University of Copenhagen PhD Short Course

ArcGIS 10 for Social Science Empirical Research

Lecture 5

Elevation

Masayuki Kudamatsu

IIES, Stockholm University

4 October, 2013

Elevation in economics

A great source of exogenous variation!

Example 1: Dinkelman (2011)

$$\Delta y_{jdt} = \alpha_1 + \alpha_2 \Delta T_{jdt} + \mathbf{X}_{jd0} + \lambda_d + \Delta \varepsilon_{jdt}$$

- Estimate impact of electrification (ΔT_{jdt}) on female labor supply Δy_{jdt} w/ 2-period panel from S. Africa
- Community-level mean land gradient as IV for electrification

Example 2: Nunn & Puga (2012)

- Estimate the impact of terrain ruggedness on per capita income
- Ruggedness is measured by the sum of squared differences in elevation between neighboring cells
    - Raster cell (30x30 arc-sec) level data is downloadable from Diego Puga's website (diegopuga.org/data/rugged/#grid)

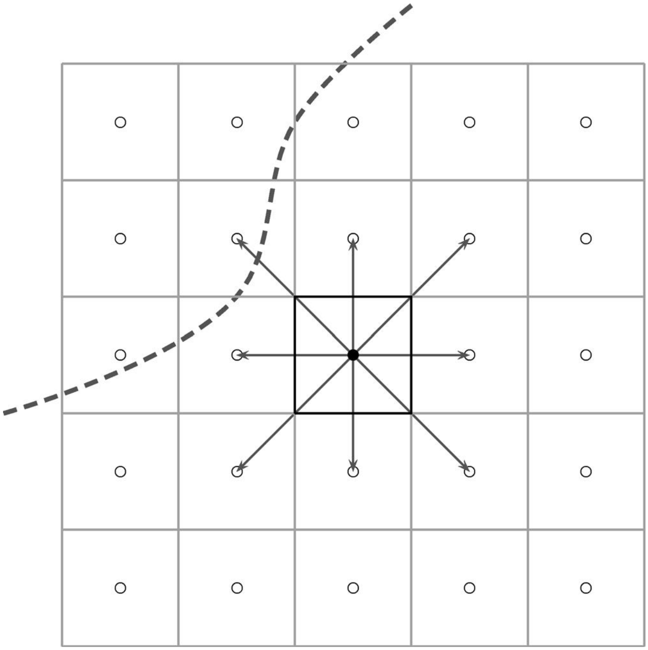


Figure 1 of Nunn & Puga (2012)

Example 2: Nunn & Puga (2012)

$$y_i = \beta_0 + \beta_1 r_i + \beta_2 r_i \cdot Africa_i + \beta_3 Africa_i + \varepsilon_i$$

- Show impact of ruggedness (r_i) on per capita income (y_i) to be
  - Positive in Africa ($\beta_1 + \beta_2 > 0$)
  - Negative outside Africa ($\beta_1 < 0$)
- Once slave export controlled for, $\beta_2 = 0$

Example 3: Qian (2008)

$$sex_{ic} = (tea_i \times post_c)\beta + \mathbf{x'}_{ic}\gamma + \psi_i + \gamma_c + \varepsilon_{ic}$$

- Estimate impact of tea production after 1979 ($tea_i \times post_c$) on sex ratio of cohorts born in year c in county i in China
- Mean slope of county i interacted w/ $post_c$: IV for ($tea_i \times post_c$)

Example 4: Olken (2009)

$$S_{vsd} = \alpha_d + \beta TV_{sd} + \mathbf{X}_{sd}\gamma + \varepsilon_{vsd}$$

- Estimate impact of # of TV channels available (TV_{sd}) on social capital in village v (S_{vsd})
- Signal strength in subdistrict s: IV for TV_{sd}

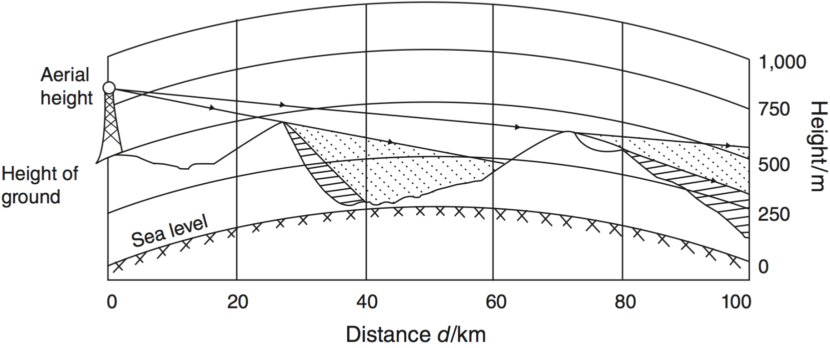


Figure 2 of Olken (2009)

