Journal of Hydrology 370 (2009) 100–108

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Assigning land use to supply wells for the statistical characterization of regional groundwater quality: Correlating urban land use and VOC occurrence

Tyler D. Johnson *, Kenneth Belitz

US Geological Survey, 4165 Spruance Rd., Suite 200, San Diego, CA 92101, USA

ARTICLE INFO

Article history:
Received 24 January 2008
Received in revised form 18 February 2009
Accepted 19 February 2009

This manuscript was handled by L. Charlet, Editor-in-Chief, with the assistance of Dammia Barcelo, Associate Editor

Keywords:
Groundwater
Land use
Water quality
VOC
Geostatistics
Autocorrelation

SUMMARY

Many national and regional groundwater studies have correlated land use "near" a well, often using a 500 m radius circle, with water quality. However, the use of a 500 m circle may seem counterintuitive given that contributing areas are expected to extend up-gradient from wells, and not be circular in shape. The objective of this study was to evaluate if a 500 m circle is adequate for assigning land use to a well for the statistical correlation between urban land use and the occurrence of volatile organic compounds (VOCs). Land use and VOC data came from 277 supply wells in four study areas in California. Land use was computed using ten different-sized circles and wedges (250 m to 10 km in radius), and three different-sized "searchlights" (1–2 km in length). We define these shapes as contributing area surrogates (CASs), recognizing that a simple shape is at best a surrogate for the actual contributing area. The presence or absence of correlation between land use and the occurrence of VOCs was evaluated using Kendall's tau (τ). Values of τ were within 10% of one another for wedges and circles ranging in size from 500 m to 2 km, with correlations remaining statistically significant ($p < 0.05$) for all CAS sizes and shapes, suggesting that a 500 m circular CAS is adequate for assigning land use to a well. Additional evaluation indicated that urban land use is autocorrelated at distances ranging from 8 to 36 km. Thus, urban land use in a 500 m CAS is likely to be predictive of urban land use in the actual contributing area.

Published by Elsevier B.V.

Introduction

National and regional assessments of groundwater quality have often relied upon statistical characterization (Alley, 1993; Gilliom et al., 1995). Statistical characterization can be used to summarize water quality at various scales and can be used to identify factors that affect water quality. Statistical characterization can be particularly useful in areas with multiple sources of contaminants and processes affecting water quality. These statistical approaches have used general linear modeling (Worrall et al., 2002), logistic regression (Squillace and Moran, 2007), and neuro-fuzzy models (Dixon, 2005), to name a few.

One variable that is often incorporated into statistical models is land use/land cover (hereafter collectively called land use). Land use has been shown to be one of the many factors that can affect water quality. For example, national studies conducted by the US Geological Survey (USGS) National Water Quality Assessment (NAWQA) program have shown a correlation between land use near a well and the occurrence of nutrients (Nolan et al., 2002), pesticides (Squillace et al., 2002), and volatile organic compounds (VOCs) (Moran et al., 2005; Squillace and Moran, 2007; Zogorski et al., 2006).

In these and many other NAWQA studies, land use was classified using a 500 m radius circle centered on each well. In other studies, land use has been assigned using radii ranging in size from 100 m to 3.2 km (Cain et al., 1989; Eckhardt and Stackelberg, 1995; Bruce and McMahon, 1996; Trojan et al., 2003; Aelion and Conte, 2004; Worrall and Kolpin, 2004; An et al., 2005; Gardner and Vogel, 2005). Still others have used variously sized wedges oriented up-gradient from the well to assign land use to a well (Cowdery, 1997; Gurdak and Qi, 2006).

The use of a 500 m circular radius for assigning land use to a well may be overly simplistic, given that contributing areas can extend up-gradient from the well, be irregularly shaped, and non-contiguous (see Barlow, 1994; Franke et al., 1998; Clark et al., 2007; Starn and Brown, 2007; Burow et al., 2008). Ideally, contributing areas are defined through detailed groundwater flow models that allow for a more accurate assignment of land use to a given well. Groundwater flow models can account for many of the factors affecting the size and location of a contributing area including regional flow dynamics, aquifer properties, pumping rates, and potential short-circuiting flow paths. In particular, models can be used to identify the fraction of water produced at a well that is modern verses pre-modern. This is important because it provides

* Corresponding author. Tel.: +1 619 225 6100; fax: +1 619 225 6101.
E-mail address: tyjohns@usgs.gov (T.D. Johnson).

0022-1694/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.jhydrol.2009.02.056

information on the influence of modern land use. However, development of detailed models can require considerable effort. It is for this reason that many national and regional studies rely on a 500 m circle for assigning land use to a well.

In this paper, we use the term "contributing area surrogate" (CAS) to describe the area within which land use is classified. The use of this term acknowledges the fact that a simple shape is, at best, a surrogate for the actual contributing area. The use of this term also acknowledges that land use is itself a surrogate for potential contaminating activities. A CAS is not meant to replace the delineation of a well head protection area (Aller et al., 1985; Paradis et al., 2007).

The objective of this study is to determine if a CAS with a 500 m circular radius is adequate for assigning land use to a well when correlating urban land use and VOC occurrences. The adequacy of the 500 m CAS was evaluated relative to CASs of other sizes and shapes. The results presented here are not intended to suggest that the percent of urban land use is the sole factor affecting the occurrence of VOCs. Other factors are clearly important and need to be considered when evaluating groundwater quality.

The data used for this study is from the Priority Basin Project (PBP), which is part of the California Water Board's Groundwater Ambient Monitoring and Assessment Program. The PBP is designed to provide a spatially-unbiased statistical assessment of raw groundwater quality used for public supply in California (Belitz et al., 2003). For the purposes of implementation, the "priority" groundwater basins have been aggregated into 35 study units (Fig. 1). Water quality in these study units is assessed using previously and newly acquired data. In total, approximately 2500 wells will be sampled over a period of eight years. The PBP was modeled after the NAWQA program. Numerous explanatory variables, one of which being land use, will be analyzed to determine their relation to groundwater quality. The primary goal is to assess the quality of groundwater used for public supply in California.

Background

The 500 m circle was chosen as the reference CAS because it has been used in numerous studies, and NAWQA protocols recommend it (Koterba, 1998). The basis for Koterba's (1998) recommendation is found in a literature review of six studies (some with multiple publications, see Table 8 in Koterba, 1998). However, these studies correlated mostly agricultural land use with either nitrate and/or pesticides, sampled relatively shallow wells, or did not test a wide range of CAS sizes or shapes. None of the studies explicitly correlated urban land use with VOC occurrence using deep production wells over a wide range of CAS sizes and shapes.

Ferrari and Ator (1995) found correlations between agricultural land use and nitrate concentrations within relatively shallow (<92 m deep) domestic wells using three CAS shapes: circles, upslope-oriented ¼ circles (wedges), and upslope-oriented semi-circles. Two radial distances were used for all shapes: 400 and 800 m. The authors found statistically significant correlations in all CAS sizes and shapes.

