UNIVERSITY OF Southampton
School of Geography

Reconstructing Indian Census Variables using Remotely Sensed Environmental Metrics

Benefiting from Earth Observation

Gary Watmough

Supervisors: Prof. P.M.Atkinson; Dr. C.W.Hutton & Dr. F.A. Johnson

4th October 2010

Contents

UNIVERSITY OF Southampton
School of Geography

Introduction

Census Uses	Census Limitations
• Allocate resources (UNSD 2008)	• 10 year enumeration
• Monitor long term trends (UNSD 2008)	• Publishing times
• Poverty reduction strategies (UNDP 2003)	• Cost
• Risk Mapping (Dobson *et al*. 2000)	• Not always possible (de Sherbinin *et al*. 2002)

- Past attempts combined census and earth observation data to map populations at risk (Dobson *et al*. 2000; Elvidge *et al*. 2009).
- Relationships between environmental factors and poverty been studied for many years. No recognition of the causal effects, but there are known links
- Ogneva-Himmelberger *et al*. (2009) it may be possible to gain understanding of "social well-being and quality of life" from a vegetation map.

UNIVERSITY OF Southampton
School of Geography

Statement of Problem

- Census an important data source in disaster risk mapping
- Limitations of the census can limit the usefulness of risk maps
- Remote sensing; increased spatial and temporal resolution may provide methods for up-dating census between enumeration
- This study: explore relationships between census variables and earth observation data.

Study Region: Assam

- Assam
- ~15 million people
- 12 Districts
- 76% rely on agriculture
- 1.25 million hectares average annual flood prone area
  - (Indian National Census 2001; Das 2005)

