PREDICTION OF SPATIAL VARIABILITY OF PHOSPHOROUS OVER THE ST-ESPRIT WATERSHED

A. SARANGI[1,*], C. A. MADRAMOOTOO[1], P. ENRIGHT[1]
and H. CHANDRASEKHARAN[2]
[1]*Agricultural and Biosystems Engineering, McGill University Macdonald Campus, MS1-019, MS building, 21111, Lakeshore Road. Ste-Anne-De-Bellevue, H9X3V9, Montreal,Quebec, Canada;*
[2]*Water Technology Centre, IARI, New Delhi-110012, India*
*(*author for correspondence, e-mail: arjamadutta.sarangi@elf.mcgill.ca, ads_wtc@rediffmail.com)*

(Received 3 February 2005; accepted 6 August 2005)

Abstract. Spatial data analysis tools for predicting the variability of non-point source pollutants minimize the time, effort and cost involved in extensive and exhaustive real field data measurements. In this study, exploratory data analysis, fitting of semivariogram models, and kriging techniques of geostatistics were used to develop the spatial variability map of soil phosphorous saturation (P_{sat}) percentage over the St-Espirit watershed (2610 ha), located in Quebec, Canada. The P_{sat} measured values for the 281 geo referenced land parcel units (LPU) within the watershed were interpreted and analyzed using the ArcGIS® tool. The geostatistical extension module of ArcGIS® was used for exploratory data analysis, semivariogram model fitting, and development of a P_{sat} prediction map using the ordinary kriging technique. Using these geostatistical procedures and adjustment of lag sizes and lag intervals representing the data sets, it was estimated that the spherical semivariogram model fitted well to represent the P_{sat} variability with residual sum square (RSS) of 0.0003 and coefficient of determination (R^2) of 0.98. Further, the developed model was used to predict the P_{sat} variability over the St. Esprit watershed using the 1605 geo-referenced LPU locations. The generated spatial variability map was geo-spatially processed with the natural drainage network and land use feature classes of the watershed to ascertain the phosphorous loading and locate vulnerable LPUs for phosphorous management. It was observed that the P_{sat} levels were higher at the up stream locations and near the drainage channels than the locations close to watershed outlet. Also, the land pockets with more than 60% agricultural land use resulted in supra-optimal P_{sat} values ($10\% > P_{sat} < 20\%$), out of which 8.5 to 16.3 ha agricultural land of the St. Esprit watershed exhibited critical agro-environmental threshold P_{sat} values ($P_{sat} > 20\%$). It was also revealed that, around 23.5% of the watersheds cropped area has reached these threshold levels which necessitate judicious P input management.

Keywords: ArcGIS®, GS+, geostatistics, kriging, soil phosphorous, semivariogram, spatial variability map

1. Introduction

Phosphorous (P) is an essential nutrient for plants and animals and is critically important to the production of crops, animals and animal products and thus essential for the food security of the burgeoning population. However, increase in the availability of soil P beyond a threshold level can give rise to excessive P losses

from agricultural fields, leading to eutrophication problems (Lu, 2003). The eutrophication caused by the excessive inputs of phosphorous (P) and nitrogen (N) is the most common impairment of surface waters in the United States and Canada (Lapp *et al.*, 1998). The affected area of surface water not suitable for designated uses such as drinking, irrigation, industry, recreation and fishing accounts for almost 50% of the impaired lake area and 60% of the impaired river reaches in the United States (Carpenter *et al*., 1998). Further, eutrophication has many negative effects on aquatic ecosystems and the major consequences are increased growth of aquatic weeds and algae causing oxygen shortages that interfere with use of water for fisheries, recreation, industry, agriculture, and drinking. However, the growth of nuisance algae is among the most pernicious effects of eutrophication, which are harmful to livestock, human, and other organisms (Carpenter *et al*., 1998). Therefore, knowledge of the spatial variability of the phosphorous and movement from the watersheds to downstream water bodies are of considerable importance today to control the P transport by implementation of best management practices.

A trend towards large-scale agricultural and livestock production during last couple of decades has increased the pressures on environmental resources and has contributed to the contamination of surrounding water bodies throughout Canada in general and the Quebec province in particular (McDowell *et al.*, 2002). In Quebec, the poor water quality in the rivers are due to intense cash crop production in the watersheds, the increased use of chemical fertilizers and pesticides along with improper manure spreading and the lack of soil conservation practices which contribute to diffuse pollution (Gangbazo *et al.*, 2002; Hamilton and Miller, 2002). It is also reported by researchers working in watersheds of Quebec province that the enhanced P export was mostly due to manured plots than the chemical fertilizers based on studies carried out under simulated rainfall conditions (Sharpley *et al.*, 2001; Michaud and Laverdière, 2004). Under these scenarios, the estimation of spatial distribution of soil P variability are needed for its effective remediation and for prioritization of watershed land units, where the control measures and best management practices (BMPs) can be implement to reduce further contamination. The spatial variability can be generated using the point data values through the deterministic or geostatistical interpolation techniques. The point samples have to be taken on grids sufficiently small enough so that data collected would relate to one another spatially and generate more realistic estimates of variability. Lauzon *et al*. (2005) observed that the spatial variability of soil test P (STP), soil test Potassium (K) (STK) and the soil pH for 23 Ontario farm fields can be accurately accessed with the sampling intensity of 11 samples ha^{-1} or less. However, the cost of this sampling intensity will be of great economic concern, because the cost of soil analysis alone will be \$132 ha^{-1}. They also revealed that higher grid sizes resulted in poor spatial variability predictions. To arrive at this conclusion, they used the GS+ (Gamma Design Software, 2005) analysis tool to develop semivariogram models and Surfer (Ver.8) (Golden Software, 2002) to map the variability of STP, STK and

Soil pH with different grid spacings. However, they did not perform any GIS based analysis to assemble the topography and natural drainage network information to quantify the area under different soil P threshold levels and ascertain P loadings for field level phosphorous management. Moreover, measurement of closely spaced point values is not always feasible due to constraints of the cost and time involved in data acquisition and analysis. As an alternative approach, uses of appropriate geostatistical prediction methods can generate accurate spatial variability maps with low sampling intensity. To substantiate this, Sarangi *et al.*, 2005, employed the ordinary kriging and co-kriging techniques of geostatistics using the ArcGIS® and GS+ tools to generate the rainfall spatial variability map of St Lucia. They could able to map the rainfall variability accurately (i.e. krigged average error (KAE) close to zero and coefficients of determination (R^2) close to 1) over these mountainous regions with very low sampling density (i.e. 40 raingauges spread over 616 km^2 area of St Lucia). Further, use of geostatistics to study the spatial variability of soil (Webster and Oliver, 2001; Goovaerts, 2000; Oliver, 1999; Borgelt *et al.*, 1997; Webster, 1985) revealed that geostatistical interpolation techniques were better than the deterministic statistics and judicious combinations of geostatistical techniques could generate spatial variability maps with acceptable accuracy. The spherical semivariogram model was observed to represent the soil spatial variability over the watersheds more accurately for point measured soil properties (Oliver, 1999). The basic assumption in using geostatistics to characterize spatial heterogeneity is that the properties on earth are not random, but have some spatial continuity which are correlated over some distance. This implies that a parameter measured at one location provides information about its probable values at other adjacent locations. Geostatistics analyzes the spatial variability of geo-referenced parameter values by generating best-fit semivariogram models and these models are used for mapping the variability through kriging technique. Besides these studies on application of geostatistics to map the spatial variability, there are several studies (Michaud and Laverdière, 2004; Preedy *et al.*, 2001; Kleinman *et al.*, 2000; Lapp *et al.*, 1998) on the hydrological responses of the watershed to phosphorous loading and phosphorous management. These studies revealed that the precipitation depths and intensities, natural drainage networks, watershed morphology, subsurface drainage installations plays a significant role in transport of the P to the downstream rivers besides the land use and soil parameters. Peak flows occurring during late winter and early spring, rain on top of melting snow and frozen soils, or an exceptional high intense storm resulted in higher P transport from the watersheds. Also, higher antecedent moisture condition in the watershed resulted in elevated stream flow leading to higher P transport and sediment load at the watershed outlet. During the flow recession stage, the artificial subsurface drainage installed in most of the watersheds acted as the main pathways for P transport to the streams. Also, the soil phosphorous saturation (P_{sat}) percentages were more for up-stream sub-watersheds than that of the sub watersheds adjacent to the outlet. Field observations indicated