Example 5: Yanagizawa (2009)

$$\ln(h_{ic}) = \beta r_{ic} + \mathbf{X}_{ic}\pi + \gamma_c + \varepsilon_{ic}$$

- Estimate impact of anti-Hutu radio station on # of genocide prosecutions per capita (h_{ic})
- r_{ic}: fraction of village i's areas w/ radio signal reception, obtained from elevation data & GIS

Example 6: Duflo and Pande (2007)

Outline

1. Duflo & Pande (2007)
2. Replicate the GIS data for Duflo & Pande (2007)

1.1 Research questions

- What’s the impact of irrigation dams on agricultural production and rural poverty in India?
- What’s the distributional consequence of building irrigation dams in India?

1.1 Research questions (cont.)

- Interesting?
  - 45,000+ dams worldwide
  - Built on nearly 1/2 of rivers worldwide
  - India: world's 3rd most prolific dam builder
  - Believed to reduce poverty, but no evidence
  - A public policy involving winners & losers
- Original?
  - Using geography to obtain credible estimates
- Feasible?

1.2 “Theory”

- Dams in upstream: beneficial
  - Irrigation
- Dams in neighborhood/downstream: costly
  - Displacement
  - Lower land productivity (due to salinity & waterlogging)
  - Water use restricted

$\Rightarrow$ Motivate 2nd-stage regression specification

1.2 “Theory” (cont.)

- Dams: easier to build if river gradient is
  - moderate (for irrigation)
  - very steep (for hydroelectricity)

$\Rightarrow$ Motivate 1st-stage regression specification

1.3 Data

- Unit of analysis: districts
- Annual agricultural production, 1971-1999, for 271 districts
- Poverty data in 1973, 83, 87, 93, 99 for 374 districts
- Dams: location (nearest city) and date of completion from World Registry of Large Dams

1.3 Data (cont.)

Fraction of river area with gradient more than 6%, 3-6%, 1.5-3%

- Data source: GTOPO30 (elevation at 30-arc second grid space) & Digital Chart of World (river drainage network)
- Identify GTOPO30 cells where rivers flow
- Calculate gradient in such cells from elevation data

1.4a Empirical strategy

- We have panel data
- But elevation is fixed over time
- How to predict dam construction over time based on elevation?

1.4a Empirical strategy

1st stage

$$D_{ist} = \sum_{k=2}^{4} \alpha_k (RGr_{ki} * \bar{D}_{st}) + \beta(\mathbf{M}_i * \bar{D}_{st}) + \sum_{k=2}^{4} \gamma_k (RGr_{ki} * I_t) + \nu_i + \mu_{st} + \omega_{ist}$$

D_{ist}: # of dams in district i of state s in year t

1.4a Empirical strategy

1st stage

$$
\begin{aligned}
D_{ist} \; = \; & \sum_{k=2}^{4} \alpha_k (RGr_{ki} * \bar{D}_{st}) + \beta(\mathbf{M}_i * \bar{D}_{st}) \\
& + \sum_{k=2}^{4} \gamma_k (RGr_{ki} * I_t) + \nu_i + \mu_{st} + \omega_{ist}
\end{aligned}
$$

RGr_{ki}: fraction of river areas with gradient falling in category k

1.4a Empirical strategy

1st stage

$$
\begin{aligned}
D_{ist} \; = \; & \sum_{k=2}^{4} \alpha_k (RGr_{ki} * \bar{D}_{st}) + \beta(\mathbf{M}_i * \bar{D}_{st}) \\
& + \sum_{k=2}^{4} \gamma_k (RGr_{ki} * I_t) + \nu_i + \mu_{st} + \omega_{ist}
\end{aligned}
$$

k: 2 for 1.5 to 3%; 3 for 3-6%; 4 for above 6%

- We will learn how to construct RGr_{ki} from elevation data, river polylines, and district polygons in ArcGIS

1.4a Empirical strategy

1st stage

$$D_{ist} = \sum_{k=2}^{4} \alpha_k (RGr_{ki} * \bar{D}_{st}) + \beta(\mathbf{M}_i * \bar{D}_{st}) + \sum_{k=2}^{4} \gamma_k (RGr_{ki} * I_t) + \nu_i + \mu_{st} + \omega_{ist}$$

$\bar{D}_{st}$: # of dams in India in year t (Figure III) multiplied by fraction of dams in state s in 1970