UNIVERSITY OF Southampton
School of Geography

Data

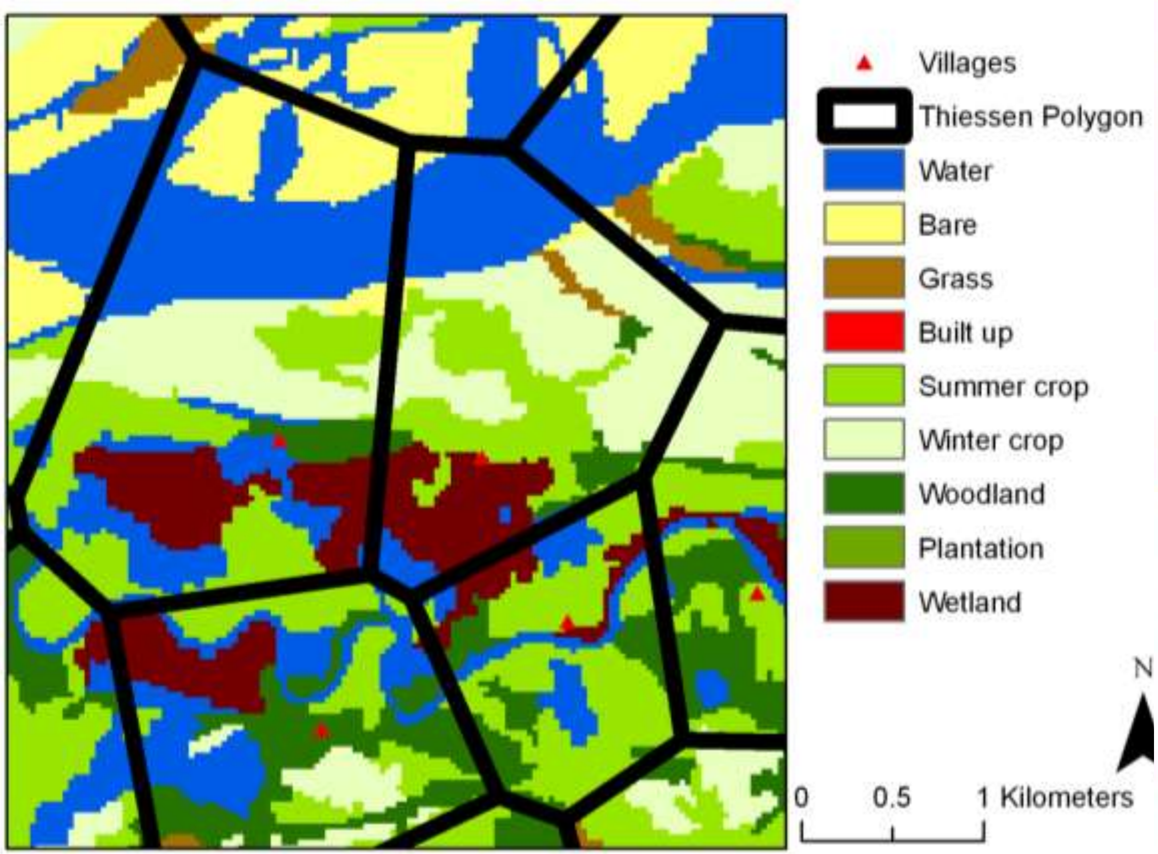
Census

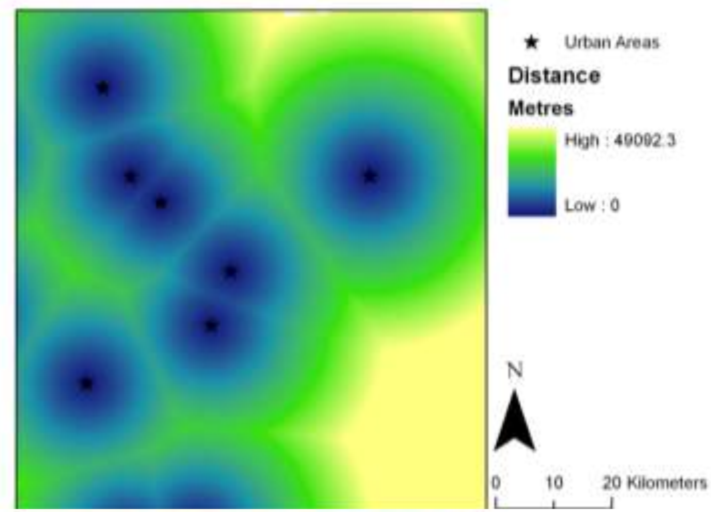
- Aggregated to Settlement level
- Geo-referenced to village centroid
- Rural only
- >14000 villages
- Variables included: literacy, population counts, employment type.

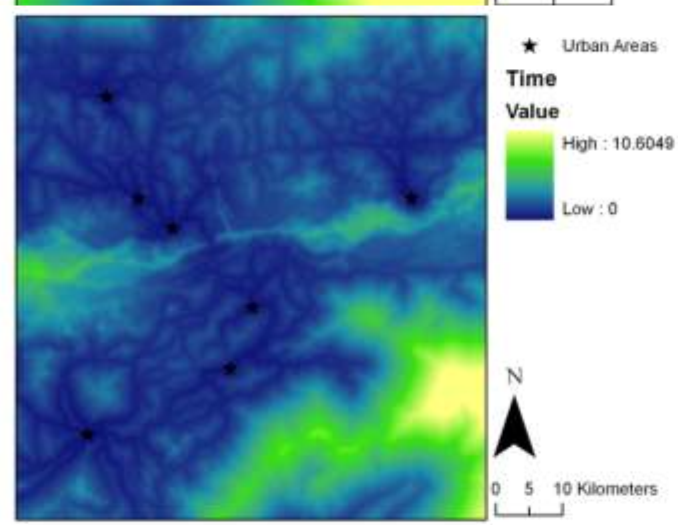
Earth Observation

- 6 Landsat ETM+ images
- Supplementary data:
  - Protected Areas
  - Flood Water extents
  - Road data

Methods: Creating the database







University of Southampton School of Geography

Methods: Logistic Regression

- Covariates split into categories
- Data split into 10% random samples
- Full model run using logistic regression and all possible covariates
- drop1 function compares model by dropping terms one at a time. If AIC changes very little the term is dropped from the model.
- Repeat 10000 times,
- Then final model run on all 14000 villages if covariate is dropped more than 500 times it is dropped from the final model.

Results: Female Literacy

University of Southampton School of Geography

- Dropped:
  - Distance to water and urban
  - Proportion of grassland and proportion of bare land
  - Road density

Negative relationships

Covariate	Coeff	SE	Z value	P Value	Odds Ratio
Winter crop (>20%)	-0.521	0.005	-105.85	<2e-16	0.59
Plantation (> 0%)	-0.502	0.003	-175.97	<2e-16	0.61
Dist main rd (>1km)	-0.285	0.003	-97.07	<2e-16	0.75
Water (>30%)	-0.263	0.006	-45.03	<2e-16	0.77
Time urban(> 90 mins)	-0.258	0.004	-66.95	<2e-16	0.77
Winter crop (>0% - 20%)	-0.126	0.002	-64.91	<2e-16	0.88
Time urban (15 - 90 mins)	-0.124	0.003	-38.34	<2e-16	0.88
Dist Flood (> 1km)	-0.076	0.003	-26.06	<2e-16	0.93
Water (>0% - 30%)	-0.067	0.002	-34.57	<2e-16	0.94