that the lower P_{sat}% values near the down stream watershed regions were due to high flow rates and accumulated water depths near the outlet regions with soil saturation and extended water logging conditions prevailing for most of the precipitation events (Kleinman *et al.*, 2000). The water logging conditions were also due to elevated water tables at the watershed outlet. Again, the upstream sub watersheds produced low runoff from the rainfall events reported in their studies. Lapp *et al.* (1998) observed that the peak pollutant concentrations from the St. Esprit watershed, Quebec, Canada were associated with high runoff producing events. Also, the spring snowmelt was identified as a significant period for export of the pollutant material. They arrived at these findings by analyzing the water quality data of two years (1994–1995) at the watershed outlet for nitrate, phosphate, suspended sediment, and atrazine. They concluded that the mean observed pollutant concentrations at the watershed outlet did not exceed drinking water quality standards. However, they did not study the spatial variability of P_{sat} over the watershed and the watershed hydrological responses on phosphorous loadings for devising possible management scenarios to minimize phosphorous outflow.

Keeping in view of the gaps in research for accounting the spatial variability of P_{sat} values and understanding the P transport through watershed hydrological behaviour, an effort was made in this study to use GIS tools to accomplish the following objectives. The objectives were to develop a semivariogram model of measured P_{sat} values with geo-referenced locations of the land parcel units (LPU) of the St. Esprit watershed using GS+ software and the geostatistical analyst module of ArcGIS® tool. Generation of the spatial variability map for the entire watershed using the best-fit semivariogram model predictions with respect to the geo-locations, and finally, to assemble the watershed topology, land use and natural drainage network feature classes of the watershed to ascertain the pathways of P transport and estimate the land area and locate the vulnerable areas of the watershed under different P_{sat} threshold levels for possible P management.

2. Study Area

The St-Espirit study watershed (2610 ha) is located in the province of Quebec, Canada, approximately 50 km north of Montreal (Figure 1). It is a part of the 210 km^2 St. Espirit river basin, which is a tributary of the L'Assomption watershed (4220 km^2). The watershed is located between 45°55′0″ and 46°0′0″ north latitude, and 73° 41′32″ and 73°36′0″ west longitude. The maximum difference in elevation from the outlet to the highest point of the watershed is about 50 m and the principal watercourse is 9 km long and there are a total of 60.3 km of watercourses on the watershed (Romero *et al.*, 2002). About 64 percent of the total area is under crop production with the majority of land use under corn crop followed by cereals, soybeans, vegetables, hay, and pastures (Table I). The slope of the cultivated land generally ranges from 0 to 3%. The upper reaches of the

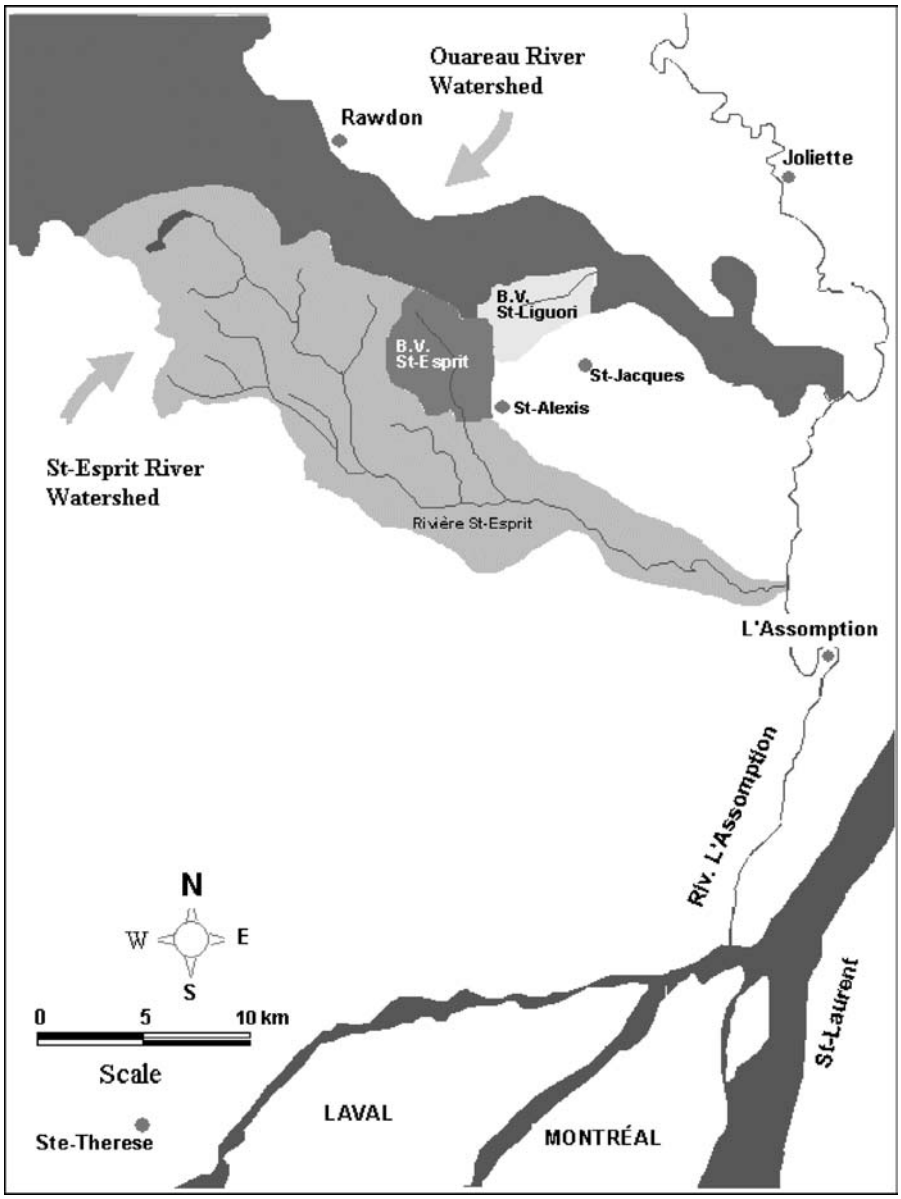

Figure 1. The location map of St. Esprit watershed.

watershed are with slope over 5% and tend to be under forest or managed sugar maple bushes. Nineteen farms on the watershed are involved in livestock production consisting of 9 dairy farms, with the reminder being swine, beef, and poultry operations. The density of livestock on the basin is 0.8 animal units per hectare.

