CENTER FOR GEOSPATIAL DATA ANALYSIS
INDIANA UNIVERSITY

Spatially distributed regional recharge-rate estimation to guide an aquifer-sensitivity assessment for mid-continental glacial environments, USA

Sally L. Letsinger, Ph.D. (sletsing@indiana.edu)
Center for Geospatial Data Analysis, Indiana Geological Survey, Indiana University, Bloomington, IN

Indiana Geological Survey

Abstract

A map showing the distribution of groundwater recharge for the glaciated portion of Indiana, USA, was facilitated by a multiple regression model that, in turn, was based on a governing conceptual model that describes recharge as controlled by topography, geology, and land cover. The proportion of baseflow to runoff in stream hydrographs (*n*=189), determined using a local-minimums approach with a recession slope test, was used as the measure of recharge in this study. Partial correlation coefficients were used to identify and rank the most statistically influential (*p*=0.001) independent variables in the study domain, as follows: the elevation of the near-surface water table, the proportion of sand in the subsurface (as determined by lithologic logs of water-well records), and the thickness of near-surface clay. Other factors that appear to explain the spatial distribution of recharge include: the cumulative upstream area occupied by water bodies (streams, lakes, ponds), the density of wetland areas, and the cumulative upstream area occupied by forest.

The model has a multiple adjusted R^2 value of 0.55. The model-estimated recharge rates vary between 1 and 34.5 cm/year (mean of 14 cm/year), averaging 12.5% of precipitation. The greatest deviations in the model residuals are overestimates in the north-northwestern part of the state where sandy soils are common. Limited pockets of this region are over-estimated by up to 10 cm/year; however, the median residual for the entire glaciated portion of Indiana (north of the Wisconsin glacial limit) is a slight underestimate of -0.04 cm/year. Groundwater-quality data collected from domestic wells, public water supplies, and commercial/industrial locations were used to identify settings that have experienced groundwater contamination. To assess the sensitivity of near-surface aquifers to contamination, locations with detectable contamination were compared to the groundwater-recharge data to classify the region into low, moderate, high, and very high aquifer sensitivity.

