2004; Jacquez et al., 2005; Meliker et al., 2005; Guo et al., 2006; Weaver et al., 2006). In contrast to traditional GIS that are based on spatial data

---

1. Address all correspondence to: Dr. Susan K. Maxwell, U.S. Geological Survey Earth Resource Observation and Science Center, 47914 252nd Street, Sioux Falls, South Dakota 57198, USA.
Tel.: +605 594-6008. Fax: 605 594-6529.
E-mail: maxwell@usgs.gov
Received 3 September 2008; accepted 8 December 2008; published online 25 February 2009

structures—the "what, where" dyad that inadequately displays changes through time—space-time technologies are based on space-time data structures (Galton, 2003) that enable characterization of the "what, where, when" triad needed for effective representation of data used to analyze health outcomes.

Addressing the second stumbling block in adopting GIS in historical exposure assessment—creating compelling space-time models of environmental measures—can be aided by remotely sensed data archives. The primary advantage of using remotely sensed data is that it allows for study at the local scale (e.g., residential level) without requiring expensive, time-consuming monitoring campaigns (if such data are even possible to collect in the historical setting). Satellite imagery and aerial photography represent a vast but largely untapped resource to significantly enhance environmental mapping and modeling applications for exposure assessment in space and time. This rich archive of data can be used to characterize environmental change over time and be used in conjunction with GIS models to re-create contaminant exposures.

The purpose of this paper is to summarize how land surface remotely sensed data are currently being used to study the relationship between cancer and environmental contaminants. This paper presents the results of our findings from a comprehensive literature review, provides an overview of aerial photograph and satellite image archives, and presents a discussion on the strengths and limitations of using remotely sensed imagery in environmental exposure assessment in cancer research, focusing primarily on agricultural chemical exposure assessment applications.

Literature review

We performed a comprehensive search of the literature to identify publications that describe how remotely sensed imagery was being used in exposure assessment for cancer research. We focused on the intersection of three topics: remote sensing (land surface sensors only), exposure assessment applications, and cancer. The search was conducted using 10 electronic reference services (e.g., SCOPUS, ScienceDirect, PubMed) covering the entire span of years available in each database. Keywords used in the search covered the major categories of cancer, human health, remote sensing, and exposure assessment, in addition to other terms such as GIS, land use, and land cover to ensure that we captured the majority of relevant publications in the literature. The search was conducted on the title, abstract, and keywords within each database. After identifying key articles, papers that cite those articles were also reviewed.

The electronic search identified 650 potential publications. We read abstracts from the original list of 650 publications to identify those that specifically addressed our topic. One-hundred and five papers were identified that used remotely sensed imagery for exposure assessment in health studies, the majority of which (68 papers) used satellite imagery for predicting land cover patterns and locations of vectors in studies of infectious disease (see Herbreteau et al., 2007 for a recent review of the literature on remote sensing and infectious disease). Fifteen papers used classified land cover maps in land use regression models for predicting air pollution (none with cancer as a health end point) (see Ryan and LeMasters, 2007 for a recent review). Seven papers used satellite imagery in exposure science applications (no health end points). These included validation studies of models derived from satellite imagery (Matthijsen et al., 2000; Maxwell et al., 2006; Rull et al., 2006; AvRuskin et al., 2008), constructing a model of atrazine in a drinking water reservoir using land cover data (Atkinson et al., 2001), estimating deterioration of asbestos cement roofing using hyperspectral imagery (Bassani et al., 2007), and using satellite imagery to document flooding in schools as an indicator of building dampness and potential respiratory risk (Guidry and Margolis, 2005). One paper used satellite imagery and GIS to estimate proximity to different crop types in a study of birth weight (Xiang et al., 2000). One paper used an existing land use map derived from ground surveys in a GIS model to estimate pesticide exposure in a study of prostate cancer (Marusek et al., 2006).

The remaining 12 papers used remotely sensed data in studies of cancer. Seven papers used remotely sensed stratospheric ozone column data (U.S. National Aeronautics and Space Administration, Total Ozone Mapping Spectrometer) for predicting tropospheric levels of ultraviolet radiation, typically in studies of skin cancer, but also as an indication of vitamin D levels in studies of other cancers (Slaper et al., 1998; Piacentini et al., 2004; Solomon et al., 2004; Garland et al., 2006; Mohr et al., 2006, 2007; Tatalovich et al., 2006). The remaining five papers used land surface remotely sensed data for environmental exposure assessment in cancer research (Table 1). These five papers are the primary focus of this review. Three papers acquired and analyzed remotely sensed data in their studies (Ward et al., 2000; O'Leary et al., 2004; Ward et al., 2006), and two used existing land cover products that were derived from remotely sensed data (Brody et al., 2002; Ayotte et al., 2006). Aerial photographs and Landsat satellite imagery were the only remotely sensed imagery applied. The studies included investigations of exposure to pesticides in relation to bladder cancer (Ayotte et al., 2006), breast cancer (Brody et al., 2002; O'Leary et al., 2004), and non-Hodgkin's lymphoma (Ward et al., 2000; Ward et al., 2006).

Studies That Used Remotely Sensed Data Directly

Of the three papers that used remotely sensed data directly, one used aerial photography (O'Leary et al., 2004), one used Landsat imagery (Ward et al., 2000), and one used a

Table 1. Summary of literature review.