Ator and Denis (1997) evaluated the correlation between land use and nitrate concentrations within relatively shallow wells (<112 m deep) using the same CAS shapes as Ferrari and Ator (1995), however the radial distance for all shapes was 400 m. The authors found that in the three of four subareas, the wedge provided more significant correlations than the circle or semi-circle, however in one subarea, all shapes correlated equally well.

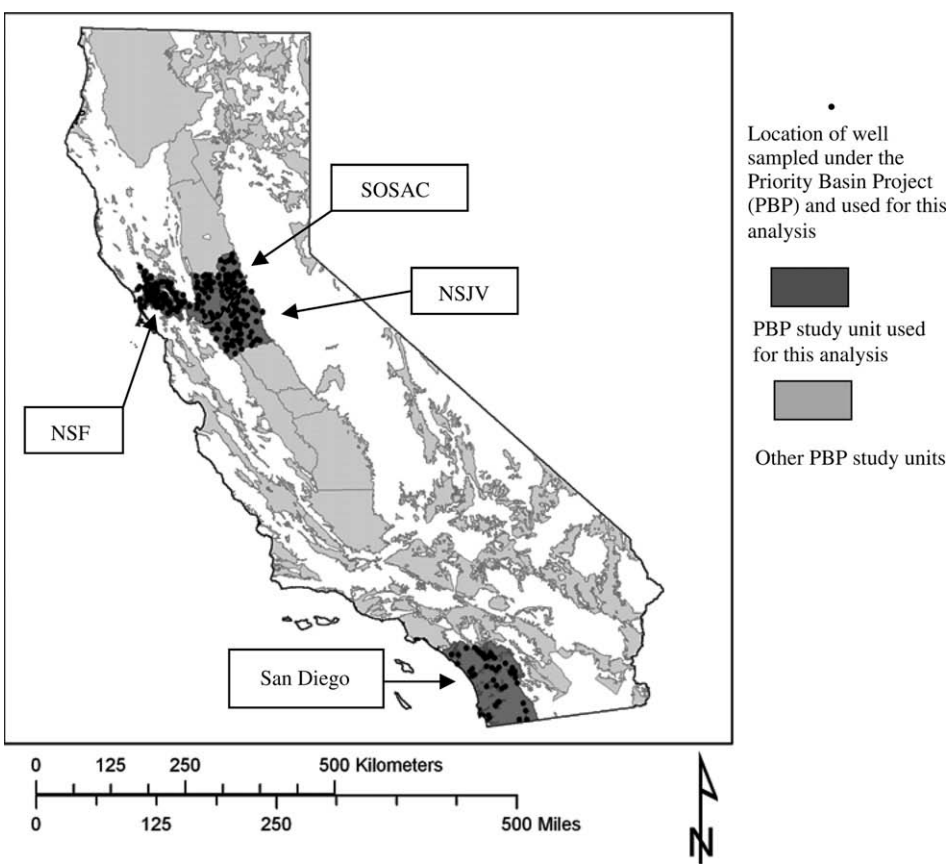

Fig. 1. Location of the four study units and the wells used for analysis.

Kolpin (1997) used perennial streams and topographic divides to delineate contributing areas within circular CASs varying in distance from 200 to 2000 m of the well; he also used the entire circle of the same radii. The author correlated the concentrations of nitrate, atrazine, and alachlor detected in shallow wells (<37 m deep) with a variety of land uses; both methods provided similar correlations.

Rupert (2003) tested 12 circular CASs (radii of 100 m to 50 km), and multiple land uses, in order to identify the most significant variables for inclusion in a logistic regression model. The author correlated the concentrations of nitrate and atrazine detected in shallow wells (median depth = 20 m) with a variety of land uses. Statistically significant correlations occurred at various CAS radii: from 100 m to 8 km for row crops (peaking at 2 km), and from 100 m to 2 km for urban land use (peaking at 500 m).

Lorenz et al. (2003) and Gurdak and Qi (2006) both used groundwater flow models to help determine the size of a wedge. Lorenz et al. (2003) identified four sizes of wedges (from 50 to 200 m) based on the type of aquifer and the water table gradient. Gurdak and Qi (2006) also identified four sizes of wedges (from 300 to 4000 m), but based on four classes of hydraulic conductivity. Lorenz et al. (2003) found a statistically significant relationship between land use and nitrate concentrations detected in shallow monitoring wells (depth to water <5 m) when using wedges, but not for circles.

In contrast to other reports, Barringer et al. (1990) evaluated the correlation between land use in circular CASs and purgeable organic compounds (POCs). Similar to other studies, correlation occurred in many CAS sizes (radii of 250 to 1000 m) and with various land uses.

This paper compliments and extends the previous work. This paper focuses on VOCs, rather than pesticides or nutrients; VOCs often originate from point sources such as leaking underground fuel or septic tanks, solvent plumes, or waste water discharge pipes, rather than non-point sources. This paper also focuses on supply wells, rather than monitoring and domestic wells; supply wells are generally deeper, have longer screened intervals, and pump larger volumes of water. Consequently, the contributing areas for supply wells are expected to be large, irregularly shaped, and not necessarily contiguous with the area around the well (Barlow, 1994; Franke et al., 1998; Kauffman et al., 2001). In addition, the contributing area of a supply well can include multiple point sources, as well as non-point sources such as street runoff or atmospheric deposition. The question we address is whether or not a 500 m circular CAS is adequate for establishing a correlation between urban land use and VOC detections in supply wells.

The approach we use is similar to the approach used in the existing literature. Non-parametric tests of significance were used to assess the strength (or absence) of correlation between land use and water quality. Data was compiled using 277 supply wells from four previously sampled PBP Study Units (Fig. 1). The data consist of the percentage urban land use and the number of VOCs detected in each well. Land use was computed using three different CAS shapes and ten different sizes for a total of up to 23 CAS shapes per well. The presence or absence of correlation provided a basis for accepting or rejecting the use of a given size and shape CAS.

Study unit descriptions

This research was conducted using wells sampled in four study units of the PBP program: San Diego (Wright et al., 2005), Northern San Joaquin Valley (NSJV) (Bennett et al., 2006), South Sacramento Valley (SOSAC) (Dawson et al., 2008), and North San Francisco Bay (NSF) (Kulongoski et al., 2006) (Fig. 1). Although all four study units have a Mediterranean climate, precipitation varies from 7 in. per year in parts of San Diego to approximately 30 in. per year in NSF. NSJV and SOSAC are located in the Central Valley where alluvial deposits are relatively deep. In contrast, San Diego and NSF are characterized by relatively narrow alluvial valleys underlain by fractured rock. The overall detection frequency for the occurrence of one of more VOCs was 38% in San Diego, 33% in NSJV, 46% in NSJV, and 61% in SOSAC.

