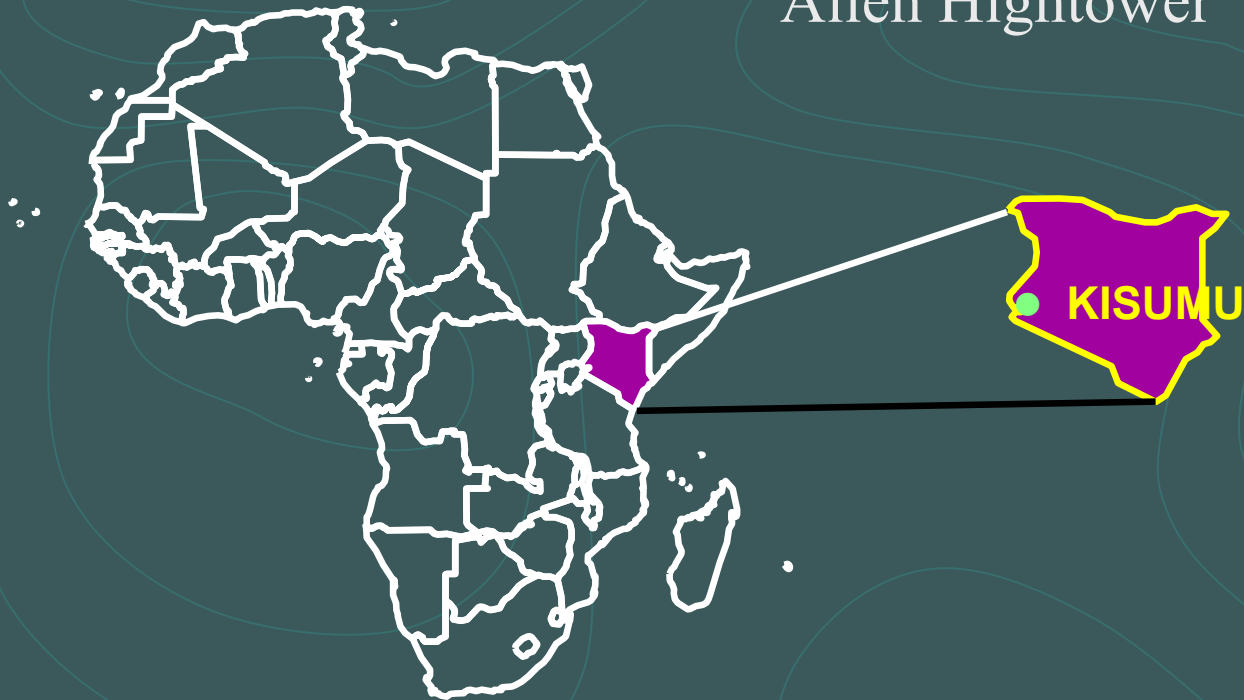# Automated Data Collection Systems in Western Kenya
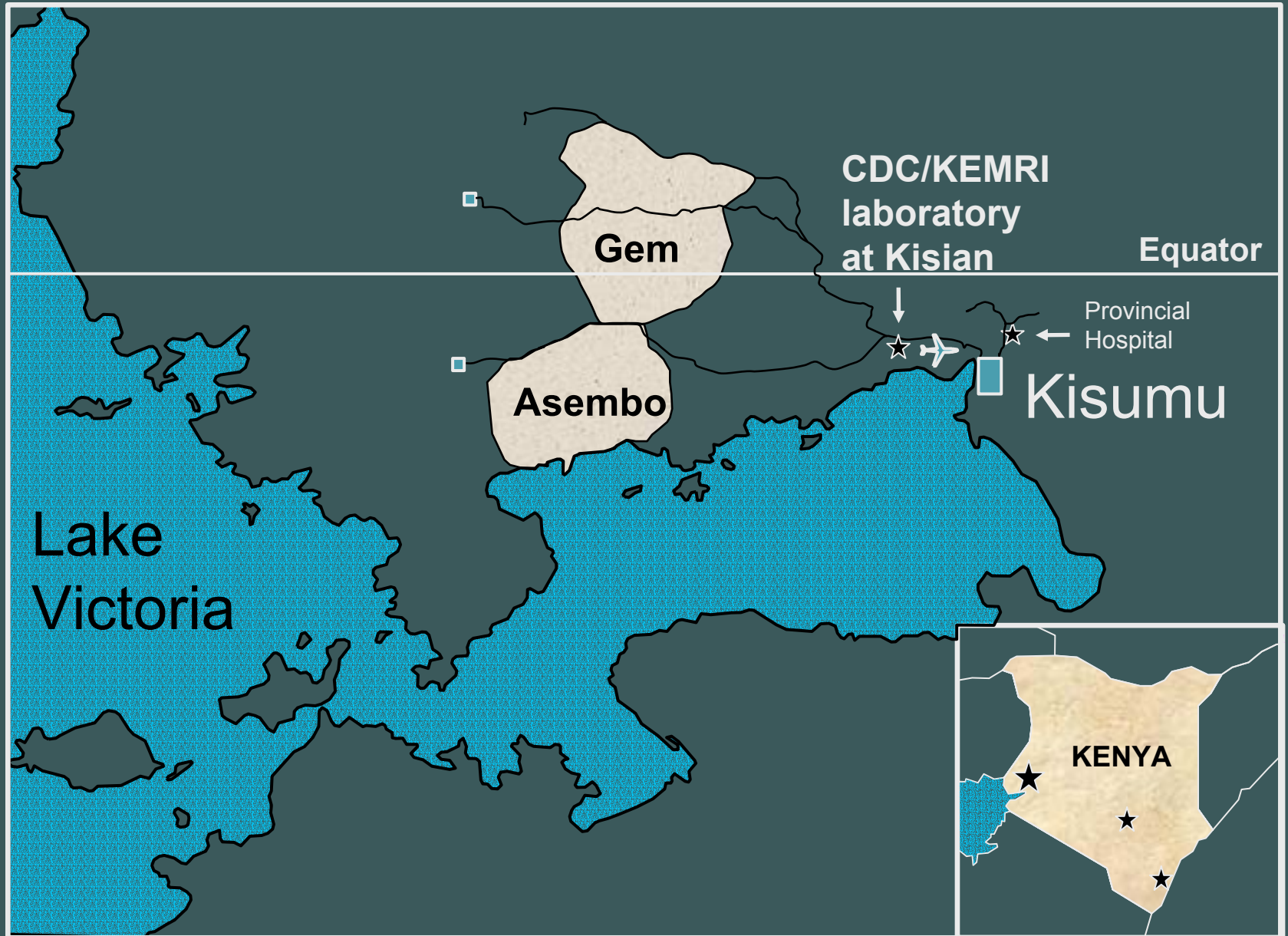
Allen Hightower