TABLE I
Agricultural land use on the St. Esprit watershed

Land use	Area (ha)	Area (%)
Corn	683	41.0
Cereals	220	13.2
Soybeans	163	9.8
Vegetables	216	12.9
Hay	313	18.7
Pastures	73	4.4
Total	1668	100.0

Approximately 15 different soil series can be found on the watershed. The topography is flat to rolling and the soils formed from glacial tills (sandy loams and loams) are located in the upland areas and occupy approximately 37% of the watershed. Soils formed from marine sediments (clay, clay loam) occupy 38% of the watershed, and the balance of the soils (sand to loamy clay) is formed mostly from alluvial deposits. The lower portion of the watershed is mostly composed of clays and clay loams including the Ste. Rosaile and St. Laurent series (Lapp *et al.*, 1998). The upper regions of the watershed are composed of loamy and sandy soils and most of the agricultural lands are on the heavy soils.

The climate of the watershed is temperate. The frost free growing season varies from 122 to 138 days and average annual precipitation is 998 mm with approximately 20 to 25% appearing as snow. The average annual potential evapotranspiration is between 400 to 560 mm. The average annual temperature is 5.2 °C and temperature in the month of July varies between 18 and 21 °C.

2.1. Data collection

Land use and land management information were collected on St. Esprit as part of an integrated watershed monitoring and management project (Enright *et al.*, 1995; Papineau and Enright, 1997). Public domain data on soils, road networks, topography and land ownership were assembled into a GIS database. A map of agricultural fields on the watershed was developed based upon cadastral information, farm plans, and air photos (Enright *et al.*, 1997). The land ownership units available from the farm plan of the watershed were termed as land parcel units (LPU) and for the entire watershed 1065 such units were digitized with the area ranging from 0.22 to 27.2 ha with the average size of 2 ha. Out of these 1065 LPUs, 803 units were under agricultural land use. The soil samples and the geo-coordinates (latitude and longitude) of 281 agricultural LPUs were acquired using a differential global positioning system (DGPS). Efforts were made to collect the soil sample from the

centroids of the LPUs through the participating owners of the land units. Further, using the ArcGIS® tool and the acquired lat-long values of 281 points, the lat-long values of the centroid of the remaining 784 LPUs were generated.

The collected soil samples were analyzed in the laboratory to estimate the soil test P and phosphorous saturation percentage (P_{sat}). The P_{sat} level, which is defined as the percentage of the P-adsorption capacity of the soil that is currently occupied by P, was determined using the method of Giroux and Tran (1996). The measurement was based on the Mehlich-III P/Aluminum (Al) ratio method (Sims *et al.*, 2002; Giroux and Tran, 1996). The reason for selecting the point P_{sat}% values for generating the spatial variability over St. Esprit watershed was due to the fact that it acts as an indicator of the dissolved P losses in surface and ground water (Sims *et al.*, 2002). For watershed based analysis, the higher value of P_{sat}% indicates that the regions are more prone to soluble P losses than the regions with lower P_{sat}%. From a management perspective, P_{sat}% values are considered for development of nutrient management plans within the study watersheds. The soil P enrichment dynamics documented for Quebec soils highlights that the P_{sat}% value below 10% are considered to be the safe limit, a range of 10 to 20% is said to be supra-optimal soil saturation ratio and P_{sat} value beyond 20% is termed as critical agro-environmental threshold (Gangbazo *et al.*, 2002; Giroux and Tran, 1996). These P_{sat} threshold levels are also applicable to most of the soils (Sims *et al.*, 2000). Therefore, in this study, these ranges are used for classifying the P_{sat} variability over the St. Esprit watershed and locating the vulnerable zones.

The natural drainage network was generated from the digital elevation model (DEM) of the St. Esprit watershed using the Watershed Morphology Estimation (WMET) tool as an interface of ArcGIS® (Sarangi *et al.*, 2004). Field specific information, such as cropping patterns was developed based upon annual aerial surveys. Land management information related to fertilizer, pesticide, manure, and tillage practices were determined based upon on-site interviews with participating agricultural producers, and were linked with the geo-referenced LPUs in the GIS database. The total quantity of organic manure and mineral (P_2O_5) phosphorous added during the growing season of the year 1997 varied from a minimum of 15 to 254 kg ha^{-1} for the agricultural LPUs. However, the higher rate of 200 kg ha^{-1} and above was applied to less number of LPUs and the average P dose was 75 kg ha^{-1}.

3. Materials and Methods

Map generation needs a continuous surface with spatial data values at various locations that can be distributed over closer intervals using a suitable interpolation technique. There are two main groups of interpolation techniques to produce a continuous surface from point measurements; (a) deterministic and (b) geostatistical. Deterministic interpolation techniques create surfaces from measured points

using mathematical functions, which are based on either the extent of similarity (e.g., Inverse Distance Weighted (IDW)) or the degree of smoothing (e.g., Radial Basis Functions (RBF)). Geostatistical interpolation techniques (e.g., kriging) utilize both the mathematical and the statistical properties of the measured points. The geostatistical techniques quantify the spatial autocorrelation among measured points and account for the spatial configuration of the sample points around the prediction location. Kriging is a stochastic interpolation technique for prediction of spatial surface. It is flexible and permits investigation of spatial autocorrelation of the data, because it uses statistical models. The basic assumption in kriging is that the data comes from a stationary stochastic process and some methods require that the data is normally distributed. Kriging is divided into two distinct tasks viz. quantifying the spatial structure of the data and producing a predicted surface. In order to predict an unknown value for a specific location, kriging will use the fitted model from variography, the spatial data configuration, and the values of the measured sample points around the prediction location. The Geostatistical Analyst extension of ArcGIS® and GS+ tool provides a number of modules to help one determine which parameters to use, and also provides reliable defaults that can be used to generate a surface quickly.

3.1. Mathematical functions for kriging

The objective of kriging is to predict parameter values at unmeasured locations, x_0, within the system domain D, using information available elsewhere in $D(x_1, x_2, \ldots, x_n)$. This can be carried out by expressing $Z(x_0)$ {where, $Z(x)$: $x \in D$} as a linear combination of the data $Z(x_1)$, $Z(x_2)$, $\ldots$, $Z(x_n)$, such that

$$\ddot{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_0) \tag{1}$$

where, λ_i is the kriging weight of the parameter value at $Z(x_0)$ for "n" number of nearby sample points to be used in estimation. The optimal weight, λ_i is calculated such that the estimation of $\ddot{Z}(x_0)$ by $Z(x_0)$ is unbiased and the sum of squares of error is minimized. Out of different kriging techniques, the ordinary kriging (OK) methods were used in the present study, because of its simplicity and prediction accuracy in comparison to other kriging methods (Isaaks and Srivastava, 1989). Moreover, it is observed (e.g. Webster and Oliver, 2001; Johnston *et al.*, 1996) that the OK of a single variable is most robust and is frequently used to account for data fluctuations and consideration of a global trend over the study region. Again, due to less number of P_{sat} measurement values (281 points in 2610 ha area), the omni directional variogram is used in the present analysis which assumes the spatial variability to be identical in X, Y and Z directions.

3.1.1. *Ordinary Kriging (OK)*

Ordinary kriging is based on two assumptions. First, the mean of the process is assumed constant and is invariant within the spatial domain. This is expressed as:

$$E[Z(x+h) - Z(x)] = 0 \tag{2}$$

where, E is the expectation and $x \in D$ and $x+h \in D$, h being the distance between two points.

Second, the variance of the difference between two values is assumed to depend only on the distance h between the two points, and not on the location x. The variance is given by:

$$\text{var}[Z(x+h) - Z(x)] = 2\gamma(h) \tag{3}$$

where, the function $\gamma(h)$ is the semivariogram.