The wells sampled were primarily in areas characterized by non-urban land use. When comparing land use within 500 m of a sampled well, SOSAC was the most urban (median = 32%). The median percentage urban land use around sampled wells in San Diego, NSF and NSJV were 5%, 12% and 8%, respectively. Although the median percentages were relatively low, wells with at least 50% urban land use occurred in all four study units.

Methods

Three CAS shapes were used for calculating the percentage of urban land use around each sampled well: circles, ¼ circles (wedges), and a "searchlight" shape (Fig. 2). Circles and wedges were chosen for their simplicity and consistency with the literature. The searchlight shape was chosen as an approximation of a groundwater contributing area for a well that is pumping from a homogeneous aquifer with a sloping water table (Dingman, 2002). For the purposes of this study, the orientation of the wedges and searchlights was based on topography rather than on water levels. To the extent that the water table is a subdued replica of the topography (Tóth, 1963; Dingman, 2002), topographic slope is a surrogate for the generalized groundwater flow direction. This assumption was made for the practical purpose of computing the orientation for hundreds of wells, where water table maps may not be readily available, and for consistency among all four study areas. Ferrari and Ator (1995) and Ator and Denis (1997) also used this method for establishing the orientation of the CAS wedges and semi-circles.

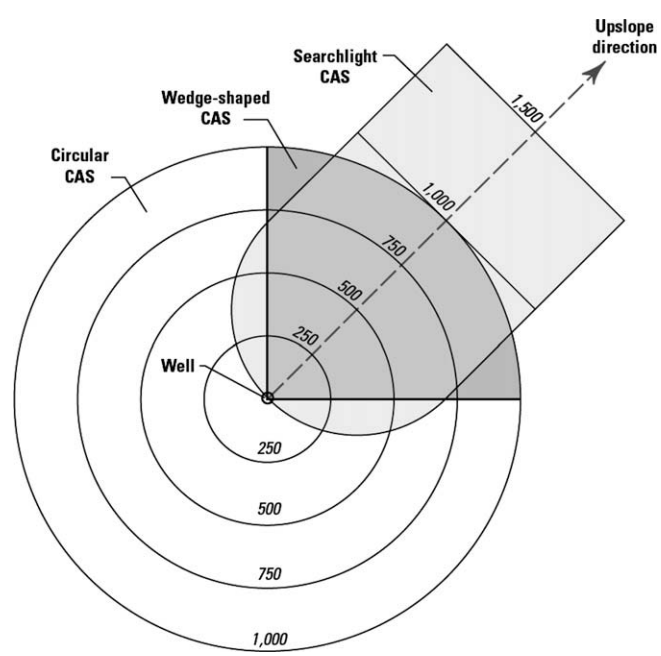

Fig. 2. Schematic of variously sized and shaped contributing area surrogates (CAS). Shown are circular CASs with a radius of 250,500,750, and 1000 m. Overlaid on these circular CASs are the upslope wedge shaped CASs using the same radii. Two searchlight shaped CASs oriented upslope are shown for the 1000 and 1500 m sizes. Not shown are the 1.5, 2, 3, 5, 7, and 10 km wedge and circular shaped CASs, as well as the 2 km searchlight shaped CAS.

For circular and wedge shaped CASs, ten radial distances were used ranging from 250 m to 10 km. Each circular CAS was centered at the well. Wedges were located with the focal point of the wedge at the well and with the direction angle orientated upslope (Fig. 2). The upslope direction was based on a USGS 30 m digital elevation model within 500 m of the well.

The searchlight apex, in a similar fashion to the wedge, was located at the well with the "beam" extending upslope. The width of the searchlight was constrained to 1000 m (derived from a 500 m circular radius), but the distance the searchlight extended was calculated at 1, 1.5, and 2 km (Fig. 2).

A total of 277 water-supply wells (municipal, domestic, and irrigation) were selected for this study: 53 from San Diego; 61 from NSJV; 74 from SOSAC; and 89 from NSF. Monitoring wells and depth dependent samples collected by GAMA were not included.

Sampling in all four study units followed standard procedures based on USGS protocols (USGS, 1999). Samples were obtained from May 18, 2004 through June 15, 2005 and were analyzed for 85 VOCs at the USGS National Water Quality Laboratory using sampling schedule "VOC 2020" as outlined by Connor et al., 1998. Quality assurance and control was routinely conducted throughout the sampling period to insure detections were indicative of groundwater and not contamination due to some other source (Wright et al., 2005; Bennett et al., 2006; Dawson et al., 2008; Kulongoski et al., 2006).

Land use was compiled from an "enhanced" version of the satellite derived (30 m pixel resolution) nationwide USGS National Land Cover Dataset (Volgelmann et al., 2001; Price et al., 2003). The data represent land use circa the early 1990s. The imagery is classified into 21 original land-cover classifications plus an additional four categories for a total of 25 categories as described in Nakagaki and Wolock (2005). In the current study, eight categories were identified as urban (Table 1), whereas Gilliom et al. (2006), recognized six categories as urban. The two additional categories included in the current study are: Transitional and Quarries. The Transitional class was included with other urban land uses because most of these areas were located adjacent to existing urban areas and had most likely completed the transition to urban by the time the wells were sampled. Landfills are also categorized as Transitional. Quarries were considered more similar to industrial/commercial land use and were therefore considered urban rather than natural or agricultural. On a CA statewide basis, these two classes represent less than 0.1% of the total land usage. All defined urban classes were reclassified to a value of 1 whereas all other non-urban classes were reclassified to 0 to speed processing times. This land-use dataset was chosen because it was consistent across all four study units, and has been used in many national and regional NAWQA reports relating land use to water quality (Gilliom et al., 2006; Moran et al., 2006; Zogorski et al., 2006).

The presence or absence of correlation between percentage urban land use and the number of VOC detections in a well was evaluated using Kendall's tau (τ). Kendall's τ is a robust rank-based test of correlation used extensively in water-quality analyses and is especially well suited for skewed relationships and/or where outliers may be present (Helsel and Hirsch, 2002). The correlation values are generally lower than typical **r** values. For example, a correlation of 0.70 for τ may be equated to 0.90 for **r** (Helsel and Hirsch, 2002). In the current study, τ values were considered statistically significant if their attained significance levels (p-values) were less than 0.05.

In addition to evaluating the correlation between urban land use and the number of VOC detections in a well, we also evaluated the spatial autocorrelation of land use. Two techniques were used: (1) non-parametric tests to identify significant differences between CASs and (2) fitting of semi-variograms to the land-use data. Two non-parametric test (Kruskal–Wallis and Wilcoxon's signed-rank test) were computed to determine if land use within the 500 m reference CAS was statistically similar to the land use within other circular CAS sizes. Values of $p < 0.05$ indicated a significant difference in land use between CASs. Semi-variogram models were computed using 510 m by 510 m blocks, which correspond to 17×17 30 m land-use pixels. The geostatistical analysis was conducted for land use within the entire study unit, not just the land use within CASs. The variogram sill indicated the range in km at which urban land use was spatially autocorrelated.