Citation	Disease	Contaminant	Sensor	Land type
Studies that used remotely sensed imagery				
O'Leary et al., 2004	Breast cancer Long Island, NY, USA	Pesticides	Aerial Photographs (~1950)	General Agriculture
Ward et al., 2000	Non-Hodgkin's Lymphoma Central NE	Pesticides	Landsat	Crop type
Ward et al., 2006	Non-Hodgkin Lymphoma Central IA	Pesticides	Landsat, Aerial photographs	Crop type Resident locations
Studies that used products derived from remotely sensed imagery				
Ayotte et al., 2006	Bladder cancer New England	Arsenic	Landsat (USGS NLCD)	Agriculture
Brody et al., 2002	Breast cancer Cape Cod, MA, USA	Pesticides	Aerial Photographs (1951, 1971, 1984, 1990)	Crop type

combination of aerial photography and Landsat satellite imagery (Ward et al., 2006). Ward et al. (2006) used digital aerial photographs only to quality control check GPS measurements that were collected at locations of residences. O'Leary et al. (2004) used aerial photographic prints from 1947 and 1950 to identify agricultural land use in an investigation of breast cancer risk and pesticide exposure in Long Island, New York. Residences were visually located on the photographs and were categorized as farmland, adjacent to farmland, residential, wooded, sand, greenhouse, open land, or not farm but may have been before 1947. Residences classified as either on farmland or adjacent to farmland were hypothesized as having higher exposure to pesticides. O'Leary et al. (2004) noted the lack of information available in aerial photographs to distinguish individual crop types and the need to use a magnifying glass to interpret the photographs.

Landsat satellite imagery was used in two studies to identify general land cover types as well as specific crop types (Ward et al., 2000; Ward et al., 2006). Both these studies used crop location to identify where pesticides were applied in relation to residential locations. Ward et al. (2000) performed a feasibility study to determine whether Landsat satellite imagery could be used to reconstruct historical crop patterns in eastern Nebraska to support pesticide exposure studies. USDA Farm Service Agency (FSA) records from 1984 were used to classify a 1984 Landsat image acquired in late summer and to validate the resulting land/crop map. General land cover types (rangeland and bare soil) and four crop types (corn, sorghum, soybeans, and alfalfa) were identified with an overall accuracy of 78%. Individual land cover accuracies ranged from 68% to 96%. They noted that the use of a second Landsat image from early spring would have improved classification accuracy of some of the crops (alfalfa, soybeans, and winter wheat). The FSA records were found to be useful, although one of the three counties they contacted had not retained the records. They found that most FSA records were only available back to the early 1980s. They concluded that historical crop maps could be constructed using a combination of Landsat imagery and FSA records, enabling epidemiologists to obtain information on possible pesticide levels in house dust.

Landsat imagery was also used by Ward et al. (2006) to determine whether crop maps were useful for predicting residential levels of crop herbicides in Iowa. Spring and late summer Landsat images for 1998, 1999, and 2000 were used to map crop types in central Iowa. A supervised classification approach was used to map corn, soybeans, and 'other' crops. A sample of crop records was obtained from USDA FSA records and used to train the classification algorithm. Validation of the maps was performed by manually viewing and interpreting slides of aerial photographs and comparing them to the digital crop map (J. Giglierano, personal communication). Accuracy ranged from 81% to 96% for corn and 87 to 93% for soybeans. Ward et al. (2006) concluded that satellite-based crop maps may be useful for estimating levels of herbicides in homes near crop fields, serving as surrogate measure of potential exposure to agricultural pesticides. The authors noted that additional research was needed to evaluate other factors, such as meteorological conditions and pesticide transport associated with wind-blown aerosols and soil (secondary drift), that might be associated with herbicide concentrations.

Studies That Used Land Cover Products Derived From Remotely Sensed Data

Two papers were found that used existing land cover maps derived from remotely sensed imagery. Brody et al. (2002) used land cover maps derived from aerial photographs for 1951, 1971, 1984, and 1990 (MacConnell et al., 1991) to study the relationship between pesticide exposure and breast cancer in Cape Cod, Massachusetts. Brody et al. (2002) used the maps to identify the source of pesticides. Twenty-six land use types associated with pesticide use were identified, including wetlands, cranberry bogs, trees, golf courses, railroad rights, and forest. Location of forest buffers was important to the study to identify potential barriers to pesticide drift. They found the only crop type identifiable from the aerial photographs was cranberry bogs; other crops could not be differentiated and were therefore lumped together as 'other agriculture.' The land use map was integrated with pesticide use in a GIS model to estimate the relative intensity of past exposures at each participant's home over a 40-year span.

Ayotte et al. (2006) used the U.S. Geological Survey (USGS) National Land Cover Dataset (NLCD) 1992 map derived from Landsat satellite imagery (Vogelmann et al., 2001) to locate agricultural land in a study investigating the relationship between bladder cancer and arsenic. The map was used in a process-based model to predict the probability of arsenic exceeding 5 mg/l in drinking water wells in New England bedrock aquifers. Arsenic was measured in water supplies of current residences and predicted in water supplies of past residences. The prediction model was developed by initially testing 50 potential explanatory variables covering geologic and anthropogenic sources of arsenic, geochemical processes, and hydrogeologic and land use factors to identify the major sources of arsenic. The authors tested several land use categories from the NLCD map including developed land, agricultural land, forest, vegetative covers, and water/wetlands. The land use data were evaluated as percentages in a 1-km radius around each well. The land cover information was not found to be useful in the prediction model, which the authors found surprising.

Discussion

We were surprised to find only five papers in the literature that used remotely sensed data or land cover maps derived from remotely sensed data for land surface environmental exposure assessment in cancer research. A vast amount of remotely sensed imagery exists to support cancer epidemiological studies (Boscoe et al., 2004), yet few cancer epidemiologists are using this technology in their research. This supports a recent literature review of 51 publications (1970–2004) by Herbreteau et al. (2007), in which 86 papers were found that used remotely sensed data for environmental epidemiology research. All the papers dealt with infectious disease research except one, which was identified as a pesticide exposure study.

