USGS
science for a changing world

National Water Quality Assessment Program
In Cooperation with
North Dakota Department of Health and Minnesota Pollution Control Agency

Structural Equation Modeling Of Phosphorus in the Red River Basin

Karen R. Ryberg, U.S. Geological Survey
2016 North Dakota Water Quality Monitoring Conference
March 4, 2016

Project Mileposts

- Literature review
- Calculation of phosphorus loads
- Determination of multiple working hypotheses
- Assembly of causal data
  - Estimation of missing values
- Structural equation model

USGS

Literature Review

What have people said about the causes of total phosphorus in the Red River of the North?

Published in: Ryberg, K.R, 2015, The impact of climate variability on streamflow and water quality in the North Central United States: Fargo, North Dakota State University, Ph.D. dissertation, 277 p.

Causal Attribution

Suggested that the higher total phosphorus (TP) concentrations in the Pembina River might be attributable to soil characteristics or agricultural practices and the topography, which is steeper than most of the Basin.

Suggested that soils or agricultural practices contributed to the high phosphorus in the Rabbit River and that effluent from Fargo and Moorhead wastewater treatment plants likely contributed to the high phosphorus in the Red River near Perley and Halstad.

Trends may have been related to livestock-management changes.

USGS

Structural Equation Modeling

SEM

Supports the testing of causal hypotheses.

USGS

Calculation of Annual Total Phosphorus Loads at 6 Sites

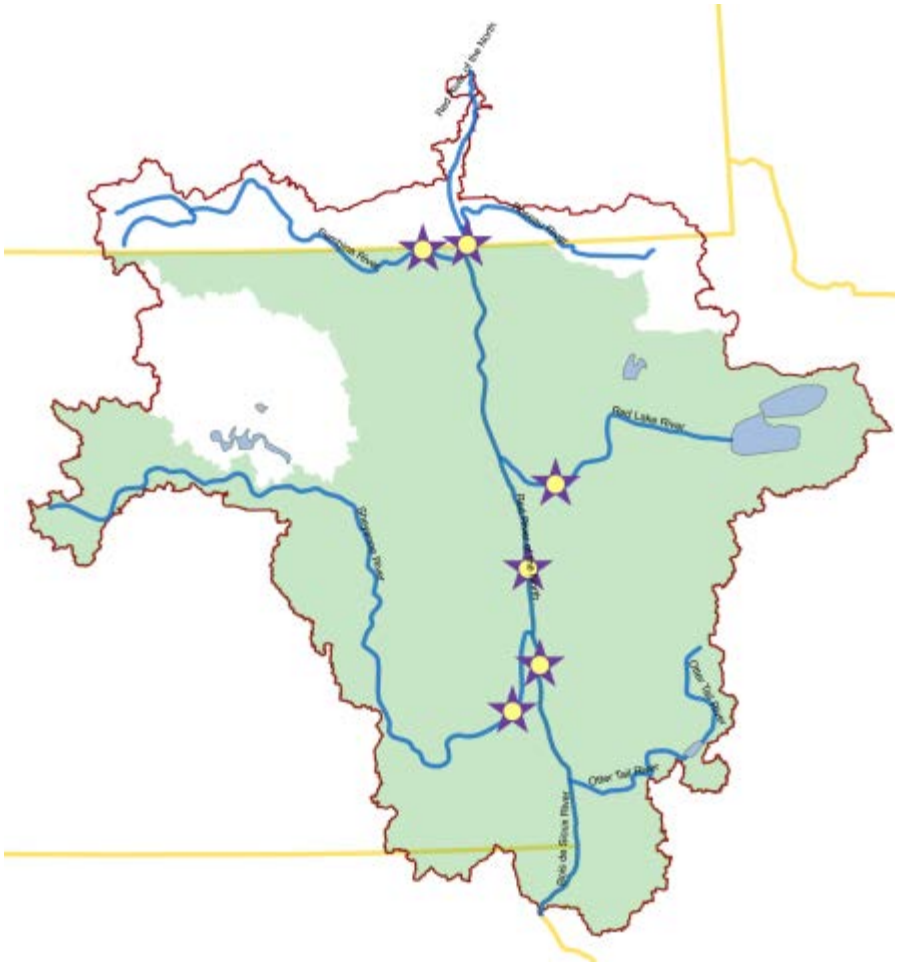
Published for Fargo/Moorhead and Emerson in:

Ryberg, K.R, 2015, The impact of climate variability on streamflow and water quality in the North Central United States: Fargo, North Dakota State University, Ph.D. dissertation, 277 p.