Based on these assumptions, the kriging equation is given as (Deutsch and Journel, 1992):

$$\gamma(h,\alpha) = \frac{1}{2N(h,\alpha)} \sum_{i=1}^{N(h)} [Z(x_i+h) - Z(x_i)]^2 \tag{4}$$

where, $\gamma(h,\alpha)$ = semivariance as a function of both the magnitude of the lag distance or separation vector (h) and its direction (α), $N(h,\alpha)$ = number of observation pairs separated by h and direction α used in each summation, and $Z(x_i)$ = random variable at location x_i.

For OK, the weighing parameter λ_i shown in Equation (1), is determined to fulfill conditions as presented in Equations (2) and (3) by solving a system of linear equations such as:

$$\begin{cases} \sum_{i=1}^{n(x)} \lambda_i(x)\gamma(x_j - x_i) - \mu(x) = \gamma(x_j - x) \quad j = 1, 2, \ldots, n(x) \\ \sum_{i=1}^{n(x)} \lambda_i(x) = 1 \end{cases} \tag{5}$$

where $\mu(x)$ is the Lagrangian parameter accounting for the constraint on the weights. The information needed for Equation (5) are the semivariogram value γ, which are estimated using the Equation (4). The spherical model is the most widely used semivariogram model and is characterized by the linear behaviour at the origin (Goovaerts, 2000). The spherical model is:

$$\gamma(h,\alpha) = \begin{cases} S\left(1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right) & h \le a \\ S & h > a \end{cases} \tag{6}$$

where $\gamma(h, \alpha)$ is the spherical semivariogram with range parameter, a, and sill, S, for lag distance h. The best semivariogram model was generated by observing the R^2 and residual sum square (RSS) values with a trial and error approach for different lag sizes and lag intervals (Goovaerts, 2000; Isaaks and Srivastava, 1989). However, the lag sizes and number of lags were varied based on a heuristic rule, generally called a "rule of thumb", in which the lag size times the number of lags should be less than one half of the largest distance in the database (Johnston *et al.*, 1996). In this study, the lag size and number of lags were selected and confined within one third to one half the maximum distances on a trial and error basis to generate an optimized semivariogram model. Then the optimum semivariogram model parameters such as sill, nugget, range and the fitted model type corresponding to the highest R^2 values were noted. The sample variance, where the semivariogram stabilizes is known as the sill, the y-axis intercept of the model is called the nugget and the separation distance along the x-axis where the model flattens out is called the range (Figure 5). The geostatitics wizard of ArcGIS® contains different empirical model fitting functions such as circular, spherical, tetra-spherical, penta-spherical, exponential, Gaussian, Rational quadratic, hole effect, k-bessel, J-bessel, and stable functions for fitting the semivariogram (Johnston *et al.,* 1996).

3.2. Geo-referencing of the P_{sat} sample points

The geo-locations (latitude and longitude) information generated for the 1065 LPUs covering the St. Esprit watershed along with the 281 geo-locations acquired through DGPS containing measured P_{sat} values were in the geographic coordinate systems. The geographic coordinates represented in the form of latitude and longitude were then transformed to projected coordinates using the North American Datum (NAD) 1983 predefined type and then to the UTM grid zone 18N, which encompasses the study region. The Arc Map module of ArcGIS® (Ver. 8.3) was used to perform these transformations, display and edit all the attribute feature classes of the St-Esprit watershed. This was done so as to represent the actual spatial distance between points on the study site and use this information for geostatistical analysis. The polygon feature classes representing the watershed boundary and the attribute information of different parameters related to fertilizer application, land use, soil type and concentration of nitrogen and phosphorous were added to the map layer and the phosphorous measurement points were identified and subjected to different data analysis steps as detailed in Figure 2.

3.3. Geostatistical analysis

The preliminary step of geostatistical analysis is Exploratory Data Analysis (EDA), in which the histogram, normality, trend of data, voroni mapping, semivariogram cloud and cross covariance cloud of the raw data were observed. These steps ensure that the data points are normally distributed and there are no global or local outliers.

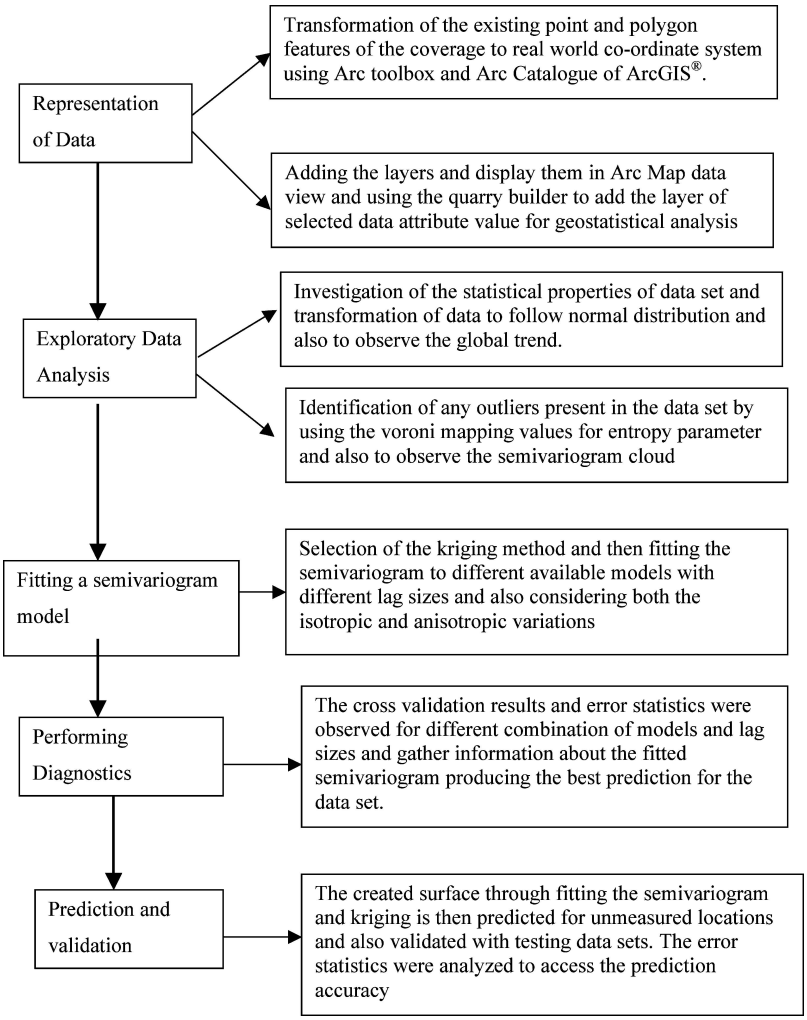

Figure 2. The schematic representation of methods used in geostatistical analysis.

The presence of outliers and skewed data distribution results in improper prediction. Therefore, data transformations are done using linear, box-cox or logarithmic techniques. Using the Arc Map module of ArcGIS®, the polygon feature class of the St-Espirit watershed along with the point feature class representing the measured P_{sat} values were analyzed. Sequentially, the exploratory data analysis was carried out on the data set to ensure that the data set is normally distributed and no outliers are present. Different transformations such as logarithmic and box-cox techniques were tried to ensure normal distribution. In order to identify presence of outliers, the entropy parameter of the voroni map for P_{sat} was developed. Voronoi polygons represent the closeness of every sample point within a polygon to other surrounding

sample points. Out of a number of different parameters associated with the voroni map, the entropy parameter reflected the possibility of local outliers. The reason is that entropy values provide a measure of dissimilarity between neighboring polygons. High entropy values of voroni map indicate the presence of local outliers in the spatial data (Jhonston *et al.*, 1996).