KISUMU

Data Management in Kenya

- Dan Rosen – Chief of Data Management and Surveillance

- James Kwach – Chief Data Manager

- Maurice Ombok – Chief, GIS/GPS Activity

CDC/KEMRI Field Station Western Kenya

Gem

Asembo

CDC/KEMRI laboratory at Kisian

Equator

Provincial Hospital

Kisumu

Lake Victoria

KENYA

The study area is rural, but densely populated.

21/4/1999

Study Area Has Very Intense Malaria Transmission

- >100 infective bites per year.
- Virtually all infants have malaria multiple times in first year of life.
- Most malaria-related mortality is in first 2 years of life.

CDC in Kenya

- Over 600 field staff, entomology and immunology labs.

The Bednet Project

- Goal: to see if insecticide-treated bednets can reduce malaria-related mortality in children.

- Study area : 450 sq km, 125,000 people.

Demographic Surveillance in Western Kenya.

- Quarterly Census of approximately 100,000 people.
- Tracks migration, marriage, mortality, and births.
- We need population figures to compute rates of disease.
- Additional components track prevalence of malaria vectors, morbidity, and mortality.

# Demographic Surveillance – data base management

- We use a modified version of the Population Council's Household Registration System.
  - www.popcouncil.org/hrs/hrs.html
  - Visual Foxpro-based
    - VF chosen due to common use in developing countries.

Demographic Surveillance

- We use a modified version of the Population Council's Household Registration System.
  - www.popcouncil.org/hrs/hrs.html
  - Visual Foxpro-based
  - Have developed scannable forms for each component.
  - Have also developed a "proof of concept" PDA-based system.

