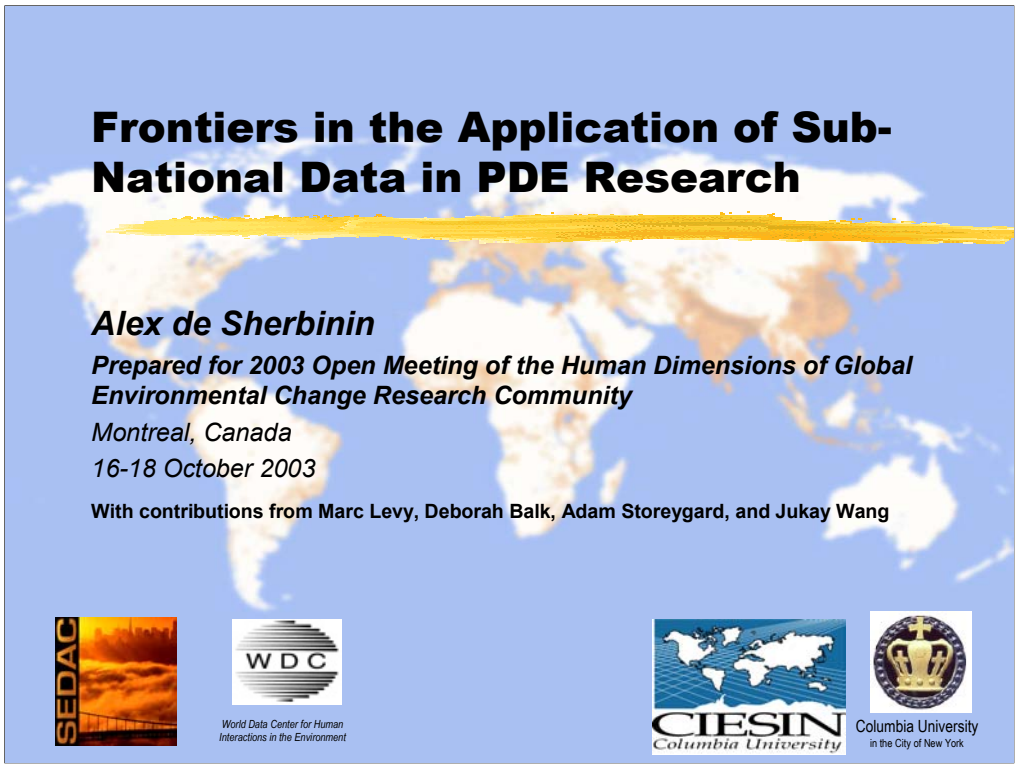


Abstract

Many early studies of population-environment linkages were essentially statistical analyses of national-level data on population size, density and growth, on the one hand, and environmental trends such as deforestation, carbon-dioxide emissions, or land degradation, on the other. Apart from relying on often questionable environmental data, these studies tended to generate weak or spurious correlations, falling easily into the ecological fallacy that if population growth coincided in the same geographic unit with environmental degradation (however construed), the one must have caused the other. This paper builds on work that CIESIN is undertaking for the United Nations Millennium Development Project. CIESIN has compiled a large collection of sub-national data on wide range of population and development indicators utilizing sources such as the Demographic and Health Surveys, national human development reports, and other national and international data collections. In addition, CIESIN has access to a large number of environmental data sets related to forest and land cover, soil quality, climatic zones, and farming systems. With richer sub-national data and more powerful GIS packages, it is possible to deepen the analysis and point out areas where the confluence of population size, density or growth; extreme poverty; and significant environmental change warrant further ground-level study. This paper will present a brief review of past efforts at population-environment national-level statistical analyses, and will then present the results of our sub-national analyses using examples from sub-Saharan Africa, Brazil, and a global analysis of infant mortality.

Structure of paper

1. Advantages of sub-national analysis
2. Take stock of P-E analyses (deforestation & biodiversity threats)
3. Presentation of sub-national Population-Development-Environment analyses
   A. Physical water availability and development indicators for Africa
   B. Predicting human development in Brazil
   C. Predicting infant mortality in Asia and Africa

This paper addresses the advantages of sub-national population-development-environment analysis, then in keeping with the session theme, it *takes stock* of past population-environment analysis using national and sub-national tabular data, and it *looks forward* by presenting new analyses using sub-national data sets developed by CIESIN for the UN Millennium Development Project (MDP).

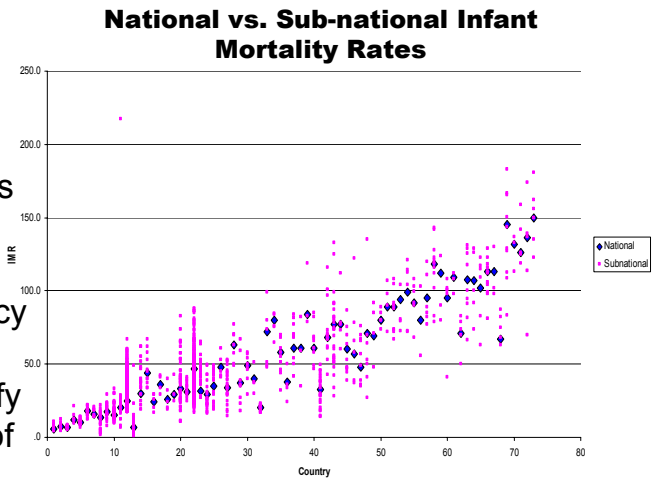
Background

- CIESIN is the "mapping arm" of the UN Millennium Development Project under Jeff Sachs
- We've collected sub-national development data from a variety of sources:
  - Demographic and Health Surveys
  - Multiple Indicator Cluster Surveys
  - National Human Development Reports
- Converting gridded data sets to sub-national indicators

