Spatial Epidemiology

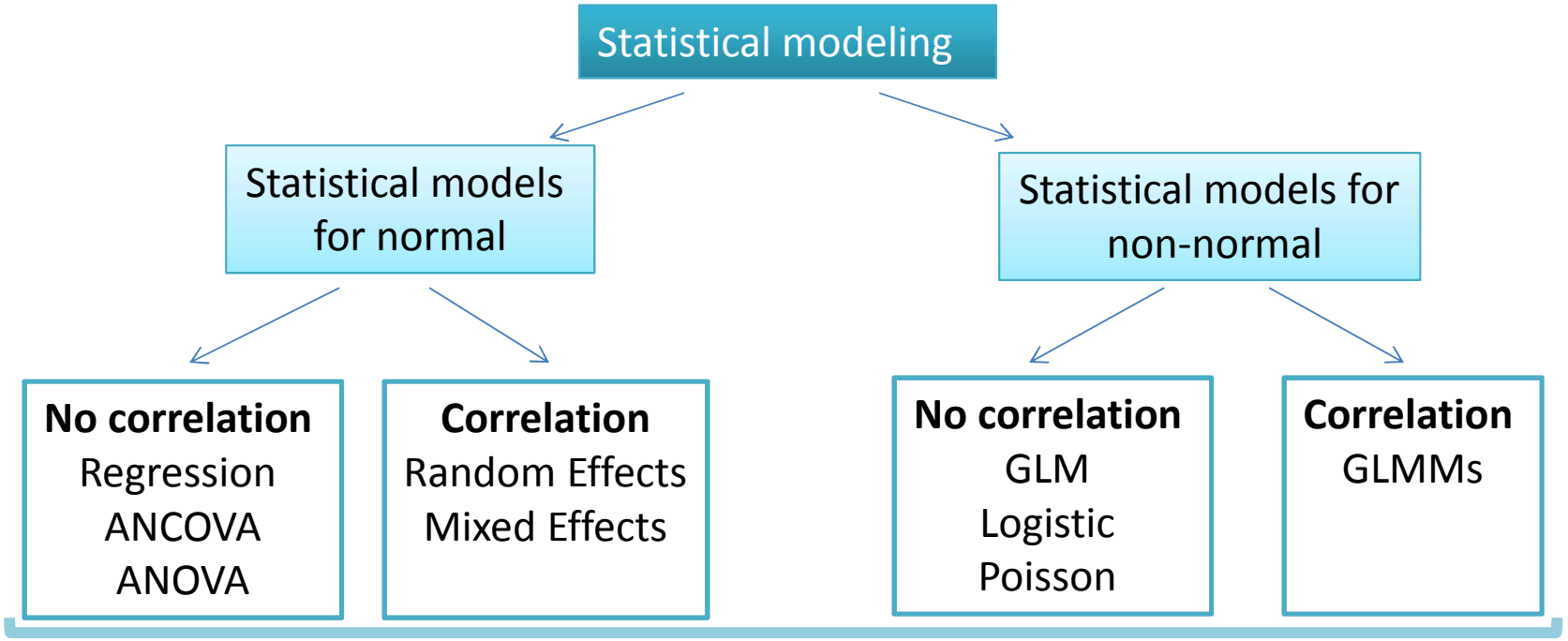
Damaris Kinyoki

27 April 2016

KEMRI | Wellcome Trust

Presentation Outline

- Background information
- Data
- Methods
- Application and discussion



Non-spatial model

Data = Pattern + residual

Explained variable

Unexplained variable

Spatial model

Data = Pattern + Spatial + residual

Explained variable

Spatial effects

Unexplained variable

Spatial autocorrelation

What is **spatial distribution**?

Arrangement of “something” on the earth surface.

First Law of Geography - Tobler’s (1970)

”everything is related to everything else, but near things are more related than distant things.”