Data Model for HRS
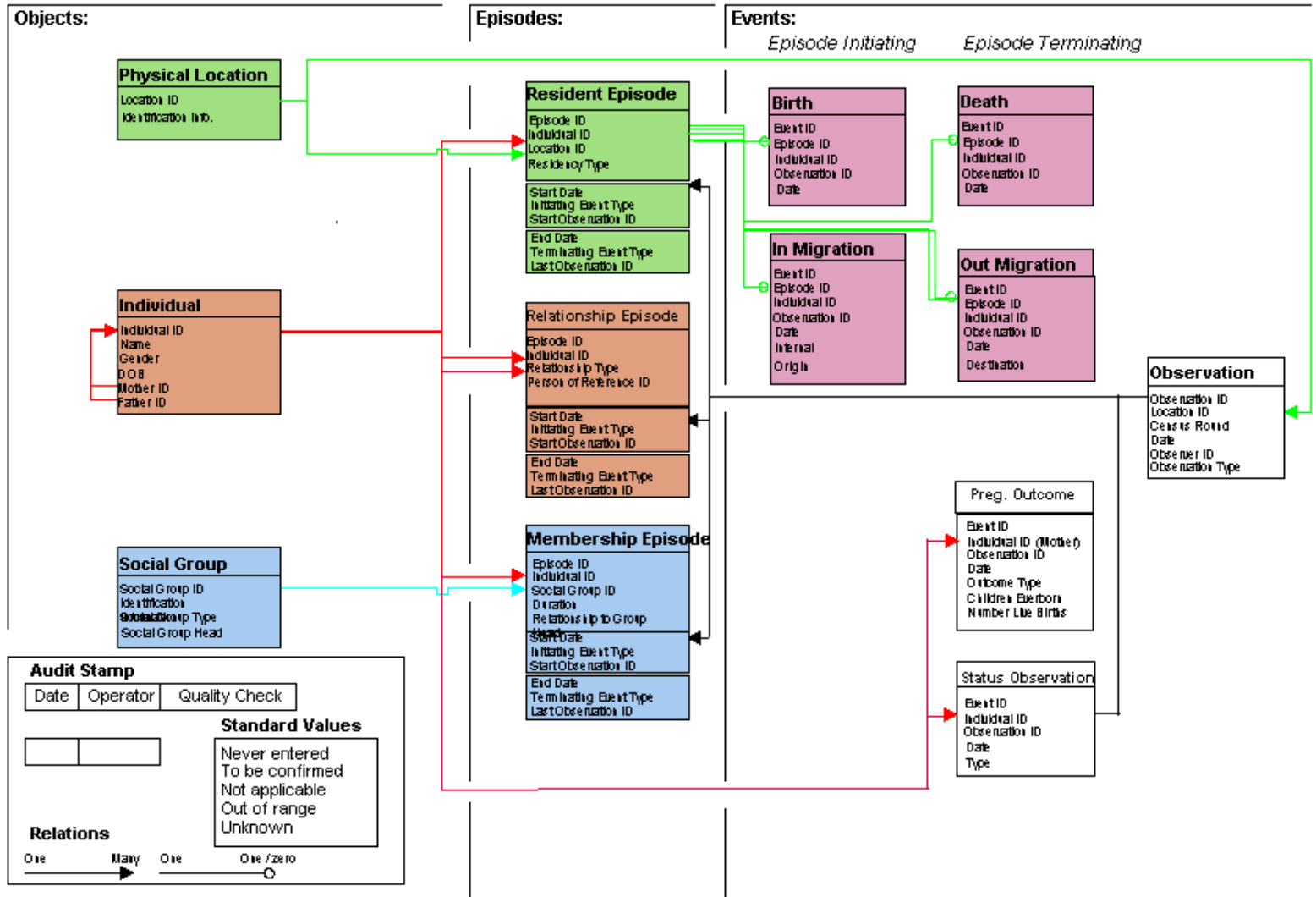
Part of the Data Entry Operation

12/2/1999

The other half of the computer room.   There is another data entry room, also.

Storage of paper forms is a big issue!

How to minimize data processing nightmares?

- Use PDA's to collect data.
- Use scan-ready forms to collect data.
- Both approaches eliminate the need for data entry.
- So, we are trying both methods.

We are using OCR for Data Collection/Entry

Scan – ready Form

High speed Scanner

Teleform (or other software)

The scan-able form

- Can scan existing forms or create new ones.
- Define the attributes of each field/column.
- Faxed forms can be used at both the form design and data entry stages.

Data Processing with OCR

- Uses neural net technology.
- More structured forms (one character/box).
- Many options – automatic coding, table lookups, q/c on scanned images.
- Forms in Demographic Surveillance are being processed in this way.

Scanning - Con's

- Larger startup cost – roughly $10,000 for the software, computer, and high-speed scanner.
- Resistance from scientific staff.
- Proliferation of form versions and associated databases.
  - Each form change = new database.
- Storage of forms is still a problem.
- Data not edited at collection phase.