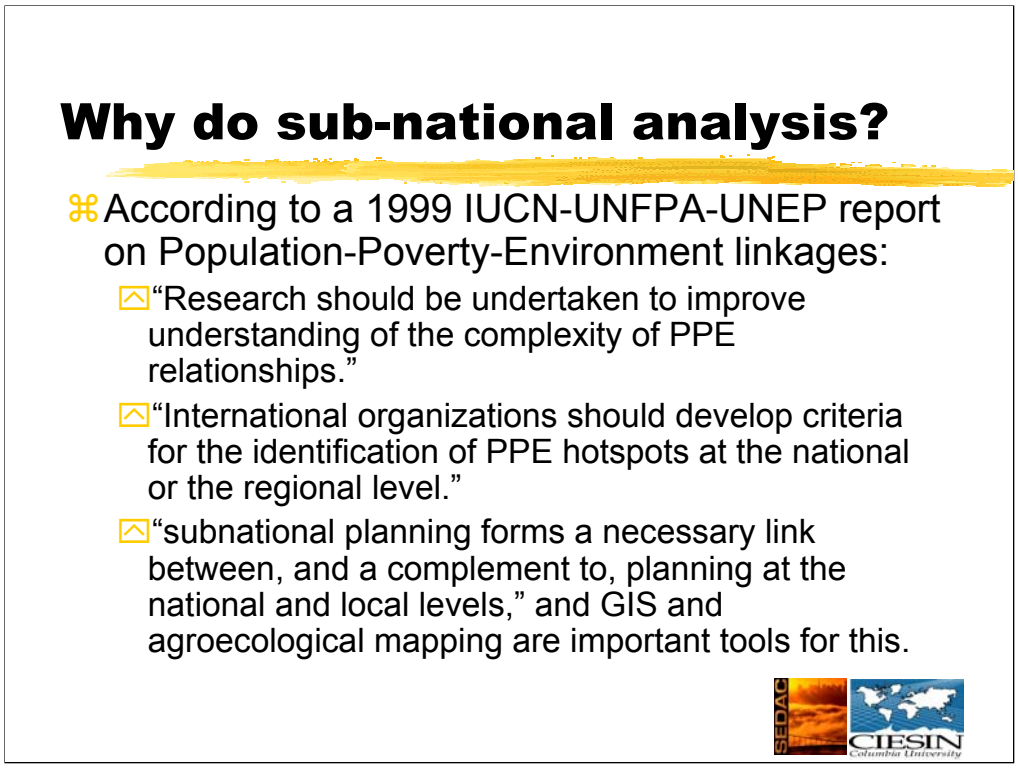
CIESIN has compiled an impressive array of population and development indicators for its work in conjunction with the MDP. We have also converted a large number of gridded data sets by aggregating up grid cell values to sub-national units.

Why do sub-national analysis?

- Sub-national analysis better captures the variation within national borders
- Less likely to commit the ecological fallacy
- Can more precisely identify areas in need of intervention



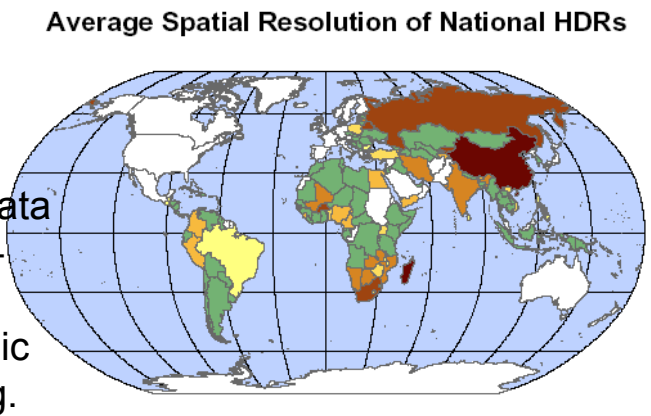
There are several reasons for do sub-national analysis. Firstly, it better captures variation within national borders, as shown by the chart at right. The blue diamonds represent national Infant Mortality Rates (IMRs), and the pink squares represent sub-national IMRs. There is obviously very wide variation in some countries, which would be lost in a purely national-level analysis. Secondly, the researcher is less likely to commit the ecological fallacy, in which it is assumed that because two things coexist in the same geographic area (e.g. population growth and deforestation), that the one necessarily caused the other. Thirdly, sub-national analysis can more precisely identify areas in need of policy intervention.



Fourthly, international bodies, as represented by the joint IUCN, UNFPA and UNEP *Report of the International Workshop on Population-Poverty-Environment Linkages* (1998), have recommended this kind of research.

