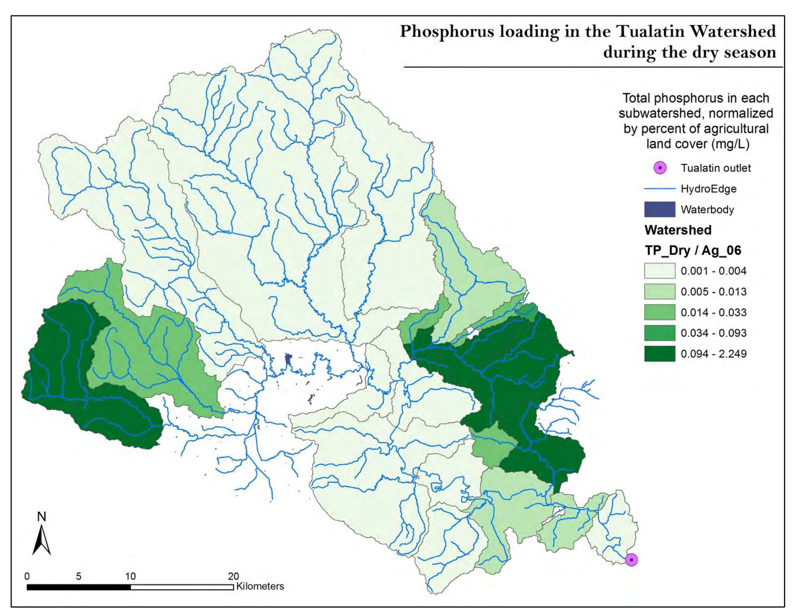
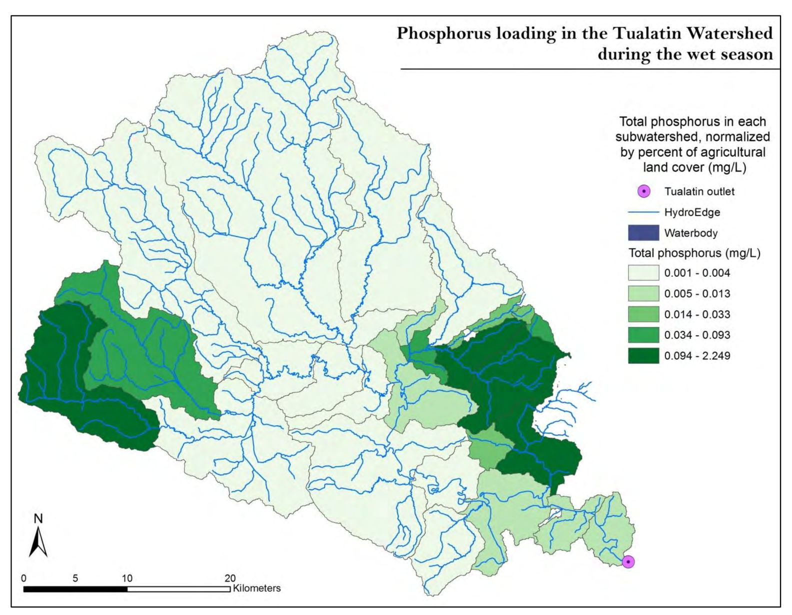
The relationship between phosphorus loading and agricultural land cover: Improved analysis with the use of an Arc Hydro stream network