Scanning - Pro's

- Rapid ramp up of projects .
- Increased speed and accuracy for processing data.
- Easier staff monitoring and reduced staffing needs for data entry.
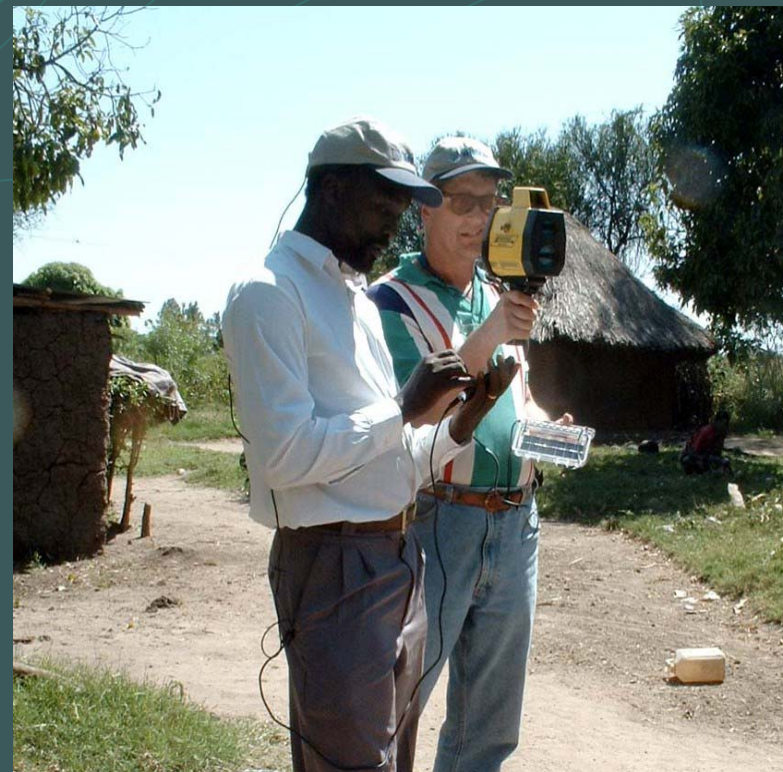- Cheaper for large number of data collectors.

A PDA-based data collection system

- Case study: the GPS/mapping system.
- Uses no paper collection forms.
- Collects home construction type, animal ownership, location information.
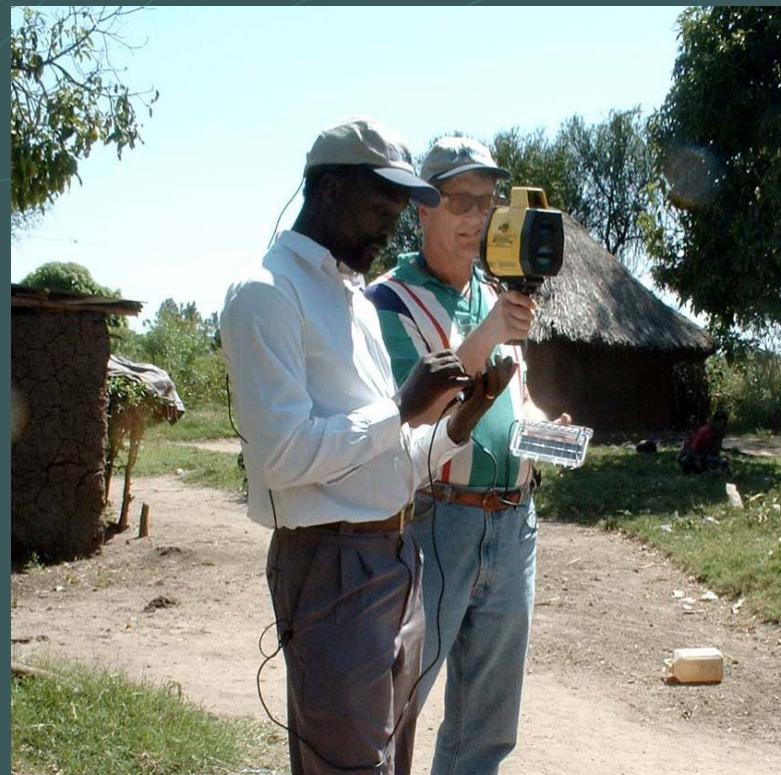
GPS mapping/census system.

- System in use since 1998.
- New hardware and upgraded software put into place Feb. 2002.

Components of the system

- The GPS unit
- The GPS antenna
- The Laser Range Finder
- The PDA/Data Collection Unit
- The Solar Battery Recharger

The GPS Unit and Antenna

- The GPS antenna
- The GPS unit
- Serial Cable connects GPS to PDA

The Laser Range Finder

- Measures distance from GPS unit to object to be mapped.