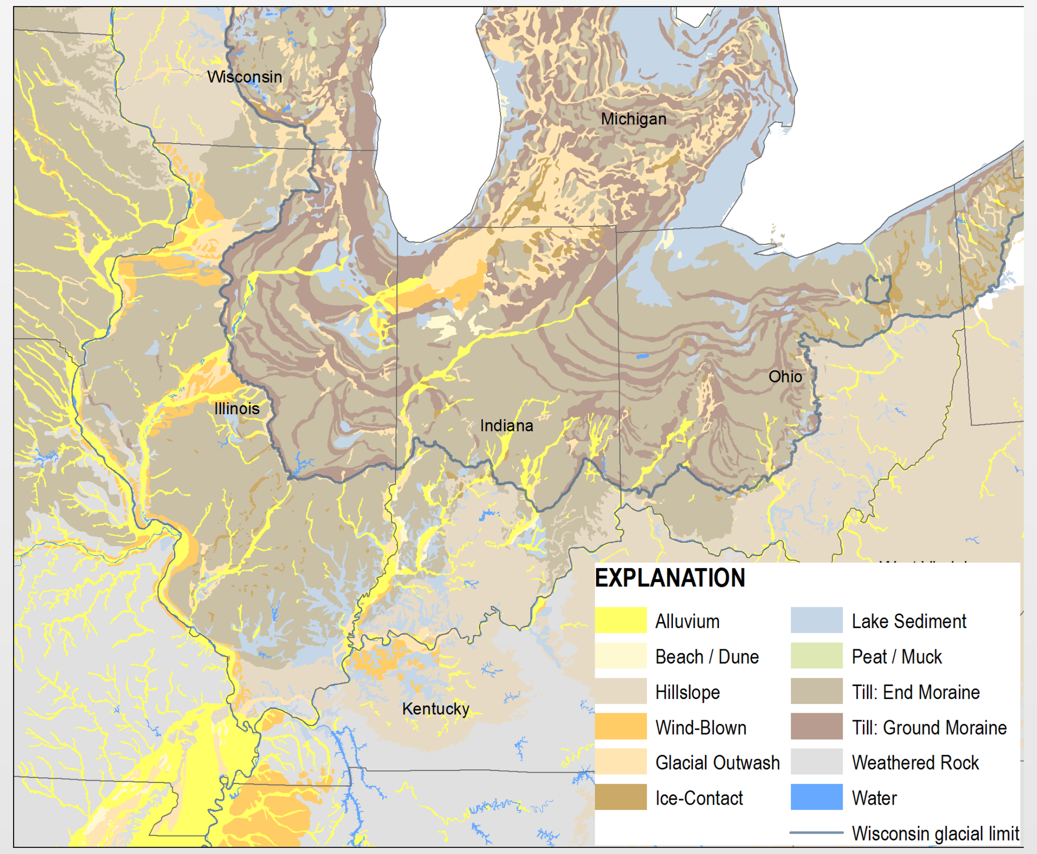

Figure 1. Glacial geology and location of the Wisconsin glacial limit in Indiana. Study area was to the north of the glacial limit in Indiana.

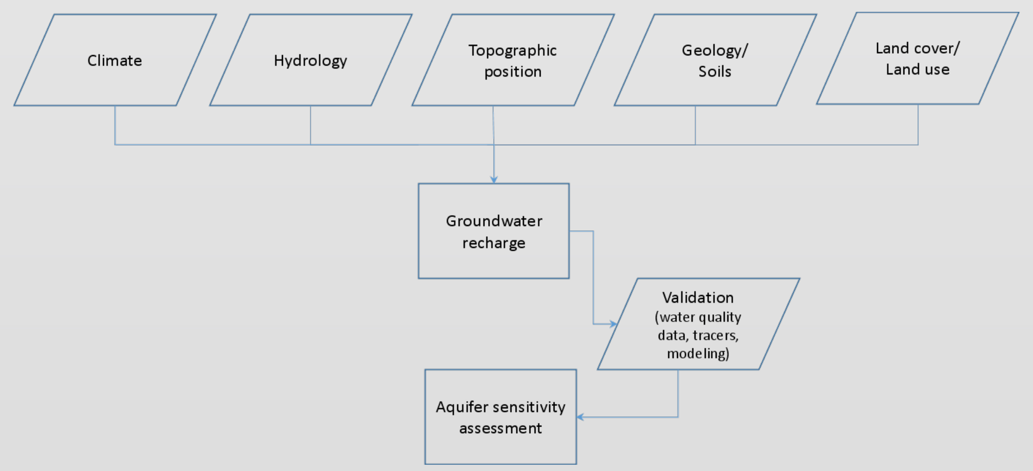

Figure 2. Conceptual model of factors that control recharge, which influences near-surface aquifer sensitivity to contamination.

Conceptual Model

The regression model is based on a conceptual model that views groundwater recharge in humid temperate climates to be controlled by near-surface processes such as climate, as well as on the topographic position, the geology and land cover, and the position of the water table (Figure 2). In the study area, the unsaturated thickness is not as great as in more arid environments, so the depth to the water table and the characteristics of the unsaturated zone are important.

The null hypothesis is that groundwater recharge is not related to landscape geologic or land-cover variables. The dependent variable in the study was groundwater recharge (cm/year), which was determined by calculating the proportion of baseflow to runoff in stream hydrographs (*n*=189), determined using a local-minimums approach with a recession slope test. For this study, the attributes that were considered are shown in Table 1, accompanied by source and scale information. The conceptual model further guiding this analysis defines aquifer sensitivity as the probability that contaminants applied on the ground surface will infiltrate and reach the water table; a concept that directly correlates with the rate of recharge.