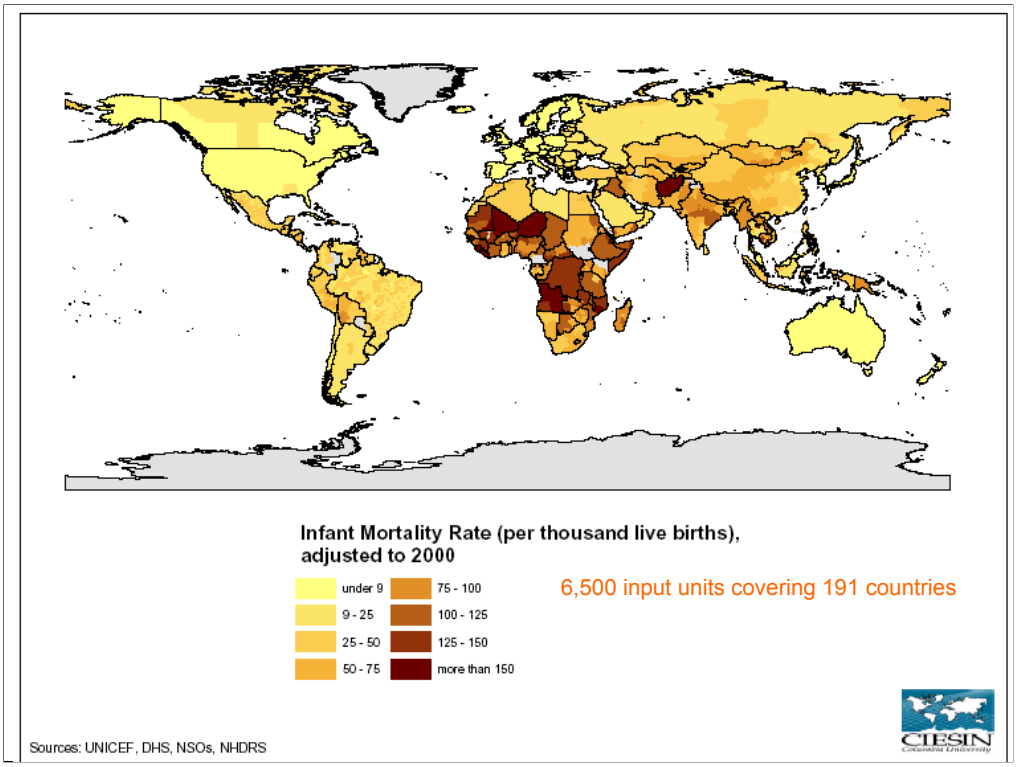
Why do sub-national analysis?

- Better availability of data – both "gridded" biophysical data sets and sub-national socioeconomic data sets (e.g. geocoded censuses)

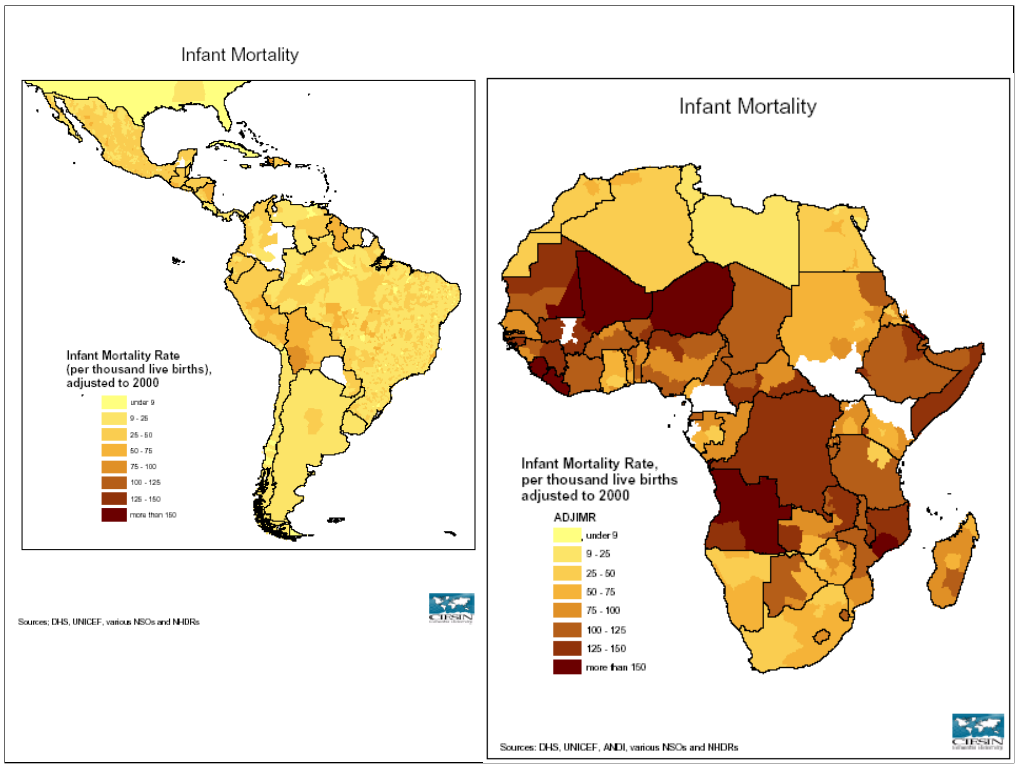


http://sedac.ciesin.columbia.edu/hdr/

Lastly, the data to undertake sub-national analyses are increasingly available. The map at right shows the availability of sub-national data from the national Human Development Reports (HDRs). The light brown and yellow countries have data available for subnational units that are, on average, below 16,000 square kilometers. A catalogue of data from the national HDRs is available at the URL shown on the slide.

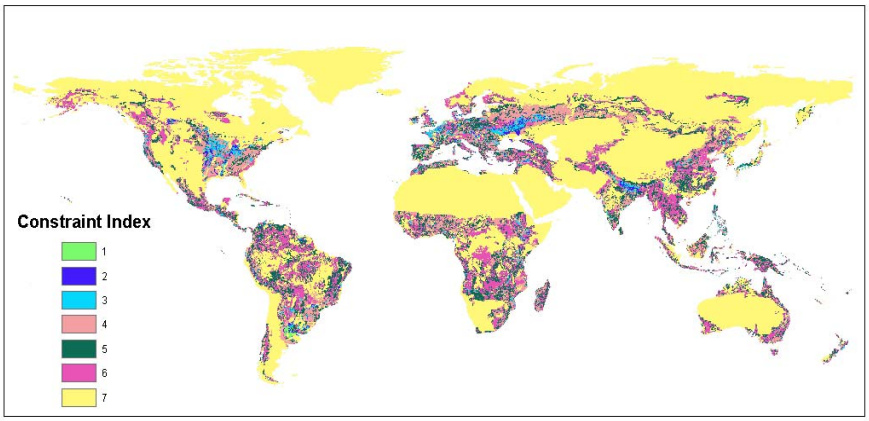


This is an example of one of the data products CIESIN developed in support of the Millennium Development Project. Instead of having 191 input units for 191 countries, it has 6,500 input units.