- Digital compass measures the direction.

- Digital level measures the angle.

- Software automatically adjusts GPS reading to account for this offset.

The Solar Battery recharger

- Keeps PDA batteries charged for all day use.
- Plugs in the base of the PDA.
- Inexpensive ($60).
- Works with cell phones and other devices.

The Windows-based PDA

- Uses Windows-based Pocket PC 2002 Operating System.
- Screen is highly visible in bright sun.
- Most Windows-based PDA's have this feature.

The PDA - 2

- Screen backlighting is automatically adjusted for both indoor and outdoor use.

- Has both numeric and alphanumeric touch-screen data entry.

The Old vs. the New

- The GPS unit/battery pack
- The Data Collection Unit
  - Old: DOS-based Proprietary System
  - New: Windows CE-based system
    - Pocket Word/Excel/Access
  - Solar Battery Recharger
    - Allows PDA to remain in the field for long periods of time.
  - Cost: New system is MUCH cheaper.


CDC
SAFER · HEALTHIER · PEOPLE

Conceptual Model

- PC software – to create applications, to store the master data set.

- PDA software– to run applications, collect field data, transfer data back to the PC.
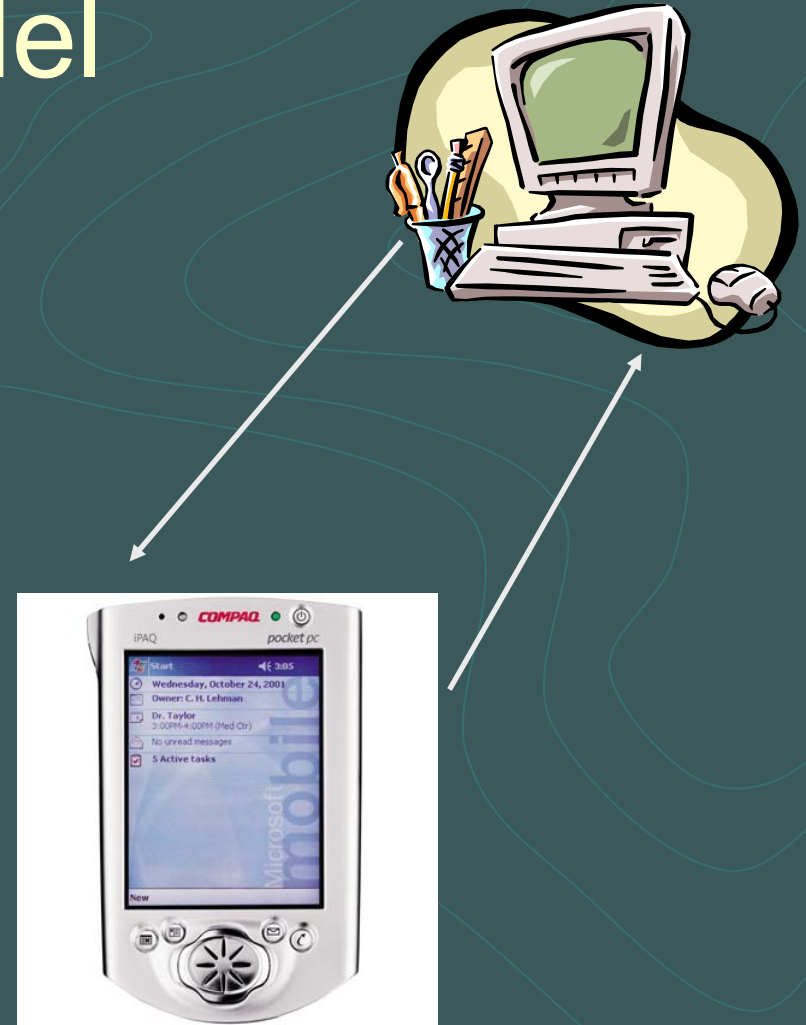
PDA to PC

