MAPPING HUMAN HEALTH RISK FROM CONTAMINATION OF DRINKING WATER SOURCES IN DEVELOPING COUNTRIES

Avit Kumar Bhowmik

June 12, 2015

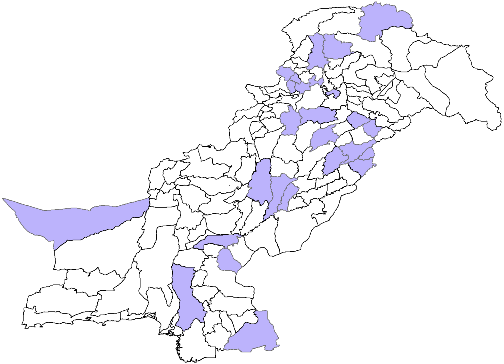
Institute for Environmental Sciences, University of Koblenz-Landau

UNIVERSITÄT KOBLENZ · LANDAU

SPATIAL STATISTICS 2015 EMERGING PATTERNS 9-12 JUNE 2015 AVIGNON, FRANCE

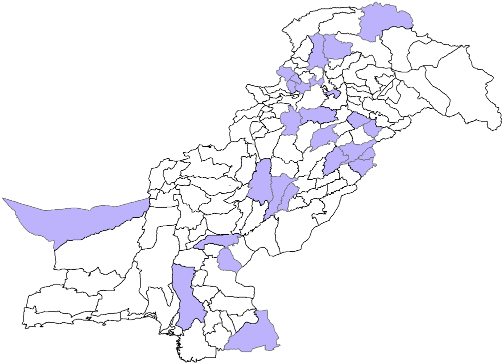
1/3 of annual mortality in Pakistan is attributed to contaminated drinking water (Azizullah et al. 2011)

1/3 of annual mortality in Pakistan is attributed to contaminated drinking water (Azizullah et al. 2011)



We compiled trace metal concentrations in ground & surface water

1/3 of annual mortality in Pakistan is attributed to contaminated drinking water (Azizullah et al. 2011)

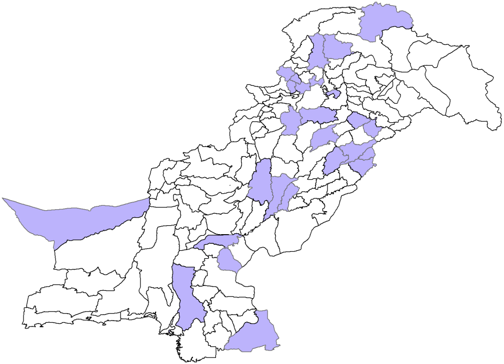


We compiled trace metal concentrations in ground & surface water

19 % coverage, max. sampled-unsampled district distance 2000 km

Low global representativeness, GCV $\geq$ 1

1/3 of annual mortality in Pakistan is attributed to contaminated drinking water (Azizullah et al. 2011)



We compiled trace metal concentrations in ground & surface water

19 % coverage, max. sampled-unsampled district distance 2000 km

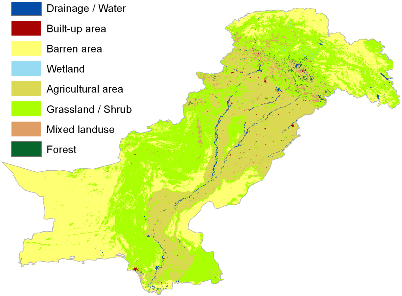
Low global representativeness, GCV ≥ 1 non-stationarity

GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) ALLOWS TO INCORPORATE LOCAL VARIATIONS (HARRIS ET AL. 2010)

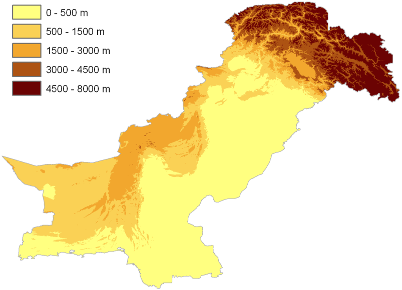
$$C_{T,z} = \delta_{z,0} + \sum_{n=1}^{m=8} \delta_{z,n} S_{z,n} + e_z$$

GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) ALLOWS TO INCORPORATE LOCAL VARIATIONS (HARRIS ET AL. 2010)

$$C_{T,z} = \delta_{z,0} + \sum_{n=1}^{m=8} \delta_{z,n} S_{z,n} + e_z$$



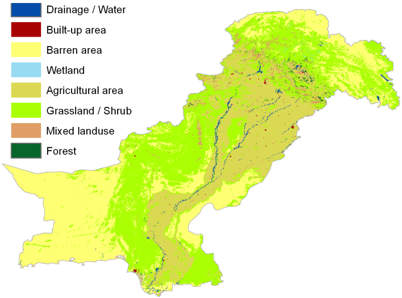
Land cover (ISCGM 2014) BL, AL, ML

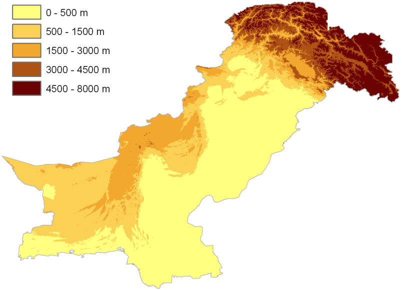


Elevation (Rodriguez et al. 2005)

GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) ALLOWS TO INCORPORATE LOCAL VARIATIONS (HARRIS ET AL. 2010)

$$C_{T,z} = \delta_{z,0} + \sum_{n=1}^{m=8} \delta_{z,n} S_{z,n} + e_z$$





Land cover (ISCGM 2014) BL, AL, ML Elevation (Rodriguez et al. 2005)

Global soil properties (Batjes 2000) SOC, SCC, WC, pH

WE CALIBRATED LOCAL GWR MODELS AND USED THEM FOR SPATIAL PREDICTION

$$C_{T,z} = \delta_{z,0} + \sum_{n=1}^{m=8} \delta_{z,n} S_{z,n} + e_z$$

- Gaussian kernel based on the spatial continuity

WE CALIBRATED LOCAL GWR MODELS AND USED THEM FOR SPATIAL PREDICTION

$$C_{T,z} = \delta_{z,0} + \sum_{n=1}^{m=8} \delta_{z,n} S_{z,n} + e_z$$

- Gaussian kernel based on the spatial continuity
- Fixed kernel bandwidths based on AIC

WE CALIBRATED LOCAL GWR MODELS AND USED THEM FOR SPATIAL PREDICTION

$$C_{T,z} = \delta_{z,0} + \sum_{n=1}^{m=8} \delta_{z,n} S_{z,n} + e_z$$

- Gaussian kernel based on the spatial continuity
- Fixed kernel bandwidths based on AIC
- Best-fit models forward entering of spatial predictors and based on AIC

e.g. Cr_SW~WC + ALU

WE CALIBRATED LOCAL GWR MODELS AND USED THEM FOR SPATIAL PREDICTION

$$C_{T,z} = \delta_{z,0} + \sum_{n=1}^{m=8} \delta_{z,n} S_{z,n} + e_z$$

- Gaussian kernel based on the spatial continuity
- Fixed kernel bandwidths based on AIC
- Best-fit models forward entering of spatial predictors and based on AIC

  e.g. Cr_SW~WC + ALU
- R package GWmodel (Gollini et al. 2013)

WE EVALUATED PREDICTION ACCURACY AND UNCERTAINTY BY LEAVE-ONE-OUT CROSS-VALIDATION

- Index of agreement (d) perfect agreement 1, no agreement 0

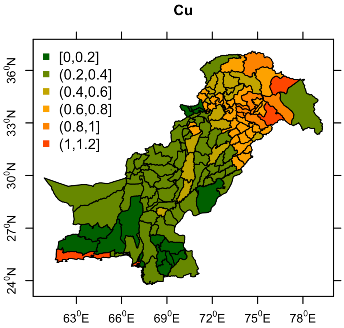
WE EVALUATED PREDICTION ACCURACY AND UNCERTAINTY BY LEAVE-ONE-OUT CROSS-VALIDATION

- Index of agreement (d) perfect agreement 1, no agreement 0
- Root mean squared deviation error (RMSDE) lower value indicates higher accuracy