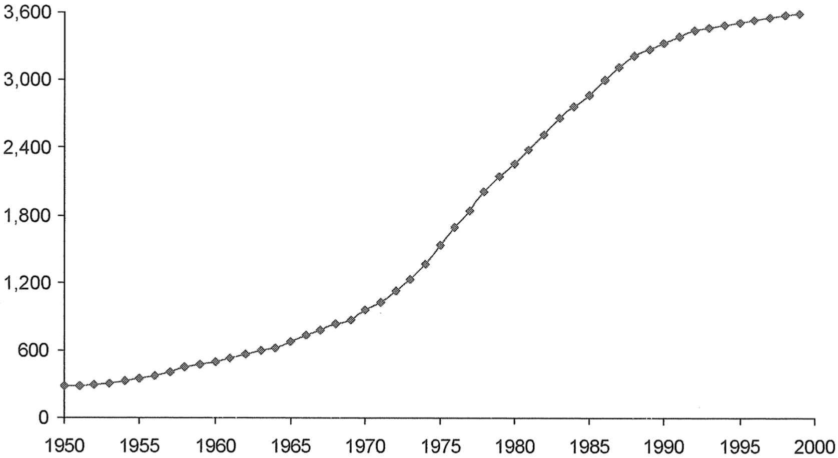


FIGURE III
Total Dams Constructed in India, ICOLD Dam Register for India

- Why not the actual # of dams in state s in year t instead of $\bar{D}_{st}$?

1.4a Empirical strategy

1st stage

$$D_{ist} = \sum_{k=2}^{4} \alpha_k (RGr_{ki} * \bar{D}_{st}) + \beta(\mathbf{M}_i * \bar{D}_{st}) + \sum_{k=2}^{4} \gamma_k (RGr_{ki} * I_t) + \nu_i + \mu_{st} + \omega_{ist}$$

ν_i: district FE

μ_{st}: state-year FE

- Why not year FE instead of state-year FE?

1.4a Empirical strategy

1st stage

$$D_{ist} = \sum_{k=2}^{4} \alpha_k (RGr_{ki} * \bar{D}_{st}) + \beta(\mathbf{M}_i * \bar{D}_{st}) + \sum_{k=2}^{4} \gamma_k (RGr_{ki} * I_t) + \nu_i + \mu_{st} + \omega_{ist}$$

$\mathbf{M}_i$: area, elevation, overall gradient, river length

Why control for $(\mathbf{M}_i * \bar{D}_{st})$?

- We will learn how to construct $\mathbf{M}_i$, in particular river length, in ArcGIS.

1.4a Empirical strategy

1st stage

$$D_{ist} = \sum_{k=2}^{4} \alpha_k (RGr_{ki} * \bar{D}_{st}) + \beta(\mathbf{M}_i * \bar{D}_{st}) + \sum_{k=2}^{4} \gamma_k (RGr_{ki} * I_t) + \nu_i + \mu_{st} + \omega_{ist}$$

I_t: year dummies

Why control for $(RGr_{ki} * I_t)$?

1st stage results

TABLE II
GEOGRAPHY AND DAM CONSTRUCTION

	Number of dams		
	Cross-section (1999)	Poverty sample	Production sample
	Not interacted	Interacted with predicted number of dams in the state	
	(1)	(2)	(3)
Fraction river gradient 1.5–3%	0.278	0.153	0.176
	(0.122)	(0.040)	(0.094)
Fraction river gradient 3–6%	−0.210	−0.191	−0.219
	(0.127)	(0.065)	(0.128)
Fraction river gradient above 6%	0.014	0.075	0.097
	(0.033)	(0.031)	(0.043)
F-test for river gradient	1.764	6.372	7.68
[*p*-value]	[0.15]	[0.000]	[0.053]
Geography controls	Yes	Yes	Yes
State*year and river gradient*year interactions	No	Yes	Yes
Fixed effects	State	District	District
N	374	1855	7743

1.4b Empirical strategy

2nd stage

$$
\begin{aligned}
y_{ist} &= \gamma_i + \eta_{st} + \delta D_{ist} + \delta^U D^U_{ist} \\
&\quad + \mathbf{Z}_{ist}\delta_{\mathbf{Z}} + \mathbf{Z^U}_{ist}\delta^{\mathbf{U}}_{\mathbf{Z}} + \varepsilon_{ist}
\end{aligned}
$$

w/ $\hat{D}_{ist}$, $\hat{D}^U_{ist}$ as excluded IVs

- y_{ist}: outcome variable
- γ_i: district FE
- η_{st}: state-year FE
- D^U_{ist}: # of dams in all upstream districts for district i

1.4b Empirical strategy

2nd stage

$$\begin{aligned} y_{ist} &= \gamma_i + \eta_{st} + \delta D_{ist} + \delta^U D^U_{ist} \\ &+ \mathbf{Z}_{ist}\delta_{\mathbf{Z}} + \mathbf{Z^U}_{ist}\delta^{\mathbf{U}}_{\mathbf{Z}} + \varepsilon_{ist} \end{aligned}$$

w/ $\hat{D}_{ist}$, $\hat{D}^U_{ist}$ as excluded IVs

- $\hat{D}_{ist}$: fitted value for D_{ist} from 1st stage
- $\hat{D}^U{}_{ist}$: the sum of fitted values for D_{ist} over all upstream districts

- Why not using $RGr_{ki} * \bar{D}_{st}$ as instruments?