We suspect the lack of applications of land surface remotely sensed data by cancer epidemiologists may be partly a result of factors such as knowledge of what remotely sensed data sets are available, data accessibility, data cost, or knowledge of the types of information that can be derived from the imagery. Our discussion below focuses on addressing some of these issues. Boscoe et al. (2004) give an overview and discussion on a wide variety of spatial data useful for geospatial analysis of cancer with only a brief discussion of remotely sensed data. We present a more detailed discussion on remotely sensed imagery most commonly used by cancer epidemiologists: aerial photography and Landsat satellite imagery. We discuss the strengths and limitations of each and highlight issues that epidemiologists should be aware of when using this imagery or land cover products derived from this imagery.

About 100 operating satellites are currently dedicated to Earth observation (Gail, 2007; accessed June 2008 http://www.asprs.org/news/satellites/satellites.html or http://directory.eoportal.org/d_art.php?type = 30700000), and image archives continue to grow at several terabytes per day (Datcu et al., 2007). Cancer studies that require land cover information generally need high-resolution (<10 m) or mid-resolution (10–100 m) imagery with a long-term image archive (20+ years). Image data sets of most interest to epidemiological research are acquired from aerial photography and sensors located on satellites (Boscoe et al., 2004). Aerial photography and the Landsat satellite data are the most common and arguably the most appropriate image products (Nuckols et al., 2004); therefore, we will focus primarily on those data in this section.

Aerial Photographs

Aerial photographs have been collected periodically since the 1930s and can be obtained from several state and federal government, university, and private organization Internet sites (Table 2). The first systematic collection of aerial photographs was performed by the USDA Farm Service Agency in 1937 for the purpose of monitoring agricultural lands. Approximately 970,000 km^2 of agricultural lands were covered in 1937, including the California Central Valley, southern Piedmont, upper Mississippi Valley, portions of the Dakotas, Nebraska, Kansas, Minnesota, Iowa, and Missouri (Monmonier, 2002). By 1947, aerial photography had been collected for 90% of the agricultural lands in the United States (Monmonier, 2002). Some aerial photography collections exist between 1950 and 1980 on a regional basis, yet systematic collection for the nation did not occur until 1980. From 1980 forward, federal and state governments combined their resources to acquire aerial photographs over the conterminous United States. Three major acquisition efforts have taken place since 1980: the National High Altitude Photography (NHAP) Program, National Aerial Photography Program (NAPP), and the National Agriculture Imagery Program (NAIP). Acquisitions were targeted for 5- to 7-year repeat coverage but may be less due to budget constraints. NHAP photograph dates range from

Table 2. Aerial photograph data.

Product	Years	Resolution	
		Spatial	Spectral
Historical Photos	1939-current	Varied	varied
National High Altitude Photography (NHAP)	1980–1987	2 m	BW, CIR
National Aerial Photograph Program (NAPP)	1987–present	1 m	BW, CIR
National Agriculture Imagery Program (NAIP)	2003–present	1m and 2m	Color, CIR

Selected websites:
Earth explorer: http://earthexplorer.usgs.gov.
Geospatial data gateway: http://datagateway.nrcs.usda.gov/GatewayHome.html.
Global Visualization Viewer (GloVis): http://glovis.usgs.gov.
Historic Aerials: http://www.historicaerials.com/.
Illinois Natural Resources Geospatial Clearinghouse: http://www.isgs.uiuc.edu/nsdihome/webdocs/ilhap/.
Iowa Geographic Map Server: http://ortho.gis.iastate.edu/.
The National Archives: http://www.archives.gov/index.html.
USDA Farm Service Agency: http://www.fsa.usda.gov/.

1978 to 1988 and are available in black and white (B&W) and color infrared (CIR) at approximately 2-m spatial resolution. NAPP photographs range from 1987 to 2003 and are available in B&W and CIR at approximately 1-m spatial resolution. NAIP photographs range from 2003 to present and are available in natural color and CIR at about 1-m and 2-m spatial resolution.

Handling and processing aerial photographs is much easier if they are obtained as georeferenced digital image files. Aerial photographs, especially older photographs, may not be available in digital georeferenced format. O'Leary et al. (2004) used printed photographs in their study, noting the difficulty of having to use a magnifying glass to identify landscape features. Georeferencing aerial photographs can be a difficult and time-consuming process. Many of the newer aerial photographs (NHAP, NAPP, and NAIP) are available as Digital Orthophoto Quads (DOQs). DOQs are aerial photographs that have been scanned, corrected for camera orientation and terrain effects (i.e., orthorectified), and georeferenced. Some of the earlier photographs are also becoming available as DOQs (e.g., Iowa at http://ortho.gis.iastate.edu/ or Minnesota at http://www.dnr.state.mn.us/airphotos/index.html).

Aerial photographs are useful for documenting historical land use change at the local level (i.e., around individual residences). Their high spatial resolution (1–2 m) makes features such as buildings, roads, and trees fairly easy to indentify (Figure 1). Using natural color or color infrared photographs can greatly enhance the ability to differentiate land cover and land use (Figure 2). Specific crop types are more difficult to identify from photographs due to the lack of spectral and temporal information. Aerial photographs are typically collected only once over the entire year and the acquisition date is not generally optimized for differentiating individual crop types. The earliest photographs were collected in leaf-off conditions to optimize soil type discrimination. The more current NAIP photographs are collected in leaf-on conditions and available in natural color and color infrared, which may allow for some distinction between crop types.