Type	Variable	Source	Scale	Units	Significance (p=0.001)
Physical Characteristics	Catchment area	National hydrography dataset (NHD)	200K (100-m)	Sq kilometers	
	Elevation	National Elevation Dataset (NED)	200K (100-m)	Meters	
	Slope	NED derivative	200K (100-m)	Gradient (dimensionless)	
	Aspect	NED derivative	200K (100-m)	Degrees	
	Compound Topographic Index (CTI)	NED derivative	200K (100-m)	dimensionless	
	Direct insolation duration	NED derivative	200K (100-m)	Watt hours per square meter - hours	
	Flow accumulation	NED derivative	200K (100-m)	Count	
	Height above local base level	NED derivative	200K (100-m)	Count	
	Downstream flow length	NED derivative	200K (100-m)	Meters	
	Stream density	NED derivative	200K (100-m)	Count/km2	
	Stream order	NED derivative	200K (100-m)	Rank	
Hydrometeorological	Hydrologic variables (modeled) • Water-table elevation • Unsaturated thickness	Indiana Geological Survey (IGS)	200K (100-m)	- Meters - Meters	* Water table elev
	Growing degree days	Climate Research Unit, Univ. of East Anglia	0.5 degree (50-km)	Days	
	Temperature, precipitation	PRISM	30 sec (800-m)	Degrees C, and mm	
Geological	Surficial geology	Indiana Geological Survey (IGS)	250K (Indiana)	---	
	Sand thickness (total borehole)	Indiana Geological Survey (IGS)	400K	Meters	* Sand thickness
	Clay thickness (total borehole)	Indiana Geological Survey (IGS)	400K	Meters	
	Surficial clay thickness	Indiana Geological Survey (IGS)	200K	Meters	* Surficial clay thickness
	Depth to bedrock	Indiana Geological Survey (IGS)	250K	Meters	
Soils	Soil Parent Material	Natural Resources Conservation Service (NRCS) SSURGO	12K – 20K	---	
	Available water storage 150cm	Natural Resources Conservation Service (NRCS) SSURGO	12K – 20K	---	
	Saturated hydraulic conductivity	Natural Resources Conservation Service (NRCS) SSURGO	12K – 20K	micrometers per second	
Land Cover	Land cover (30 classes evaluated; local and cumulative upstream area)	National Land Cover Dataset (NLCD)	24K (30-m)	Area, sq meters	* Wetland density * Upstream water * Upstream forest

Table 1. Variables considered. Only variables that demonstrated statistical significance at *p*=0.001 (denoted by *) were included in the final analysis.

Statistical Model Background

The spatial data were prepared for input into a multivariate regression model, intended to investigate the predictive power of the landscape variables on groundwater recharge, and therefore aquifer sensitivity to contamination. This method is appropriate for this analysis because errors in the measurement of the dependent variable, groundwater recharge, are probable, and the multiple regression appropriate because random error computation is incorporated into the parameter estimates and significance tests.

The regression model had the form: $Log\ Y = b0 + b1\ x1 + b2\ x2 + b3\ x3 + b4\ x4 + \dots bm\ xm + e$

Where Y is recharge (cm/year), $x1 – xm$ are independent variables that are used to improve the prediction of recharge, and $b0 – bm$ are best-fit regression coefficients estimated using ordinary least squares, and e is an error term.

Multiple Regression Analysis

Each of the six landscape variables was included in the regression analysis as an independent variable. The student's *t* statistics (Table 2) indicate that all of the variables were contributing significantly to the prediction at the 99.9% confidence level. Figure 3 is a scatterplot of the observed versus predicted recharge values, as calculated from the regression equation. The scatter around the 1:1 line is well distributed for the bulk of observed recharge values; however, the predicted values are under-estimated relative to observed values at the high end of the range, suggesting that the current model is conservative when used for aquifer sensitivity assessment. Figure 4 shows the independent variables in order of significance. The null hypothesis is that the regression parameter equals zero is rejected at the 99.9% confidence level when the calculated t-ratio exceeds the critical value of 2.6 for degrees of freedom = 182.The magnitude of *t* indicates the relative importance of each independent variable to the prediction of groundwater recharge. The best-fit model combined variables that elucidated details from the geology and vadose zone, with lesser influence from land-cover variables. Although some variables appear to be describing the same effect on groundwater recharge (e.g., water-table elevation, wetland density), the correlation matrix does not show interdependence, and the predictive power of the regression equation is diminished when either variable is removed. The resultant spatially distributed groundwater recharge map (cm/year) is shown in Figure 5.

Variable ID	Variable name	Parameter Estimate, b_i	Standard Error (SE)	t-ratio ($t=b_i/SE_{bi}$)	Confidence Interval Lower 99.9%	Upper 99.9%
V_1	Clay thickness	-0.032	0.007	-4.62	-0.055	-0.009
V_2	Sand proportion	0.037	0.006	6.01	0.016	0.057
V_3	Upstream water	0.262	0.070	3.75	0.028	0.494
V_4	Wetland density	0.155	0.043	3.59	0.010	0.300
V_5	Water-table elevation	-0.049	0.008	-6.31	-0.075	-0.023
V_6	Upstream forest	0.013	0.013	3.56	0.003	0.087

Table 2. Parameter estimates, standard errors, and t-ratios. Total sample size, *n* = 189. R^2 = 0.55. Multiple R = 0.74. RMSE = 1.5.