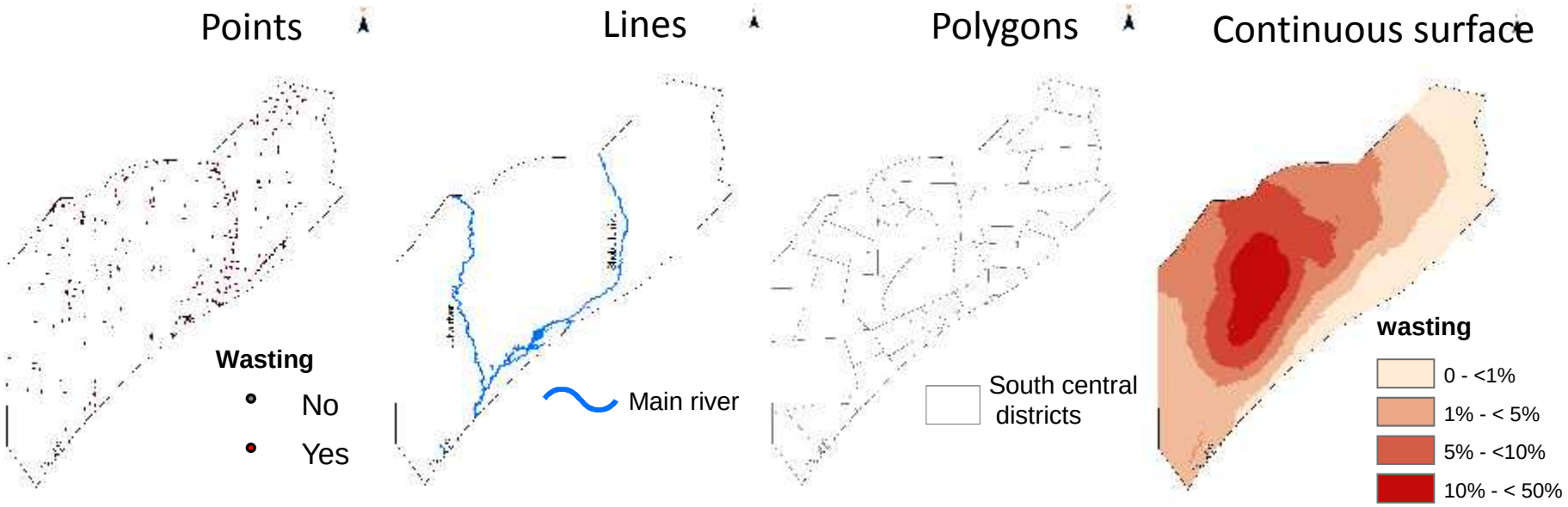
Values of outcomes variables at locations close together are more likely to be a like than those further apart

Similar values for a variable will tend to occur in nearby locations, leading to spatial clusters.

Where independent assumption is violated

1. Spatial autocorrelation - Data collected over a space/location
2. Serial autocorrelation - Data collected over time

Examples of Spatial Data



Types of spatial features

- Point - location, s, in space; a dot on a map.
- Line - a sequential collection of connected points. Roads, rivers, and geographical boundaries
- Area/Polygon - a region enclosed by lines. Counties, states, and census.
- Continuous surface - a region which a phenomenon in a continuous manner.

Main question

What is the geographical and temporal (time) variation of malaria prevalence among children age 2 – 10 years in Kenya?

Key questions:

1. How is malaria prevalence distributed geographically in Kenya?
2. How is malaria prevalence distributed in different time points annually?
3. What is the intra-annual or seasonal fluctuation of malaria prevalence ?
4. How is malaria prevalence distributed over space and time (space-time)?
5. What are the main determinants of these distributions?
6. What are the policy implications of this work?

Data processing: Malnutrition data

Data assembly

Entire dataset on malnutrition 2007 - 2010 (n=75,543)

↓

Exclusion of clusters without cluster names (n=652)

↓

Exclusion of clusters without coordinates (n=859)

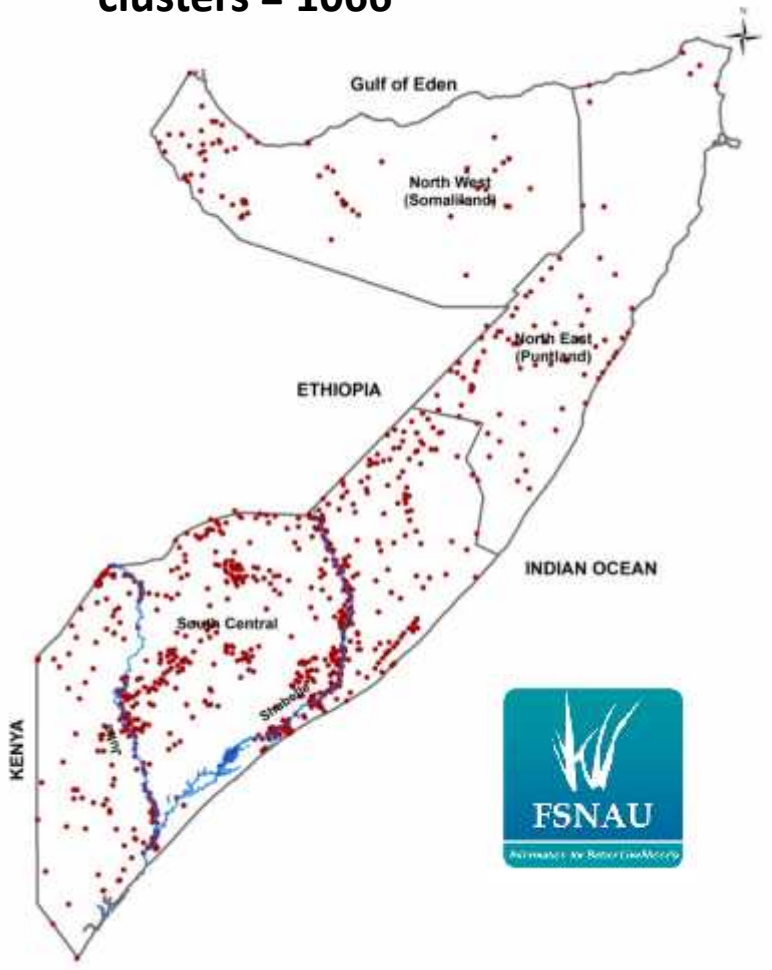
↓

Exclusion of clusters with missing dates (n=254)