Ryberg, K.R., Akyüz, F.A., and Lin, W., 2015, Changes in total phosphorus concentration in the Red River of the North Basin, 1970–2012: : American Society of Agricultural and Biological Engineers ASABE/CSBE North Central Intersectional Meeting, Fargo, N. Dak., April 10-11, 2015, paper number RRV15-054, 9 p. http://dx.doi.org/10.13031/rrv2015054.

Sites with Sufficient Data for Load Calculation

Loads will be published in a USGS Data Release



USGS

Analyzed using WRTDS

Weighted Regressions on

- Time,
- Discharge (streamflow),
- and Season.

Smoothing model that computes estimates of concentration and flux (load) for every day in the study period.

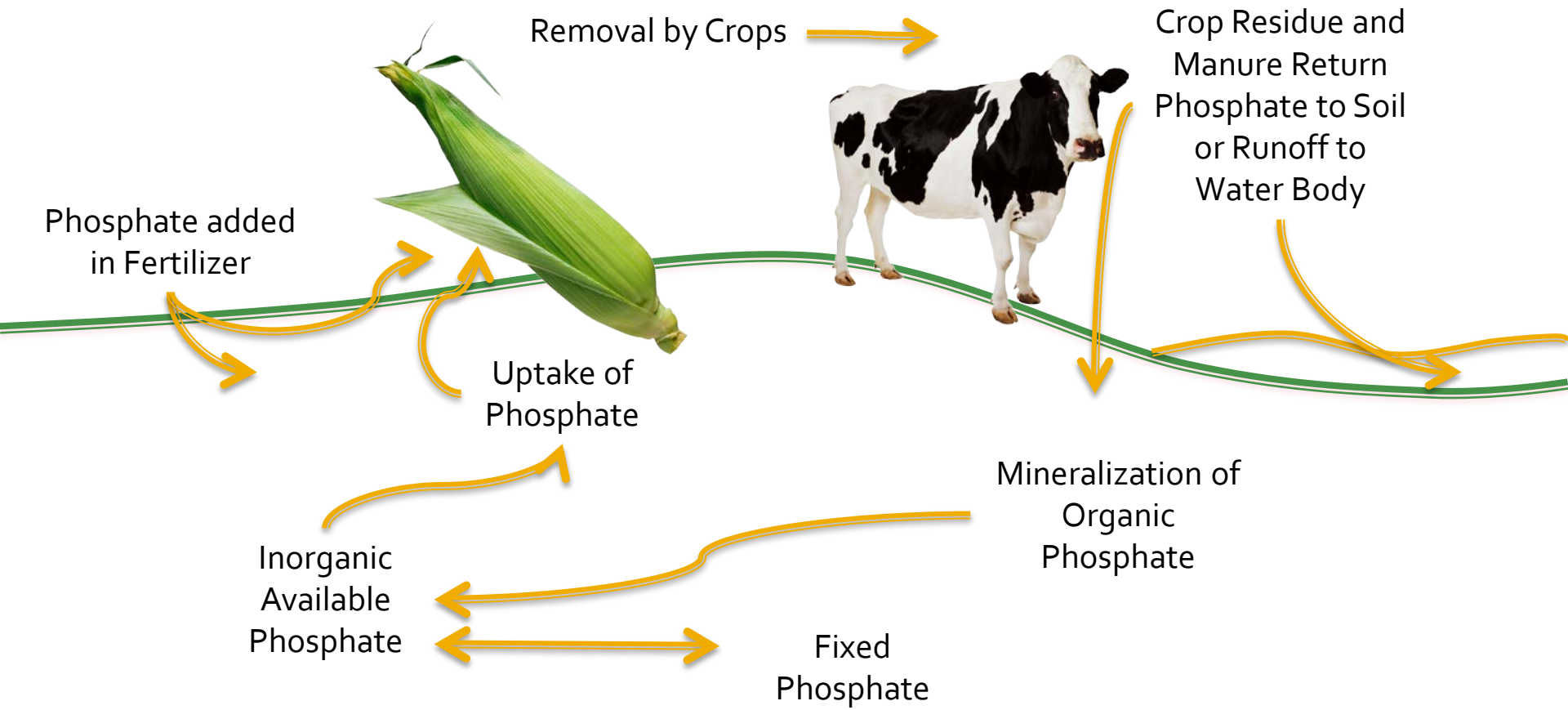
USGS

Hirsch, R.M, Moyer, D.L., and Archfield, S.A., 2010, Weighted Regressions on Time, Discharge, and Season (WRTDS), with an Application to Chesapeake Bay River Inputs, Journal of the American Water Resources Association p. 857-880.