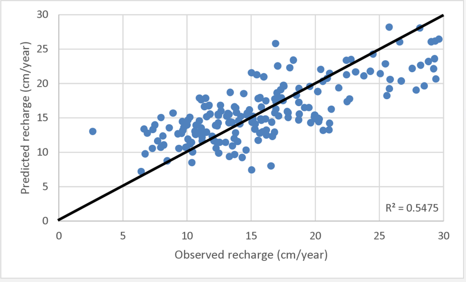

Figure 3. Scatter plot of observed versus predicted values of recharge (cm/year) using the multiple regression equation. Black line is 1:1 line.

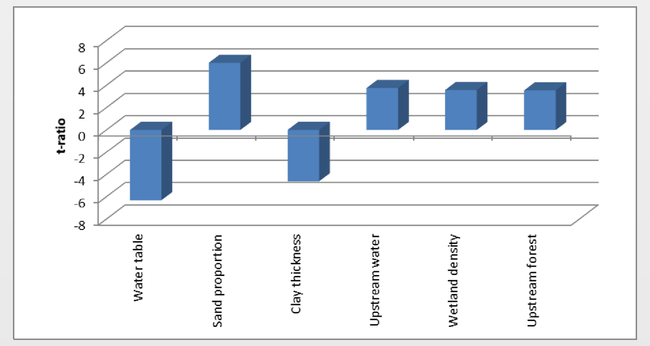

Figure 4. Computed *t*-ratios of significant parameters used in the multiple regression analysis.

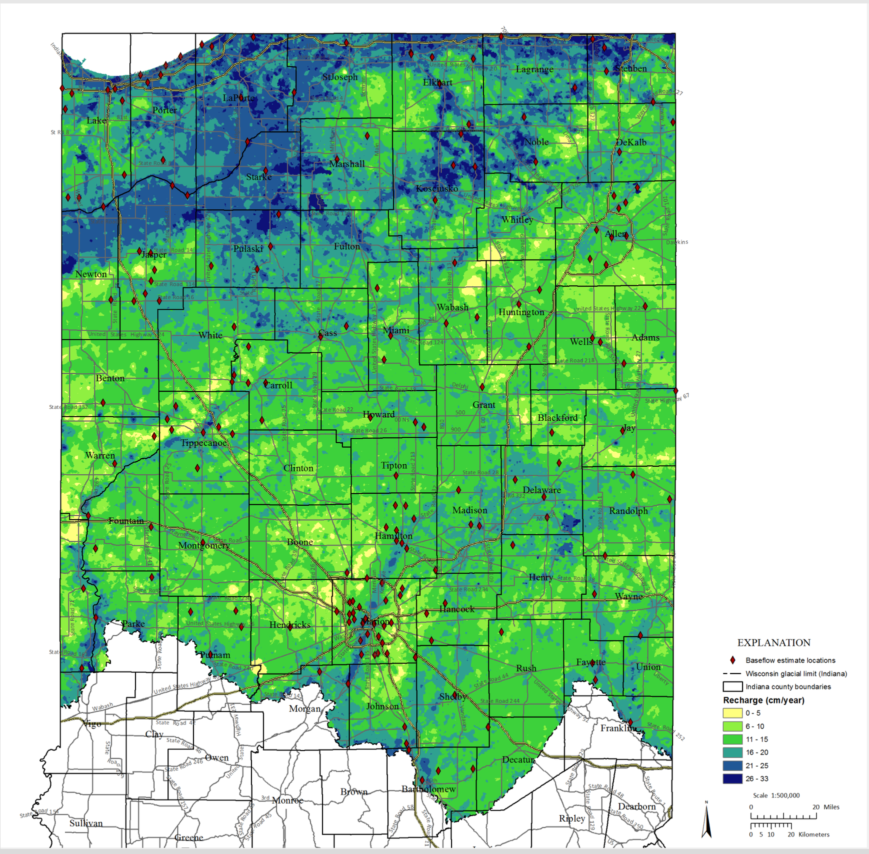

Figure 5. Predicted groundwater recharge rates (cm/year) based on multiple regression analysis.

Relating groundwater recharge to near-surface aquifer sensitivity to contamination

As the transport of most potential groundwater contaminants from the surface to the water table occurs in the aqueous phase, recharge represents a significant control on aquifer sensitivity to contamination (Robins, 1998). Maps of aquifer sensitivity identify locations where groundwater is exposed to rapid infiltration of potential contaminants applied at or near the ground surface and are often based on physical characteristics of surficial deposits. Water-table recharge is controlled by the coupled processes of soil-water infiltration, unsaturated-zone storage, and vertical percolation. The spatial distribution of recharge, due to characteristics of the land surface, soil and unsaturated zone, land cover, and so on, can be significant and needs to be accounted for in regional aquifer sensitivity studies (Batelaan and De Smedt, 2007). Areas of focused recharge allow contaminants to migrate rapidly through the unsaturated zone to underlying aquifers (Scanlon et al., 2002). Therefore, knowing where recharge rates are likely to be higher or lower within a larger watershed could lead to an improved basis for evaluating aquifer sensitivity to contamination.