Results and discussion

CAS size and shape analysis

The results from correlating urban land use and VOC detections using all 277 supply wells are shown in Fig. 3. Each point on the graph represents the Kendall's tau (τ) result when regressing 277 wells and their corresponding urban land use percentage with the number of VOC detections in the well. Therefore, each line on the graph represents 2770 calculations of land use. In all cases, p-values were less than 0.0001, far below the statistically significant threshold of 0.05. The maximum value of τ occurred for a 500 m wedge. However, the value of τ for the wedge was only 8% larger than the value obtained from using the 500 m reference circle (Table 2). Overall, τ values were within 10% of one another for wedges and circles ranging in size from 500 m to 2 km. These results are similar to results from previous studies, which showed statistically significant correlation over a range of CAS sizes.

At the study unit scale, the size and shape of a CAS that maximized τ varied (Fig. 4 and Table 2). For San Diego and SOSAC, τ was largest when using a 500 m wedge. For NSJV, τ was largest when using a 1.5 km circle. With the exception of a 10 km circle

Table 1
NLCDe categories used to designate urban land use.

Land use class	Description
21	Low intensity residential
22	High intensity residential
23	Commercial/industrial/transportation
25	GIRAS enhanced residential
26	GIRAS enhanced residential
32	Quarries/strip mines/gravel pits
33	Transitional
85	Urban/recreational grasses

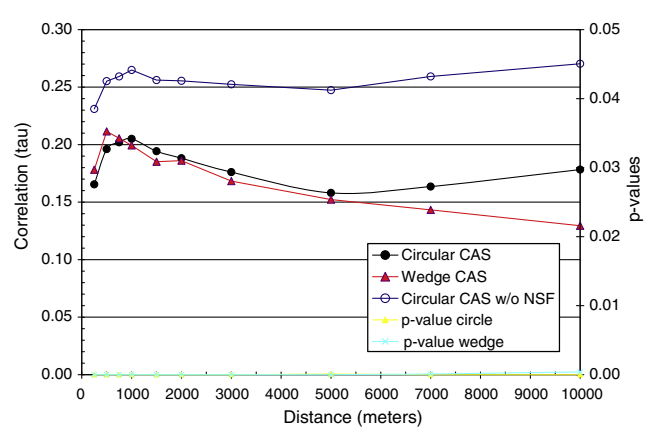

Fig. 3. Correlations between percentage urban land use within circular and wedge shaped CASs and VOC detections expressed in Kendal's tau for all study units. Results without the NSF study unit are also shown. Distance refers to the radial distance for each CAS used.

Table 2
Maximum correlation values for circular and wedge shaped CASs and the distance at which it occurred for each study unit and all study units collectively. Tau values are compared to a 500 m circular CAS.

Study unit	Number of wells	Size of CAS that maximized tau (m)	Tau	*p*-Value	% increase in correlation (tau) over 500 m circle
Circular shaped CASs					
San Diego	53	2000	0.29	0.0004	9.9
NSJV	61	1500	0.28	0.0005	26.5
SOSAC	74	3000	0.24	0.0012	6.1
NSF	89	ns	ns	ns	ns
All	277	1000	0.21	<0.0001	4.5
Wedge shaped CASs					
San Diego	53	500	0.30	0.0001	13.2
NSJV	61	1000	0.24	0.0083	8.4
SOSAC	74	500	0.26	0.0005	13.7
NSF	89	ns	ns	ns	ns
All	277	500	0.21	<0.0001	8.1

ns = Not significant ($\alpha = 0.05$).

in NSJV ($p = 0.09$) the correlations were statistically significant ($p < 0.05$) for all shapes and sizes in all three of the study units.

The only study unit that did not indicate any statistically significant correlation between land use and VOC occurrence was NSF. Moreover, the absence of a correlation was independent of CAS size and shape. This finding is important because it shows if a correlation exists at some other shape/size, the 500 m CAS did not fail to identify it. In contrast, where a correlation did exist at some other CAS size or shape, as in the other study units, it existed at the 500 m CAS as well.

To test the effect that a non-correlating study unit has on the overall correlation, data from NSF was removed. Using the three remaining study units (188 wells), τ was recalculated at all distances for circular CASs. The results show τ increased for all radii, and also show that all values were within 10% of the value computed for the 500 m circle (Fig. 3).

In NSJV a 1.5 km circle produced the largest value of τ, indicating that the wedges may not have been correctly orientated in that study unit. One possible cause could be due to the method for determining the upslope direction in the relatively flat NSJV study unit. To address this possibility, slope aspect was determined using elevation data within 1.5 km of the well, as opposed to utilizing elevation data within 500 m of the well (Fig. 4B); overall, τ values did not improve. For NSJV the circle was a better CAS than the wedge.

A searchlight shaped CAS might be viewed as a better approximation of the contributing area for a producing well in a regional flow field (Javandel et al., 1984; Dingman, 2002). Searchlight shaped CASs were evaluated in the three study units showing correlation between land use and the number of VOCs detected in a well (Fig. 4). In San Diego and SOSAC, the searchlight did not provide τ values as high as the 500 m wedge. In NSJV, the searchlight did not produce τ values as high as the 1.5 km circle. Therefore, the searchlight was not an improvement over other shaped CASs, and does not appear warranted give the added complexity in generating the shape.

There are many confounding factors that could affect a correlation between land use and VOC occurrence, such as groundwater age, pumping rate, well construction, etc. The availability of these datasets will vary depending on each study area. As an example of how the CAS correlations may vary due to a confounding factor, we examine the depth to the top of the well perforation (ToP).

The depth of a well screen is one of many potential explanatory factors affecting water quality, particularly in vertically-dominated flow systems (Koterba et al., 1993; Kolpin, 1997; Koterba, 1998; Hamlin et al., 2002). Approximately 75% of the 277 wells analyzed had well screen information. For each study unit, the deepest well (based on the ToP) with a VOC detection determined the "threshold" depth. Wells with a ToP deeper than this threshold were removed from that study unit's dataset, all others were included. In two of the four study units (San Diego, NSJV), the well with the deepest ToP had a VOC detection, and therefore no wells were removed. In the other two study units (NSF, SOSAC) the deepest well with a detection occurred at 114 m for NSF and 121 m for SOSAC; therefore, wells with a ToP deeper than these depths were removed. Correlations between land use and VOC detections were subsequently recalculated for these two study units. Although there was an increase in τ, there was no change in the size of the circular CAS that maximized τ in SOSAC (Fig. 4C). Similarly, in the NSF study unit, τ increased, but it was still insignificant at all circular CAS sizes. Therefore, removing wells that had a ToP deeper than the deepest well with a detection served to strengthen the correlation, but did not necessarily change the optimum size at which the strongest correlation occurred.

Examination of spatial autocorrelation of urban land use: why do 500 m circles work?