- PDA Data Entry Screens can have edit checks.
  - Menu – pick systems.
- Data Can be Sent from the PC to the PDA for further updates.
  - Example: Census files that are updated quarterly.
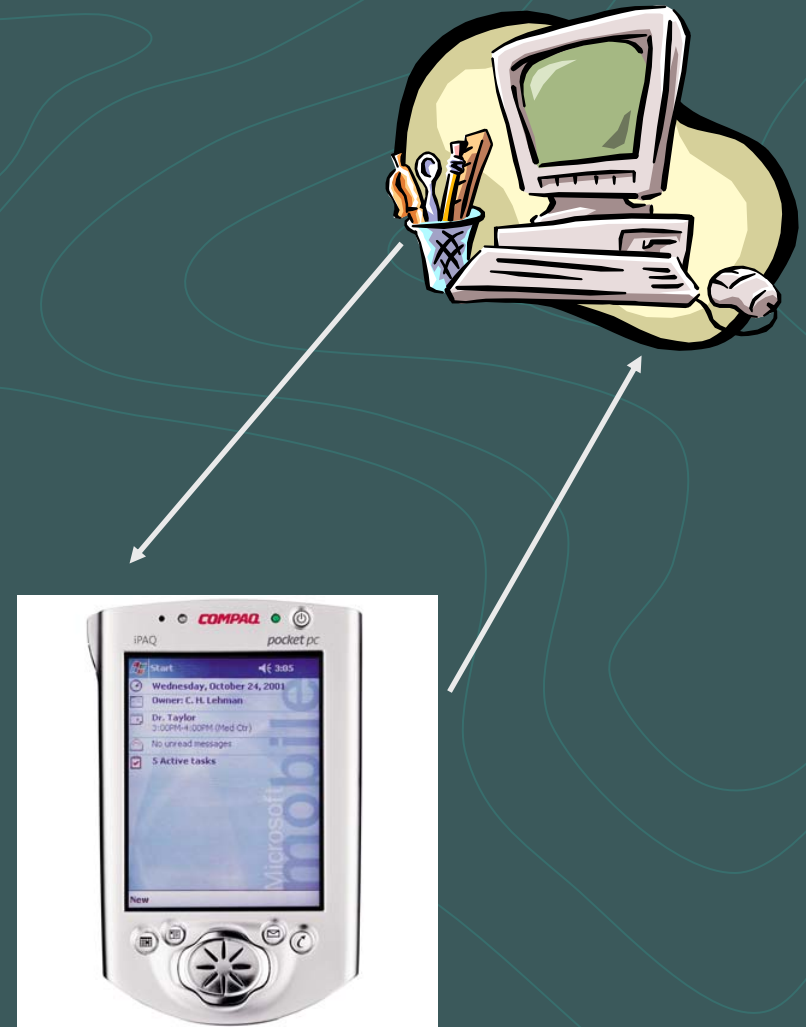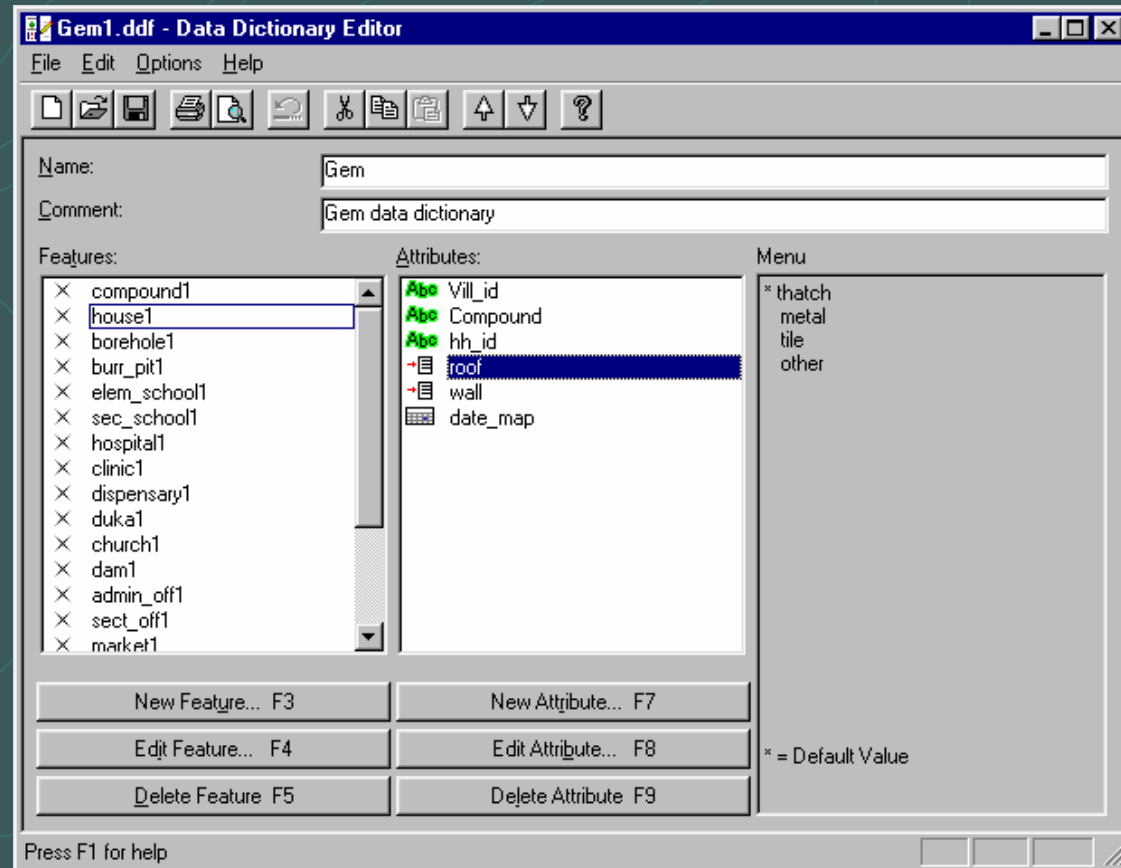
PC Client: Automated Data Dictionary

