**Assessing interannual variability in nitrogen sourcing and retention through hybrid Bayesian watershed modeling**

**Jonathan W. Miller, Kimia Karimi,** **A. Sankarasubramanian, and Daniel R. Obenour**

Department of Civil, Construction and Environmental Engineering (Miller, Sankarasubramanian, Obenour) and Department of Geospatial Analytics (Karimi), North Carolina State University

**Text S1**: RStan code used for pre and post-1980 model

data {

int nl; //3600 subwatersheds separated by year

int nr; //483 incremental watershed-years

int nd; //985 dischargers associated with “nr”

vector [nr] load; //WRTDS load at LMS

vector [nr] SD; //SD calculated for WRTDS load

vector [nl] chick; //chickens in subwatershed

vector [nl] cow; //cow in subwatershed

vector [nl] hog; //hogs (swine) in subwatershed

int wsd [nr]; //count variable for LMSs

int wshed\_size; //# of LMS watersheds

vector [nr] increm\_area; //Incremental area for each loading station

vector [nr] av\_prec; //normalized precipitation

vector [nr] av\_prec2; //scaled precipitation

vector [nr] up\_t\_load1; //upstream TN loading for nested wsds

vector [nr] up\_t\_load2; //upstream TN loading for nested wsds

vector [nr] up\_t\_load3; //upstream TN loading for nested wsds

vector [nr] str\_loss\_load1; //stream losses of upstream TN loading

vector [nr] str\_loss\_load2; //stream losses of upstream TN loading

vector [nr] str\_loss\_load3; //stream losses of upstream TN loading

vector [nr] res\_loss\_load1; //reservoir losses of upstream TN loading

vector [nr] res\_loss\_load2; //reservoir losses of upstream TN loading

vector [nr] res\_loss\_load3; //reservoir losses of upstream TN loading

vector [nd] d\_loss\_str; //stream losses of dischargers to LMSs

vector [nd] d\_loss\_res; //reservoir losses of dischargers to LMSs

vector [nd] d\_vals; //dischargers TN loadings

vector [nl] l\_loss\_str; //stream losses for subwatersheds

vector [nl] l\_loss\_res; //reservoir losses for subwatersheds

vector [nl] ag; //agriculture in subwatershed

vector [nl] devpre; //pre-1980 urban in subwatershed

vector [nl] devpost; //post-1980 urban in subwatershed

vector [nl] wild; //undeveloped in subwatershed

vector [nl] tot\_l; //total size of subwatershed

int l\_start [nr]; //count variables to link subwatersheds to LMSs

int l\_end [nr]; //count variables to link subwatersheds to LMSs

int d\_start [nr]; //count variables to link dischargers to LMSs

int d\_end [nr]; //count variables to link dischargers to LMSs

}

transformed data{

}

parameters {

real<lower =0> Be\_a; //Agriculture export

real<lower =0> Be\_d\_pre; //pre-1980 developed export

real<lower =0> Be\_d\_post; //post-1980 developed export

real<lower =0> Be\_w; //Undeveloped export

real<lower =0, upper = 1> Be\_ch; //Chickens export

real<lower =0, upper = 1> Be\_h; //Hogs (swine)

real<lower =0, upper = 1> Be\_cw; //Cows

real <lower =0, upper = 1> Sn; //Stream retention

real <lower =1, upper = 60> Sn2; //Reservoir retention

real <lower =0, upper = 0.40> PIC\_q; //PIC for stream flow

vector<lower =0, upper = 10> [7] pic\_p; //PIC for land classes

real <lower=0> Be\_dch; //Point source discharge coefficient

real<lower=0, upper = 2> sigma\_res; //Model residual sigma

real <lower=0, upper = 5> sigma\_w; //Random effect sigma

vector [wshed\_size] alpha; //# of watersheds

real<lower = 0, upper = 2> sigma\_B1; //PIC sigma

real<lower = 0, upper = 3> Bp\_mean; //PIC mean

vector [nr] ly; //log unknown true loads

}

transformed parameters {

}

model {

vector [nr] tot; // Sum of all loadings from all sources

vector [nr] sigma; //sigma for watershed random effects

vector [nr] y\_hat; //total loadings plus random effects minus waterbody losses

vector [nr] A; //To compile Agriculture with PIC

vector [nr] Dpr; //To compile pre-1980 urban with PIC

vector [nr] Dpt; //To compile post-1980 urban with PIC

vector [nr] W; //To compile undeveloped urban with PIC

vector [nr] Dch; //To compile dischargers with stream/reservoir losses

vector [nr] tot\_loss;