SAMANTHA HAMLIN AND ERIC WATSON

Introduction

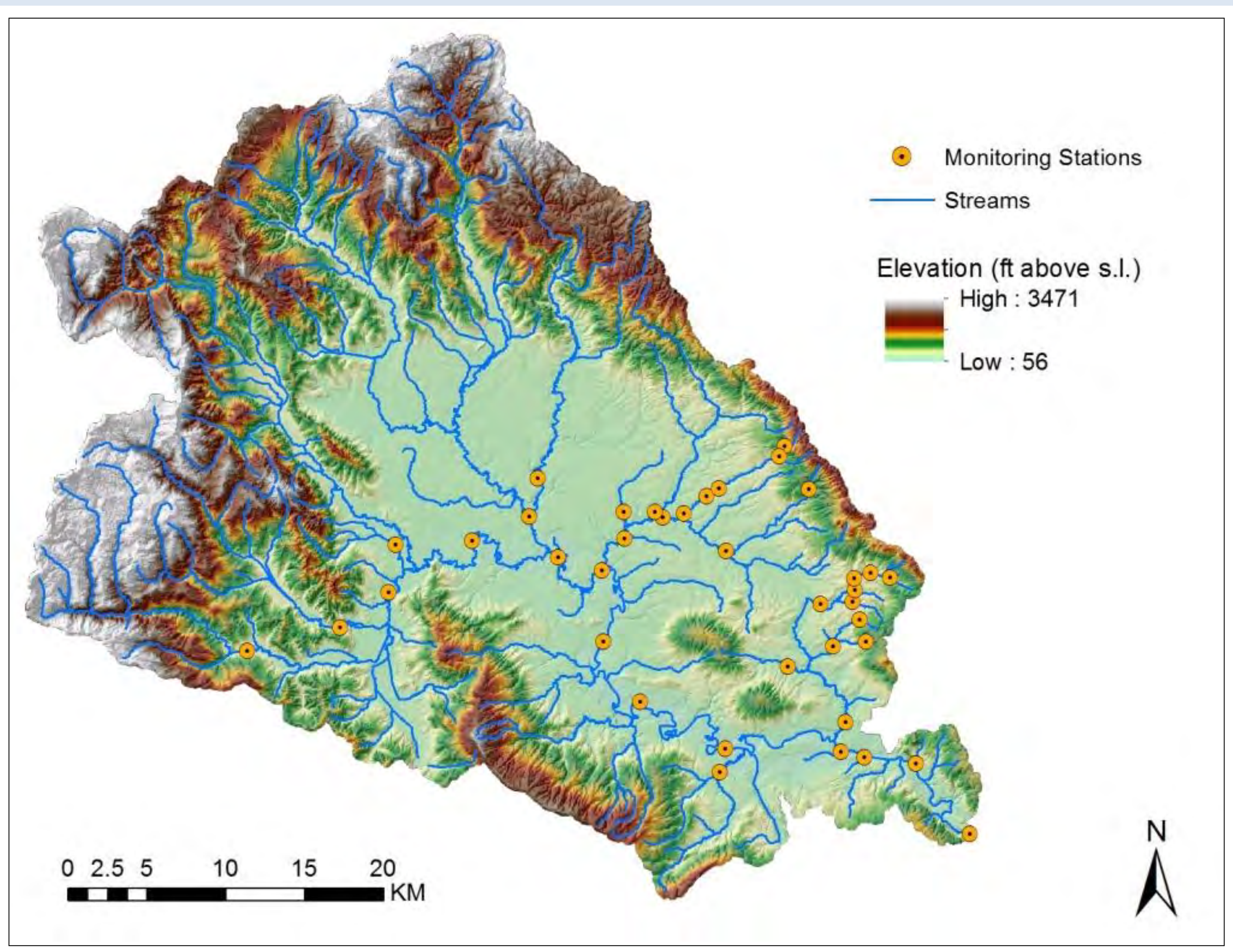
There is substantial evidence linking land use/land cover and nutrient runoff, particularly in agricultural areas (see examples below. Lowicki 2012; Uuemaa, Mander, and Marja 2013; Uuemaa, Roosaare, and Manden 2007). Statistically evaluating this land cover/water quality relationship requires several assumptions be met, most notably that the different explanatory variables be independent—a wholly inappropriate assumption for hydrological networks. Statistical analysis frequently addresses this issue by applying different weighting schemes to the independent variables to correct for violating the assumption of independence. We opted to build a hydrological network in our study area to inform this weighting scheme. Our study aims to evaluate whether the use of network distance with a spatial weighting coefficient can improve analyses that examine this relationship between landcover type and water quality parameters—specifically, agricultural land cover and phosphorus loading.