Part of GPS software

Creates a table for each feature.

Menus, other data entry features

Databases export directly to GIS with location data fields.

# Linking Map features to Databases

- The Village, Compound, and House ID are painted on each study house.

- The ID is recorded with the location data.

- Data from any other study that uses these household ID numbers can then be linked to map locations.

21/4/1999

Transfer of Data from PDA to Desktop Computer

- GPS to PC transfer
- Calibration of GPS readings
- Export data to GIS software
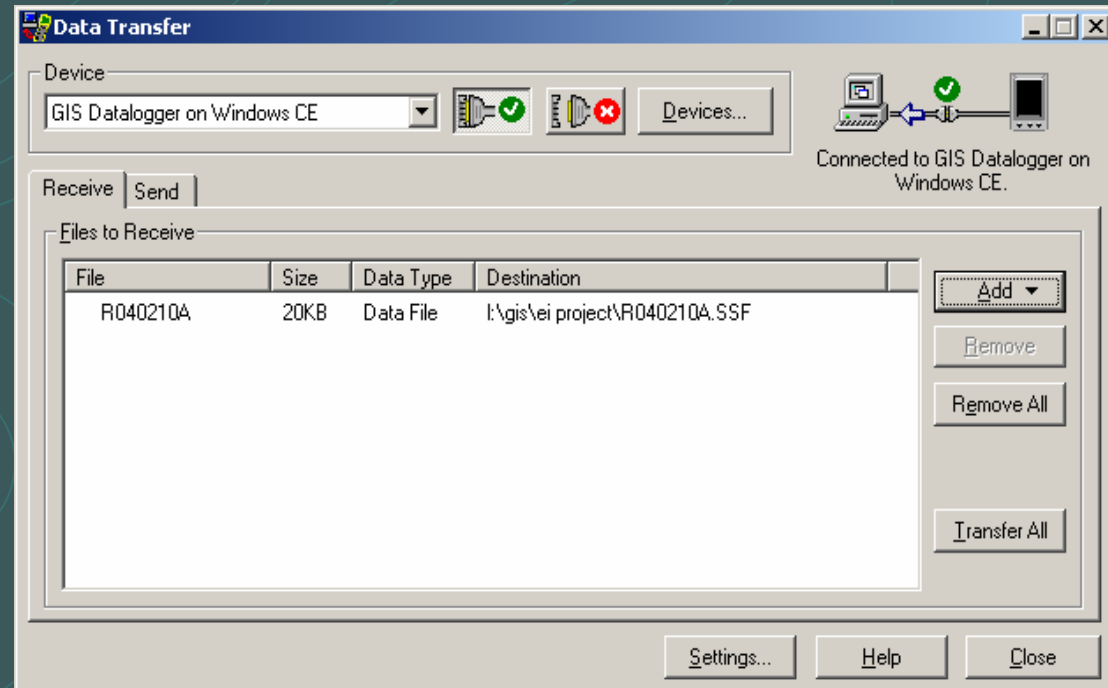
Survey Data Collection

- You can also use this system for collecting survey data.
- All that is needed is the PDA and the solar battery recharger.
  - No GPS or Laser unit.