1.4b Empirical strategy

2nd stage

$$
\begin{aligned}
y_{ist} &= \gamma_i + \eta_{st} + \delta D_{ist} + \delta^U D^U_{ist} \\
&\quad + \mathbf{Z}_{ist}\delta_{\mathbf{Z}} + \mathbf{Z^U}_{ist}\delta^{\mathbf{U}}_{\mathbf{Z}} + \varepsilon_{ist}
\end{aligned}
$$

w/ $\hat{D}_{ist}, \hat{D}^U_{ist}$ as excluded IVs

- $\mathbf{Z}_{ist}$: vector of $\mathbf{M}_i * \bar{D}_{st}, RGr_{ki} * I_t$
- $\mathbf{Z^U}_{ist}$: vector of $\mathbf{M}_i * \bar{D}_{st}, RGr_{ki} * I_t$ for upstream districts (summed for river length & district areas, averaged for elevation & overall gradient)

1.4c Empirical strategy

Estimation method

- For agricultural outcomes, Feasible optimal IV with S.E. robust to arbitrary covariance of the residual w/i state (see ft. 15 for how to implement this)
- Why?
  - Autocorrelation at state level
  - Feasible GLS: more efficient than OLS with S.E. clustered

  $\Rightarrow$ Small effect more likely to be detected (Power of test $\nearrow$)

1.5a Impact on agriculture

	Area			
	Gross irrigated area		Gross cultivated area	
	Level	Log	Level	Log
	(1)	(2)	(3)	(4)
	Part A. FGL			
Dams				
Own district	14.528	0.131	114.493	0.094
	(13.300)	(0.156)	(47.838)	(0.059)
Upstream	17.830	0.198	77.641	0.028
	(12.639)	(0.162)	(48.233)	(0.054)
	Part B. Feasible Op			
Dams				
Own district	232.092	0.728	325.358	0.875
	(235.847)	(1.002)	(263.509)	(0.590)
Upstream	49.754	0.328	58.602	0.088
	(22.339)	(0.154)	(35.674)	(0.062)
N	4,536	4,536	4,522	4,522
First stage *F*-statistic (own district)	8.48	8.48	8.51	8.51

1.5a Impact on agriculture

	Agricultural production			
	Production	Yield	Production	
	All crops		Water-intensive crops	Non–water-intensive crops
	(6)	(7)	(8)	(9)
Dams				
Own district	0.184	0.152	0.063	0.640
	(0.334)	(0.196)	(0.334)	(0.585)
Upstream	0.530	0.227	0.569	0.801
	(0.155)	(0.141)	(0.243)	(0.307)
Dams				
Own district	0.085	−0.033	0.366	−0.105
	(0.699)	(0.451)	(0.782)	(1.349)
Upstream	0.341	0.193	0.470	0.181
	(0.118)	(0.097)	(0.154)	(0.307)
N	7,078	7,077	7,143	6,786
First stage *F*-statistic (own district)	9.22	9.22	9.03	9.14

- In average district, # of dams in upstream increased from 3.6 to 13.9 (Table I)

1.5b Impact on poverty

	Per-capita expenditure	Headcount ratio: Original	Headcount ratio: Assume poor in-migrants	Headcount ratio: Assume rich in-migrants
	(1)	(2)	(3)	(4)
Part A. OLS / FGLS				
Dams				
Own district	−0.289	0.273	0.407	0.174
	(0.115)	(0.084)	(0.083)	(0.081)
Upstream	0.093	−0.083	−0.079	−0.082
	(0.057)	(0.039)	(0.038)	(0.038)
Part B. 2SLS / Feasible Optimal IV				
Dams				
Own district	−0.457	0.772	0.879	0.651
	(0.467)	(0.324)	(0.314)	(0.315)
Upstream	0.142	−0.154	−0.149	−0.150
	(0.084)	(0.068)	(0.066)	(0.066)
N	1,799	1,799	1,799	1,799
First stage F-statistic (own district)	7.71	7.71	7.71	7.71

LATE

- The estimated impact DOES NOT capture the impact of dams constructed for, say, political reasons

Example 7: Lipscomb, Mobarak, and Barham 2012

- Estimate the impact of electricity availability on poverty in Brazil
- The instrument is constructed by
  - Regress (probit) having hydropower plants on topography measures (river, gradient, etc.) in Brazil (or USA)
  - Use estimated coefficients to rank locations by its geographic suitability for hydropower plants
  - Obtain total # of hydropower plants constructed nationwide in each period

- If n plants were constructed in period 1, turn on the instrument for the n most suitable locations
- If m plants were additionally constructed in period 2, turn on the instrument for the next m most suitable locations
- And so forth...