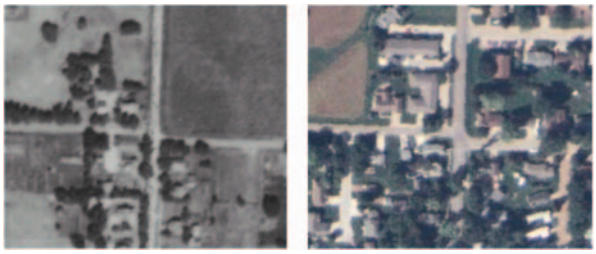
1930s 2006

Figure 1. Comparison of 1930s black and white aerial photograph and 2006 natural color aerial photograph in northwest Iowa.

Landsat Satellite Images

Landsat images have been continually collected by Earth-orbiting satellites from space since 1972. Landsat sensors collect data on a 16- to 18-day repeat cycle and constitute the longest continuous record of the Earth's surface as seen from space (Table 3). Landsat sensor characteristics have changed as technology has improved. Early satellites in the series (Landsats 1–5) carried only a 4-band Multispectral Scanner sensor with a 57-m spatial resolution, whereas later satellites carried the 6-band Thematic Mapper (TM) and 7-band Enhanced TM Plus (ETM+) sensors with 30-m spatial resolution. Each Landsat scene spans an area about 170 km by 185 km.

Landsat data are easily accessible through several websites, such as the USGS Global Visualization Viewer (http://glovis.usgs.gov/) and the Michigan State University Global Observatory for Ecosystem Services (http://www.landsat.org).


2002 Black and White 2007 Natural Color 2007 Color Infrared

Figure 2. Comparison of black and white, natural color, and color infrared aerial photographs in northwest Iowa.

Table 3. Landsat satellite data.

Satellite	Sensor	Years	Spatial resolution	Spectral bands
Landsat 1	MSS	1972–1978	57 m	2 visible, 2 near infrared
Landsat 2	MSS	1975–1982	57 m	2 visible, 2 near infrared
Landsat 3	MSS	1978–1983	57 m	2 visible, 2 near infrared
Landsat 4	TM	1982–2001	30 m (reflective bands) 120 m (thermal band)	3 visible, 1 near infrared, 2 middle infrared, 1 thermal
Landsat 5	TM	1984–	30 m (reflective bands)	3 visible, 1 near infrared, 2 middle infrared, 1 thermal
Landsat 7	ETM +	1999–	15 m (panchromatic) 30 m (reflective bands) 60 m (thermal bands)	3 visible, 1 near infrared, 2 middle infrared, 2 thermal 1 panchromatic

Selected websites:
Earth Explorer: http://earthexplorer.usgs.gov.
Global Visualization Viewer (GloVis): http://glovis.usgs.gov.
Michigan State Observatory for Ecosystem Services: http://www.landsat.org/.
Multi-Resolution Land Characteristics Consortium: http://www.mrlc.gov.

Also, organizations have joined together to purchase Landsat imagery and have set up websites to make the data freely available, for example, the Multi-Resolution Land Characteristics Consortium (http://www.mrlc.gov/) and the individual websites of the AmericaView program (e.g., http://sdview.sdstate.edu/ for South Dakota Landsat data). The cost of Landsat data has varied widely over its lifetime (up to several thousand dollars per image). USGS recently announced that all Landsat imagery (current and archived) will be available at no cost by early 2009 (USGS, 2008), thus eliminating one of the barriers to using this data for cancer research.

Landsat data are ideal for many epidemiological applications due to the substantial archive of imagery now available back to 1972. Issues with cost and image preparation will soon be minimized due to the new USGS policy to offer the data at no cost in a standard georectified format beginning in January 2009. The multispectral and multitemporal features of Landsat data allow landscape features to be distinguished by their spectral and phenological characteristics compared with aerial photographs, which have limited spectral and temporal information. Crop type mapping using Landsat imagery has been performed since the mid-1970s (Landgrebe, 1997). Identifying where specific crops were grown was needed for several of the studies we reviewed (Ward et al., 2000; Xiang et al., 2000; Brody et al., 2002; Meyer et al., 2006; Ward et al., 2006) to determine the source of agricultural contaminants. The multispectral/multitemporal capability of the Landsat data allows for effective crop type mapping in many regions of the United States, although distinguishing crops may still be difficult in some areas depending on crop management practices, such as in regions where a wide variety of crops are grown or multiple crops are harvested over the growing season.

Visually interpreting Landsat imagery can be more challenging than aerial photographs. Small features, such as houses and roads, are not easily recognized by visual examination of the image due to the larger spatial resolution of Landsat (30–60 m) (Figure 3). Converting Landsat imagery to land cover maps can be a complex process and is primarily performed by remote sensing experts using specialized computer software.

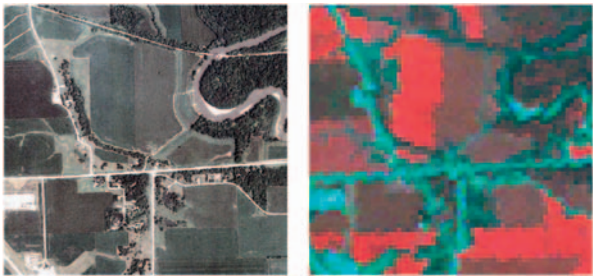

Figure 3. Comparison of features identifiable in an aerial photograph and Landsat image. The aerial photograph was collected during the summer of 2005 and is shown in natural color. The Landsat image was collected on 9 August 2005, and is shown in color infrared. Bright red tones in the agricultural fields in the Landsat image indicate soybeans and medium red tones indicate corn fields.