The Phosphorus Cycle

A Basis for Multiple Working Hypotheses

The Phosphorus Cycle



Modified from U.S. Environmental Protection Agency, http://www.epa.gov/agriculture/ag101/impactphosphorus.html

USGS

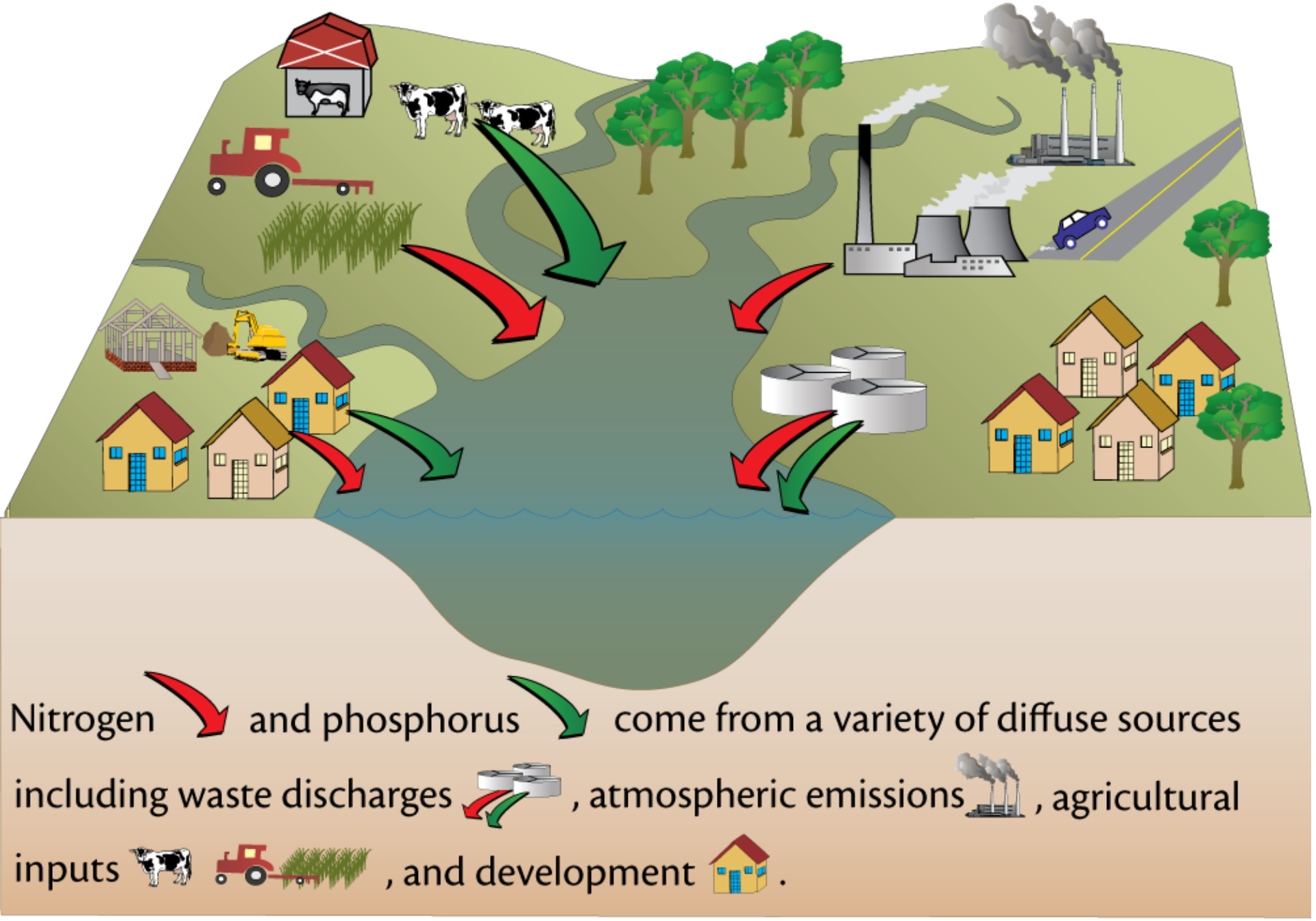
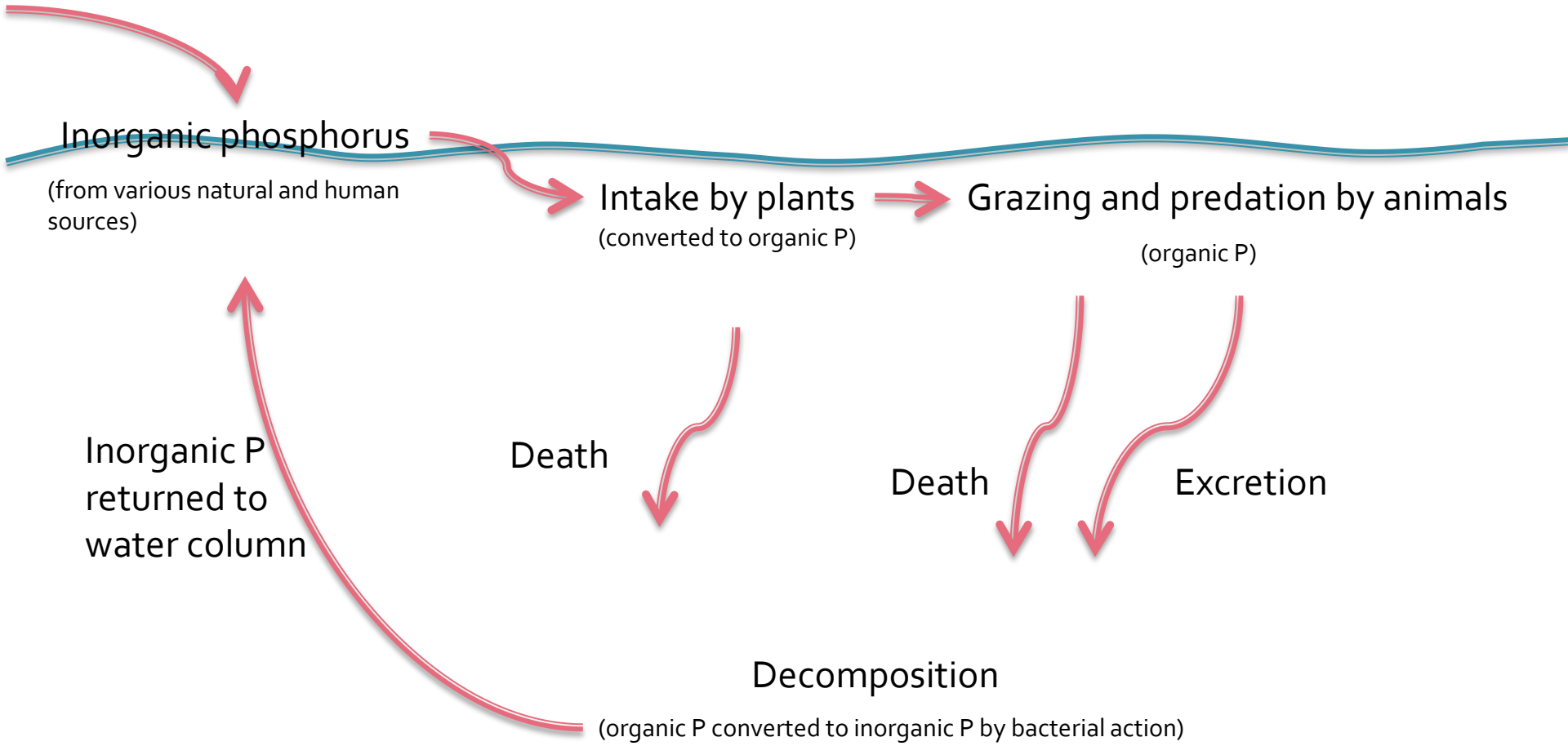


Diagram courtesy of the Integration and Application Network (ian.umces.edu), University of Maryland Center for Environmental Science. Source: Lane, H., J.L. Woerner, W.C. Dennison, C. Neill, C. Wilson, M. Elliott, M. Shively, J. Graine, and R. Jeavons. 2007. Defending our National Treasure: Department of Defense Chesapeake Bay Restoration Partnership 1998-2004. Integration and Application Network, University of Maryland Center for Environmental Science, Cambridge: MD.