These slides show the same data for Latin America and Africa. The spatial resolution for most of Latin America is, on average, much higher than for Africa. But even the Africa sub-national data represents a vast improvement over national level averages. In countries like Nigeria and Zimbabwe it is possible to identify significant sub-national variation in infant mortality rates.

Combined Climate, Soil, and Terrain Slope Constraint



FAO-IIASA Global Agroecosystem Zone Assessment

This is an example of one gridded data set that CIESIN has aggregated to sub-national boundaries that correspond with the human development data (like the IMRs). In this case we developed a metric that represents the average level of soil, terrain and slope constraints in each sub-national unit.

Deforestation analyses: Allen & Barnes (1985)

- National-level analysis based on FAO production yearbooks, small sample size (n = 39), 1968-1978
- Controlling for 1968 GNP p.c., wood fuels and wood exports, and % land area under plantations, found a positive relationship between both 1968 pop density and total forest area change (R^2 of .353, P<.10)

Allen, J., and D. Barnes (1985). "The Causes of Deforestation in Developing Countries," *Annals of the American Association of Geographers* 75(2): 163-182.

This and the next two slides provide summaries of population-environment studies that used statistical approaches to try to understand the determinants deforestation. The first two utilized national-level data sets, and the third utilized sub-national data. The full citations are at the bottom of each slide.

Deforestation analyses: Mather & Needle 2000

- Another national-level analysis based on *Faostats* data, n = 111 (tropical to temperate), 1980-1990
- Percent change in pop 1970-80 and 1980-90 negatively related to % change in forest area (R^2 of .215 and .189 respectively, P< .0001)
- Mention weakening relationship over time; inversion in developed countries

Mather, A.S., and C.L. Needle (2000). "The Relationships of Population and Forest Trends," *Geographical Journal* 166(1): 2-13.

Deforestation analyses: Uusivouori, Lehto & Palo 2002

- Sub-national analysis based on *FORIS* data, n = 477 (67 countries), single year data for the period 1970-1991
- Controlling for GDP p.c., data reliability and ecological zones, they found sub-national pop density to be significantly negatively related to the ratio of forest to non-forest area (R^2 of .42, P<.01)
- Results seem to be of higher validity

Uusivouori, J., E. Lehto, and M. Palo (2002). "Population, Income and Ecological Conditions as Determinants of Forest Area in the Tropics," *Global Environmental Change* 12: 313-323.

The results of this study seem more robust, if for no other reason than the N is significantly higher owing to the larger number of units analyzed.

Biodiversity Threats: McKee *et al.* 2003

- National-level analysis based on IUCN Redlist and WCMC data, n = 114 , 2000
- Pop density significantly correlated with threatened species per unit area (R^2 of .402, P<.001)
- Pop density and species richness are the most effective combined predictors of threatened species per area (R^2 of .879, P<.001)

McKee, J.K., PW. Scuilli, C.D. Fooce, and T.A. Waite (2003). "Forecasting Global Biodiversity Threats Associated with Human Population Growth," *Biological Conservation* (Forthcoming).

This study, which utilized national-level data, found that population density and species richness were very good predictors of the the number of threatened species per unit area.

Presentation of sub-national PDE analyses

1. Physical water availability and development indicators for Africa
2. Determinants of Human Development in Brazil
3. Predicting infant mortality in Asia and Africa

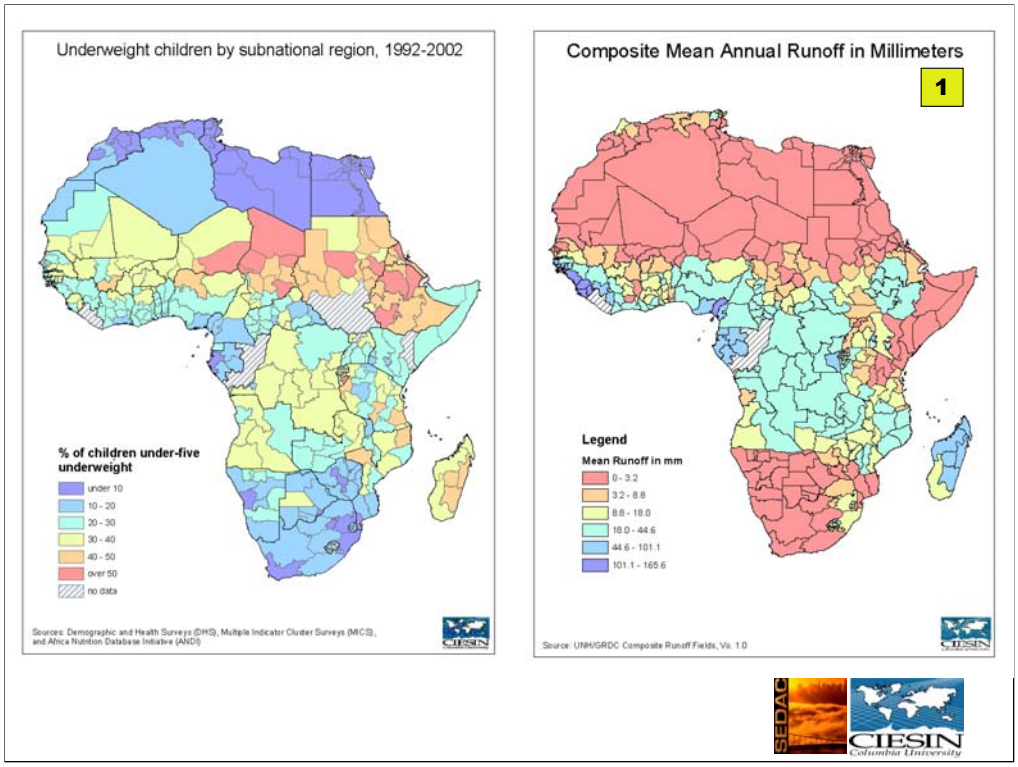
In the remainder of this presentation I will provide three examples of sub-national analyses conducted here at CIESIN.