WE EVALUATED PREDICTION ACCURACY AND UNCERTAINTY BY LEAVE-ONE-OUT CROSS-VALIDATION

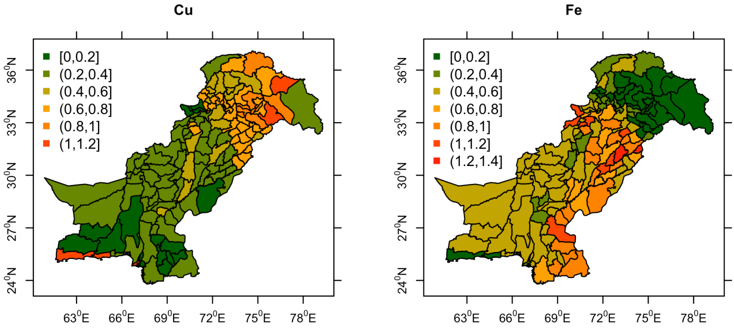
- Index of agreement (d) perfect agreement 1, no agreement 0
- Root mean squared deviation error (RMSDE) lower value indicates higher accuracy
- Standard deviation of prediction z-scores (SDZ) unity indicates high certainty

WE ESTIMATED AREA AND POPULATION AT RISK BY COMPARING TO "WHO" GUIDELINE THRESHOLDS (WHO 2011)



$$RQ_{T,z} = \frac{\hat{C}_{z,0}}{C_{WHO-T}}$$

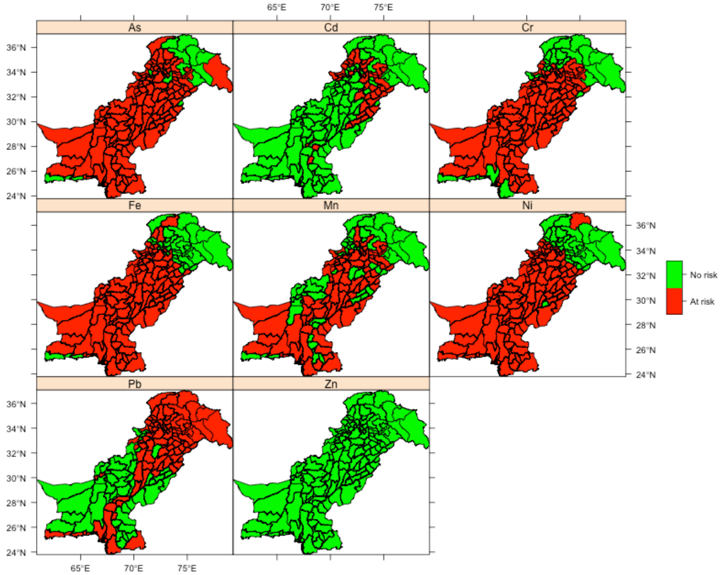
WE ESTIMATED AREA AND POPULATION AT RISK BY COMPARING TO "WHO" GUIDELINE THRESHOLDS (WHO 2011)



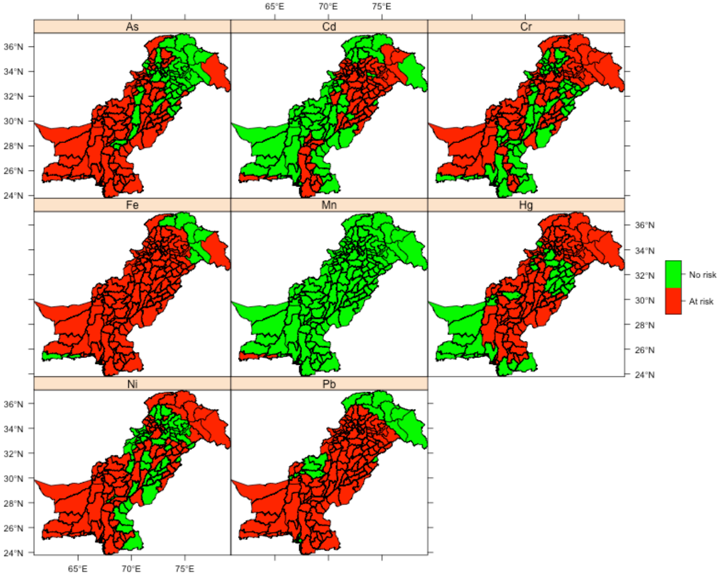
$$RQ_{T,z} = \frac{\hat{C}_{z,0}}{C_{WHO-T}}$$