The Phosphorus Cycle



USGS

Modified from U.S. Environmental Protection Agency, http://water.epa.gov/type/rsl/monitoring/vms56.cfm

Subset Multiple Working Hypotheses

Causal factors	Potential influence	Comments (Additional Information/Mechanism)	Data	Decision and Justification
Agriculture				
Fertilizer	Phosphate fertilizer makes its way to the stream in crop residue or adsorbed to sediments	Phosphates "not very mobile in soils and sediments" (Hem, 1985). However, the Red River Basin is subject to overland flooding (visualized in Ryberg et al., 2007). Overland flooding can transport phosphate adsorbed to sediment to the [illegible].	National phosphorus fertilizer data compiled by the USGS	Use
Animal waste	Animal [illegible] particularly in the form of manure [illegible] as fertilizer contributes [illegible] to streams	Waste could be mobilized [illegible] directly by overland flow or be part of soils eroded to streams	Manure data [illegible] by the USGS. Poultry data, would have to interpolate between years	Maybe.
Soil erosion (CRP and percent clay soil as surrogates)	Increased soil erosion would bring more phosphorus to the river	The conservation reserve program (CRP) is a surrogate for soil erosion as it represents acreage taken out of agricultural production. CRP related to other climatic and economic factors.	Data compiled by the USGS	Use CRP data.
Percent agricultural land	Increased agricultural land-use would represent increases on fertilizer and potential soil erosion, both of which contribute to fertilizer	Correlated with fertilizer and CRP.	Data compiled by the USGS	No. The percent agricultural land has been very high, and stable, in the Red River Basin for the entire study period and, therefore, would not provide much explanatory information.

USGS

Data for Potential Causal Factors

The BIGGEST Challenge

Will be published in a USGS Data Release

Data Determined Feasible to Use

- Annual precipitation
  - Including a number of measures of annual seasonal precipitation
- Land use
  - Percent land developed, semi-developed, in agricultural production
- Percentage of cropland in Conservation Reserve Program
- Crop types
  - Percent of agricultural land in harvested wheat, corn, and soybeans
- Phosphorus from fertilizer and from manure
- Total phosphorus load from wastewater treatment plants in the basin

USGS

The Challenge

Widely varying periods of record

Some annual series, some every 5 years, every 4 years, or varying measurement intervals

Annual estimates needed

At right is an example of the missing data problem for select years during the study period.

Year	Land Use	CRP	Crop Type	P from Fertilizer	P from Manure	WWTP Load
1972	✗	NA	✗	✗	✗	✗
1985	✗	NA	✗	✓	✗	✗
1986	✗	✓	✓	✓	✗	✓
1987	✗	✓	✗	✓	✓	✗
1988	✗	✓	✗	✓	✗	✓
1989	✗	✓	✗	✓	✗	✗
1990	✗	✓	✗	✓	✗	✓
1991	✗	✓	✗	✓	✗	✗
1992	✓	✓	✓	✓	✓	✓
2012	✓	✓	✓	✗	✗	✗

USGS

Missing Years Estimated

- In some cases, such as agricultural land use, there was little change over multi-year periods, so missing years were filled in with linear interpolation.
- Many other variables have much more year-to-year variability. They were estimated using imputation of missing values with principal components analysis.