Africa - Water Availability and Development Indicators

1

- Using UNH-GRDC's Composite Runoff data set; DHS & MICS data; GPW
- Hypothesized negative relationship between water availability and malnutrition
- Hypothesized positive relationship between water availability and improved water sources

The first example is a study that looked at physical water availability and development indicators in Africa. The study drew on data from the University of New Hampshire-Global Runoff Data Center's composite runoff grid. Runoff can be thought of as the proportion of precipitation that is left *after* evapotranspiration and *after* the soil moisture deficit is satisfied. It is typically reported in units of depth (e.g., in millimeters) just like precipitation, and is an areally-averaged quantity (i.e., average runoff depth over a basin). I also used subnational data compiled from Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS), and CIESIN's own Gridded Population of the World (GPW). In Africa, a high proportion of agricultural households are dependent on rainfed lands. I hypothesized that drier areas with low runoff would have high proportions of underweight children. I also hypothesized that there would be a positive relationship between water availability and improved water sources.



The map at left shows the composite picture of underweight status (low weight-for-age) in Africa. The sub-national units represent the smallest geographic units for which the survey results are still statistically valid. The map at right simply aggregates the runoff data to the same geospatial units as the underweight (hunger) data so that statistical analyses can be performed.

In SSA, negative relationship between runoff & malnutrition

1

	Unstandardized Coefficients
(Constant)	34.179
Average Runoff (in mm)	-0.045 *
GDP per Capita (PPP, 1998)	-0.003 ***
North Africa Dummy Variable	-8.220 ***

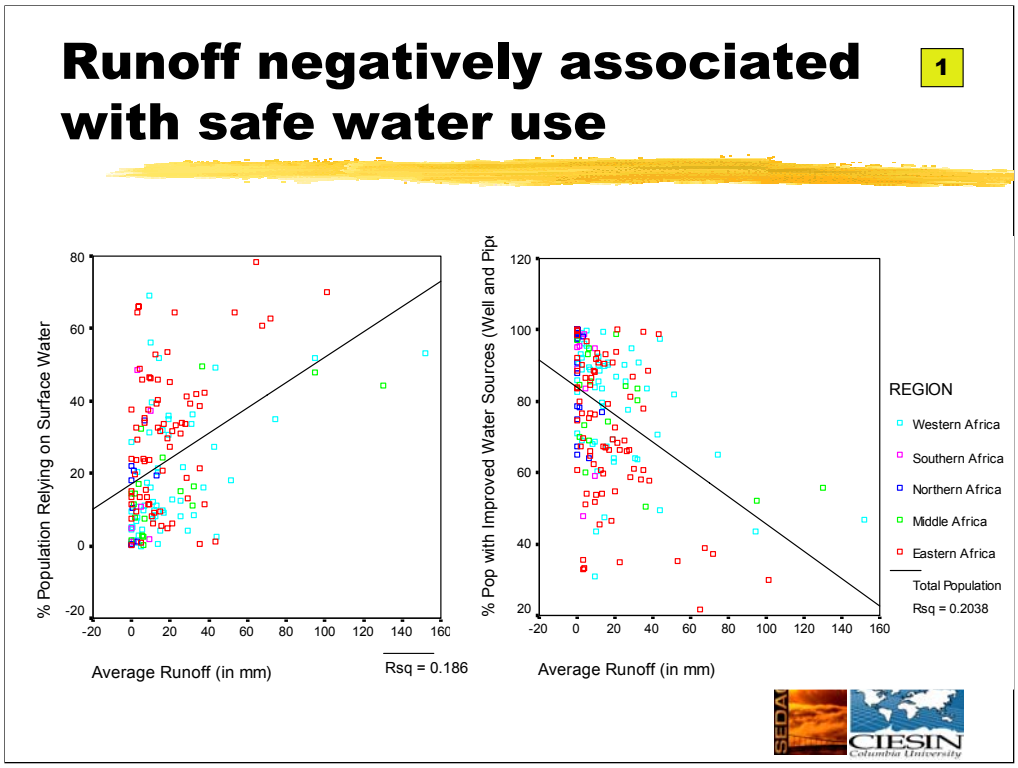
Dependent Variable: Percent of Children Underweight

Adjusted R^2 = 0.386 ***

N = 340, *** P < .001, ** P < .01, * P < .05

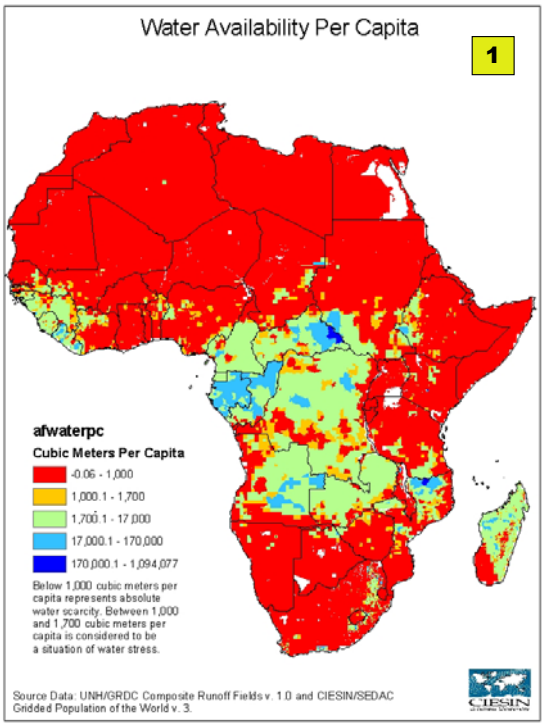
Note: The North Africa region includes Algeria, Egypt, Libya, Morocco, and Tunisia

I hypothesized a negative relationship between runoff levels and malnutrition, and this relationship was confirmed. In Sub-Saharan Africa there appears to be a negative relationship between annual water availability and child malnutrition (left). Because per capita income also reduces child malnutrition, in North Africa where development levels are higher per capita income strongly mediates the negative effect of low water availability.