Other Remotely Sensed Data

There are many other types of aerial photograph and satellite data available. We only discussed a few of the most common imagery data. Other Landsat-like satellite data (i.e., 10–60 m) are also available, such as Terra Advanced Spaceborne Thermal Emission and Reflection Radiometer and Indian Remote Sensing (IRS) Advanced Wide Field Sensor (AWiFS). These data have limited temporal and spatial coverage, either in terms of historical availability or repeat time over the growing season. These data may be useful to fill in gaps in areas where additional data are needed to support land cover determination or if Landsat data are unavailable due to cloud cover or equipment failure. Coarser spatial resolution imagery (e.g., >250 m) may also be useful in some circumstances, such as identifying general soil or vegetation properties over the growing season, as suggested by Ward et al., 2006, or monitoring the spatial distribution of smoke plumes. These data are typically collected once or twice daily and some archives extend back to 1980. The Advanced Very High Resolution Radiometer (1 km) and Moderate Resolution Imaging Spectroradiometer (250 m to 1 km) data are examples of this type of imagery.

Characterizing structural information may be important for some epidemiological studies, such as those incorporating fate and transport GIS modeling (e.g., Brody et al., 2002). Most remotely sensed imagery collects data only in two-dimensional space (e.g., ground area). The LiDAR sensor collects data vertically (http://lidar.cr.usgs.gov/) allowing for characterization of height and density of land surface objects (e.g., trees, buildings). Lidar data are currently being collected over many areas of the United States through aircraft and are accessible from the USGS (http://lidar.cr.usgs.gov/).

Land Cover Maps Derived from Remotely Sensed Data

Existing land cover maps can be very useful for epidemiological research. The maps are generally free and can be downloaded from many Internet sites. Considerable expense and effort has gone into the production of the maps; the process of converting raw image data to useful information is complex and generally requires remote sensing expertise (Herbreteau et al., 2007; Jenson, 2007). The quality of maps derived from remotely sensed imagery depends on the interpreter's experience, the interpreter's knowledge of the landscape features, and the quality of the images being used. The greatest expense in using remotely sensed data is the need for a well-trained image analyst, not the cost of the data (Jenson, 2007).

In general, maps produced from aerial photographs, such as used by Brody et al. (2002), have been derived by a manual interpretation process in which a human directly views the image. Manual interpretation was also used by O'Leary et al. (2004) and Ward et al. (2006) to identify individual residences and general land use classes. Maps produced from aerial photographs are generally regarded as being a fairly accurate representation of landscape features at the local scale (i.e., residential level). In fact, manual interpretations of aerial photographs are often used to validate land cover maps produced from lower resolution satellite imagery. For example, aerial photographs were interpreted manually by remote sensing analysts and used to assess the accuracy of the NLCD map, which was produced with 30-m Landsat imagery (Zhu et al., 2000).

Products such as the NLCD and CDL apply a digital interpretation process as opposed to a manual method generally applied to high-resolution remotely sensed data. Landsat imagery contains multiple spectral bands of image data, and several acquisitions of imagery over the growing season may be required to discriminate land cover types making visual interpretation impractical. A person can view an aerial photograph and identify features using characteristics such as shape, size, texture, color, shadow, and association. However, programming a computer to recognize these features is difficult. Several websites offer tutorials on image classification for readers who are interested in more detailed discussions on the topic (e.g., http://rst.gsfc.nasa.gov/, http://www.ccrs.nrcan.gc.ca/resource/tutor/fundam/chapter4/07_e.php).

Two of the papers from our literature review used digital land cover maps derived from aerial photographs (Brody et al., 2002) or Landsat satellite imagery (Ayotte et al., 2006). Brody et al. (2002) used a series of high-resolution maps derived from aerial photographs for 1951, 1971, 1984, and 1990 for Massachusetts (http://www.mass.gov/mgis/lus.htm; accessed June 2008). Historical maps such as these are relatively rare due to the large effort necessary to collect, process, and interpret the photographs. MacConnell (1975) estimated that for each photograph, 150 skilled man-hours were required for photo interpretation, cartography, and area measurement calculations. With 189 USGS maps needed to cover Massachusetts, about 14 man-years were required to produce one statewide map. Maps derived from aerial photographs have been produced by the federal government for national surveys of soil (USDA Natural Resources Conservation Service, Soil Survey Geographic database; http://soils.usda.gov/survey/geography/ssurgo/; accessed July 2008) and wetlands (USFWS National Wetlands Inventory; http://www.fws.gov/nwi/; accessed July 2008).

Land cover maps produced from Landsat satellite imagery are more common than those derived from aerial photographs. Maps derived from Landsat imagery are available at the regional (e.g., Platte River Basin; Dappen and Tooze, 2001), state (e.g., Gap Analysis Program; USGS National Gap Analysis Program, 2004; http://gapanalysis.nbii.gov/), multistate (e.g., Southwest Regional Gap Analysis; http://ftp.nr.usu.edu/swgap/landcover.html and USDA CDL; http://www.nass.usda.gov/research/Cropland/sarsexample3.html), and national (e.g., USGS NLCD; http://edc.usgs.gov/products/landcover.html) level. The USGS produced a series of national (1992 and 2001) land cover maps for the conterminous United States from Landsat imagery. The 1992 NLCD map was used by Ayotte et al. (2006) to study the relationship between land use near well locations (1 km radius) and arsenic concentrations. These maps have about 20 general land cover types such as urban, deciduous forest, conifer forest, and grassland. Agricultural land was classified as row crops, small grains, and fallow in the 1992 map, with agricultural crops being lumped into one 'cultivated cropland' class in the 2001 map. These maps were developed to provide a consistent land cover product for the entire United States for regional and national scale applications (Homer et al., 2007). Although it may be tempting to use the NLCD for local-scale studies such as Ayotte et al. (2006), the NLCD authors state that the map was not designed for county-level and other localized applications (Homer et al., 2007).