USGS

Structural Equation Modeling

Preliminary

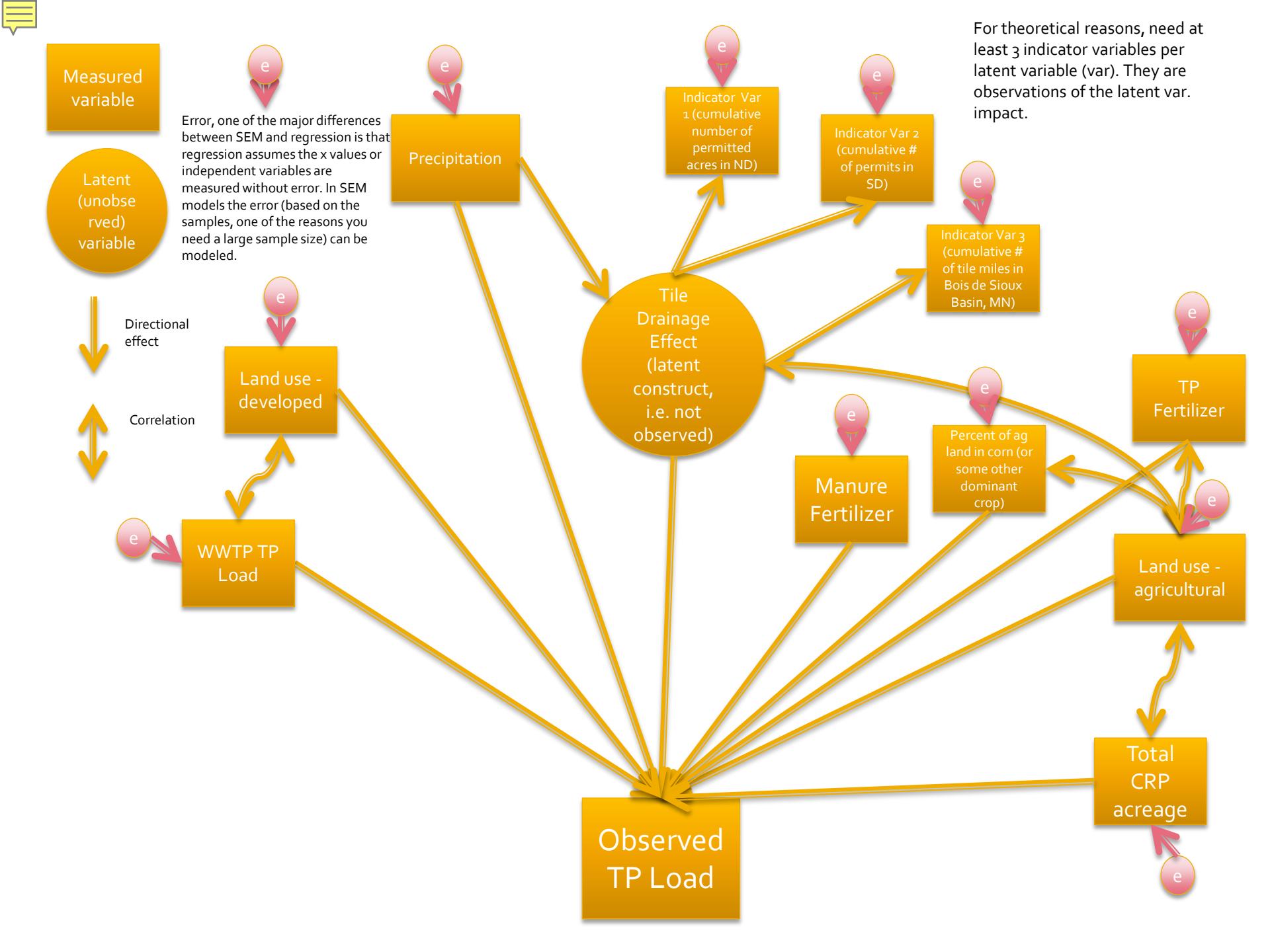
Will be published in a journal article

Structural Equation Modeling

"Structural equation models (SEMs), also called simultaneous equation models, are multivariate (i.e., multi-equation) regression models. Unlike the more traditional multivariate linear model, however, the response variable in one regression equation in an SEM may appear as a predictor in another equation; indeed, variables in an SEM may influence one-another reciprocally, either directly or through other variables as intermediaries. These structural equations are meant to represent causal relationships among the variables in the model" (Fox, 2002).

Fox, John, 2002, Structural Equation Models: Appendix to an R and S-PLUS companion to applied regression, http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-sems.pdf

USGS


Measured variable
Latent (unobserved) variable
Directional effect
Correlation
e
Error, one of the major differences between SEM and regression is that regression assumes the x values or independent variables are measured without error. In SEM models the error (based on the samples, one of the reasons you need a large sample size) can be modeled.
For theoretical reasons, need at least 3 indicator variables per latent variable (var). They are observations of the latent var. impact.
Precipitation
Indicator Var 1 (cumulative number of permitted acres in ND)
Indicator Var 2 (cumulative # of permits in SD)
Indicator Var 3 (cumulative # of tile miles in Bois de Sioux Basin, MN)
Tile Drainage Effect (latent construct, i.e. not observed)
Land use - developed
WWTP TP Load
Manure Fertilizer
Percent of ag land in corn (or some other dominant crop)
TP Fertilizer
Land use - agricultural
Total CRP acreage
Observed TP Load

Structural Equation Modeling

In pursuit of knowledge, every day something is acquired;
in pursuit of wisdom, every day something is dropped.