The chart at left shows that as water runoff increases, the proportion of the population that relies on surface water as their sole water source also increases. The chart at right shows that as runoff increases, the proportion of the population relying on improved water sources (piped water and wells) decreases. Thus, paradoxically, those countries with the greatest physical water availability are more likely to have unimproved water sources. This suggests that countries with an abundance of water resources do not invest as much in improved water supply technologies, or that if those supplies exist, the population still finds it easier to obtain its supply from unimproved sources.

- The map at right depicts the extent of water scarcity in Africa. It represents available runoff divided by population for 30 minute grid cells.[1]
- The red areas suffer absolute water scarcity, whereas the brown areas suffer from water stress. Green and blue areas are relatively water abundant.



[1] This map does not take into account water that flows between grid cells in the form of river corridor discharge. This is why major river basins such as the Nile and Niger show up as being water scarce.

Africa Water & Development Hotspots

1

Sahel and savannah dominate

		% Underweight				Runoff (mm)		
	GNI PPP	High	Low	Mean	Nat'l Avg.	High	Low	Mean
Ethiopia	$800	52.5	13.9	40.2	47.1	35.2	0.0	16.5
Eritrea	$1030	51.2	23.4	41.1	43.7	5.5	0.0	1.6
Sudan	$1750	51.3	34.2	43.0	-	6.9	0.0	2.1
Madagascar	$820	44.1	31.8	38.1	40.0	101.1	9.3	61.3
Mali	$770	40.7	25.8	34.2	40.0	13.8	0.0	5.1
Niger	$880	51.0	33.2	39.5	39.6	1.4	0.0	0.6
Burkina Faso	$1120	36.8	32.9	34.8	37.6	6.2	1.2	4.0
Tanzania	$520	48.2	14.2	30.8	30.9	26.2	0.0	9.3
Chad	$1060	70.6	24.6	43.2	27.6	31.8	0.0	8.8
Mozambique	$1050	49.8	5.7	27.2	26.1	38.0	0.0	13.8
Mauritania	$1940	37.1	21.4	30.5	23.0	0.4	0.0	0.1

Note: High, Low and Mean for % Underweight and Runoff relate to the figures for the subnational units in each country. GNI PPP per capita, 2001 (US$) is the gross national income in purchasing power parity (PPP) divided by midyear population, from the World Bank, *World Development Indicators 2002*.

SEDAC CIESIN Columbia University

Water and Development Hotspots are countries with low per capita income, high malnutrition, and low levels of runoff. In these countries a high proportion of GDP is derived from agriculture. Therefore, water scarcity is likely to be a brake to development.

Brazil – Examining the determinants of HDI

2

- Brazil's Human Development data set for 4,492 municipalities, 1991
- CIESIN added:
  - Total area and % area in various biomes
  - % area within 100 km of the coast
  - Average growing season length
  - Crop constraints (from GAEZ)
  - % territory in top three crop suitability classes
  - Average human footprint

Brazil represents a "gold standard" of sorts for the availability of sub-national data. The 1991 Atlas for Human Development has data on a variety of development indicators for approximately 4,500 municipalities. CIESIN complemented this data source with a variety of environmental data sets, all of which were gridded data that were aggregated up to municipal boundaries.

2

Brazil – Predictors of HDI

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.547	.008		68.244	.000
	% of municipio's area covered in deserts and xeric shrublands	-1.56E-03	.000	-.415	-40.419	.000
	Perctage of Mun. falling within 100 km of coast	-3.75E-04	.000	-.114	-11.459	.000
	Crop Contraints	-2.07E-02	.001	-.155	-15.584	.000
	% of Population Living in Urban Areas, 1991	2.921E-03	.000	.475	46.369	.000

a. Dependent Variable: HDI-M: 1991

$R^2 = .567$

This shows that more than 50 percent of the variation in municipal-level Human Development Index scores can be predicted by variations in the following variables:

1. Percent of the municipality's area that is covered in deserts and xeric shrublands.
2. Percentage of the municipality falling within 100 km of the coast.
3. The average degree of crop constraints (soil, climatic, and slopes)
4. The percent of the population living in urban areas.

The municipalities with less area in arid lands, more inland, lower crop constraints, and higher proportions of the population living in urban areas have higher levels of human development. All of the variables were in the expected direction except the distance from the coast. Perhaps this is because of the impact of large metropolitan areas such as Sao Paulo, Brazilia, Belo Horizonte and Curitiba that are further than 100 km from the coast and well developed.