As previously shown, the 500 m circle produces a τ value comparable to τ values produced by other CAS shapes and sizes (Fig. 3, Table 2). This result seems counterintuitive given the likelihood that the actual contributing area will be outside and up-gradient of a 500 m circle. One possible reason for the persistence of a correlation at all shapes and sizes may be due to the presence of land use autocorrelation. We utilized two approaches to help explore this possibility. In both approaches we evaluated the distribution of land use: (1) non-parametric statistics were used to assess the differences between percentages of urban land use within circular CASs and (2) by fitting semi-variograms to the land use within the entire study unit for three study units that exhibited a correlation with VOCs.

The Kruskal–Wallis rank-sum test, a non-parametric equivalent of an ANOVA test (Helsel and Hirsch, 2002), was used to determine if there was a statistically significant difference among the land-use data computed using 10 different circular CASs. The results indicate that there was no significant difference between groups (p-value = 0.907). This result helps to explain why the correlation between urban land use and VOC detections is relatively insensitive to CAS size.

Another non-parametric test, the Wilcoxon signed-rank (Wsr) test, was used to determine if the percentage of urban land use within the 500 m CAS was statistically different than the percentage of urban land use in other sized CASs. The Wsr test is a matched-pair test that determines whether the median difference between paired observations is different from zero (Helsel and Hirsch, 2002). The land use for each well in the 500 m CAS was paired with land use for the same well in another CAS size. An example of the 500 m CAS vs the 1000 m CAS is shown in Fig. 5.

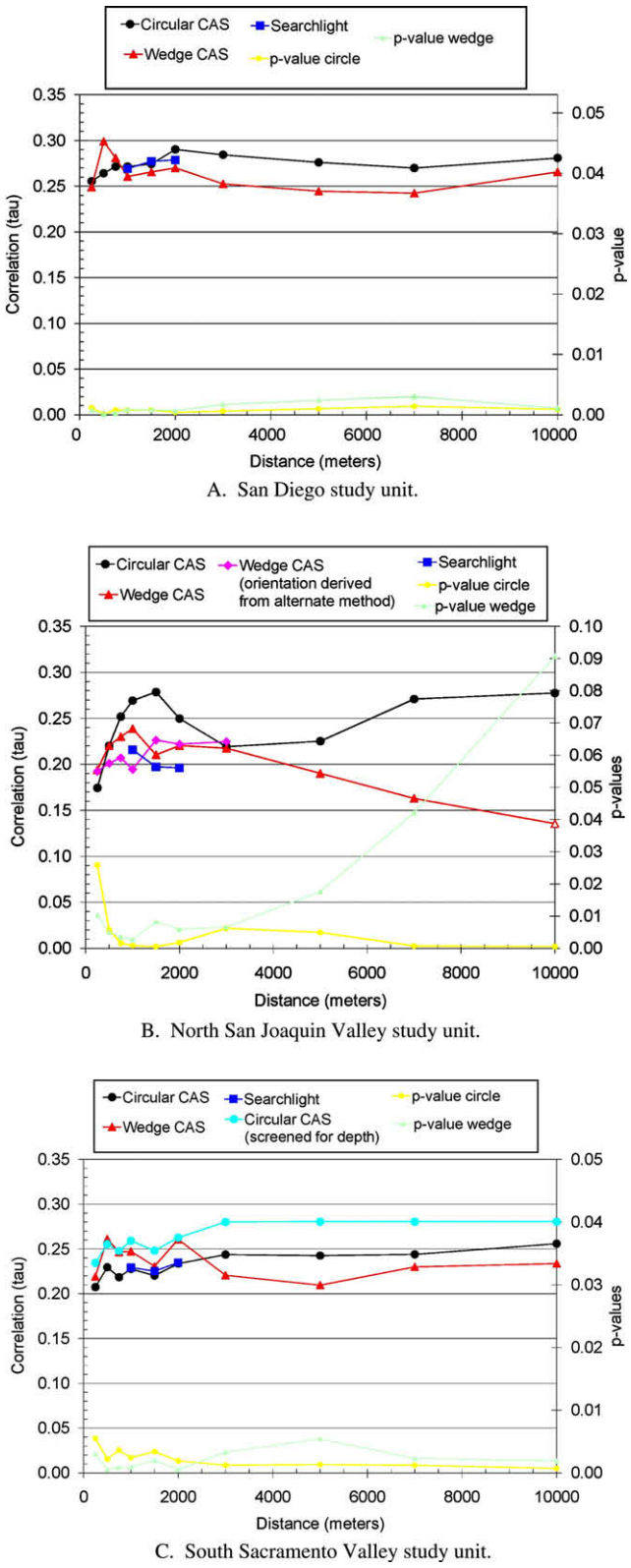

Fig. 4. (A–C) Correlations between percentage urban land use within circular, wedge, and searchlight shaped CASs and VOC detections expressed in Kendal's tau for three study units. Distance refers to the radial distance for each CAS used. (NSF study unit not shown because of the lack of correlation at any distance for circular or wedge shaped CASs; consequently, searchlight CASs were not calculated).

The results show that there is no statistical difference between land use within the 500 m circular CAS and land use within the 250 m, 750 m, 1000 m, and 1500 m circular CASs (Table 3). The results from the Wsr test, like the results from the Kruskal–Wallis tests help explain why a correlation between urban land use and VOC detection is relatively insensitive to CAS size.

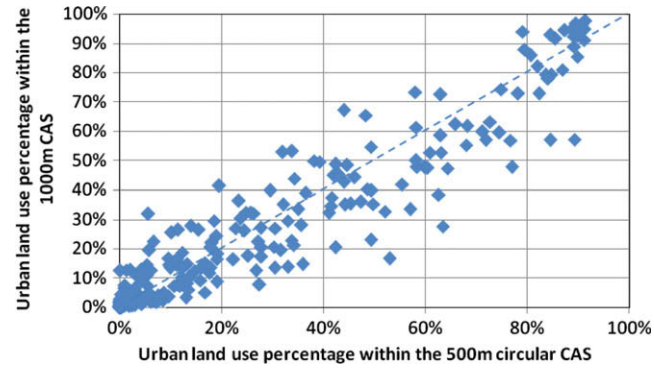

Fig. 5. Matched-pair plot of the percentage of urban land use for wells within the 500 m circular CAS vs. the percentage of urban land use for the same wells within the 1000 m circular CAS. Dashed line represents the X = Y line.

Table 3
Non-parametric test results comparing the percentage of urban land use within the 500 m circular CAS and other circular CASs.

