Available online at www.sciencedirect.com

SCIENCE DIRECT®

Advances in Water Resources 29 (2006) 181–193

Advances in Water Resources

www.elsevier.com/locate/advwatres

Defining probability of saturation with indicator kriging on hard and soft data

Steve W. Lyon [a], Arthur J. Lembo Jr. [b], M. Todd Walter [a], Tammo S. Steenhuis [a,*]

[a] Biological and Environmental Engineering, Cornell University, Ithaca, New York 14853, USA
[b] Crop and Soil Sciences, Cornell University, Ithaca, New York 14853, USA

Received 28 February 2005; accepted 28 February 2005

Abstract

In humid, well-vegetated areas, such as in the northeastern US, runoff is most commonly generated from relatively small portions of the landscape becoming completely saturated, however, little is known about the spatial and temporal behavior of these saturated regions. Indicator kriging provides a way to use traditional water table data to quantify probability of saturation to evaluate predicted spatial distributions of runoff generation risk, especially for the new generation of water quality models incorporating saturation excess runoff theory. When spatial measurements of a variable are transformed to binary indicators (i.e., 1 if above a given threshold value and 0 if below) and the resulting indicator semivariogram is modeled, indicator kriging produces the probability of the measured variable to exceed the threshold value. Indicator kriging gives quantified probability of saturation or, consistent with saturation excess runoff theory, runoff generation risk with depth to water table as the variable and the threshold set near the soil surface. The probability of saturation for a 120 m × 180 m hillslope based upon 43 measurements of depth to water table is investigated with indicator semivariograms for six storm events. The indicator semivariograms show high spatial structure in saturated regions with large antecedent rainfall conditions. The temporal structure of the data is used to generate interpolated (soft) data to supplement measured (hard) data. This improved the spatial structure of the indicator semivariograms for lower antecedent rainfall conditions. Probability of saturation was evaluated through indicator kriging incorporating soft data showing, based on this preliminary study, highly connected regions of saturation as expected for the wet season (April through May) in the Catskill Mountain region of New York State. Supplementation of hard data with soft data incorporates physical hydrology of the hillslope to capture significant patterns not available when using hard data alone for indicator kriging. With the need for water quality models incorporating appropriate runoff generation risk estimates on the rise, this manner of data will lay the groundwork for future model evaluation and development.

Keywords: Indicator kriging; Indicator semivariogram; Geostatistics; Binary logistic regression; Risk assessment

1. Introduction

To limit the pollution in waterways associated with agricultural land use (i.e., fertilizer, pesticide, and other chemicals coming from field applications), watershed managers need objective methods to identify high-risk, contaminant-contributing areas from agricultural fields. Pollution transport models are often used to rank vulnerability of contaminant sources with field-scale properties ignoring landscape position and current theories in hydrology on how and where runoff is generated [18]. Many studies demonstrate that this approach is not appropriate when considering chemical transport [9,11,17,23,42,31,38]. Quantifying probability of satura-

---

[*] Corresponding author. Tel.: +1 607 225 2489; fax: +1 607 255 4080.

E-mail addresses: sl336@cornell.edu (S.W. Lyon), ajl53@cornell.edu (A.J. Lembo), mtw5@cornell.edu (M.T. Walter), tss1@cornell.edu (T.S. Steenhuis).

doi:10.1016/j.advwatres.2005.02.012

tion to represent runoff production and nutrient transport is imperative to the development of tools based on physical hydrological mechanisms. For example, in humid, well-vegetated areas, such as in the northeastern US and many other locations, most runoff is generated by saturation excess near stream channels. Since these hydrologically active areas vary in size seasonally and during individual storm events, they are often referred to as variable source areas (VSA) [12,16]. Simulation of VSA provides the groundwork for risk-assessment tools usable by watershed planners to limit pollution that incorporate current theories in hydrology. Risk assessment tools predicting VSAs based on topographic index distributions similar to that in TOPMODEL [3] are beginning to emerge [13,45,18,34,25,26,1]. The extremely high spatial and temporal variability of soil moisture and groundwater significantly impacts many hydrological processes making development of such tools a challenge for researchers. Without methods to objectively quantify regions of runoff production, it is impossible develop risk-assessment tools incorporating the appropriate hydrological mechanisms.