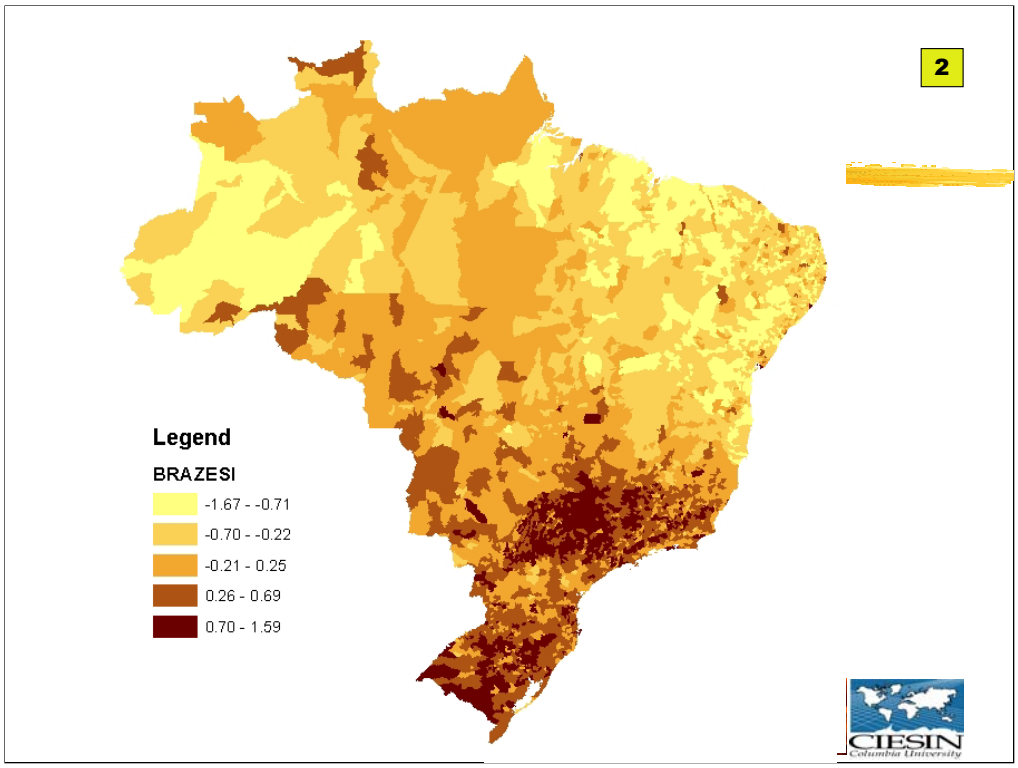
Brazil – Env'tal and Human Development Potential

2

This is measured as an unweighted average of the z-scores for:

- Average level of human impact on the environment
- Proportion of territory in the top 3 crop suitability classes
- Average level of climatic, soil, and terrain slope constraints
- Human Development Index
- % of children 7-14 who do not attend school
- % of population >25 years with more than 11 yrs of schooling
- Adult illiteracy rate
- % of domiciles with adequate water supply
- % of domiciles with adequate sewerage

In a very preliminary manner, I set out to create a measure of environmental and human development potential based upon available and comparable data at the municipality level, and assumptions regarding pre-requisites for rural sustainable development. I considered these to be human capital, a supply of adequate water and sanitation services, and agricultural potential. Note that I did not have ready access to data on market access or roads and other infrastructure, which in an ideal index would also be included. The variables used for Human Capital and Supply of Adequate Services were the following, all obtained from the Atlas of Human Development for Brazil (1991). **Human Capital:** Human Development Index; Percent of children 7-14 who attend school; Percent of population >25 years with more than 11 years of schooling; and Adult literacy rate. **Supply of Adequate Services:** Percent of domiciles with adequate water supply; and Percent of domiciles with adequate sewerage. I then added some of our own, CIESIN-generated variables to measure agricultural potential. All of these were data sets on a 1 km square grid. Values for municipalities represent some aggregation of the values of the grid cells within that municipality. The variables include: **Agricultural Potential:** The proportion of the territory in the top 3 crop suitability classes (from the FAO/IIASA Global Agro-ecosystem Zone Assessment); The average level of climatic, soil and terrain slope constraints (from the same Assessment); and The average level of human impact on the environment (from CIESIN's Human Footprint data set).



So, what are the results? This slide provides a map of scores for Brazil. The darker municipalities represent those with higher environmental and human development potential. It is not terribly surprising that the southern most parts of Brazil are the ones that have the highest potential. The index reflects the fact that these are the regions that people have historically found most suitable for agriculture and human industries, and therefore they have been settled longest, and are also the most densely settled. Nevertheless, there are several municipalities in Amazonia and the northeast that have high levels of potential.

Predicting infant mortality using sub-national data

3

- Sub-national data on IMR from DHS, MICS, ANDI, and NHDRs
- Hypothesized the following physical, biological, and epidemiological correlates to IMR levels
  - Malaria (+ Correlation)
  - Precipitation (- Correlation)
  - Growing Season Length (- Correlation)
  - Agro-ecosystem Constraints (+ Correlation)
  - Rainfed Crop Suitability (- Correlation)
  - Ecosystem/Levels of Aridity (+ Correlation)

This example comes from work by Marc Levy and Juju Wang at CIESIN. The first step was the compilation of the sub-national data on infant mortality (see earlier slides) from survey data and the national human development reports. Levy and Wang hypothesized several physical, biological and epidemiological correlates to infant mortality rates (IMRs), and compiled the data from gridded data sources at the same level of sub-national geography.

3

Best models predicting IMR

- For Africa, the best model predicting IMR includes Malaria and GDP p.c. (R^2 = .312)

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	264.96	21.107		12.553	.000
	MALARIA	20.513	7.968	.142	2.575	.011
	GDPLOG	-56.87	6.277	-.501	-9.061	.000

a. Dependent Variable: ADJIMR

- For Asia, the best model includes the length of growing season and GDP p.c. (R^2 = .436)

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	237.790	10.564		22.509	.000
	GDPLOG	-52.207	2.879	-.614	-18.133	.000
	GROWSEA	-22.756	2.747	-.280	-8.284	.000

a. Dependent Variable: ADJIMR

They found that for Africa, the best predictors of IMR were malaria and gross domestic product (GDP) per capita. In Asia, the best predictors were the length of the growing season (longer growing seasons permit double-cropping) and GDP per capita.