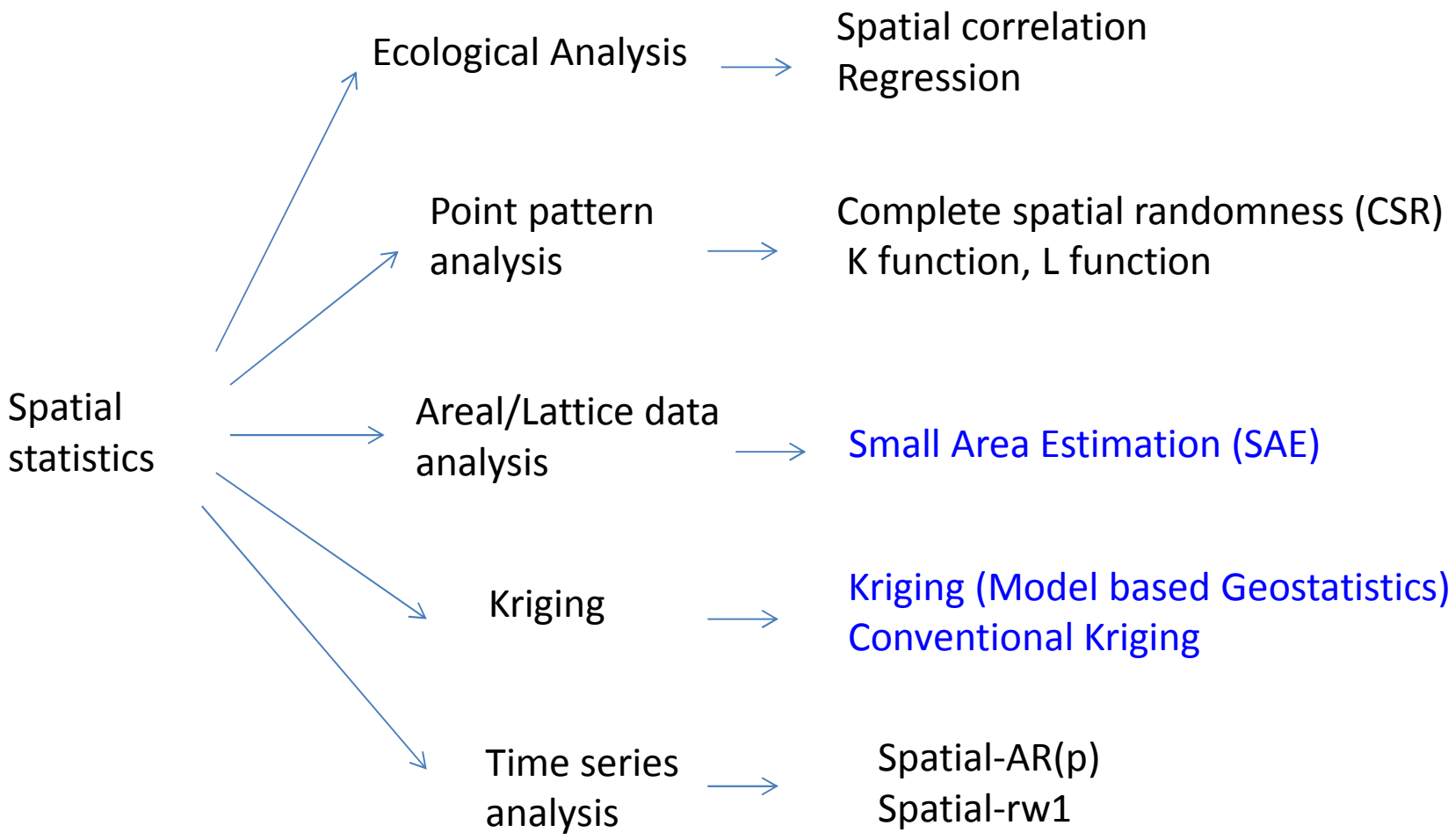
↓

Final dataset on malnutrition 2007 - 2010 (n= 73,778)

Data: 4 years of malnutrition data 2007-2010 (n = 73,778); number of clusters = 1066



Classes of Spatial Statistics



Bayesian Small Area Estimation

Small Area Estimation (Rao, 2003) is a statistical models that provide reliable estimates of a target variable in a set of small geographical areas.