Study Area

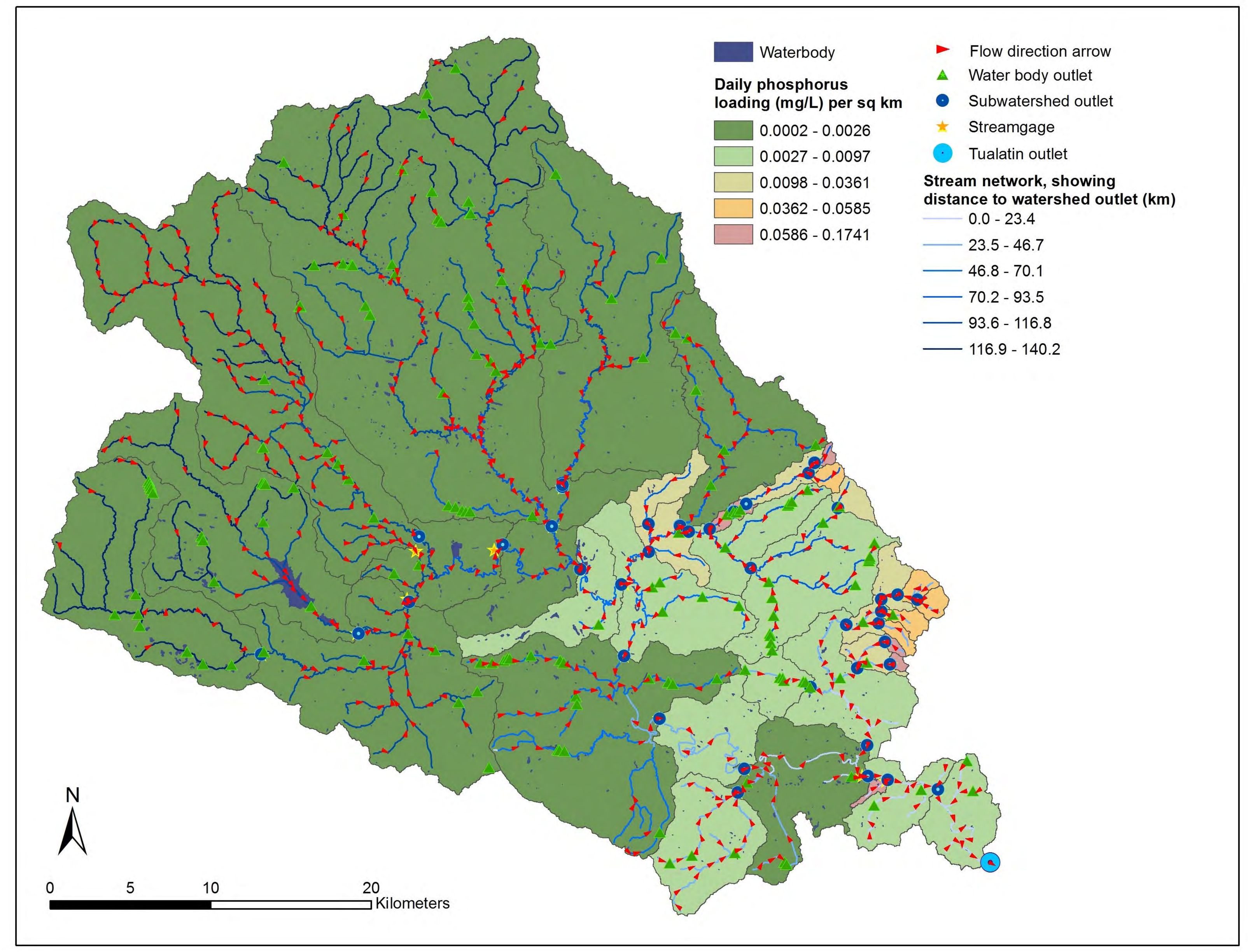
Tualatin Watershed, northwest Oregon: approximately 1,844 km²



Data Sources

USGS: Streamgage locations and water quality data
Oregon Geospatial Enterprise Office:
Stream reaches, water bodies, and watersheds
Multi-Resolution Land Characteristics Consortium: 2006 land cover data

Stream Network



Statistical Methods & Results

Data and Methods:

Streams within the TRB are monitored by Clean Water Services and Portland's Bureau of Environmental Services. Datasets from both agencies were included if they provided monthly measurements from 2004 to 2008, a temporal restriction chosen to coincide with the 2006 National Land Cover Database (NLCD) survey.

Data from two years before and after the NLCD survey were included to account for between-year variability. This temporal restriction reduced the number of available monitoring sites that provide total phosphorus data to 39. Because water quality metrics are not independent, a geometric seasonal mean was used to calculate each parameter.