Positive relationships

Covariate	Coeff	SE	Z value	P Value	Odds Ratio
Dist IUCN (>40km)	0.514	0.007	73.46	<2e-16	1.67
Woodland (>50%)	0.412	0.009	45.84	<2e-16	1.51
Dist IUCN (>1km - 40km)	0.372	0.007	53.73	<2e-16	1.45
Woodland (>0% - 50%)	0.34	0.009	38.53	<2e-16	1.4
Summer crop (>40%)	0.275	0.005	55.45	<2e-16	1.32
Dist Flood (>0m - 1km)	0.149	0.003	55.49	<2e-16	1.16
Summer crop (>0% - 40%)	0.115	0.005	24.71	<2e-16	1.12

- Residual Deviance: 855572
- AIC: 937894
- Null deviance: 1027143

Results: Economic alternatives to agriculture

UNIVERSITY OF Southampton
School of Geography

- Dropped:
  - Winter cropland

Negative relationships

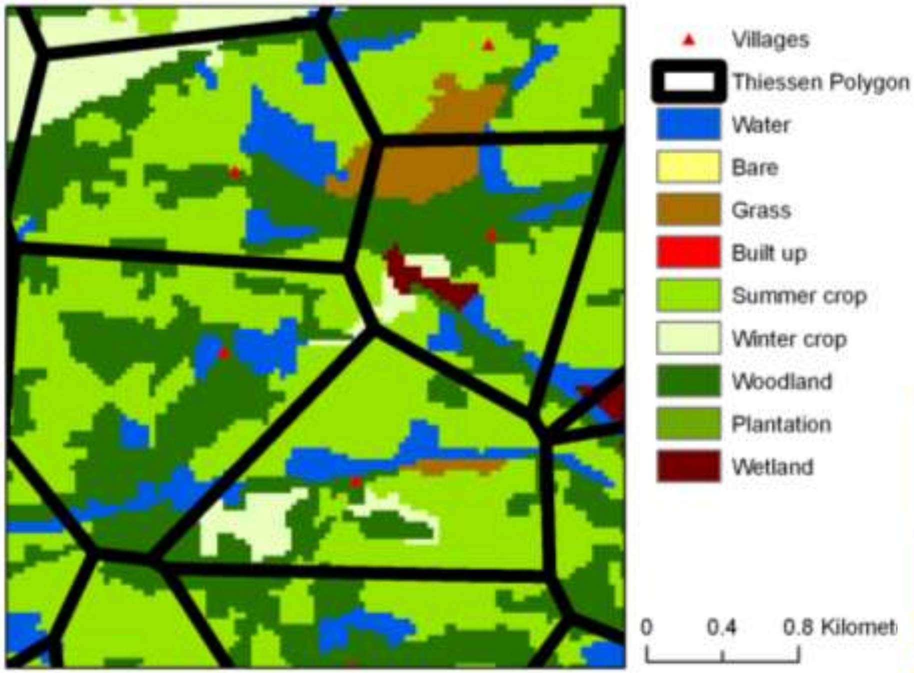
Covariate	Coeff	SE	Z value	P Value	Odds Ratio
Protected (100%)	-1.259	0.232	-5.427	0	0.28
Dist main rd (>1km)	-0.622	0.007	-88.137	<2e-16	0.54
Time urban(> 90 mns)	-0.258	0.004	-63.146	<2e-16	0.77
Water (>30%)	-0.232	0.006	-38.695	<2e-16	0.79
Dist Urban (>10km)	-0.177	0.003	-65.363	<2e-16	0.84
Dist main rd (>250m - 1km)	-0.111	0.003	-32.004	<2e-16	0.89
Water (>0% - 30%)	-0.101	0.002	-42.65	<2e-16	0.9
Dist Urban (>5 - 10km)	-0.094	0.003	-33.614	<2e-16	0.91
Time urban (>15 - 90 mns)	-0.031	0.003	-10.28	<2e-16	0.97