3.3.1. *Development of Semivariogram Model and Kriged Map*
Selection of a best semivariogram model using ArcGIS® was achieved with the visual interpretation and confirmation of the model by applying cross validation statistics. However, ArcGIS® geostatistical analyst is not equipped with the functionality for determination of semivariogram model fitting statistics along with fitting parameters. To achieve this, the GS+ software was used for semivariogram analysis and estimation of the active lag distance and lag class sizes for getting a best-fit semivariogram model with the fitting statistics. The variogram model was generated and different combinations of active lag distance and lag sizes were tried to arrive at a better fitted variogram model with highest R^2 and minimum RSS value. The best semivariogram model obtained for the measured locations were used to predict the P_{sat} values for 1065 locations using the geo-referenced points over the St. Esprit watershed. The kriged maps generated using the measured locations and the predicted locations were generated for comparison and subsequent geo-processing analysis. The GIS thematic coverage feature classes of land use, LPUs, P_{sat} variability, and natural drainage network were carried out to ascertain the P transport and locate the vulnerable zones for P management.

3.4. Data overlay and P_{sat} threshold analysis

The kriged surface was classified under three different sub-classes of P_{sat} values to locate the zones of safe range ($0\% < P_{sat} < 10\%$), supra-optimal range ($10\% < P_{sat} < 20\%$) and critical agro-environmental range ($P_{sat} > 20\%$). The generated surface was then converted to polygon feature classes to estimate the area of different limiting P_{sat} ranges. The geo-processing capability of the ArcGIS® tool was used to join, union, intersect, and clip the land use, spatial P_{sat} variability and LPU feature classes to estimate the number of LPUs falling under different zones and the corresponding cropped area affected by these P_{sat} ranges. Further, the generated drainage network line feature class was superimposed over the P_{sat} variability map to observe the possible hydrologic pathways of P transport.

4. Results and Discussions

The assembled coverage feature classes of the LPUs, delineated St. Esprit watershed and the generated drainage network of the watershed locations are shown in Figure 3. The stream channels of the natural drainage network were ordered using Strahler's

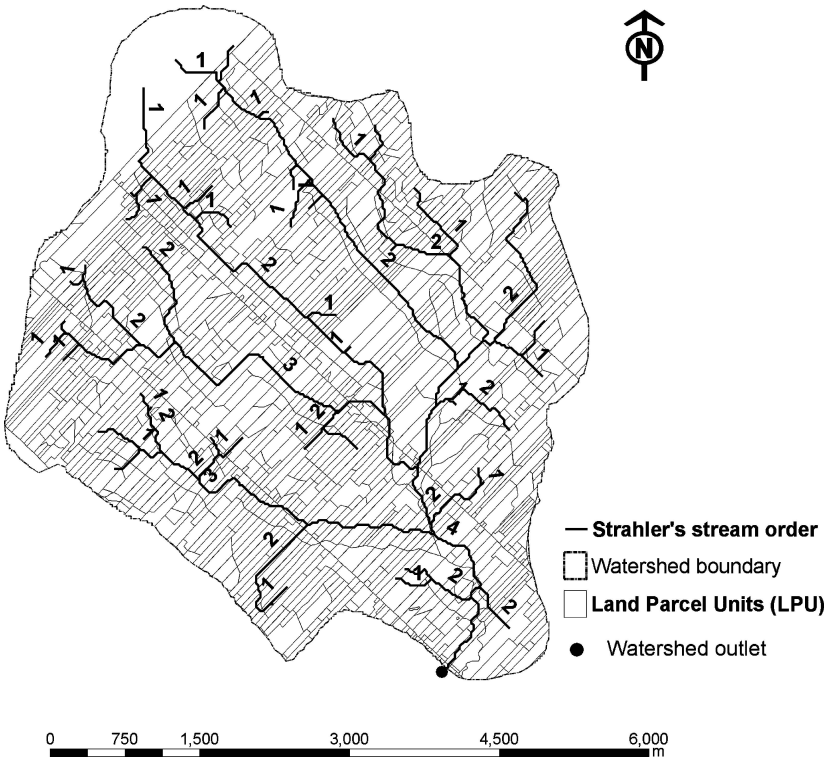

Figure 3. The geo-processed map of the Strahler's stream network and the LPUs of the St Esprit watershed.

ordering scheme (Strahler, 1964), where the joining of two lower order streams give rise to the next higher order and this sequence is carried out towards the watershed outlet. It is observed from Figure 3 that the St. Esprit watershed is a 4th order watershed with 51.42 km long streams of all orders and the drainage density was estimated to be 1.97 km^{-1}. This higher drainage density value of the St. Esprit indicate a well defined drainage network over the watershed and quicker translation of runoff towards the outlet from a given rainfall event (Sarangi *et al.*, 2004). Also the depth and intensity of precipitation events affects peak and volume of runoff at the watershed outlet. It is also observed at watershed outlet that, the precipitations with higher intensity resulted in higher sediment yield. These findings ascertain that the LPUs with higher P_{sat} values would lead to elevated P transport for high intensity precipitation events.

Exploratory data analysis of the measured P_{sat} values revealed that the logarithmic transformation could convert the data points to a normal distribution for subsequent analysis. Also, the trend line depicting the data variability was steeper towards the east-west direction than the north-south direction, which indicated that

the variation of P_{sat} is more prominent in east-west direction of the St-Esprit watershed. This could be due to less number of P_{sat} measured values in the northern part of the watershed, which are mainly forested lands. Further, to locate the presence of any global or local outliers, the entropy parameter of voroni map was analyzed and the highest entropy range (2.07 to 2.34) was identified and these were represented by only 5 points. So, these outliers were excluded, leaving the 276 P_{sat} values for semivariogram analysis and model development.

The semivariogram model was fitted using the GS+ with different lag sizes and number of lags. Keeping in view of the "rule of thumb" for selection of lag sizes and number of lags, the optimal semivariogram was obtained with lag size of 400 m with 9 numbers of lags. As per the "rule of thumb", the lag sizes multiplied with the number of lags (3600 m), were within the one-third (2700 m) to one-half (4050 m) of the maximum separation distance (h) between the geo-referenced data points (8100 m) of the watershed. The best fitted semivariogram model ($R^2 = 0.98$ and RSS $= 0.0003$) obtained through this analysis was spherical type with range (a) of 314 m, nugget of 0.0002 units (% P_{sat} squared) and sill (S) values 0.16 units (%P_{sat} squared) (Figure 4). The semivariogram model was used to generate the krigged surface as shown in Figure 5 along with the P_{sat} measured point locations. The observed P_{sat} values of 276 point samples and the corresponding kriged surface predicted data points were plotted (Figure 6). The 1:1 line representing the data points depicted a closer match with R^2 of 0.99. Further, substituting the best fitted spherical model parameter (i.e. "$S = 0.16$", "$a = 0.32$" and "$h = 3.6$") in Equation (6), the variogram model representing the P_{sat} variability of the St. Esprit watershed was obtained. This model was used to predict the P_{sat} values for the 1065 geo-referenced locations over the watershed. This was achieved using the prediction command button control, which is available in the geostatistical analyst extension

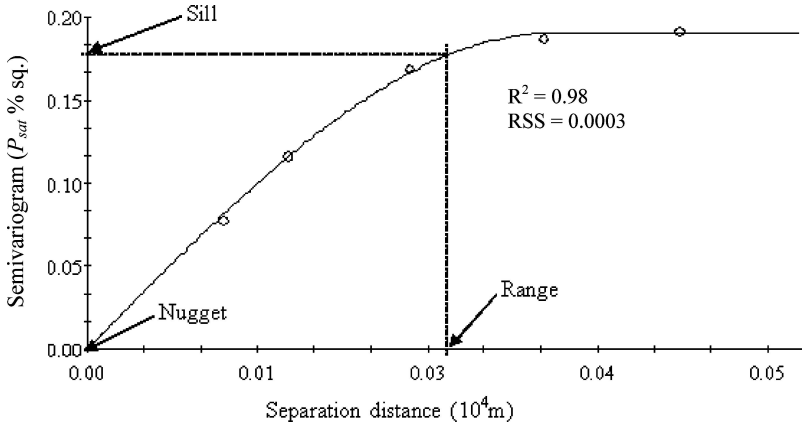

Figure 4. The spherical semivariogram model with range, sill and nugget values for the measured P_{sat} values of the St. Esprit watershed.