Benefits

- The possibility of borrowing information from neighboring areas when estimating spatially-correlated random effects.
- Improving estimation in non-sampled areas.

Disadvantages

- Masks the variability within unit of analysis

Types of Bayesian Hierarchical SAE Models

Unit level model

Case in which individual level information on the target variable and covariates from the survey sample is available in all areas.

$$«_{ij} = \sim + x_{ij} \quad + u_i + v_i$$

Area Level model

Use aggregated data

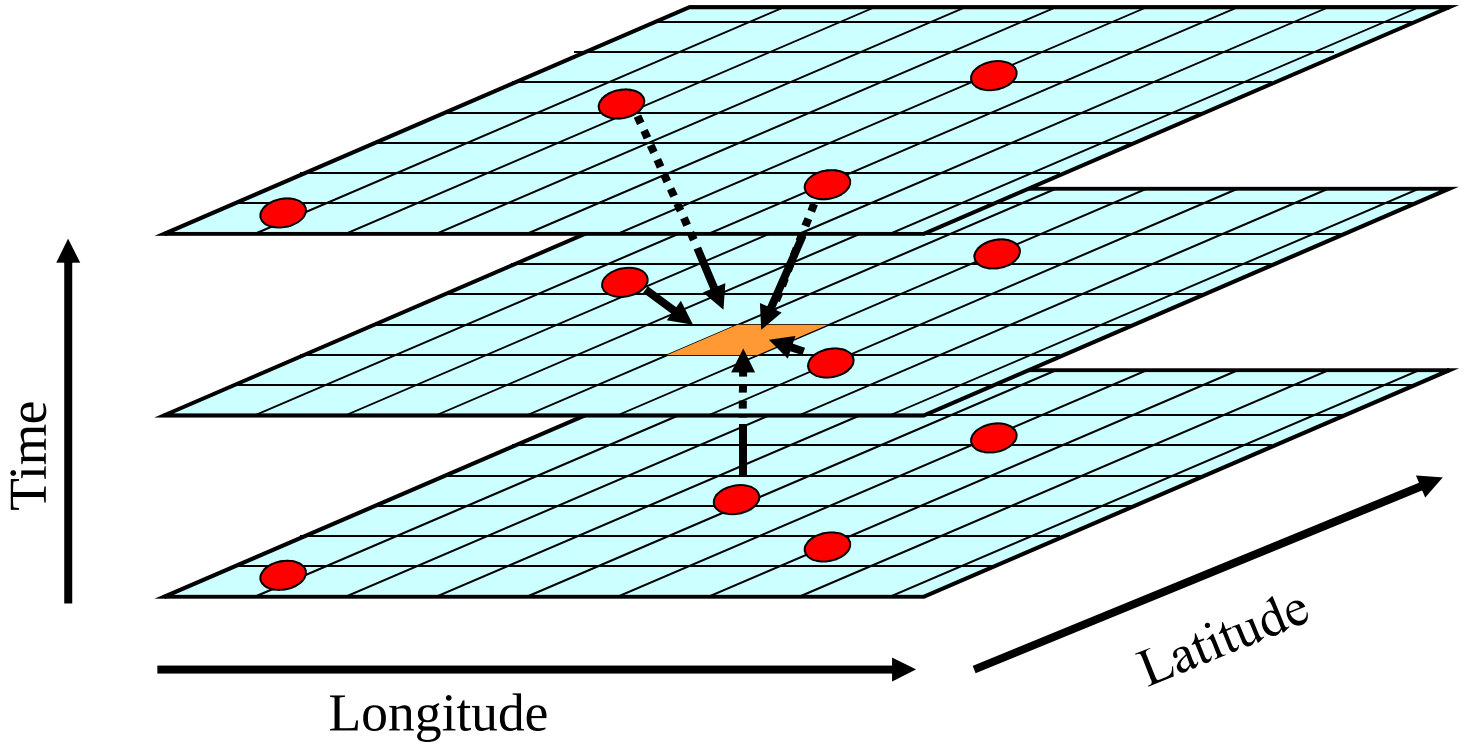
$$«_{B,i} = \sim^{*} + \overline{X}_i \quad ^{*} + u_i^{*} + \hat{v}_i^{*}$$