Subbasins for each monitoring point were delineated using a DEM. Percent agricultural land cover within each subbasin was calculated from the 2006 NLCD.

An ordinary least-squares (OLS) regression was compared with two spatial lag models. The two spatial lag regressions differed in how they defined distances between monitoring points. The first model used Euclidean distances while the second model used the distances determined by the stream network created within the ArcHydro environment.

Results:

For dry season geometric means, the OLS technique produced an R-squared value of 0.25 (F = 12.66). The inclusion of spatial autocorrelation into the model greatly increased the observed relationship between agriculture and dry season total phosphorus. The two spatial lag models, using Euclidean and network distances, both performed equally well, resulting in R-squared values of 0.53 (F = 73.98).

Similarly, the distance-based spatial lag models outperformed the OLS model when predicting wet season total phosphorus concentration. In fact, no significant relationship was detected between percent agricultural land and wet season total phosphorus by the OLS model ($R^2 = 0.005$, F = 0.15). In contrast, when either Euclidean or network distances were included within a spatial lag model, the R-squared value increased to 0.48 (F = 82.91).

Arc Hydro Stream Network Creation

The stream network was generated using four base data shapefiles: stream reaches, water quality monitoring stations, water bodies, and subwatersheds.

The first step was to create a HydroJunction feature representing all water body and subwatershed outlet points. To do so, some basic GIS processing was done as a preliminary step to building some geometric networks. Creating the network places network junctions at the intersections of all network edges (i.e., Streams)—these network junctions are required at the outlet of each water body and subwatershed.

To do this, stream reaches were first intersected with subwatersheds, creating a line output file with 3,645 stream segments; this line feature was then intersected with water bodies, reducing the number of stream segments that intersect with both subwatershed boundaries and water bodies to 547. At this point, two intermediate geometric networks were built from these files using the intersect tool (and switch selection when needed) to determine both water body and watershed outlets. Manual editing ensured that each watershed had only one outlet point, watersheds didn't share the same outlet point, no outlet point was located at the intersection of streams, each outlet point was located on the primary stream of its subwatershed, and each outlet point was located inside or on the border of the subwatershed it represents. Hydro IDs were then assigned to all junctions and stream segments within these networks with Arc Hydro.

A new geometric network was created—the final stream network, containing stream segments and all junction points. Network connectivity was checked and manually corrected as needed using the Geometric Network Editor. Once full connectivity was confirmed, Arc Hydro was used to set and store flow direction throughout the Basin. Flow patterns were checked using the upstream trace function in the Geometric Network Editor and flow arrows were displayed for visual confirmation (see Stream Network figure).

Arc Hydro was then used to "compute length downstream for edges [streams]" and for junctions [outlets]. This final step assigned relative network distances to the stream network.

Discussion

The distance-weighted spatial lag models greatly outperformed the standard OLS regression technique. This reflects the inherent spatiality that exists within stream networks. The conditions observed at points along the stream are connected through flow and cannot be analyzed as independent samples. This effect was observed most dramatically when analyzing wet season total phosphorus.

Interestingly, while spatial models greatly surpassed the aspatial OLS technique, we found no difference in predictive strength when comparing Euclidean and network distance schemes. This might be explained by the scale of our study. On a different scale, the distinction between Euclidean and network distances may be more significant.

Additionally, the datasets we included for streamlines and waterbodies were obtained from a statewide geospatial clearinghouse. Our results may have differed if we had delineated our own streamlines.

Future research should consider the effects of symmetry within the distance weighting scheme. Symmetry defines whether upstream and downstream neighbors effect one another to the same degree. The unidirectional nature of flow within a stream network suggests that distances should be expressed asymmetrically.

References

Łowicki, Damian. 2012. "Prediction of flowing water pollution on the basis of landscape metrics as a tool supporting delimitation of Nitrate Vulnerable Zones." *Ecological Indicators* 23:27–33.

Uuemaa, Evelyn, Ülo Mander, and Riho Marja. 2013. "Trends in the use of landscape spatial metrics as landscape indicators: A review." *Ecological Indicators.*

Uuemaa, Evelyn, Jüri Roosaare, and Ülo Manden. 2007. "Landscape metrics as indicators of river water quality at catchment scale." *Nordic Hydrology* 38:125–138.