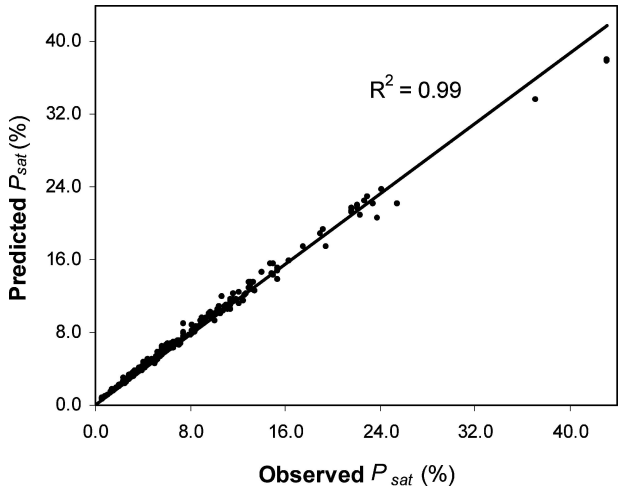

Figure 5. The 1:1 line of the observed and predicted P_{sat} values (276 locations) using the spherical semivariogram model.

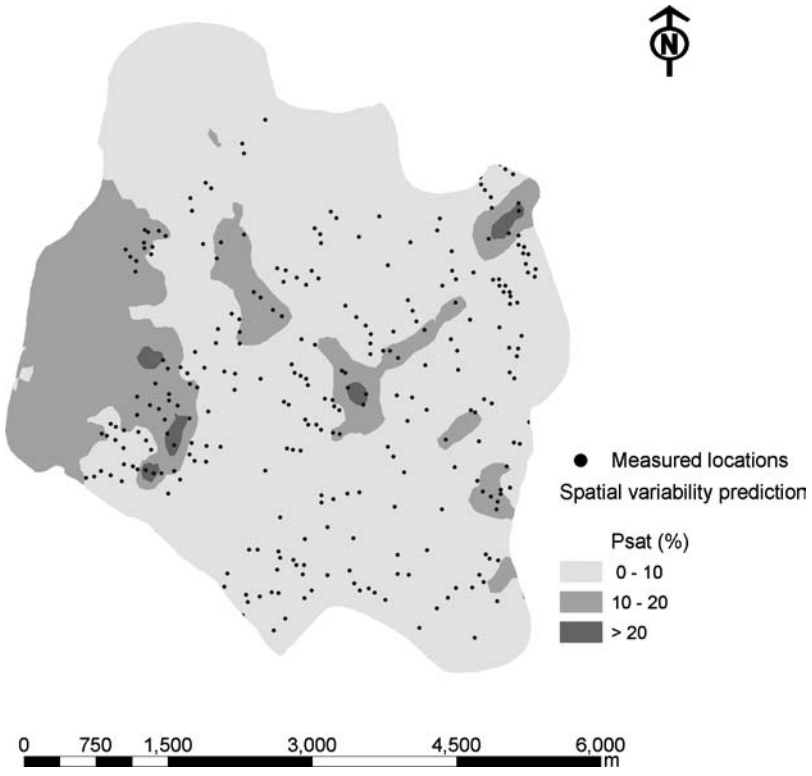

Figure 6. The kriged map of the P_{sat} range values of the measured locations over St Esprit watershed.

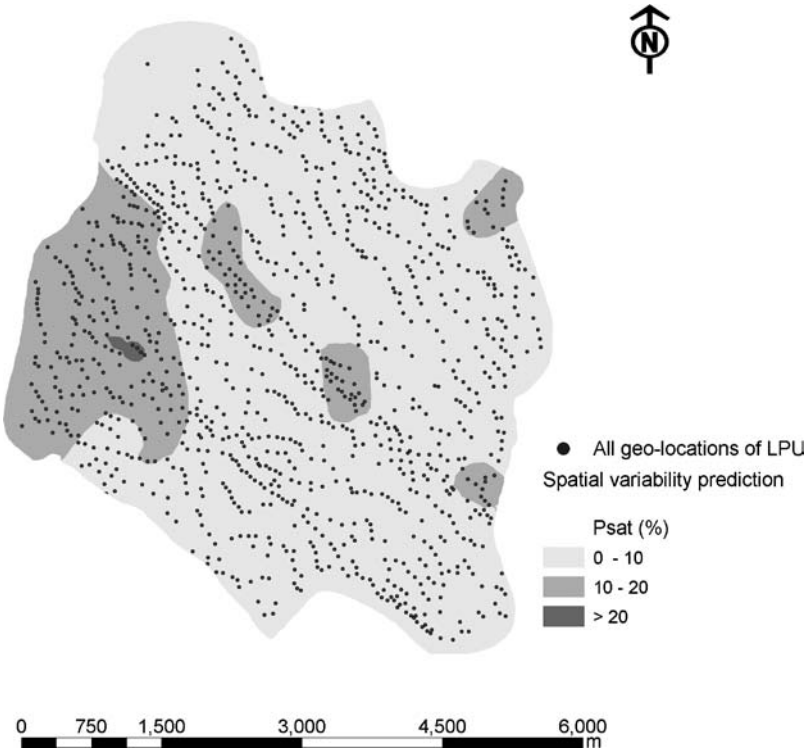

Figure 7. The predicted P_{sat} variability map with threshold values of all the geo-referenced locations over the St Esprit watershed.

module of ArcGIS®. The generated krigged layer (Figure 7) was classified under different P_{sat} threshold ranges and geo processed with the thematic layers of land use, drainage network, LPUs and the delineated St. Esprit watershed for ascertaining P transport pathways and locating vulnerable zones for P management. The cross validation statistics of the generated surface was with error mean of 0.004 and R^2 of 0.935.

4.1. P_{sat} VARIABILITY AND IDENTIFICATION OF THRESHOLD P_{sat} ZONES IN THE WATERSHED

The generated krigged maps of measured locations (Figure 6) and all geo-referenced locations (Figure 7) classified under three major ranges of P_{sat} variability was transformed to polygon feature classes and clipped over the LPU coverage to generate the areas under different land uses. It was revealed that, 23.5% of the watersheds cropped area has reached the soil P_{sat}% of more than 10% (Supra-optimal range) and about 1% of the cropped area has reached P_{sat} level of more than 20% (critical

agro-environmental range). Also, about 58 to 60% of the LPUs under agricultural production are within the safe soil P_{sat} level ($0\% < P_{sat} < 10\%$); 83 to 92% of LPUs under agricultural production are within supra-optimal range ($10\% < P_{sat} < 20\%$) and 74 to 79% of LPUs under critical threshold levels (Table II). These values indicate that the LPUs under agricultural production are more prone to P accumulations. It was also revealed that in the forested up stream reaches of the watershed and non-agricultural LPUs, the P_{sat} variability was within the safe limit. Phosphorous applied to the LPUs under agricultural crops at 120 Kg ha^{-1} and above resulted in higher soil P_{sat} values. Overall, the 2182 ha area of the watershed were considered safe and the rest 428 ha were observed to be of soil P_{sat} values exceeding 10%, are vulnerable to subsequent P transport, leading to contamination of the downstream water bodies.