RQ>1 at risk $RQ \leq 1$ negligible risk

MORE THAN 53 % OF TOTAL AREA AND 74 MILLION PEOPLE ARE AT RISK FROM ARSENIC, CHROMIUM, IRON, NICKEL AND LEAD



MORE THAN 53 % OF TOTAL AREA AND 74 MILLION PEOPLE ARE AT RISK FROM ARSENIC, CHROMIUM, IRON, NICKEL AND LEAD



OUR GWR MODEL PREDICTIONS EXHIBITED A HIGH ACCURACY AND A LOW UNCERTAINTY

- $d \geq 0.8$ except for iron in ground water and zinc in surface water

OUR GWR MODEL PREDICTIONS EXHIBITED A HIGH ACCURACY AND A LOW UNCERTAINTY

- $d \geq 0.8$ except for iron in ground water and zinc in surface water
- $RMSDE \leq 0.45$ 0.97 for iron in ground water and 0.8 for zinc in surface water

OUR GWR MODEL PREDICTIONS EXHIBITED A HIGH ACCURACY AND A LOW UNCERTAINTY

- $d \geq 0.8$ except for iron in ground water and zinc in surface water
- $RMSDE \leq 0.45$ 0.97 for iron in ground water and 0.8 for zinc in surface water
- $SDZ \geq 0.7$ except for cadmium and zinc in ground water, and cadmium and manganese in surface water

OUR GWR MODEL PREDICTIONS EXHIBITED A HIGH ACCURACY AND A LOW UNCERTAINTY

- $d \geq 0.8$ except for iron in ground water and zinc in surface water
- $RMSDE \leq 0.45$ 0.97 for iron in ground water and 0.8 for zinc in surface water
- $SDZ \geq 0.7$ except for cadmium and zinc in ground water, and cadmium and manganese in surface water
- Calibrated spatial predictors are in agreement with the causes of trace metal contamination e.g. As_SW~SOC + pH (Husain et al. 2012)

THE PREDICTIONS INFORM WATER RESOURCES MANAGEMENT ON POTENTIAL HOT SPOTS

- Predictions may not reflect true trace metal concentration variation 453 km $\leq$ kernel bandwidth $\leq$ 1335 km (Bhowmik and Costa 2014)

THE PREDICTIONS INFORM WATER RESOURCES MANAGEMENT ON POTENTIAL HOT SPOTS

- Predictions may not reflect true trace metal concentration variation $453\text{ km} \leq \text{kernel bandwidth} \leq 1335\text{ km}$ (Bhowmik and Costa 2014)
- Predictions complement field monitoring although require field validation

THE PREDICTIONS INFORM WATER RESOURCES MANAGEMENT ON POTENTIAL HOT SPOTS

- Predictions may not reflect true trace metal concentration variation 453 km $\leq$ kernel bandwidth $\leq$ 1335 km (Bhowmik and Costa 2014)
- Predictions complement field monitoring although require field validation
- Predictions with low accuracy and high uncertainty need to be treated with caution e.g. zinc

THE PREDICTIONS INFORM WATER RESOURCES MANAGEMENT ON POTENTIAL HOT SPOTS

- Predictions may not reflect true trace metal concentration variation $453 \text{ km} \leq \text{kernel bandwidth} \leq 1335 \text{ km}$ (Bhowmik and Costa 2014)
- Predictions complement field monitoring although require field validation
- Predictions with low accuracy and high uncertainty need to be treated with caution e.g. zinc
- Perhaps over-prediction for areas with access to water purification more relevant for rural areas (Khan et al. 2012)

THE PREDICTIONS INFORM WATER RESOURCES MANAGEMENT ON POTENTIAL HOT SPOTS

- Predictions may not reflect true trace metal concentration variation 453 km $\leq$ kernel bandwidth $\leq$ 1335 km (Bhowmik and Costa 2014)
- Predictions complement field monitoring although require field validation
- Predictions with low accuracy and high uncertainty need to be treated with caution e.g. zinc
- Perhaps over-prediction for areas with access to water purification more relevant for rural areas (Khan et al. 2012)
- May also indicate indirect effects e.g. via consumption of vegetables irrigated with contaminated water (Amin et al., 2012)

For more: Bhowmik et al. 2015, Sci. Tot. Env. *under review*

Collaborators: Ambreen Alamdar, Ioannis Katsoyiannis, Heqing Shen, Nadeem Ali, Syeda Maria Ali, Habib Bokhari, Ralf B. Schäfer, Syed Ali Musstjab Akber Shah Eqani

Thank you

Questions?