$\Rightarrow$ 1st-stage F-stats: > 20

- DOWNLOAD T:/Economics/Lecture6
- Unzip
  - g2009_1990_2.zip (subnational district polygons)
  - 10m-rivers-lake-centerlines.zip (river polylines)
- Browse them and SRTM30/e060n40c (elevation raster)

2. Replicating geography variables for Duflo & Pande (2007)

- Exercise 1: Each district’s fraction of river areas w/ gradient 1.5-3%, 3-6%, >6%
- Exercise 2: Each district’s total length of rivers

2.1 Overview of Exercise 1

- Use Zonal Statistics as Table (cf. Lec 5) with two inputs:
  a. Multi-part river polylines by district
     - All the rivers within a district are dissolved as one feature
  b. Raster of each gradient category
     - Raster value is 1 if gradient is, for example, 1.5-3% and 0 otherwise

$\Rightarrow$ Mean within each zone is the fraction of river areas in each gradient category

2.1 Overview of Exercise 1 (cont.)

Therefore we need:

a. Indian district polygons with unique positive integer ID
b. Multi-part river polylines for each district
c. Raster for each gradient category

2.1a India district polygons

- Start w/ GAUL (v2009) second level admin boundaries for 1990
- Copy Features
- Define Projection (as WGS 1984 $\Leftarrow$ see meta data)
- Select (India)
- Dissolve (by district)
- Add Field
- Calculate Field (by !FID! + 1)

Select

(cf. Lec 5 exercises 1-2)

- Before adding the Select tool to the Model Builder, it's always a good practice to run the model to create the input file.

$\Rightarrow$ Easier to write the expression for selection criteria

Select (cont.)

- Expression:
  "ADM0_NAME" = 'India'

- Field name must be enclosed by double quotation marks
- Just double-click it from the list of fields at the top (if the input file is already created)

Select (cont.)

- Expression:
  "ADM0_NAME" = 'India'

- If field value is a string, must be enclosed by single quotation marks
- Click "Get Unique Values" and double click the field value (if the input file is already created)
  - cf. If numeric, no need to enclose (In Lec. 5, it was "FID_GREG" = -1)

- Now run the model. Browse the output's attribute table. (right-click ADM2_NAME (the field for district names) & click "Sort Ascending")
- Notice each district consists of more than one polygon.
- But we need one polygon for each district, to calculate fraction of rivers in each gradient category by district.
- The tool to do this is...

Dissolve

(cf. Lec 5 exercise 1)

- Dissolve Field: ADM1_NAME (state name), ADM2_NAME (district name)
  - ⇐ Some districts may have the same name but in different states (at least one such case, indeed)
- Check "Create multipart features"

- We are going to use Zonal Statistics as Table later to calculate each district's fraction of rivers in different gradient categories

$\Rightarrow$ Need to assign the unique positive integer ID to each district (cf. Lec 5 exercise 1)

- The tools to do this are...

Add Field

- Field Name: dist_id (or whatever)
- Field Type: SHORT
  - $\Leftarrow$ # of districts less than 32767

Calculate Field

- Field Name: dist_id (or the name you chose for Add Field)
- Expression: !FID!+1
- Expression Type: PYTHON 9.3

2.1 Overview of Exercise 1

We need to create:

a. Indian district polygons with unique positive integer ID
b. Multi-part river polylines for each district
c. Raster for each gradient category

2.1b Multi-part river line feature for each district

- Duflo & Pande (2007) use dnnet from the Digital Charts of the World (DCW)
- DCW appears to be no longer available due to its inaccuracy
- www.maproom.psu.edu/dcw recommends Natural Earth

2.1b Multi-part river line feature for each district (cont.)

- So we start w/ Natural Earth 1:10m river + lake centerlines (v1.4)
- Intersect (w/ district polygons) (cf. Lec 4)
- Dissolve (by district)
  - ⇐ Some districts have more than one river.
  - We need to keep the "dist_id" field so that Zonal Statistics as Table will create a dBASE table with "dist_id"

Dissolve

- Dissolve Field(s): dist_id

If you prefer keeping state/district names...

- Statistics Field(s)
  - Choose ADM1_NAME & ADM2_NAME
  - Then choose FIRST (or LAST) as Statistics Type

    ⇐ This is how to keep string variables after Dissolve.