- Lau Tzu

USGS

Actual Working Model

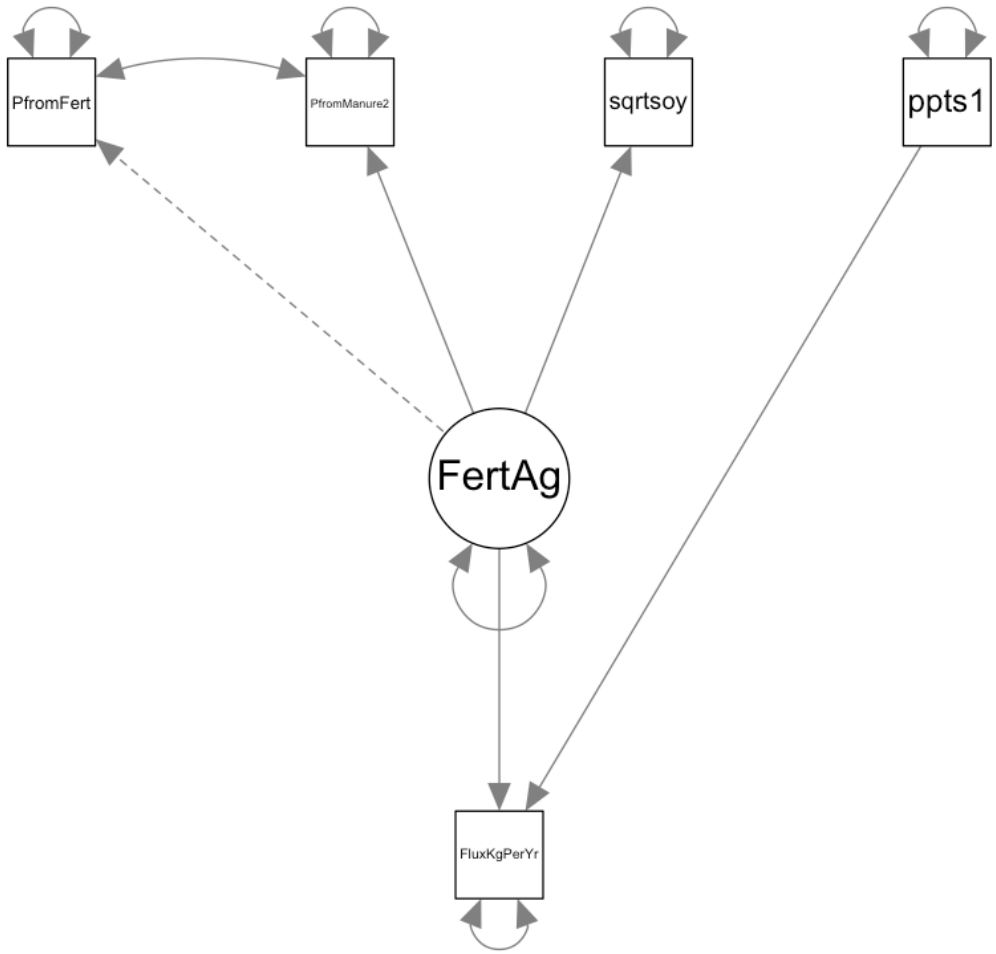
Explains about 60% of the variability in total phosphorus flux in the Red River at Emerson, Manitoba.

Measured variable

Latent variable

Directional effect

Correlation



USGS

More to come….

Karen R. Ryberg, Ph.D.
Statistician
USGS ND Water Science Center

kryberg@usgs.gov

Thanks to James Falcone, USGS National Water Quality Assessment Program, for GIS processing of data to study basins.

Thanks to Tammy Ivahnenko, USGS Water-Use Data and Research Program, and the ND Department of Health, for wastewater treatment plant data.

USGS