Challenges (1)

- Finding sub-national data on all relevant PDE variables is challenging (e.g. income)
- Time series analysis is challenging due to changing sub-national administrative boundaries (e.g. Brazil), lack of historical depth
- Sub-national boundaries are not consistent across data sources (e.g. DHS versus admin1)

Although the promise for sub-national analysis is great, there are also many challenges.

- Finding sub-national data on all variables of interest is often difficult. A partial solution is to apply the same value for every sub-national unit in a country, as was done for income in the infant mortality analysis. However, we know that income varies significantly over space, so failing actual measures, some modeling might be required to redistribute income through assumptions about night-time lights, for instance.
- Even using national-level data in global time series analyses can be challenging, because national borders change over time as countries merge or divide. This problem is multiplied significantly in sub-national analyses, as the administrative divisions within a country can change significantly over time. Between 1991 and 2001, close to 1,000 municipalities were created in Brazil by dividing the larger municipalities.
- Sub-national boundaries are often inconsistent between data sources. For instance demographic and health survey results are not valid at the admin 1 level for all countries, so they have created their own boundaries which represent something more like sub-national regions comprised of more than one state or province. If your other socioeconomic variables are collected at differing sub-national scales, they either have to be broken down into smaller units, or aggregated to larger units, or gridded and re-aggregated.

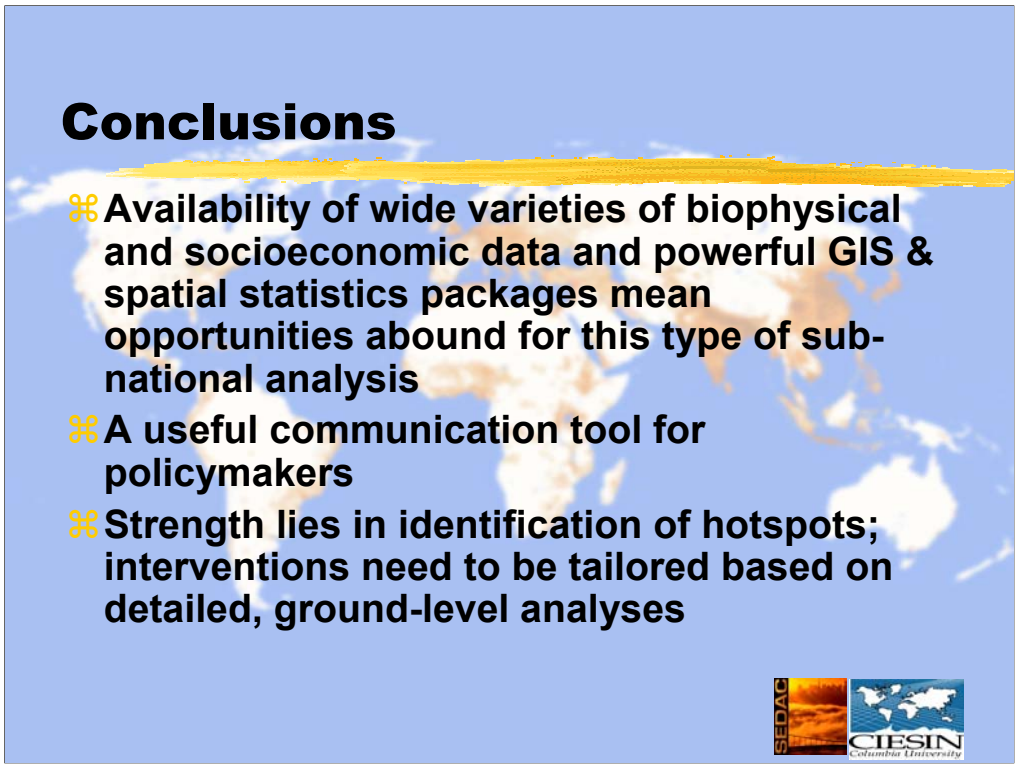
Challenges (2)

- Danger of uncritical use of questionable input data quality
- Spatial autocorrelation
- Boundary matching non-trivial

•In the rush to perform interesting analyses, there is a temptation to utilize data sources uncritically. This may represent a particular problem for social scientists using environmental data sets, and for physical scientists using socioeconomic data sets.

•In the real world, phenomena co-vary over space. Spatial autocorrelation is the term used to describe this. Non-spatial independence suggests many statistical tools and inferences are inappropriate. Correlation coefficients or ordinary least squares regressions (OLS) to predict a consequence assumes that the observations have been selected randomly. If the observations, however, are spatially clustered in some way, the estimates obtained from the correlation coefficient or OLS estimator will be biased and overly precise. A lot of research is going into this problem, but there remain differences of opinion as to how important it is, and to what degree statistical analyses using spatial data are invalid if they do not attempt some kind of corrective factor. The paper by Mageean and O'Conner at this conference has addressed this issue in considerably more detail.

•Finally, for global analyses of sub-national data, one must often rely on different sources. Matching the national-level boundaries between sources is a non-trivial task that can take days if not weeks, depending on the number of different data sets used.



In conclusion:

•The availability of wide varieties of biophysical and socioeconomic data and powerful GIS & spatial statistics packages mean opportunities abound for this type of sub-national analysis.

•One of the strengths of this kind of analysis is that it is a useful communication tool for policymakers, and particularly local policy makers who find little use for national-level analyses. The maps generated can be a powerful tool for understanding the patterns of population, development and environmental issues.

•There is a real strength in being able to identify PDE hotspots. However, interventions still need to be tailored based on detailed, ground-level analyses.