CAS radius in meters	Wilcoxon signed-rank matched-pair test *p*-values
250	0.1623
500	–
750	0.1346
1000	0.2516
1500	0.1884
2000	0.0131
3000	0.0052
5000	<0.0001
7000	<0.0001
10,000	<0.0001

The spatial autocorrelation of urban land use was evaluated by calculating and fitting semivariogram functions for the three study units exhibiting a statistically significant correlation between land use and VOCs: San Diego, NSJV, and SOSAC (Table 4 and Fig. 6). In all three study units, the range of autocorrelation was relatively large, exceeding 8 km. This result indicates that a 510 m square block of land use in one location is predictive of land use at other locations within 8 km or more (depending on the study unit and search direction). Alternatively stated, a 510 m square CAS centered on the well can be representative of the urban land use in the actual contributing area, provided that the actual contributing area is within 8 km (or more depending on the study and search direction) of the well. If the range exceeds 8 km, for example in San Diego and SOSAC, then the predictive value can extend further. To illustrate, the value of the semi-variogram at 2 km in these two study units was about 15% of the sill value (Fig. 6), signifying that two points within 2 km of one another were relatively similar as compared to two points further apart.

The results presented in this paper suggest that the insensitivity of Kendal's τ to CAS size and shape is due to the spatial autocorrelation of land use. This finding is consistent with a regional study across 10 states in the Midwest US by Worrall and Kolpin (2004): they found that there was no significant difference between the land use within a 3.2 km radius around the well and the land use found within the contributing area.

Summary and conclusions

Statistical characterization of groundwater quality can provide valuable information at the national and regional scale. One of

Table 4
Semi-variogram model parameters. The original land use grid with a cell size of 30 m was aggregated (using the mean) into 510 m cells (seventeen 30 m pixels) and converted into points.

Study unit	Number of points	Trend	Model type	Sill	Direction 1		Direction 2	
					Azimuth (degrees)	Range (km)	Azimuth (degrees)	Range (km)
San Diego	37,983	None	Exponential	0.074	–	36	–	–
NSJV	21,241	3rd Order	Exponential	0.042	345	16	75	8
SOSAC	21,378	3rd Order	Spherical	0.093	45	25	315	18

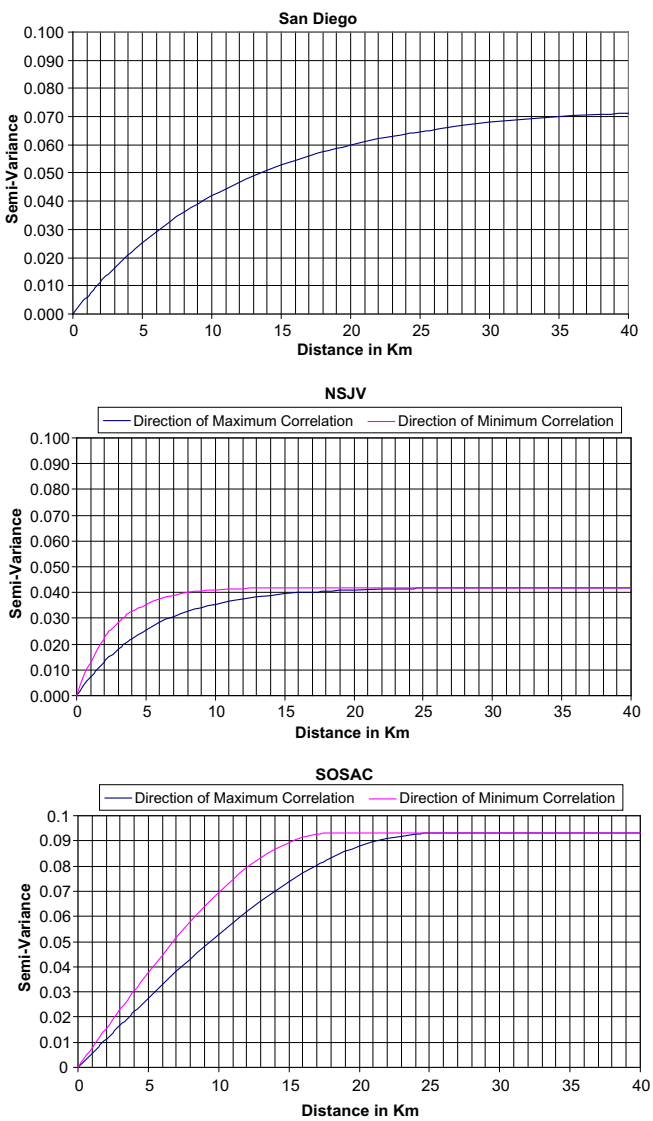

Fig. 6. Semi-variogram models for three study units.

the many potential factors affecting water quality is land use. Typically, land use is characterized using a 500 m radius circle centered on the well. However, the use of a 500 m circle appears overly simplistic given that contributing areas for wells can extend up-gradient from the well and can be irregularly shaped and non-contiguous.

In this study we address the issue of whether a 500 m circle is appropriate when assigning land use to a supply well for evaluating regional-scale water quality. We use the term "contributing area surrogate" (CAS) to describe the area within which land use is classified. The use of this term acknowledges the fact that a simple shape is, at best, a surrogate for the actual contributing area. The use of this term also acknowledges that land use is itself a surrogate for potential contaminating activities. We also define the 500 m circle as the "reference" CAS, given its extensive use by the USGS NAWQA program (Koterba, 1998; Gilliom et al., 2006; Moran et al., 2006; Zogorski et al., 2006).

In this study, we examined the correlation between percentage urban land use and VOC detections in 277 supply wells located in four study units in California. Land use was computed using ten different-sized circles and ten different-sized wedges (radii ranging from 250 m to 10 km), and three different-sized "searchlights." The presence or absence of correlation for each CAS size and shape was evaluated using Kendall's tau (τ).

The results indicate, for the four study units collectively, the correlation was statistically significant ($p < 0.05$) for all sizes and shapes. Overall, the 500 m wedge (oriented upslope) maximized τ. However, relative to the 500 m circular CAS, the increase was 8.1% (τ increased from 0.196 to 0.212). In addition, τ values were within 10% of one another for wedges and circles ranging in size from 500 m to 2 km. Similarly, for three of the four study units, the 500 m circular CAS produced τ values comparable to those produced by other CAS sizes and shapes. No correlation was found in one study unit, regardless of CAS size or shape. These results suggest that the 500 m circular CAS is robust: if a correlation exists at some other shape/size, the 500 m CAS did not fail to identify it, conversely when a relationship did not exist at the 500 m CAS, a relationship did not exist at any other sized or shaped CAS.

We utilized two approaches to help explain why the correlation between urban land use and VOC occurrence was relatively insensitive to CAS size and shape. In both approaches we evaluated the distribution of land use: (1) non-parametric statistics (Kruskal–Wallis and Wilcoxon rank sum) demonstrated that the percentage of urban land use within the 500 m reference CAS was not statistically different than the percentage of urban land use within other CASs; (2) fitting of semi-variograms to land-use data at the study unit scale demonstrated that land use was spatially autocorrelated to distances exceeding 8 km. Spatial autocorrelation helps to explain why land use in the reference CAS correlated with land use in CASs of other sizes and shapes. Spatial autocorrelation also helps to explain why the statistical relationship between urban land use and VOC occurrence is relatively insensitive to CAS size and shape.