int w;

vector [nr] alpha\_vals; // Watershed indicator

real t;

vector [nr] A; //Agriculture vector

vector [nr] D\_lc\_pre; //pre-1980 vector

vector [nr] D\_lc\_post; //post-1980 vector

vector [nr] W\_lc; //Undeveloped vector

vector [nr] Disch; //point source dischargers

vector [nr] C; //chickens for adding PIC

vector [nr] H; //hogs for adding PIC

vector [nr] Cw; //cows for adding PIC

vector [nr] C\_r; //chickens for aggregating subwatersheds

vector [nr] H\_r; //swine for aggregating subwatersheds

vector [nr] Cw\_r; //cows for aggregating subwatersheds

vector [nr] ly\_hat; // log of TN load with transformation and offset

vector [nr] y; //TN load

// Loop through all export groups including stream and reservoir losses

for(i in 1:nr){

//Looping to aggregate subwatersheds and corresponding stream and reservoir retention.

A\_lc[i]= sum((ag[l\_start[i]:l\_end[i]]) .\* (exp((-Sn/(1+PIC\_q\*av\_prec[i])) \* l\_loss\_str[l\_start[i]:l\_end[i]])) .\* (exp((-Sn2/(1+PIC\_q\*av\_prec[i])) ./l\_loss\_res[l\_start[i]:l\_end[i]])));

D\_lc\_pre[i]= sum((devpre[l\_start[i]:l\_end[i]]) .\* (exp((-Sn/(1+PIC\_q\*av\_prec[i])) \* l\_loss\_str[l\_start[i]:l\_end[i]])) .\* (exp((-Sn2/(1+PIC\_q\*av\_prec[i])) ./l\_loss\_res[l\_start[i]:l\_end[i]])));

D\_lc\_post[i]= sum((devpost[l\_start[i]:l\_end[i]]) .\* (exp((-Sn/(1+PIC\_q\*av\_prec[i])) \* l\_loss\_str[l\_start[i]:l\_end[i]])) .\* (exp((-Sn2/(1+PIC\_q\*av\_prec[i])) ./l\_loss\_res[l\_start[i]:l\_end[i]])));

W\_lc[i]= sum((wild[l\_start[i]:l\_end[i]]) .\* (exp((-Sn/(1+PIC\_q\*av\_prec[i])) \* l\_loss\_str[l\_start[i]:l\_end[i]])) .\* (exp((-Sn2/(1+PIC\_q\*av\_prec[i])) ./l\_loss\_res[l\_start[i]:l\_end[i]])));

Disch[i]=sum((d\_vals[d\_start[i]:d\_end[i]]) .\* exp((-Sn/(1+PIC\_q\*av\_prec[i])) \* d\_loss\_str[d\_start[i]:d\_end[i]]) .\* exp((-Sn2/(1+PIC\_q\*av\_prec[i])) ./ d\_loss\_res[d\_start[i]:d\_end[i]]));

C\_r[i]= sum((chick[l\_start[i]:l\_end[i]]) .\* (exp((-Sn/(1+PIC\_q\*av\_prec[i])) \* l\_loss\_str[l\_start[i]:l\_end[i]])) .\* (exp((-Sn2/(1+PIC\_q\*av\_prec[i])) ./l\_loss\_res[l\_start[i]:l\_end[i]])));

H\_r[i]= sum((hog[l\_start[i]:l\_end[i]]) .\* (exp((-Sn/(1+PIC\_q\*av\_prec[i])) \* l\_loss\_str[l\_start[i]:l\_end[i]])) .\* (exp((-Sn2/(1+PIC\_q\*av\_prec[i])) ./l\_loss\_res[l\_start[i]:l\_end[i]])));

Cw\_r[i]= sum((cow[l\_start[i]:l\_end[i]]) .\* (exp((-Sn/(1+PIC\_q\*av\_prec[i])) \* l\_loss\_str[l\_start[i]:l\_end[i]])) .\* (exp((-Sn2/(1+PIC\_q\*av\_prec[i])) ./l\_loss\_res[l\_start[i]:l\_end[i]])));

//Adding precipitation impact coefficient to land export

A[i]= Be\_a \* pow(av\_prec2[i],pic\_p[1]) .\* A\_lc[i];

Dpr[i] = Be\_d\_pre \* pow(av\_prec2[i],pic\_p[2]) .\* D\_lc\_pre[i];