2.1 Overview of Exercise 1

We need to create:

a. Indian district polygons with unique positive integer ID
b. Multi-part river polylines for each district
c. Raster for each gradient category

2.1c Raster for each gradient category

- Duflo & Pande (2007) use GTOPO30 for elevation
- SRTM30 (v2.1) now supersedes GTOPO30
- Global (excl. Antarctica) raster data w. spatial resolution 30 x 30 arc seconds (roughly 1km x 1km)
- We use E060N40 tile (which covers whole India)

Assignment 6z

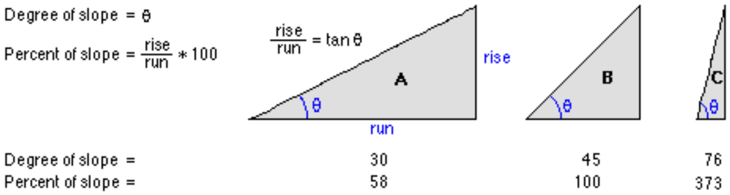
- Clean the SRTM30 data
  - Elevation ranges from -46 to 8685 in the data
  - To geoprocess raster data, you need to convert it into ESRI Grid format
  - When using Copy Raster to convert, by default, negative values are not allowed and 65536 will be added
  - We need to correct this by Raster Calculator
- See the Assignment6z folder for detail.

2.1c Raster for each gradient category

- Start w/ the properly converted SRTM30
- Slope tool to convert elevation into gradient in percentage
- Reclassify tool to convert gradient into categories

Slope

Returns θ if degree; $\tan\theta \times 100$ if percent_rise



Slope (cont.)

$$\tan\theta = \sqrt{(dz/dx)^2 + (dz/dy)^2}$$

$$dz/dx = \left[\frac{c + 2f + i}{4} - \frac{a + 2d + g}{4}\right]/2$$

$$dz/dy = \left[\frac{a + 2b + c}{4} - \frac{g + 2h + i}{4}\right]/2$$

a	b	c
d	e	f
g	h	i

Slope (cont.)

Choose z-factor (ie. # of (x,y) units in one z unit)

- By default, it's 1, ie. units are the same between (x,y) and z
- If (x,y) are in degrees & z is in meters (our case), choose 0.000009
  - $1^\circ \approx$ 111,120 meters
- This may not be accurate.
  - For middle- to high- latitude areas, 1° in longitude < 1° in latitude

Project Raster?

- If we look at high-latitude countries, we need to project the raster into UTM first before using the slope tool
- Project Raster tool does this
- It requires two arguments
- Resampling algorithm
  - Nearest / Majority for integer raster
  - Bilinear / Cubic for continuous raster
- Cell size
  - Pick the number in meters close to the original raster data spatial resolution

Project Raster? (cont.)

- Whatever choice of resampling algorithm & cell size causes some distortions
- For low-latitude areas, projection adds little for Slope
- We proceed without projecting raster below

cf. In Lec. 7, we will project raster.

Slope (cont.)

- Output measurement: PERCENT RISE
- Z factor: 0.000009

Now run the model. Next tool requires the input to be already created

Reclassify

- Transforms input raster into categorical raster
- Can be used for creating a dummy variable from original raster data
- Here we want to create a dummy which equals 1 if gradient is 1.5-3%, 3-6%, or >6%

Dummy for >6%

- Delete Entries (only if you see default categorization of raster values into 9 groups)
- Click Add Entry twice
- Then type as follows:

Old values	New values
0 - 6	0
6 - 193.229706	1
NoData	NoData

Dummy for 3-6%

Old values	New values
0 - 3	0
3 - 6	1
6 - 193.229706	0
NoData	NoData

Reclassify (cont.)

Similarly, reclassify the slop raster into 1.5-3%.

Assignment 6a

Create elevation category raster files (250-500m, 500-1000m, >1000m)

2.1d Fraction of river by gradient

- Now we're ready to obtain each district's fraction of river areas w/ gradient 1.5-3%, 3-6%, and >6%
- We can use Zonal Statistics as Table for this purpose (cf. Lec 5)
- The mean statistics gives the fraction
  - Raster cells of 30 by 30 arc seconds: roughly same size w/i each district

- India is located between 8°& 37°North
- 1°in longitude at $15^\circ = 107.551$km
- 1°in longitude at $30^\circ = 96.486$km
- Districts in India are at most 3° wide in latitude
- 30 seconds in longitude can differ up to 18m if 3° apart

$\Rightarrow$ Treating 30-second cells as the same size within district does not seem too bad

Zonal Statistics as Table

- Zone field: dist_id
- Statistics Type: MEAN

Assignment 6b

Produce dBASE tables for

- Fraction of district area w/ gradient 1.5-3%, 3-6%, >6%
- Fraction of district area w/ elevation 250-500m, 500-1000m, >1000m