or

$$«_i = \sum_j \frac{«_{ij}}{N_i} = \sim + \overline{X}_i \quad + u_i + v_i$$

Analysis Framework

Spatial mapping with covariates at different time points



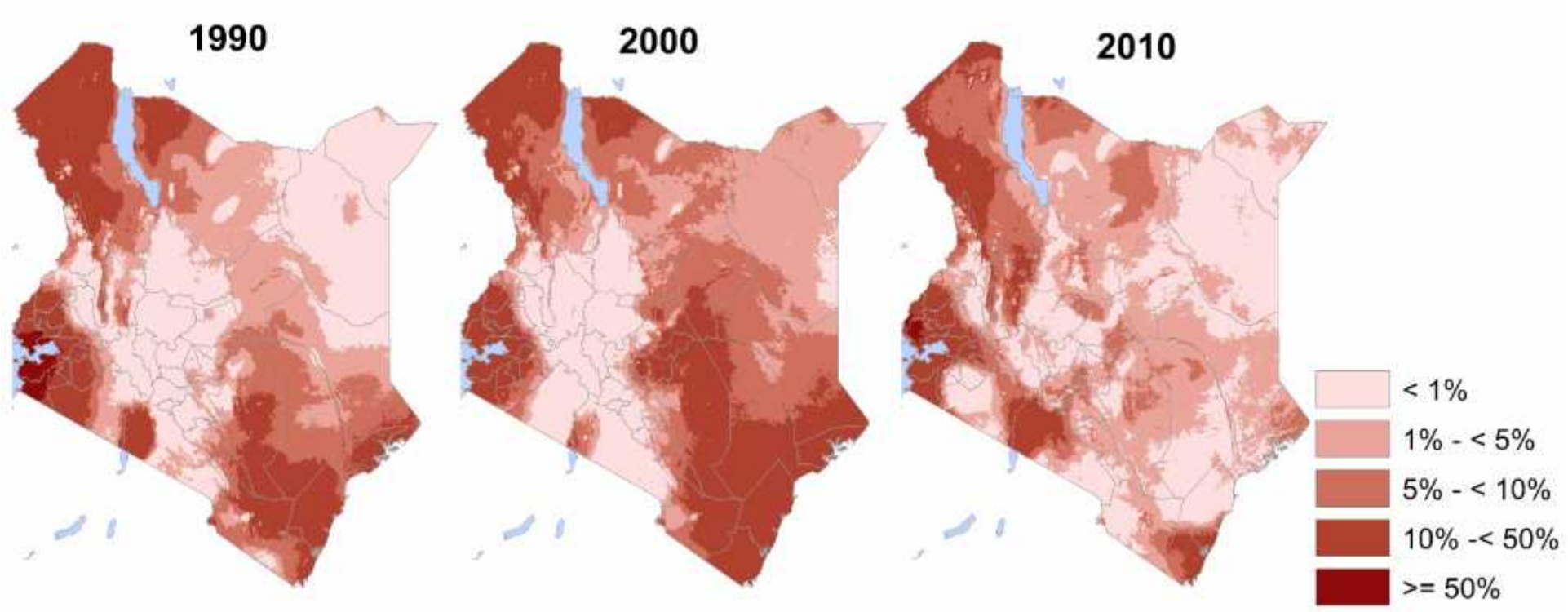
Formula: $y_{(s_i,t_j)} = z_{(s_i,t_j)} + ¶_{(s_i,t_j)} + £_{(s_i,t_j)}$

- $y_{(s_i,t_j)}$: A realization of spatial-temporal process (risk)
- $z_{(s_i,t_j)}$: Covariates effects
- $¶_{(s_i,t_j)}$: Latent process
- $£_{(s_i,t_j)}$: Measurement error