The strength of the correlation between the percentage of urban land use and VOC occurrence was not particularly large in this paper ($\tau \sim 0.20$–0.30). Higher values of tau could possibly be obtained using physically-based methods, such as detailed groundwater flow models. Given the presence of spatial autocorrelation of land use, however, an important question remains to be addressed: How much stronger would the correlation be if a CAS were identified using a physically-based approach rather than a circle centered on the well? Similarly, if one were to develop a map for the probability of VOC occurrence, what would be the sensitivity of the predicted probabilities to the method of identifying a CAS. One could compare that sensitivity to the uncertainties associated with the predicted values.

Based upon the results presented in this paper, a 500 m circular CAS appears sufficient for identifying a correlation between urban

land use and VOC detections in supply wells at the regional scale. The use of a 500 m circle can be appropriate when it is impractical to use a more physically-based approach.

Acknowledgements

The authors wish to thank our cooperators, the California State Water Resources Control Board for funding the Groundwater Ambient Monitoring and Assessment program and the Priority Basin Project from which this research is based. We would also like to thank our colleagues at the USGS, including M. Landon and M. Wright for their technical support, and K. Burow and S. Qi for their suggestions and comments. We also thank the anonymous reviewers for their helpful comments and especially one reviewer who suggested the term "Contributing Area Surrogate."

References

Aelion, C.M., Conte, B.C., 2004. Susceptibility of residential wells to VOC and nitrate contamination. Environment Science and Technology 38 (6), 1648–1653.

Aller, L., Bennett, T., Lehr, J.H., Petty, R.J., 1985. DRASTIC – A Standardized System for Evaluating Ground Water Pollution Potential Using Hydrogeologic Settings. US Environmental Protection Agency report EPA/600/2-85/018, 163p.

Alley, W. (Ed.), 1993. Regional Ground-Water Quality. Springer, New York. p. 634.

An, Y.J., Kampbell, D.H., Jeong, S.W., Jewell, K.P., Masoner, J.R., 2005. Impact of geochemical stressors on shallow groundwater quality. Journal of Hydrology 348, 257–266.

Ator, S.W., Denis, J.M., 1997. Relation of Nitrogen and Phosphorus in Ground Water to Land Use in Four Subunits of the Potomac River basin. USGS Water-Resources Investigations Report 96-4268. USGS, Reston, Virginia.

Barlow, P.M., 1994. Two- and three-dimensional pathline analysis of contributing areas to public-supply wells of Cape Cod, Massachusetts. Ground Water 32 (3), 399–410.

Barringer, T., Dunn, D., Battaglin, W., Vowinkel, E., 1990. Problems and methods involved in relating land use to ground-water quality. Water Resources Bulletin 26 (1), 1–9.

Belitz, K., Dubrovsky, N.M., Burow, K., Jurgens, B., Johnson, T., 2003. Framework for a Ground-Water Quality Monitoring and Assessment Program for California. USGS Water-Resources Investigations Report 03-4166. USGS, Reston, Virginia.

Bennett V, G.L., Belitz, K., Dawson, B.J.M., 2006. California GAMA Program – Ground-Water Quality Data in the Northern San Joaquin Basin Study Unit, 2005. USGS Data Series 196. USGS, Reston, Virginia.

Bruce, B.W., McMahon, P.B., 1996. Shallow ground-water quality beneath a major urban center: Denver, Colorado, USA. Journal of Hydrology 186, 129–151.

Burow, K.R., Jurgens, B.C., Kauffman, L.J., Phillips, S.P., Dalgish, B.A., Shelton, J.L., 2008. Simulations of Ground-Water Flow and Particle Pathline Analysis in the Zone of Contribution of a Public-Supply Well in Modesto, Eastern San Joaquin Valley, California. USGS Scientific Investigations Report 2008-5035. USGS, Reston, Virginia.

Cain, D., Helsel, D.R., Ragone, S.E., 1989. Preliminary evaluations of regional ground-water quality in relation to land use. Ground Water 27 (2), 230–244.

Clark, B.R., Landon, M.K., Kauffman, L.J., Hornberger, G.Z., 2007. Simulation of Ground-Water Flow Transport, Age, and Particle Tracking Near York, Nebraska, for a Study of Transport of Anthropogenic and Natural Contaminants (TANC) to Public-supply Wells. USGS Scientific Investigations Report 2007-5068.

Connor, B.F., Rose, D.L., Noriega, M.C., Murtagh, L.K., Abney, S.R., 1998. Methods of Analysis by the US Geological Survey National Water Quality Laboratory – Determination of 86 Volatile Organic Compounds in Water by Gas Chromatography/Mass Spectrometry, Including Detections Less than Reporting Limits. USGS Open File Report 97-829. USGS, Reston, Virginia.

Cowdery, T.K., 1997. Shallow Ground-Water Quality Beneath Cropland in the Red River of the North Basin Minnesota and South Dakota, 1993–1995. USGS Water Resources Investigations Report 97-4001. USGS, Reston, Virginia.

Dawson, B.J.M., Bennett V, G.L., Belitz, K., 2008. Ground-Water Quality in the Southern Sacramento Valley, California, 2005, Results from the California GAMA program. USGS Data Series 285. USGS, Reston, Virginia.

Dingman, S.L., 2002. Physical Hydrology, second ed. Prentice-Hall, Upper Saddle River, New Jersey. p. 646.

Dixon, B., 2005. Applicability of neuro-fuzzy techniques in predicting ground-water vulnerability: a GIS-based sensitivity analysis. Journal of Hydrology 309, 17–38.

Eckhardt, D.A.V., Stackelberg, P.E., 1995. Relation of ground-water quality to land use on Long Island, New York. Ground Water 33 (6), 1019–1033.

Ferrari, M.J., Ator, S.W., 1995. Nitrate in Ground Water in the Great Valley Carbonate Subunit of the Potomac River Basin. USGS Water-Resources Investigations Report 1995-4099. USGS, Reston, Virginia.

Franke, O.L., Reilly, T.E., Pollock, D.W., LaBaugh, J.W., 1998. Estimating areas Contributing Recharge to Wells. USGS Circular 1174. USGS, Reston, Virginia.

Gardner, K.K., Vogel, R.M., 2005. Predicting ground water nitrate concentration from land use. Ground water 43 (3), 343–352.

Gilliom, R.J., Alley, W.M., Gurtz, M.E., 1995. Design of the National Water-Quality Assessment Program: Occurrence and Distribution of Water-Quality Conditions. USGS Circular 1112. USGS, Reston, Virginia.

Gilliom, R.J., Barbash, J.E., Crawford, C.G., Hamilton, P.A., Martin, J.D., Nakagaki, N., Nowell, L.H., Scott, J.C., Stackelberg, P.E., Thelin, G.P., Wolock, D.M., 2006. The Quality of Our Nation's Waters – Pesticides in the Nation's Streams and Ground Water, 1992–2001. USGS Circular 1291. USGS, Reston, Virginia.