Dpt[i] = Be\_d\_post \* pow(av\_prec2[i],pic\_p[3]) .\* D\_lc\_post[i];

D[i] = Dpr[i]+Dpt[i];

W[i] = Be\_w \* pow(av\_prec2[i],pic\_p[4]) .\* W\_lc[i];

C[i] = Be\_ch \* pow(av\_prec2[i],pic\_p[5]) .\* C\_r[i];

H[i] = Be\_h \* pow(av\_prec2[i],pic\_p[6]) .\* H\_r[i];

Cw[i] = Be\_cw \* pow(av\_prec2[i],pic\_p[7]) .\* Cw\_r[i];

Dch[i] = Be\_dch \* Disch[i];

}

//Loop to determine random effect for each watershed

for (i in 1:nr){

w= wsd[i];

sigma[i] = sqrt(pow(SD[i],2)+pow(sigma\_res,2)); //with regular sigma values

alpha\_vals[i] = alpha[w];

}

//Sum loadings from all sources

tot = A + D + W + C + H + Cw + Dch;

//Add random effects to source loadings and subtract losses from upstream loads

y\_hat = tot + 10000 \* alpha\_vals - up\_t\_load1 .\* (1-(exp((-Sn ./ (1+PIC\_q\*av\_prec)) .\* str\_loss\_load1) .\* exp((-Sn2 ./ (1+PIC\_q\*av\_prec)) ./ res\_loss\_load1))) - up\_t\_load2 .\* (1-(exp((-Sn ./ (1+PIC\_q\*av\_prec)) .\* str\_loss\_load2) .\* exp((-Sn2 ./ (1+PIC\_q\*av\_prec)) ./ res\_loss\_load2))) - up\_t\_load3 .\* (1-(exp((-Sn ./(1+PIC\_q\*av\_prec)) .\* str\_loss\_load3) .\* exp((-Sn2 ./ (1+PIC\_q\*av\_prec)) ./ res\_loss\_load3)));

//Priors for the model

Be\_a ~ normal(900,700); //Prior for agriculture, TN/km2/yr

Be\_d\_pre ~ normal(800,300); //Prior for pre-1980 development, TN/km2/yr

Be\_d\_post ~ normal(800,300); //Prior for post-1980 development, TN/km2/yr

Be\_w ~ normal(200,200); //Prior for undeveloped, TN/km2/yr

Be\_ch ~ normal(0.001,0.0003); //Prior for chickens (TN/count)

Be\_h ~ normal(0.04,0.02); //Prior for swine (TN/count)

Be\_cw ~ normal(0,5); //Uninformed Prior for cow

Be\_dch ~ normal(1,.10); //Prior for point source delivery, unitless

sigma\_res ~ normal(0,20); //st error of the model

sigma\_w ~ normal(0,300); //st. deviation of random effect hyperdistribution

alpha ~ normal(0,sigma\_w); //watershed random effects

sigma\_B1 ~ normal(0,1); //st. deviation for PIC hyperdistribution

Bp\_mean ~ normal(1,1); //mean for PIC hyperdistribution

pic\_p[1] ~ normal(Bp\_mean,sigma\_B1); //PIC ag distribution

pic\_p[2] ~ normal(Bp\_mean,sigma\_B1); //PIC pre distribution

pic\_p[3] ~ normal(Bp\_mean,sigma\_B1); //PIC post distribution

pic\_p[4] ~ normal(Bp\_mean,sigma\_B1); //PIC undeveloped distribution

pic\_p[5] ~ normal(Bp\_mean,sigma\_B1); //PIC chicken distribution

pic\_p[6] ~ normal(Bp\_mean,sigma\_B1); //PIC hog distribution

pic\_p[7] ~ normal(Bp\_mean,sigma\_B1); //PIC cow distribution

Sn ~ normal(.14,.05); //Prior for stream retention rate, 1/d

Sn2 ~ normal(11,2); //Prior for waterbody retention rate, m/y

PIC\_q ~ normal(0,1); //PIC for retention

// with log transformation

ly\_hat=log((y\_hat ./10000) + 10); // vector to get log of TN load (y\_hat)

ly ~ normal(ly\_hat,sigma\_res); //parameter that calibrates ly\_hat (log of TN load) with ly ()

y=exp(ly)-10; // parameter that calibrates to load

load ~ normal(y,SD); // load = WRTDS estimate, SD = WRTDS sd

}