As new risk-assessment tools appear, the question is raised, "How can a high-risk region in a field be objectively quantified?" This question requires new methods of collecting and interpreting hydrologic data to evaluate these management tools. In recent years, progress has been made in collecting snap shots of soil moisture using various remote sensing techniques [6,4,14,39,40] and field measurements [46,29,28,41,52]. These methods are powerful, but can be limited by satellite coverage, accessibility to watershed, and watershed size. Also, while spatially robust, they are often too temporally sparse (i.e., low frequency of sampling) to encompass concepts of VSAs integrated into new risk-assessment tools. High temporal resolution measurements of depth to water table are becoming readily available due to inexpensive, self-contained, water level data loggers (e.g., TruTrack, Inc). Since water table height is crucial in runoff generation, especially for saturation excess theory and during the wet season, these measurement techniques provide new information for assessing risk at the field scale. The point nature of these datasets leads naturally to geostatistical techniques to characterize spatial patterns for objective measures saturation during a rain event. With appropriate spatial interpolation techniques, temporally rich, discrete point systems generate spatial measures for hillslopes capable of capturing the dominant characteristics of VSA hydrology.

Water table elevations have been estimated through various forms of kriging and cokriging [32,10,2,7]. At the heart of any kriging methodology is the semivariogram. The semivariogram model provides a quantitative estimate of the structure required to characterize a spatial pattern from a series of point data [47]. The major structural parameters of a traditional semivariogram model include the range, sill, and nugget. The range provides a measure of the maximum distance over which spatial correlation affects the variable of interest. The sill, if it exists, represents the spatial variance of two distant measurements. The nugget represents the variance between two relatively close measurements. The nugget gives the variance in the measurement due to the inherent variability of the sampling device and the occurrence of spatial patterns smaller than the sampling interval. Within the realm of semivariogram and kriging techniques, indicator kriging provides a manner to capture extreme values [24]. Traditionally, indicator semivariograms and the resulting kriging assess risk of contamination in various constituents such as heavy metals [44,37,19] and assess uncertainty in soil properties [33,27,21]. In the most basic form, indicator semivariograms treat data as a binary indicator with respect to a threshold value (i.e., 1 if threshold is exceeded; 0 if threshold is not exceed). For a more complete discussion of indicator semivariograms and indicator kriging along with many possible derivatives in algorithms and methodology, see Goovaerts [20], Deutsch and Journel [8], and Chilès and Delfiner [5].

Western et al. [48] examined soil moisture patterns through indicator semivariograms and showed good spatial structure for high soil moisture conditions. They point out that there are shortcomings in the technique for distinguishing between connected and unconnected patterns in soil moisture. That is the semivariogram does not fully characterize the spatial structure of the pattern. Nevertheless, indicator semivariograms still have promise when applied directly to "hard data" or the actual measurements of the variable of interest, in the form of depth to water table for distributing runoff source areas. Areas prone to saturation either have a high ground water table or hard pan (fragipan) at shallow depth indicating landscape factors such as soil depth (i.e., available water storage capacity), upland-watershed area, and local topography as important factors determining whether or not a particular area saturate. For wet periods, interflow will be higher and often expand the extent of saturation around saturation-prone areas; conversely, dry periods will decrease interflow and extent of saturation. Thus, if the threshold value is near the soil surface, indicator semivariograms and kriging quantify the probability of saturation for an area. Another option available with indicator kriging is the incorporation of "soft data", or local information that is a proxy to the variable of interest and need not relate directly [20], to supplement hard data. Soft data can help compensate for a lack of spatially exhaustive observations by providing information about predictions where no hard data is available using prior probabilities [50]. With saturation excess theory and