Assignment 6c

Replicate Figure IV (average river gradient by district)

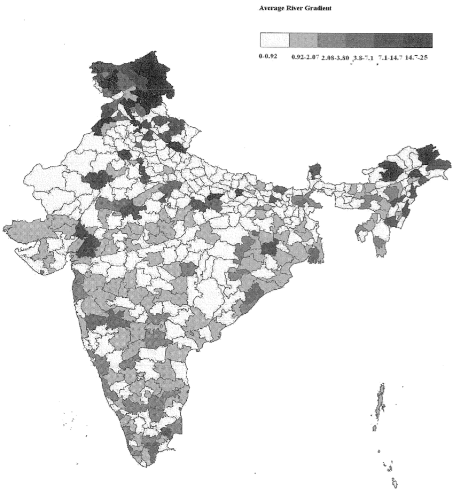


FIGURE IV

Assignment 6c (cont.)

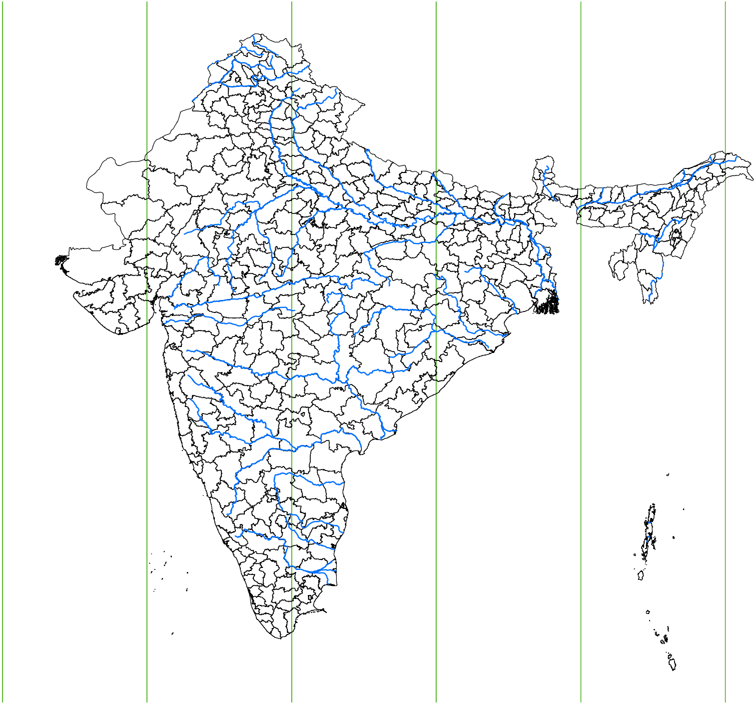
- You need to attach the zonal statistics table to the district polygon shapefile
- First, Copy Features to create a copy of the district polygon shapefile
- Then, Join Field (we will learn this tool later in this lecture)

2.2 Exercise 2: River length

- No single projection for large areas such as India preserves distance in every dimension...
- Stata's globdist can only be used for distance between two points, not for polyline length

$\Rightarrow$ Need to project each district into the appropriate UTM projection (cf. Lec 3)

$\Rightarrow$ Need to loop over districts



Steps to obtain river length

a. For each rivers-by-district, assign UTM zone number

b1. Extract a subset of rivers-by-district whose centroid is within the UTM zone

b2. Project it to the UTM

b3. Calculate the length

c. Make steps b1-b3 as a loop over UTM zone numbers in Python

Before starting...

- Create a new Model.
- We will use the river-district intersections (created in Exercise 1) as one of the inputs of the Model.

2.2a Assign UTM zone number

- Create centroid points for each rivers-by-district by Feature To Point
- Spatial Join them w/ UTM zone polygons available at "C:/Program Files (x86)/ArcGIS/Desktop10.1/Reference Systems/utm.shp"
- Use Join Field to attach UTM zone number to the original river polyline features

Spatial Join

- Join Operation: JOIN ONE TO ONE
- Check Keep All Target Features
- Field Map of Join Features:
  - keep everything if the tool is run by the Model Builder
  - delete everything if the tool is run by Python
- Match Option: INTERSECT or WITHIN

Join Field

- This tool essentially does what Stata's merge command does.
- Except that the variable name used for merging can be different between two datasets
- Overwrites the master dataset

$\Rightarrow$ Use Copy Features before using Join Field

Join Field (cont.)