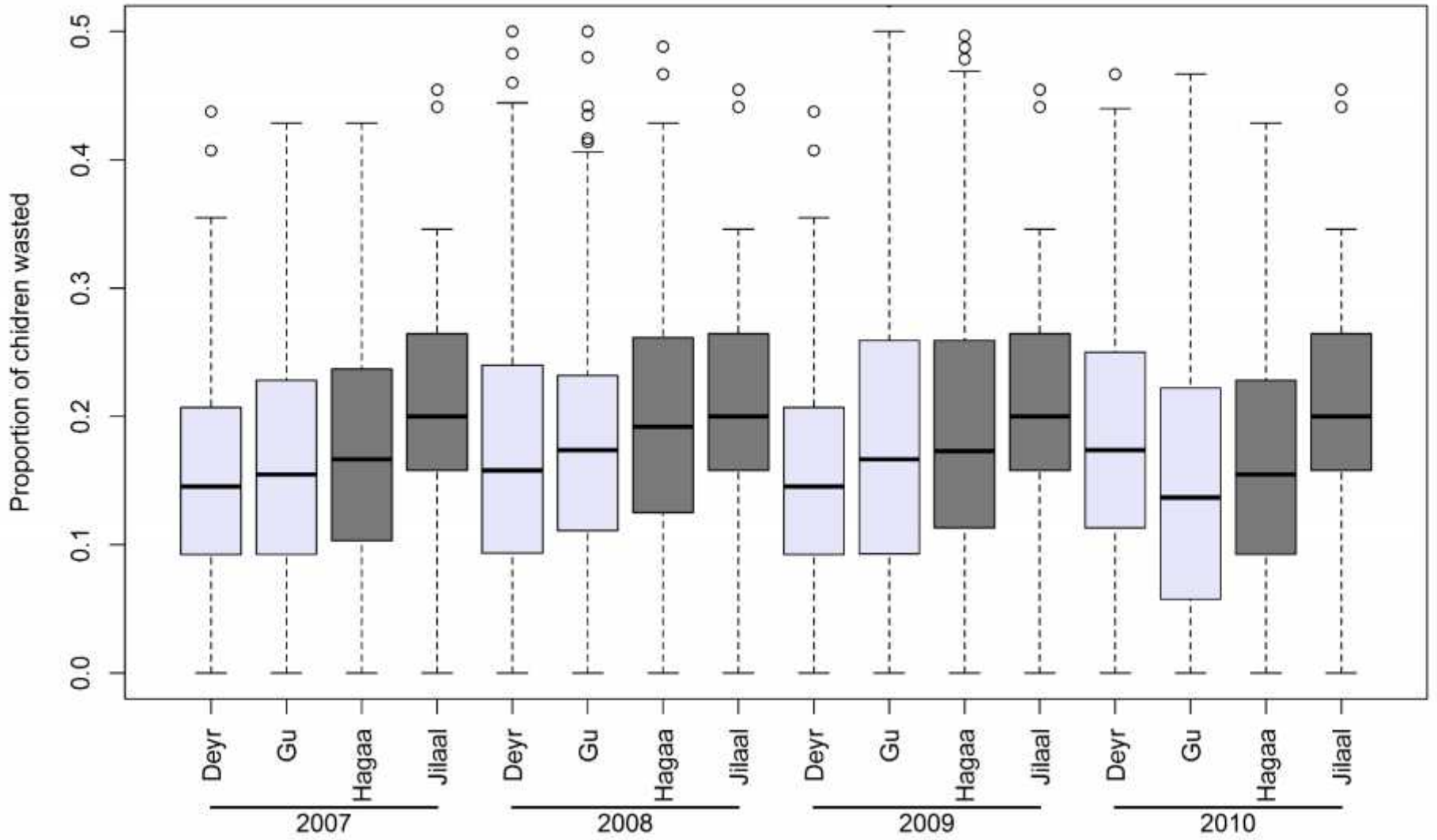
Application

- Pattern
- Relationship
- Trend
- Numbers / prevalence / counts

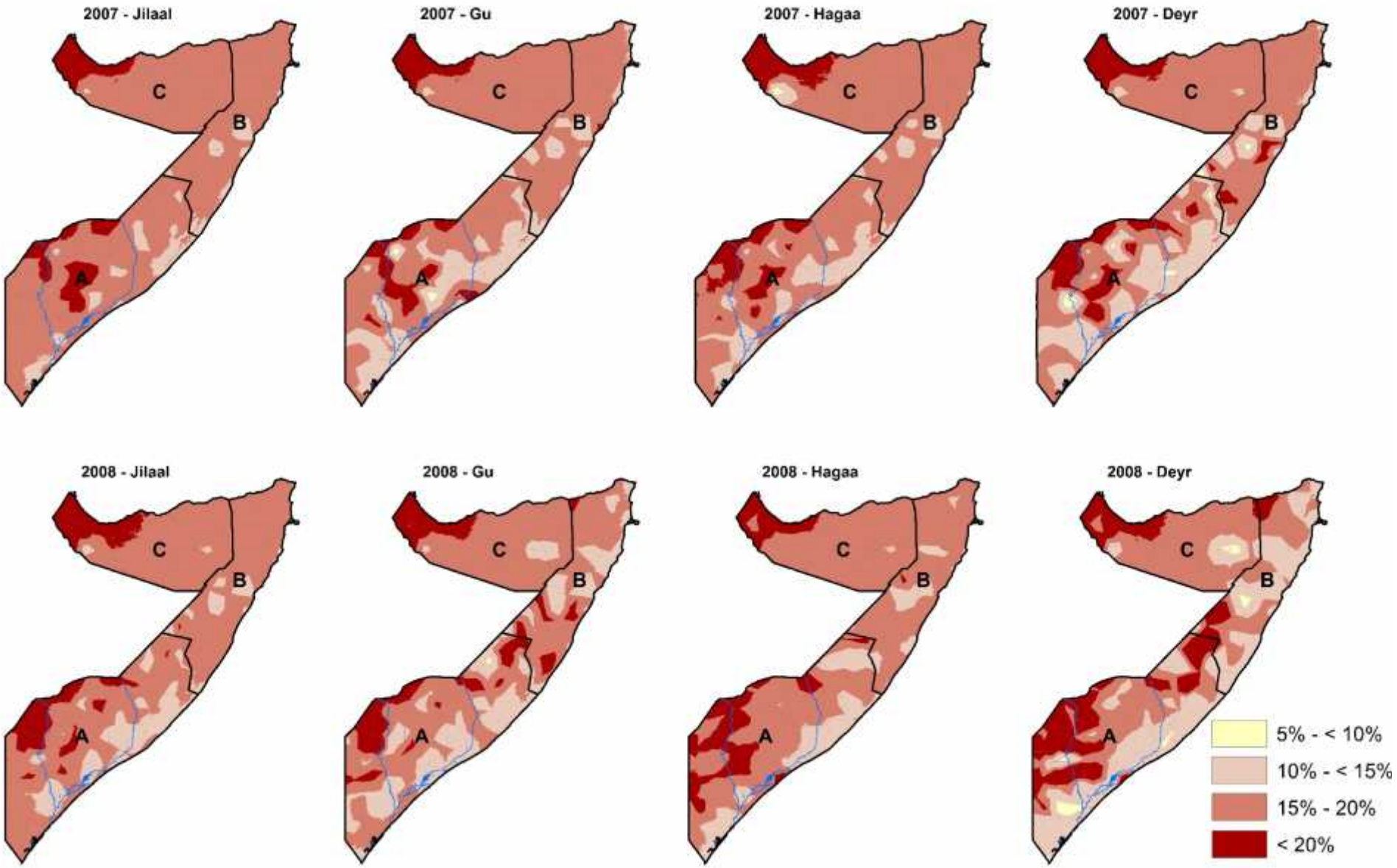
Maps of malaria prevalence for 1990 – 2010 at 1×1 km grid location in Kenya