For example, topography influences whether precipitation and contaminants applied at the surface will run off (often the case with steeper slopes) or will remain on the surface long enough to infiltrate and ultimately contribute to recharge (Toth, 1963). The role of soil media is important in the infiltration process, as texture controls the ability of a soil to retain or to drain soil moisture in both saturated and unsaturated conditions (Robins, 1998). If the soil has a very fine texture, water may stay on the surface until it has been evaporated or transpired back into the atmosphere. Similarly the impact of the vadose zone can promote recharge through the presence of fractures and zones of coarsely textured sediment, or it can retard recharge due to the presence of restrictive layers (Cuthbert et al., 2010). Depth to water has a strong influence on recharge, as shallow water tables will generally exhibit greater recharge due to shorter flow paths (Batelaan and De Smedt, 2007). As these authors note, processes that control the migration of water into the subsurface can also control the migration of pollutants into the subsurface.

The resultant map of aquifer sensitivity, a direct product of the spatially distributed estimates of groundwater recharge, was derived by classifying the recharge map into sensitivity categories by standard deviation and then comparing the classes to known areas of contamination (by water-soluble pollutants such as nitrates, sodium, chloride, as well as dissolved metals, pesticides, herbicides, and volatile compounds). The groundwater pollutants served as a validation data set to determine if the aquifer sensitivity classes could be predicted using groundwater recharge. For example, the highest contaminant concentrations and highest density of pollutants above the maximum contaminant level (MCL) occur in the highest standard deviation category ($> 1.5\ \sigma$, very high aquifer sensitivity) and cover 7% of the study area, whereas the lowest concentrations and lowest density of contaminants occur in the lowest standard deviation category ($< 0.5\ \sigma$, low aquifer sensitivity) and cover 25% of the study area. Only the dissolved metals occurred throughout the entire domain, without correspondence to recharge rate. Further investigation shows that these are likely the results of leaking underground storage tanks, which circumvent the natural flow from the surface to the water table. This highlights a flaw in considering a conceptual model only of diffuse recharge to the water table; obviously, lateral flow and focused vertical recharge can occur through other mechanisms. Point sources of contamination directly into the subsurface are a known concern.

Recharge rates (cm/year)	Standard Deviation Range	Area (km²)	% Area	Aquifer Sensitivity Category
0.0 – 11.5	< 0.5	14,815	25%	Low
11.5 – 16.5	-0.50 – 0.50	28,227	48%	Moderate
16.5 – 21.5	0.5 – 1.50	11,157	19%	High
21.5 – 34.5	> 1.5	4,174	7%	Very high

Conclusion

Although the multiple regression analysis did not explain all of the variability in the characteristics correlated to groundwater recharge, over half of the variability in the spatial distribution of diffuse water-table recharge can be explained by the landscape-level variables selected for the analysis. This suggests that the conceptual model of recharge by geologic, topographic, and land cover (including soils) factors is fundamentally correct at the scale of the analysis. Additional predictive power could likely be gained by a more complex model. One goal of this work was to develop the simplest highest quality (*p*=0.001) model to elucidate the governing variables for recharge, and therefore, diffuse aquifer sensitivity in heterogeneous glacial materials in a humid environment. Local or large-scale variables or processes might explain the remaining variability in the data.