- Input Table: the river polyline features created in Exercise 1
- Input Join Field: dist_id
- Join Table: the shape file created by Spatial Join
- Output Join Field: dist_id
- Join Fields: ZONE (ie. the field in the UTM zone shapefile indicating zone number)

2.2b Loop over UTM zones

- Select rivers-by-district for each UTM zone
- Project each file to the UTM
- Then use Add Field & Calculate Field to calculate length of river within each district
- Use the Model Builder to export the script within each loop
- Then use Python editor to make the loop

Select

- Expression: "ZONE" = 43

Project

- Output Coordinate System: Projected Coordinate Systems > UTM > WGS 1984 > Northern Hemisphere > WGS 1984 UTM Zone 43N.prj

Add Field

- Field Name: length (or whatever)
- Field Type: FLOAT

Calculate Field

- Field Name: length (or the name chosen for Add Field)
- Expression: float(!shape.length@kilometers!)
- Expression Type: PYTHON 9.3

!shape.length@kilometers!

- Without "@kilometers", the unit will be meters
- See Desktop Help for Calculate Field, for other units of measurement

- Now export this model into a Python script.
- Then create a loop over UTM zones from 43 to 47
  - It happens to be the case that there is no river in zone 42 (though India does cover zone 42)

2.2c Create a loop over values

- You can create a loop over UTM zone numbers, by typing

```
for zone in range(43,48):
```

- This command assigns 43 to the variable zone.
- Execute all the indented commands that follow
- Then assign 44 to the variable zone, and so forth.
- Repeat until 47.

Other syntax in Python for looping over values

- While loop

```
zone = 43
while zone < 48:
    commands
    zone = zone + 1
```

- List loop

```
zoneList = [43,44,45,46,47]
for zone in zoneList:
    commands
```

Scripting tip #1

- We may want to use the UTM zone number for output file names.
- The variable *zone* is numeric. We need to convert this into string
- Type

```
str(zone)
```

```
for zone in range(43,48):
    print "Processing UTM Zone "+str(zone)
    # Process: Select
    print "Extracting UTM Zone"
    arcpy.Select_analysis(river2_shp, "xxriver"+str(zone)+".shp", "\"ZONE\" = "+str(zone))
```

Scripting tip #2

- 3rd argument for Project is coordinate system
- Model Builder exports a very long string for this
- Which is not convenient for looping
- Alternatively, we could use the projection factory code & SpatialReference method

Coordinate systems' factory code

- factory codes for UTM projection for zone 43N to 47N are: 32643 to 32647
  - Each projection's factory code can be obtained from the help document for the SpatialReference method

(search "SpatialReference" at

resources.arcgis.com/en/help/main/10.1/index.html)

$\Rightarrow$ Define the factory code as a numerical variable (named, say, csfile)

$$\text{csfile} = 32600 + \text{zone}$$

- Then use SpatialReference method (case sensitive!) to create a variable that can be used for the Project tool

```
cs = arcpy.SpatialReference(csfile)
```

- Then use Project

```
arcpy.Project_management("xxrl"+str(zone)+".shp",
"rl"+str(zone)+".shp", cs)
```

Read dBASE output tables in Stata

```
#delimit ;
set more off;
clear all;
set debug on; /* Otherwise the error message doesn't make sense. */
cd Z:/Documents2/TEACHING/2010gis/L8review/datacreated;
foreach n of numlist 42(1)47 {;
  odbc load, table("river`n'.dbf") dsn("dBASE Files") lowercase clear;
  if `n' ~= 42 {;
    append using temp;
    };
  save temp, replace;
  };
```

Assignment 6d

Calculate district area

- Use the UTM projection as in river length
- Obtain the area in km^2

3. What we've learned for ArcGIS

- Create...
  - A shapefile for a specific area (e.g. a country) out of the global data
  - Multipart polyline features by district
  - Slope raster
  - A dummy variable raster
- Calculate the length of polyline features
- Loop over values

References for Lecture 6

Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." American Economic Review 101(7): 3078–3108.

Duflo, Esther, and Rohini Pande. 2007. "Dams." Quarterly Journal of Economics 122(2): 601-646.

Hansen, Christian B. 2007. "Generalized Least Squares Inference in Panel and Multilevel Models with Serial Correlation and Fixed Effects." *Journal of Econometrics* 140(2): 670-694.

Lipscomb, Molly, Ahmed Mushfiq Mobarak, and Tania Barham. 2012. "Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil." American Economic Journal: Applied Economics forthcoming.

Nunn, Nathan, and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." Review of Economics and Statistics 94(1): 20–36.

Olken, Benjamin A. 2009. "Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages." *American Economic Journal: Applied Economics* 1(4): 1-33.

Qian, Nancy. 2008. "Missing Women and the Price of Tea in China: The Effect of Sex-specific Income on Sex Imbalance." *Quarterly Journal of Economics* 123 (3).

Yanagizawa, David. 2009. "Propaganda and Conflict: Theory and Evidence From the Rwandan Genocide." http://www.hks.harvard.edu/fs/dyanagi/Research/RwandaDYD.pdf