4.2. Interpretation of the P transport and possible P management in the watershed

Phosphorous losses from soils to water bodied occur mainly by erosion and surface runoff (particulate and dissolved P) and as leaching or surface runoff of P (dissolved P). All forms of soil P estimates have been shown to be susceptible to transport from land to water. Within watershed systems, the transport is mainly carried out through the natural drainage networks or constructed irrigation channels. Moreover, in St. Esprit watershed, the main water supply to the farmers LPUs are through the natural drainage networks (Figure 3). The hydrological responses of the watershed to generate runoff and sediment losses plays a significant role as most of P gets transported from the soils by surface processes and pathways such as overland flow and erosion. The mechanism involved in soluble P transport are straightforward, which starts with initial desorption or dissolution of P bound by soil particles, followed by water movement from source soil to a stream, river, or drainage ditch that later discharges into a water body. The pathways of P transport were ascertained by overlaying the generated drainage network thematic map of the watershed over the P_{sat} variability map of the St. Espirit watershed (Figure 8). It was revealed from the figure, that the higher concentrations of P_{sat} values were observed in LPUs, that were adjacent to channel reaches with intensive agricultural practices. Also, the LPUs adjacent to lower order stream channels showed more P_{sat} concentrations than the higher order channels, because the P-transport from LPUs near the higher order channels were facilitated due to concentrated overland flows leading towards the watershed outlet. Therefore, the adjoining areas of the watershed outlet were having low P_{sat} percentage as compared to the upstream locations. These findings were in line with the studies reported by Kleinman *et al.* (2000). Further, the slope variation of only 0.5 to 1% in the western part of the watershed along with existence of lower order stream channels resulted in higher soil P_{sat}% values in comparison to the northern part of the watershed where the slope exceeds 5% (Figure 8). Because, the

TABLE II

Effect of agricultural land use on the variability of soil phosphorous saturation % over the land parcel units (LPUs)

P_{sat}% variation ranges (%)	Spatial variability predictions for measured locations			Spatial variability predictions for all geo location of the watershed		
	LPUs	Area (ha)	Area under agriculture (ha) [% of LPU area (%)]	LPUs	Area (ha)	Area under agriculture (ha) [% of LPU area (%)]
0–10	1406	2233.7	1349.2 [60.4]	1389	2182.1	1276.5 [58.5]
10–20	185	365.7	302.5 [82.7]	208	416.4	383.0 [91.9]
> 20	14	20.6	16.3 [79.1]	8	11.5	8.5 [73.9]
Total	1605	2610.0	1668.0	1605	2610.0	1668.0

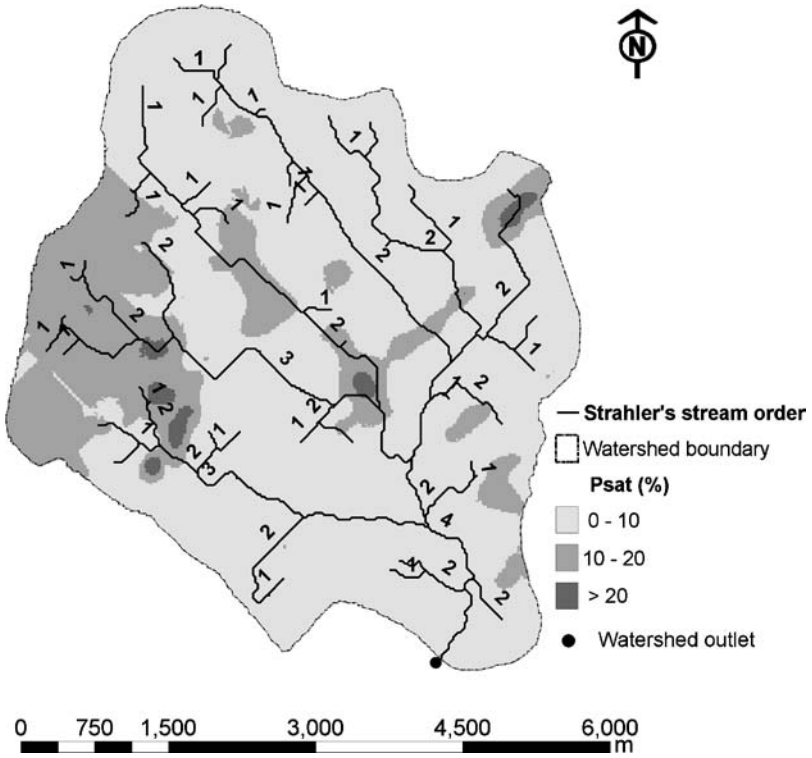

Figure 8. The geo processed map of the drainage network and P_{sat} variability with threshold values and stream orders over the St. Esprit watershed.

higher slope ranges accelerate the P removal through the stream channels towards the watershed outlet. The P-transport mechanism of the St. Esprit watershed was observed to be governed by the well defined hydrological responses.

Agricultural P management has focused on identifying, through soil testing the amount of P for addition to enrich the deficit soil or to adopt best management practices (BMPs) to reduce P loads reaching downstream for P rich soils. In the present study, phosphorous management is targeted towards reducing the soil P_{sat} values through BMPs. Whereas, the BMPs include selection of appropriate agricultural practices that avoid excess P applications in the fields; regular testing of the soil to reduce potential P losses to water and attain economically optimal crop yields; implementation of soil conservation practices that prevent soil P losses; and addition of manure and organic P sources in timely manner for sustainable crop growth. Also, to accomplish the BMPs within the St. Esprit watershed, the on-site discussion with the farmers revealed that the law obliges the farmers of Quebec, to calculate their P fertilizer applications based upon the crop requirements and the soil test P and P_{sat} percentage. While some farmers have undertaken the training necessary to do these recommendations, most rely upon professional consulting

services. The maps developed in Figures 7 and 8 allows for individual producers to identify if they are in a priority zone for soil P richness, and voluntarily adopt the BMPs to address the question of elevated P losses. Beyond the safe P_{sat} levels, risks of excessive P losses to the environment from agricultural fields are elevated.

5. Conclusions

This study described the geostatistical approach for mapping the spatial variability of soil P_{sat} levels, which is an important soil nutrient for site specific soil management. The watershed based P management strategies not only mitigate the eutrophication problems of downstream water bodies, but also assists in sustainable management of land, water and vegetation resources to enhance the biomass production. Overall, the use of ArcGIS® modules to develop a digital data base of the land use, LPUs, natural drainage networks and P_{sat} sample values, perform the spatial analysis and generate geo processing assisted composite thematic maps, resulted in meaningful interpretation of the P_{sat} variability. It can be concluded that, the geostatistics can be efficiently used to predict the spatial variability with minimum sampling intensity, which reduces the cost and effort involved in acquisition of closer data samples. So, geostatistical prediction techniques can be a substitute of higher data sampling intensity, when there is a little compromise of the prediction accuracy. The spherical semivariogram model developed in this study to predict the P_{sat} variability was observed to predict the variability more accurately as observed from the prediction error statistics. The generated kriged maps were used to locate the vulnerable land parcel units under different P_{sat} threshold levels. The generated spatial variability maps will be useful for the farmers of the watershed to adopt BMPs to reduce the P loadings and enhance the productivity with optimal field level P applications.