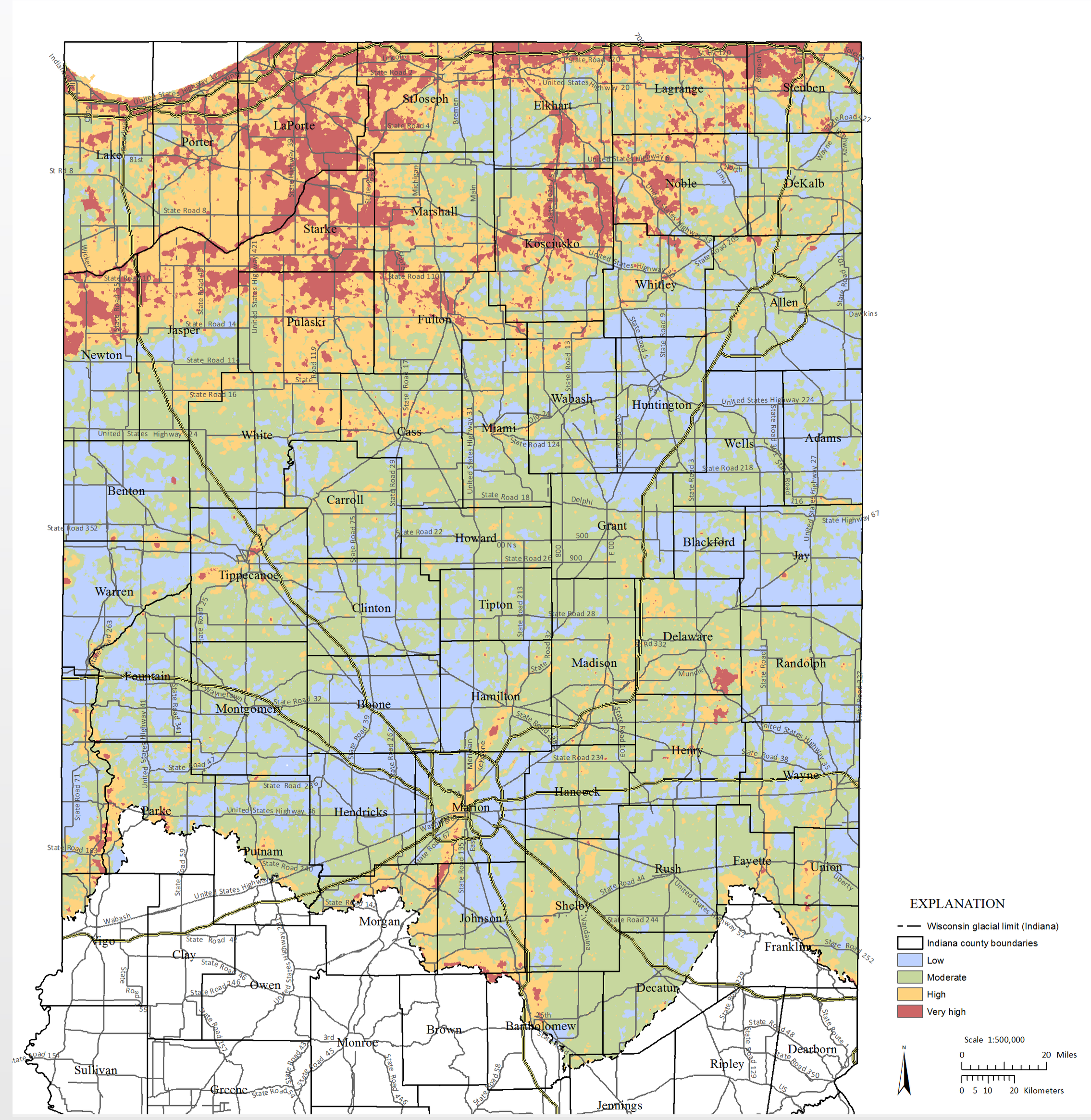

Figure 6. Aquifer sensitivity assessment based on multiple regression analysis of recharge rates in near-surface aquifer materials.

References

Batelaan, O., and De Smedt, F., 2007. GIS-based recharge estimation by coupling surface–subsurface water balances. Journal of Hydrology 337, p. 337-355.

Cuthbert, M.O., Mackay, R., Tellam, J.H., Thatcher, K.E., 2010. Combining unsaturated and saturated hydraulic observations to understand and estimate groundwater recharge through glacial till. Journal of Hydrology 391, p. 263-276.

Fullerton D.S., Bush C.A., Pennell J.N., 2003. Surficial deposits and materials in the eastern and central United States (east of 102 degrees west longitude). Geological Survey Geologic Investigations Series, I-2789, U.S. Geological Survey, Denver, CO.

Robins, N.S., 1998. Recharge: the key to groundwater pollution and aquifer vulnerability. Geological Society of London, Special Publications 130, p. 1-5.

Scanlon, B.R., Healy, R.W., Cook, P.G., 2002. Choosing appropriate techniques for quantifying groundwater recharge. Hydrogeology Journal 10, p. 18-39.

Toth, J., 1963. A theoretical analysis of groundwater flow in small drainage basins. Journal of Geophysical Research 63(16), p. 4795-4812.