The USDA CDL maps are also targeted for county level and larger applications. These crop type maps have been produced by USDA NASS for many states since 1997 for the purpose of estimating crop acreages at the county level (http://www.nass.usda.gov/research/Cropland/sarsexample3.html#Newer; accessed June 2008). The most recent maps for 2007 are available for 22 states. The maps produced from 1997 to 2006 were derived from Landsat imagery, whereas IRS AWiFS imagery (60 m) was used to produce more recent CDL maps (i.e., 2007 and forward). Meyer et al. (2006) used CDL maps from 1997 to 2002 maps in eastern Arkansas in a GIS-based exposure model to identify specific crop types near residential locations (500-m buffer) in a study of hypospadias. Meyer et al. (2006) noted that limitations in their methods may have produced non-differential exposure misclassification and biased the risk estimates most likely toward the null. The authors noted that the use of CDL maps did, however, provide a means for estimating pesticide use on a scale smaller than state or county boundaries.

Cancer epidemiologists need to be cautious when using existing land cover maps for point location analysis. All of the epidemiological papers we reviewed were studies at the residential level. Specific point location and immediately surrounding landscape (e.g., <1 km) were the most critical in terms of identifying sources of contaminants and modeling fate and transport. In applications such as individual exposure assessment, it is particularly important to create the highest classification accuracy at the specific location where the exposure is to be predicted. Land cover maps, such as the NLCD and CDL, may be useful in some cases, but they should be closely evaluated at selected locations within the study to ensure accuracy is acceptable (Thogmartin et al., 2004; Maxwell et al., 2008). As a general rule, validating historical exposure measures created from remotely sensed data is a challenge because historical validation data seldom exist. Validating current predictions may serve as a substitute, but epidemiologists must be aware that the validity of the presumed exposures may depend on the use of monocultures, crop rotations, second season plantings, and variability in pesticide use patterns across regions over time. Remotely sensed data should not be viewed as a panacea to challenges of exposure assessment; effort will still be required to substantiate the exposure model.

Summary

Satellite data providers recognize the importance of remotely sensed imagery for human health and have identified the link between environmental factors and human health and well-being as one of the initial areas to address (GEO, 2005). The vast array of satellite imagery data clearly is an insufficiently tapped resource for applications in exposure assessment and long-term cancer studies. Use of these data sets is impeded by multiple technical challenges including cost, accessibility, quality control, and a number of technical details including formats, georegistration, mosaicing, ortho-rectification, cloud cover, large volume data management, and access to the tools necessary for managing, processing, and analyzing these tremendously large multitemporal multiresolution data sets consisting of dozens to thousands of images. Although some of these issues are currently being addressed, such as the increased availability of aerial photographs and Landsat satellite imagery through Internet providers at no cost, challenges still exist in converting the raw data to useful information. The most important advances in remote sensing may be in how we derive knowledge from sensor data, not from advances in sensor technology (Gail, 2007). In addition, developing appropriate GIS models to generate individual-level exposure estimates from a combination of land use maps, environmental contaminants (e.g., pesticides, nitrates in drinking water), and other environmental factors, such as atmospheric conditions, topography, and soil properties, presents further challenges to the effective application of remotely sensed data to environmental exposure assessment applications.

Acknowledgements

We thank Carol Deering and Eric Wood for assisting with the library searches and collecting the literature for this study. This work was performed under BioMedware contract HHSN261200700061C and USGS contract 03CRCN0001.

References

Atkinson S.F., Waller W.T., and Crooks T.J. Assessing atrazine pollution potential to a drinking water reservoir using remote sensing and geographic information system modeling. *Aquatic Ecosystem Health and Management* 2001: 4: 327–338.

AvRuskin G.A., Jacquez G.M., Meliker J.R., Slotnick M.J., Kaufmann A.M., and Nriagu J.O. Visualization and exploratory analysis of epidemiologic data using a novel space time information system. *Int J Health Geogr* 2004: 3: 26.

Avruskin G.A., Meliker J.R., and Jacquez G.M. Using Satellite Derived Land Cover Information for a Multi-Temporal Study of Self-Reported Recall of Proximity to Farmland. *J Expo Sci Environ Epidemiol* 2008: 18: 381–391.

Ayotte J.D., Nolan B.T., Nuckols J.R., and Cantor K.P. Modeling the probability of arsenic in groundwater in New England as a tool for exposure assessment. *Environ Sci Technol* 2006: 40: 3578–3585.

Bassani C., Cavalli R.M., Cavalcante F., Cuomo V., Palombo A., Pascucci S., and Pignatti S. Deterioration status of asbestos-cement roofing sheets assessed by analyzing hyperspectral data. *Remote Sensing of Environment* 2007: 109: 361–378.

Beyea J., and Hatch M. Geographic exposure modeling: a valuable extension of geographic information systems for use in environmental epidemiology. *Environ Health Perspect* 1999: 107(suppl 1): 181–190.

Boscoe F.P., Ward M.H., and Reynolds P. Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer. *Int J Health Geogr* 2004: 3: 28.

Brody J.G., Vorhees D.F., Melly S.J., Swedis S.R., Drivas P.J., and Rudel R.A. Using GIS and historical records to reconstruct residential exposure to large-scale pesticide application. *J Expo Anal Environ Epidemiol* 2002: 12: 64–80.

Dappen P., and Tooze M. *Delineation of Land Use Patterns for the Cooperative Hydrology Study in The Central Platte River Basin*, University of Nebraska-Lincoln, Center for Advanced Land Management Information Technologies Report 2001 73p.

Datcu M., D'Elia S., King R.L., and Bruzzone L. Introduction to the special section on image information mining for Earth observation data. *IEEE Trans Geosci Remote Sens* 2007: 45: 795–798.

Elliott P., Briggs D., Morris S., de Hoogh C., Hurt C., Jensen T.K., Maitland I., Richardson S., Wakefield J., and Jarup L. Risk of adverse birth outcomes in populations living near landfill sites. *BMJ* 2001: 323: 363–368.