The P transport problems over the watershed systems are governed by the hydrologic responses. Therefore, understanding the watershed hydrology is critical to reduce the P loss and develop efficient P management practices in the watershed. This study revealed that the geo processing operations to assemble the GIS feature classes of the natural drainage network, land uses and the surface topological features provided a coherent explanation of P accumulation and transport over the watershed system. Further, the P transport and accumulation can be accurately explained by considering the runoff and sediment generation ability of the LPUs through hydrological and water quality prediction models for different precipitation events over the watershed. Nonetheless, considering the study objectives, the modeled P_{sat} values over the geo-reference points of the watershed were considered reasonable and the hydrological behaviour of P transport through hydrologic pathways were acceptable with respect to the recorded information of the study watershed. The developed methodology of geostatistical analysis and geo processing

of the GIS feature classes within the ArcGIS® environment can be applied to any watershed regions for understanding the spatial variability of pollutant loads and migration to downstream water bodies.

References

Borgelt, S. D., Wieda R. E. and Sudduth K. A.: 1997, 'Geostatistical analysis of soil chemical properties from nested grids', *Applied Engineering in Agriculture* **13**(4), 477–483.

Carpenter, S. R., Caraco N. F., Correll D. L., Howarth R. W., Sharpley A. N. and Smith V. H.: 1998, 'Non-point pollution of surface waters with phosphorous and nitrogen', *Ecological Applications* **8**(3), 559–568.

Deutsch, C. V. and Journel A. G.: 1992, *GSLIB Geostatistical Software Library and User's Guide*, Oxford Univ. Press, New York.

Enright, P., Papineau, F. and Madramootoo, C. A.: 1997, Water quality and pollutant concentrations on paired agricultural watersheds in Quebec. Proceedings of the 1997 Annual Conference of the Canadian society of Agricultural Engineering, 142–151.

Enright, P., Papineau F., Madramootoo, C. A. and Leger, E.: 1995, The impacts of agricultural production on water quality in two small watersheds. CSAE Paper #95-101. CSAE Annual Meeting, Ottawa, Canada.

Gamma Design Software: 2005, GS+ version 5.03 beta. Gamma Design Software, Plainwell, MI.; http://www.gammadesign.com.

Gangbazo, G., Cluis, D. and Buon, E.: 2002, 'Suspended sediments and phosphorus transport in an agricultural watershed', *Vecteur Environnement* **35**, 44–53.

Giroux, M. and Tran, T. S.: 1996, 'Critères agronomiques et environnementaux liés à la disponibilité, la solubilité et la saturation en phosphore des sols agricoles du Québec', *Agrosol* **IX**, 51–57.

Golden Software: 2002, Surfer, version 8. User's Guide. Golden Software Inc., Golden, CO; http://www.goldensoftware.com.

Goovaerts, P.: 2000, 'Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall', *Journal of Hydrology* **228**, 113–129.

Hamilton, P. A. and Miller, T. L.: 2002, 'Lessons from the National Water-Quality Assessment', *J. Soil and Water Consv*. **57**, 164-A–24A.

Isaaks, E. H. and Srivastava, R. M.: 1989, *An introduction to Applied Geostatistics*, Oxford Univ. Press, New York.

Johnston, K, Hoef, J. M.V., Krivoruchko, K. and Lucas, N.: 1996, *Using ArcGIS Geostatistical Analysis*, GIS user Manual by ESRI, NY, 120–187.

Kleinman, P. J. A., Bryant, R. B., Reid, B. S., Sharpley, A. N. and Pimentel, D.: 2000. 'Using phosphorus behavior to identify environmental thresholds', *Soil Science* **165**, 943–950.

Lapp, P., Madramootoo, C. A., Enright, P., Papineau, F. and Perrone, J.: 1998, 'Water quality of an intensive agricultural watershed in Quebec', *J. of Am. Water Resources Association* **34**(2), 427–437.

Lauzon, J. D., Halloran, I. P. O., Fallow, D. J., Bertoldi, A. P. V. and Aspinall, D.: 2005, 'Spatial variability of soil test phosphorous, potassium and pH of Ontario soils', *Agronomy Journal* **97**, 524–532.

Lu, J.: 2003, 'Characteristics of phosphorous components in drainage water', *Bulletin of Environmental Contamination and Toxicology* **71**, 951–960.

McDowell, R. W., Sharpley A. N. and Kleinman, P. J. A.: 2002, 'Integrating phosphorus and nitrogen decision management at watershed scales, *J. Am. Water Resources Association* **38**(2), 479–491.

Michaud, A. R. and Laverdière, M. R.: 2004, 'Effects of cropping, soil type and manure application on phosphorus export and bio-availability', *Can. J. Soil Sci.* **84**(3), 295–305.

Oliver, M. A.: 1999, 'Exploring soil spatial variability geostatistically', in: J. V. Stafford (ed.), *Precision Agriculture '99. Proc. Eur. Conf. On Prcision Agric.*, 2nd, Odense, Denmark. 11–15 July 1999. Part 1. Sheffield Academic Press, Sheffield, UK, pp. 3–17.

Papineau, F. and Enright, P.: 1997, Gestion de l'Eau dans le Bassin Versant de la Partie Superieure du Ruisseau St-Esprit. Project 61-13008. Charactérisation de la problématique environnementale. McGill University, Montreal, Canada.

Preedy, N., McTiernan, K., Matthews, R., Heathwaite, L. and Haygarth, P.: 2001. 'Rapid incidental phosphorus transfers from grassland', *J. Environ. Qual*. **30**, 2105–2112.

Romero, D., Madramootoo, C. A. and Enright, P.: 2002, 'Modelling the hydrology of an agricultural watershed in Quebec using SLURP', *Canadian Biosystems Engineering* **44**, 1.11–1.20.

Sarangi, A., Madramootoo, C. A., Singh, D. K. and Singh, A. K.: 2004, 'Performance of GIS interfaces in watershed delineation and stream network generation from DEM', *Journal of Indian Society of Agricultural Engineering* **41**(2), 41–48.

Sarangi, A., Cox, C. A. and Madramootoo, C. A.: 2005, 'Geostatistical methods for prediction of spatial variability of rainfall in a mountainous region', *Transactions of ASAE* **48**(3), 943–954.

Sharpley, A. N., McDowell, R. W., Weld, J. L. and Kleinman, P. J. A.: 2001. 'Assessing site vulnerability to phosphorus loss in an agricultural watershed', *J. Environ. Qual.* **30**, 2026–2036.

Sims, J. T., Edwards, A. C., Schoumans, O. F. and Simard, R. R.: 2000, 'Integrating soil phosphorous testing into environmentally-based agricultural management practices', *J. Environ. Qual.* **29**, 60–72.

Sims, J. T., Maguire, R. O., Leytem, A. B., Gartley, K. L. and Pautler, M. C.: 2002, 'Evaluation of Mehlich III as an agri-environmental soil phosphorous test for the Mid-Atlantic United State of America', *Soil Sci. Soc. Am. J*. **66**, 301–319.

Strahler, A. N.: 1964. Quantitative geomorphology of drainage basins and channel networks; Section 4-2, in *Handbook of Applied Hydrology*, ed. Ven te Chow, McGraw-Hill, New York.

Webster, R.: 1985, 'Quantitative spatial analysis of soil in the field', *Advances in Soil Sci*. **3**, 1–70.

Webster, R. and Oliver, M. A.: 2001, *Geostatistics for Environmental Scientists*, John Willey and Sons Ltd, U.K., 271 pp.