Positive relationships

Covariate	Coeff	SE	Z value	P Value	Odds Ratio
Plantation (>50%)	0.723	0.005	136.415	<2e-16	2.06
Plantation (> 0% - 50%)	0.486	0.003	192.708	<2e-16	1.63
Woodland (>50%)	0.417	0.011	37.396	<2e-16	1.52
Dist IUCN (>40km)	0.326	0.011	29.618	<2e-16	1.39
Built Up (>5%)	0.289	0.005	52.999	<2e-16	1.33
Dist Flood (> 1km)	0.252	0.003	73.458	<2e-16	1.29
Dist IUCN (>1km - 40km)	0.225	0.011	20.513	<2e-16	1.25
Woodland (>0% - 50%)	0.183	0.011	16.773	<2e-16	1.2
Dist Flood (>0m - 1km)	0.16	0.003	49.125	<2e-16	1.17
Dist Water (>1.5km)	0.144	0.006	22.395	<2e-16	1.15
Bareland (>10%)	0.127	0.004	28.611	<2e-16	1.14
Road Density (>0m/km^2)	0.122	0.002	60.837	<2e-16	1.13
Dist Water (>0m - 1.5km)	0.106	0.006	17.629	<2e-16	1.11
Bareland (>0% - 10%)	0.095	0.002	47.225	<2e-16	1.1
Protected (>0 - 99%)	0.083	0.002	43.978	<2e-16	1.09
Summer crop (>0% - 40%)	0.064	0.005	13.206	<2e-16	1.07
Summer crop (>40%)	0.031	0.005	6.213	0	1.03

- Residual Deviance: 659718
- AIC: 731623
- Null Deviance: 880465

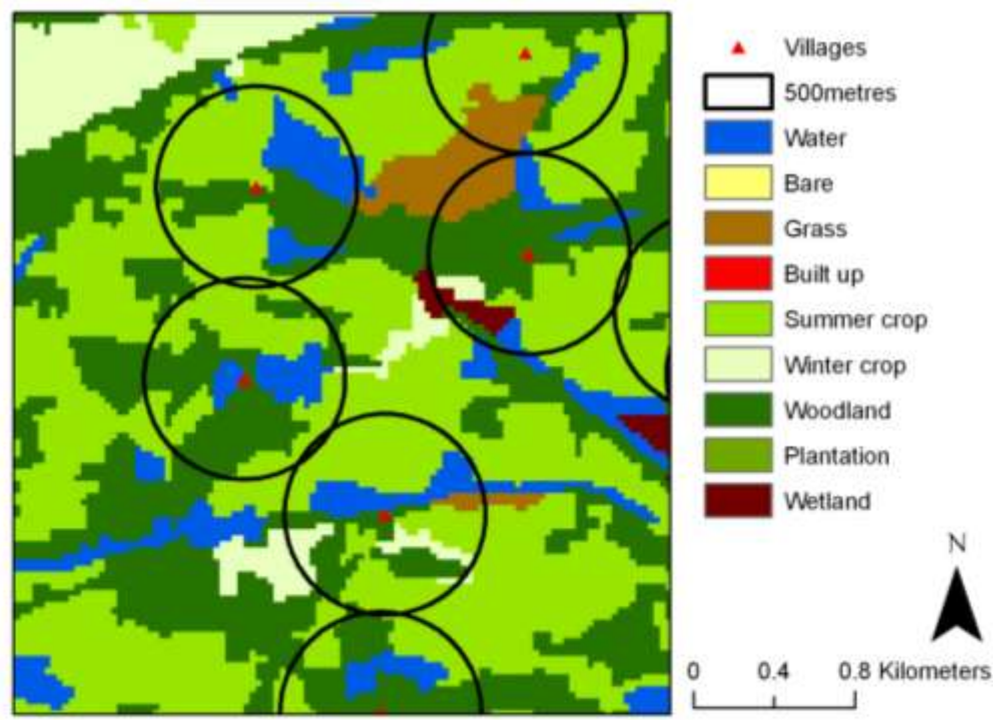
Discussion



s

t with the literature.