Gail W.B. Remote sensing in the coming decade: the vision and the reality. *J Applied Remote Sensing* 2007: 1: 1–19.

Galton A Desiderata for a spatio-temporal geo-ontology. In: Kuhn W., Worboys MF., Timpf S., (Eds) *Spatial Information Theory: Foundations of Geographic Information Science, Lecture Notes in Computer Science* Volume 2825/2003 Springer-Verlag: Berlin, 2003, 1–12.

Garland C.F., Mohr S.B., Gorham E.D., Grant W.B., and Garland F.C. Role of ultraviolet B irradiance and vitamin D in prevention of ovarian cancer. *Am J Prev Med* 2006: 31: 512–514.

GEO. 2005. *Global Earth Observing System of Systems (GEOSS) - 10-year Implementation Plan Reference Document*, Tech. Rep. GEO 1000R/ESA Publications Division, The Netherlands, 210p; http://earthobservations.org/documents/10-Year%20Plan%20Reference%20Document.pdf; accessed July 2008).

Guidry V.T., and Margolis L.H. Unequal respiratory health risk: Using GIS to explore hurricane related flooding of schools in Eastern North Carolina. *Environ Res* 2005: 98: 383–389.

Guo D., Chen J., MacEachren A.M., and Liao K. A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP). *IEEE Trans Vis Comput Graph* 2006: 12: 1461–1474.

Herbreteau V., Salem G., Souris M., Hugot J., and Gonzalez J. Thirty years of use and improvement of remote sensing, applied to epidemiology: from early promises to lasting frustration. *Health Place* 2007: 13: 400–403.

Homer C., Dewitz J., Fry J., Coan M., Hossain N., Larson C., Herold N., McKerrow A., VanDriel J.N., and Wickham J. Completion of the 2001 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* 2007: 73: 337–341.

Jacquez G.M., Greiling D., and Kaufmann A. Design and implementation of space time information systems. *J Geogr Syst* 2005: 7: 7–24.

Jenson F.R. *Remote sensing of the Environment: an Earth Resource Perspective*. 2nd edn. Prentice Hall: Upper Saddle River, NJ, 2007.

Landgrebe D. The evolution of Landsat data analysis. *Photogrammetric Engineering and Remote Sensing* 1997: 63: 859–867.

MacConnell W.P., Goodwin D.W., and Jones K.M.L. Land-use update for Cape Cod with area statistics for 1971, 1984 and 1990. *Research Bulletin Number 737*. Massachusetts Agricultural Experiment Station, University of Massachusetts at Amherst, College of Food and Natural Resources, 1991.

MacConnell W.P. Remote sensing 20 years of change in Massachusetts 1951/52–1971/72. *Research Bulletin Number 630*. Massachusetts Agricultural Experiment Station, University of Massachusetts at Amherst, College of Food and Natural Resources, 1975.

Marusek J.C., Cockburn M.G., Mills P.K., and Ritz B.R. Control selection and pesticide exposure assessment via GIS in prostate cancer studies. *Am J Prev Med* 2006: 30(2 Suppl): S109–S116.

Mather F.J., Whited L.C., Langlois E.C., Shorter C.F., Swalm C.M., Shaffer J.G., and Harley W.R. Statistical methods for linking health, exposure and hazards. *Environ Health Perspect* 2004: 112: 1440–1445.

Matthijsen J., Slaper H., Reinen H.A.J.M., and Velders G.J.M. Reduction of solar UV by clouds: a comparison between satellite-derived cloud effects and ground-based radiation measurements. *J Geophys Res-Atmospher* 2000: 105: 5069–5080.

Maxwell S.K., Wood E.C., and Janus A. Comparison of the USGS 2001 NLCD to the 2002 USDA Census of Agriculture for the Upper Midwest United States. *Agric Ecosyst Environ* 2008: 127: 141–145.

Maxwell S.K., Nuckols J.R., and Ward M.H. A method for mapping corn using the US Geological Survey 1992 National Land Cover Dataset. *Computers and Electronics in Agriculture* 2006: 51: 54–65.

McNamee R., and Dolk H. Does exposure to landfill waste harm the fetus? *BMJ* 2001: 323: 351–352.

Meliker J.R., Slotnick M.J., AvRuskin G.A., Kaufmann A., Fedewa S.A., Goovaerts P., Jacquez G.M., and Nriagu J.O. Individual lifetime exposure to inorganic arsenic using a Space-Time Information System. *Int Arch Occ Env Health* 2007: 80: 184–197.

Meliker J.R., Slotnick M.J., AvRuskin G.A., Kaufmann A., Jacquez G.M., and Nriagu J.O. Improving exposure assessment in environmental epidemiology: Application of spatio-temporal visualization tools. *J Geogr Syst* 2005: 7: 49–66.

Meyer K.J., Reif J.S., Veeramachaneni D.N.R., Luben T.J., Mosley B.S., and Nuckols J.R. Agricultural pesticide use and hypospadias in Eastern Arkansas. *Environ Health Perspect* 2006: 114: 1589–1595.

Mohr S.B., Garland C.F., Gorham E.D., Grant W.B., and Garland F.C. Is ultraviolet B irradiance inversely associated with incidence rates of endometrial cancer: an ecological study of 107 countries. *Prev Med* 2007: 45: 327–331.

Mohr S.B., Gorham E.D., Garland C.F., Grant W.B., and Garland F.C. Are low ultraviolet B and high animal protein intake associated with risk of renal cancer? *Int J Cancer* 2006: 119: 2705–2709.