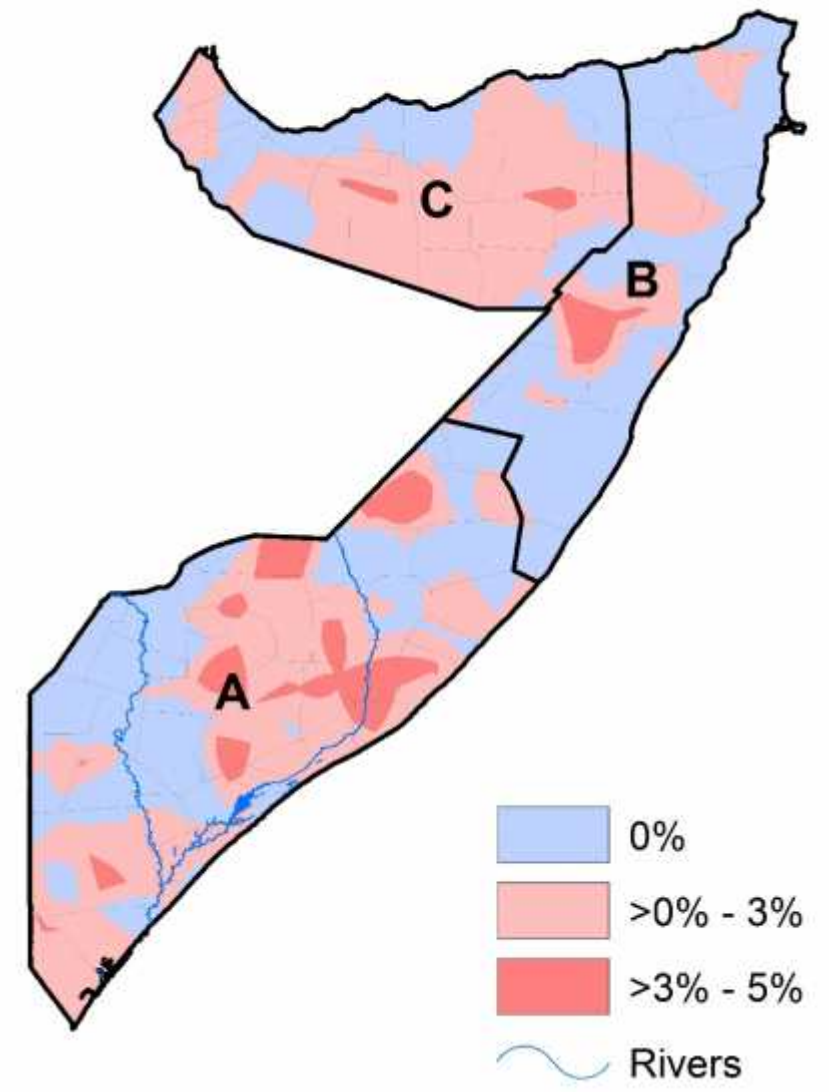
Data distribution

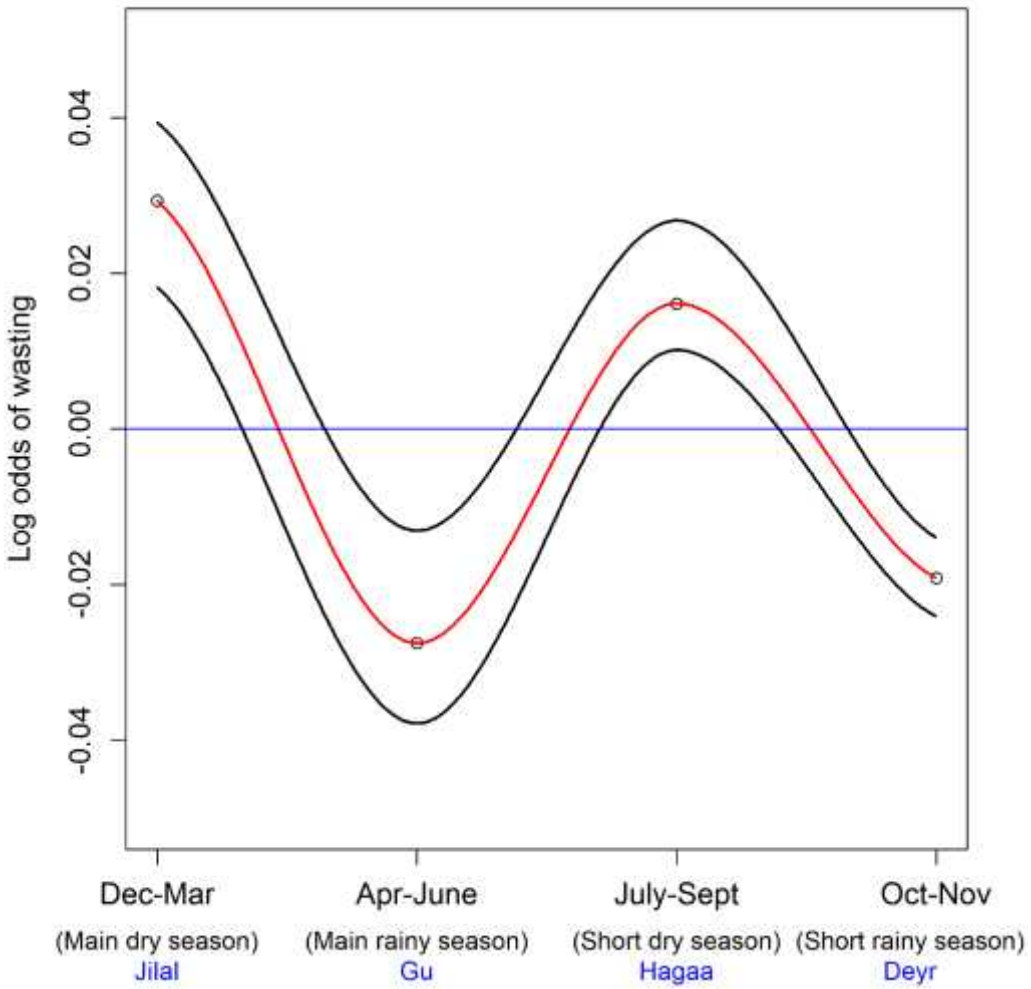


Maps of annual mean wasting for 2009 & 2010 at 1×1 km grid location in Somalia



Mean difference of the prevalence of wasting during the dry and wet seasons





Estimate of the number of children wasted during the four seasons from 2007 to 2010 in Somalia

WHO prevalence class	Year of survey	Season	Zone name		
			South Central	North East	North West
Alert (5% - <10%)	2007	Jilaal	< 50	< 50	< 50
		Gu	< 50	< 50	< 50
		Hagaa	< 50	< 50	713
		Deyr	696	< 50	< 50
	2008	Jilaal	< 50	< 50	< 50
		Gu	< 50	< 50	< 50
		Hagaa	< 50	< 50	< 50
		Deyr	774	< 50	< 50
	2009	Jilaal	< 50	< 50	< 50
		Gu	2,902	< 50	< 50
		Hagaa	< 50	< 50	< 50
		Deyr	< 50	< 50	< 50
	2010	Jilaal	< 50	< 50	< 50
		Gu	5449	142	< 50
		Hagaa	< 50	< 50	< 50
		Deyr	1,580	493	< 50
Serious (10% - < 15%)	2007	Jilaal	10,432	2,469	723
		Gu	27,995	2,841	1,082
		Hagaa	36,033	2,763	11,157
		Deyr	62,488	4,178	3,451

Key challenges

- Resolving the level of analysis (individual, cluster, district level??)
- Different naming of the locations or change of names with time
- Some of the coordinates are missing from publications
- Use of same names for different locations
- Displacement of coordinates for confidential reasons - DHS
- Coordinates not clearly defined - decimal points
- Unspecified coordinate systems – Global geographic vs Local coordinate systems

Thank you.

KEMRI | Wellcome Trust