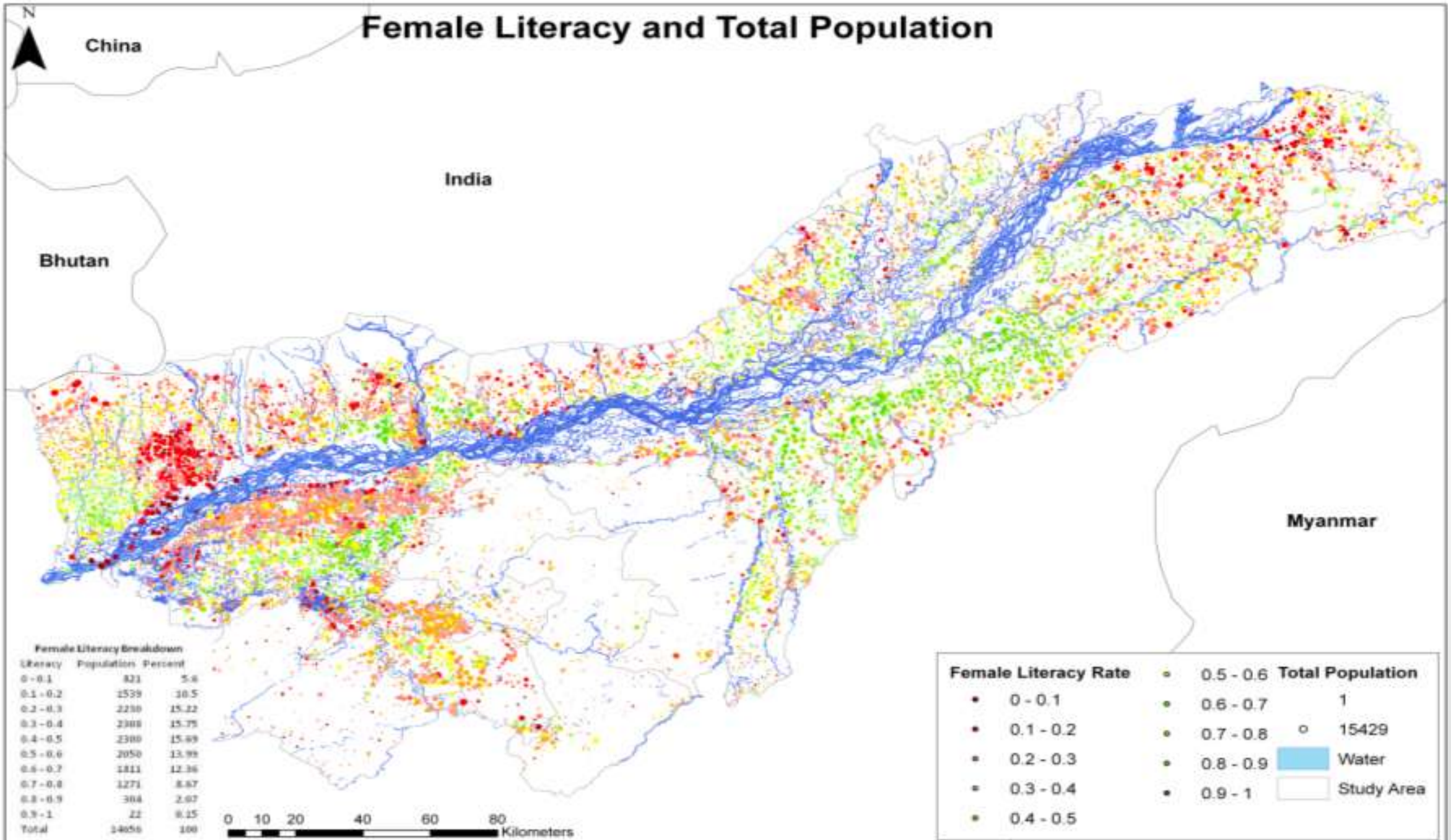
there is a better model for explaining



- Village boundary polygons will be exp

UNIVERSITY OF Southampton
School of Geography

Discussion: Further work



UNIVERSITY OF Southampton
School of Geography

Conclusion and wider significance

- Some variance in census variable can be explained using only remotely sensed environmental metrics.
- Future work;
  - Explore other census variables
  - Explore other linking parameters (radial buffer zones)
  - Explore use of spatial statistics
- Could in the future use similar methods to provide updates to census data during the intercensal period.

References

- De Sherbinin, A; Balk, D; Yager, K; Jaiteh, M; Pozzi, F; Giri, C; Wannedo, A (2002) “A CIESIN Thematic guide to social science application of remote sensing”. *CIESIN Thematic Guides*. Paisades; Centre for Int. Earth Sci. Info. Network.
- Dobson, JE; Bright, EA; Coleman, PR; Durfee, RC; Worley, BA (2000) “A global population database for estimating populations at risk” *Photogra. Eng. and Remote Sensing* 66; 849-857.
- Elvidge, CD; Sutton, PS; Baugh, KE; Tuttle, BT; Howard, AT; Erwin, EH; Bhaduri, B; Bright, E (2009) “A global poverty map derived from satellite date” *Computer and Geoscience* 35: 1652-1660.
- UNDP (2003) “Indicators for monitoring the Millennium Development Goals”. New York: United Nations Development Programme.
- UNSD (2008) “Principles and recommendations for population and housing censuses revision 2”. *United Nations Statistics Division* ST/ESA/STAT/SER.M/67/Rev.2.

Acknowledgements

- Royal Geographical Society (with IBG) Slawson Awards and The Gilchrist Educational Trust helped to fund the field work in India.
- Indian Institute of Technology Guwahati, Prof. Arupjyoti Saikia and Dr. Partha das for fieldwork support
- GeoData Institute for providing data and Ph.D. Support
- Dr. Guy Abel, School of Social Sciences, University of Southampton for computation statistics support and advice.