Monmonier M. Aerial photography at the agricultural adjustment administration: acreage controls, conservation benefits, and overhead surveillance in the 1930s. *Photogrammetric Engineering and Remote Sensing* 2002: 68: 1257–1261.

Nieuwenhuijsen M., Paustenbach D., and Duarte-Davidson R. New developments in exposure assessment: The impact on the practice of health risk assessment and epidemiological studies. *Environ Intl* 2006: 32: 996–1009.

Nuckols J.R., Ward M.H., and Jarup L. Using Geographic Information Systems for Exposure Assessment in Environmental Epidemiology Studies. *Environ Health Persp* 2004: 112: 1007–1015.

O'Leary E.S., Vena J.E., Freudenheim J.L., and Brasure J. Pesticide Exposure and Risk of Breast Cancer: a nested Case-Control Study of Residentially Stable Women Living on Long Island. *Environ Res* 2004: 94: 134–144.

Piacentini R.D., Cede A., Luccini E., and Stengel F. Determination of the UV solar risk in Argentina with high-resolution maps calculated using TOMS ozone climatology. *Adv Space Res* 2004: 34: 2215–2220.

Pickle L., Waller L., and Lawson A.B. Current Practices in cancer spatial data analysis: a call for guidance. *Int J Health Geogr* 2005: 4: 3.

Pickle L.W., Szczur M., Lewis D.R., and Stinchcomb D.G. The crossroads of GIS and health information: a workshop on developing a research agenda to improve cancer control. *Int J Health Geog* 2006: 5: 51.

Reif J.S., Burch J.B., Nuckols J.R., Metzger L., Ellington D., and Anger W.K. Neurobehavioral effects of exposure to trichloroethylene through a municipal water supply. *Environ Res* 2003: 93: 248–258.

Reynolds P., Hurley S.E., Gunier R.B., Yerabati S., Quach T., and Hertz A. Residential proximity to agricultural pesticide use and incidence of breast cancer in California, 1988–1997. *Environ Health Perspect* 2005: 113: 993–1000.

Riggs P.D Assessing multiple geospatial modeling techniques of assigning pesticide exposure in the California central valley. [PhD Thesis]. Colorado State University: Fort Collins, CO, 2007.

Rull R.P., Ritz B., and Shaw G.M. Validation of self-reported proximity to agricultural crops in a case-control study of neural tube defects. *J Expo Sci Environ Epidemiol* 2006: 16: 147–155.

Ryan P.H., and LeMasters G.K. 2007. A review of land-use regression for characterizing intraurban air models pollution exposure. *Inhal Toxicol* 2007: 19: 127–133.

Slaper H., Velders G.J.M., and Matthijsen J. Ozone depletion and skin cancer incidence: A source risk approach. *J Hazard Mater* 1998: 61: 77–84.

Solomon C.C., White E., Kristal A.R., and Vaughan T. Melanoma and lifetime UV radiation. *Cancer Causes Control* 2004: 15: 893–902.

Tatalovich Z., Wilson J.P., Mack T., Yan Y., and Cockburn M. The objective assessment of lifetime cumulative ultraviolet exposure for determining melanoma risk. *J Photochem Photobiol B* 2006: 85: 198–204.

Thogmartin W.E., Gallant A.L., Knutson M.G., Fox T.J., and Suarez M.J. Commentary: A cautionary tale regarding use of the National Land Cover Dataset 1992. *Wildlife Society Bulletin* 2004: 32: 970–978.

US Geological Survey. Imagery for everyone… timeline set to release entire USGS Landsat archive at no charge, Technical Announcement April 21, 2008. Available: http://landsat.usgs.gov/images/squares/USGS_Landsat_Imagery_Release.pdf [accessed July 2008]. 2008.

US Geological Survey National Gap Analysis Program. *Provisional Digital Land Cover Map for the Southwestern United States*. Version 1.0. RS/GIS Laboratory, College of Natural Resources, Utah State University, 2004. Available: http://ftp.nr.usu.edu/swgap/swregap_landcover_report.pdf [accessed July 2008].

Vogelmann J.E., Howard S.M., Yang L., Larson C.R., Wylie B.K., and Van Driel J.N. Completion of the 1990's National Land Cover Data Set for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* 2001: 67: 650–662.

Ward M.H., Lubin J., Giglierano J., Colt J.S., Wolter C., Bekiroglu N., Camann D., Hartge P., and Nuckols J.R. Proximity to crops and residential exposure to agricultural herbicides in Iowa. *Environ Health Perspect* 2006: 114: 893–897.

Ward M.H., Nuckols J.R., Weigel S.J., Maxwell S.K., Cantor K.P., and Miller R.S. Identifying populations potentially exposed to agricultural pesticides using remote sensing and a geographic information system. *Environ Health Perspect* 2000: 108: 5–12.

Weaver C., Fyfe D., Robinson A., Holdsworth D., Peuquet D., and MacEachren A.M. Visual Analysis of Historic Hotel Visitation Patterns. *IEEE Trans Vis Comput Graph* 2006: 12: 35–42.

Xiang H.Y., Nuckols J.R., and Stallones L.A. Geographic information assessment of birth weight and crop production patterns around mother's residence. *Environ Res* 2000: 82: 160–167.

Yu C.L., Wang S.F., Pan P.C., Wu M.T., Ho C.K., Smith T.J., Li Y., Pothier L., and Christiani D.C. Residential exposure to petrochemicals and the risk of leukemia: Using Geographic Information System Tools to estimate individual-level residential exposure. *Am J Epidemiol* 2006: 164: 200–207.

Zhu Z., Yang L., Stehman S.V., and Czaplewski R.L. Accuracy assessment for the US Geological Survey regional land cover mapping program: New York and New Jersey Region. *Photogrammetric Engineering and Remote Sensing* 2000: 66: 1425–1435.