Gurdak J.J., Qi, S.L. 2006. Vulnerability of Recently Recharged Ground Water in the High Plains Aquifer to Nitrate Contamination. USGS Scientific Investigations Report 2006-5050. USGS, Reston, Virginia.

Hamlin, S.N., Belitz, K., Kraja, S., Dawson, B., 2002. Ground-Water Quality in the Santa Ana Watershed, California: Overview and Data Summary. USGS Water-Resources Investigations Report 02-4243. Reston, Virginia: USGS.

Helsel, D.R., Hirsch, R.M., 2002. Statistical Methods in Water Resources: U.S. Geological Survey Techniques of Water-Resources Investigations. Book 4, Chapter A3, USGs, Reston, Virginia, p. 510.

Iraj, Javandel, Doughty, C., Tsang, C.F., 1984. Groundwater Transport: Handbook of Mathematical Models. Water Resources Monograph Series 10. American Geophysical Union, Washington, DC.

Kauffman, L.J., Baehr, A.L., Ayers, M.A., Stackelberg, P.E., 2001. Effects of Land Use and travel Time on the Distribution of Nitrate in the Kirkwood-Cohansey Aquifer System in Southern New Jersey. USGS Water-Resources Investigations Report 01-4117. USGS, Reston, Virginia.

Kolpin, D.K., 1997. Agricultural chemicals in groundwater of the Midwestern United States: relations to land use. Journal of Environmental Quality 26, 1025–1037.

Koterba, M.T., Banks, W.S.L., Shedlock, R.J., 1993. Pesticides in shallow groundwater in the Delmarva Peninsula. Journal of Environmental Quality 22, 500–518.

Koterba, Michael T., 1998. Ground-Water Data-Collection Protocols and Procedures for the National Water-Quality Assessment Program: Collection, Documentation, and Compilation of Required Site, Well, Subsurface, and Landscape Data for Wells. USGS Water Resources Investigations Report 98-4107. USGS, Reston, Virginia.

Kulongoski, J.T., Belitz, K., Dawson, B.J., 2006. Ground-Water Quality in the North San Francisco Bay Hydrologic Provinces, California, 2004: Results from the California Ground-Water Ambient Monitoring and Assessment (GAMA) Program. USGS Data Series Report 167. USGS, Reston, Virginia.

Lorenz, D.L., Goldstein, R.M., Cowdery, T.K., Stoner, J.D., 2003. Comparison of Two Methods for Delineating Land Use Near Monitoring Wells Used for Assessing Quality of Shallow Ground Water. USGS Water-Resources Investigations Report 03-4067. USGS, Reston, Virginia.

Moran, M.J., Zogorski, J.S., Squillace, P.J., 2005. MTBE and gasoline hydrocarbons in ground water of the United States. Ground water 43 (4), 615–627.

Moran, M.J., Zogorski, J.S., Rowe, B.L., 2006, Approach to an Assessment of Volatile Organic Compounds in the Nation's Ground Water and Drinking-Water Supply Wells: US Geological Survey Open-File Report 05-1452. USGS, Reston, Virginia.

Nakagaki, N., Wolock, D.M., 2005. Estimation of Agricultural Pesticide Use in Drainage Basins Using Land Cover Maps and County Pesticide Data. USGS Open-File Report 05-1188. USGS, Reston, Virginia.

Nolan, B.T., Hitt, K.J., Ruddy, B.C., 2002. Probability of nitrate contamination of recently recharged groundwaters in the conterminous United States. Environmental Science & Technology 36 (10), 2138–2145.

Paradis, D., Martel, R., Karanta, G., Lefebvre, R., Michaud, Y., Therrien, R., Nastev, M., 2007. Comparative study of methods for WHPA delineation. Ground Water 45 (2), 158–167.

Price, C.V., Nakagaki, N., Hitt, K.J., Clawges, R.M., 2003. Mining GIRAS – Improving on a National Treasure of Land Use Data. In: Proceedings of the 23rd ESRI International Users Conference, July 7–11, 2003, Redlands, Calif.

Rupert, M.G., 2003. Probability of Detecting Atrazine/Desethyl-Atrazine and Elevated Concentrations of Nitrate in Ground Water in Colorado. USGS Water-Resources Investigations Report 03-4269. USGS, Reston, Virginia.

Starn, J.J. Brown, C.J., 2007. Simulations of Ground-Water Flow and Residence Time near Woodbury, Connecticut. USGS Scientific Investigations Report 2007-5210. USGS, Reston, Virginia.

Squillace, P.J., Scott, J.C., Moran, M.J., Nolan, B.T., Kolpin, D.W., 2002. VOCs, pesticides, nitrate, and their mixtures in groundwater used for drinking water in the United States. Environmental Science & Technology 36 (9), 1923–1930.

Squillace, P.J., Moran, M.J., 2007. Factors associated with sources, transport, and fate of volatile organic compounds and their mixtures in aquifers of the United States. Environmental Science & Technology 41 (7), 2123–2130.

Tóth, J., 1963. A theoretical analysis of groundwater flow in small drainage basins. Journal of Geophysical Research 68, 4795–4812.

Trojan, M.D., Maloney, J.S., Stockinger, J.M., Eid, E.P., Lahtinen, M.J., 2003. Effects of land use on ground water quality in the Anoka sand plain aquifer of Minnesota. Ground water 41 (4), 482–492.

US Geological Survey, 1999. National field Manual for The Collection of Water Quality Data: USGS Techniques of Water Resources Investigations, Book 9. USGS, Reston, Virginia.

Volgelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K., Van Driel, N., 2001. Completion of the 1990s National Land Cover Data Set for the conterminous Unites States from Landsat Thematic Mapper data and ancillary data sources. Photogrammetric Engineering & Remote Sensing 67, 650–662.

Worrall, F., Kolpin, D., 2004. Aquifer vulnerability to pesticide pollution – combining soil, land-use and aquifer properties with molecular descriptors. Journal of Hydrology 293, 191–204.

Worrall, R., Besien, T., Kolpin, D.W., 2002. Groundwater vulnerability: interactions of chemical and site properties. The Science of the Total Environment 299, 131–143.

Wright, M.T., Belitz, K., Burton, C.A., 2005. California GAMA Program – Ground-Water Quality in the San Diego Drainages Hydrogeologic Province, California, 2004. USGS Data Series 129. USGS, Reston, Virginia.

Zogorski, J.S., Carter, J.M., Ivahnenko, T., Lapham, W.W., Moran, M.J., Rowe, B.L., Squillace, P.J., Toccalino, P.L., 2006. The Quality of Our Nation's Waters – Volatile Organic Compounds in the Nation's Ground Water and Drinking-Water Supply Wells. USGS Circular 1292. USGS, Reston, Virginia.