Survey Data Collection

- Can use a variety of other database development tools for the Pocket PC
- All produce Access-compatible databases.
- Can also use Visual Studio with Embedded tools for CE.
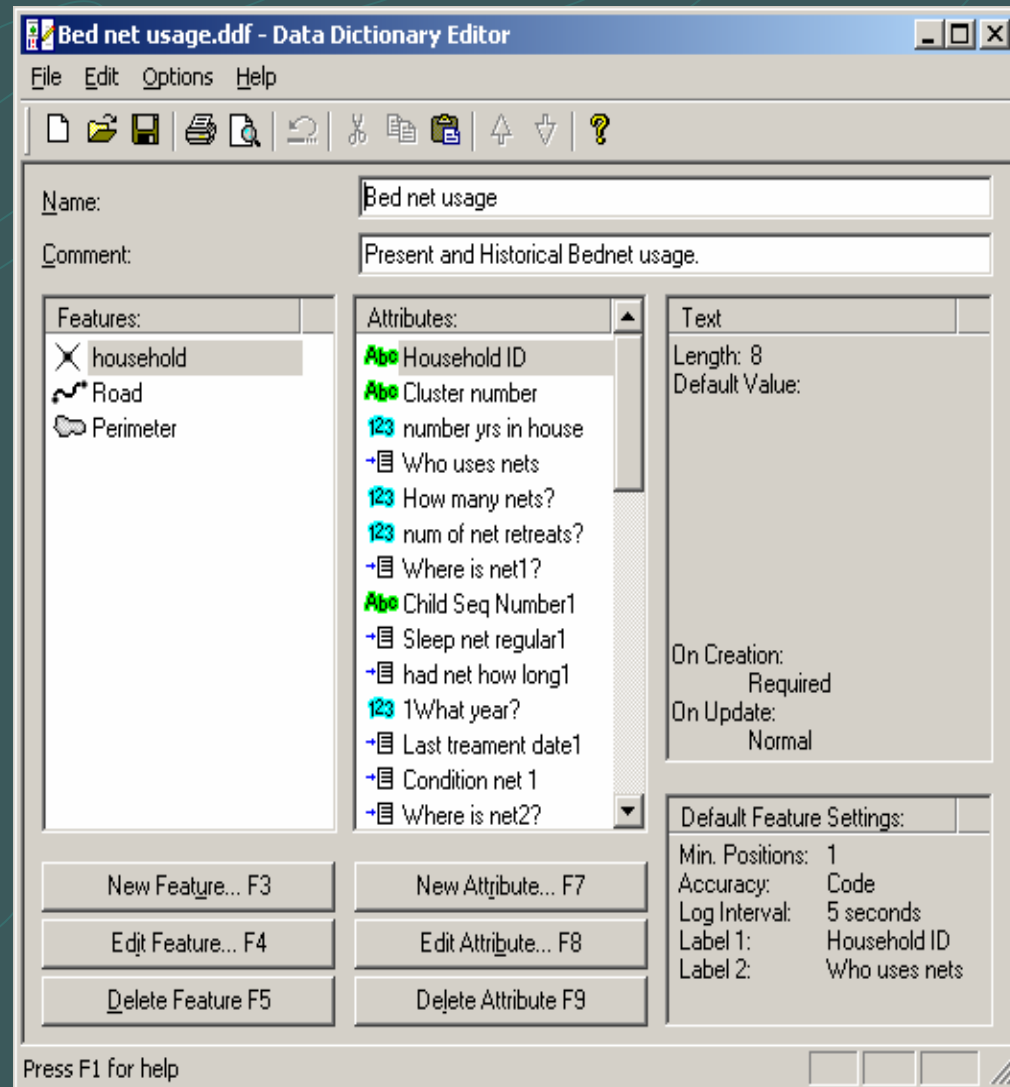
Survey Data Collection

- Only the CF card needs to be brought back from the field.
  - ID files on CF identifies the source of data to the Data Manager.
  - Software on PDA prevents incorrect CF installation.
  - ID written on CF identifies card to transport team.
  - CF card slides into a slot on the PDA.

Survey Data Collection

- Solar battery recharging allows unlimited field use.

Projects Using PDA's in Kenya:

- GPS/Mapping of Asembo, and Gem.
- Mapping/Census of Kisian for wash-durable net study.
- Survey of 45 villages on bednet usage habits and condition.
- Mapping/attributes of mosquito larval habitat in Asembo/Gem.

Projects Using PDA's in other countries:

- <u>Haiti</u> : GPS mapping for filariasis project.
- <u>Guatemala</u>: Data collection for malaria surveillance in 26 villages.
- <u>American Samoa</u>: GPS mapping/census for filariasis project.
- <u>Tanzania</u> – Demographic surveillance – will be tested in 3 villages in early '03.

Pros for PDA-based data collection

- Users really seem to like the system
- Easy to train users – no nightmares to date.
- Simple applications are very easy to develop and modify using a number of products.
  - Complex applications developed via Visual Studio
- Has been used for approximately 10 projects of varying size and complexity.
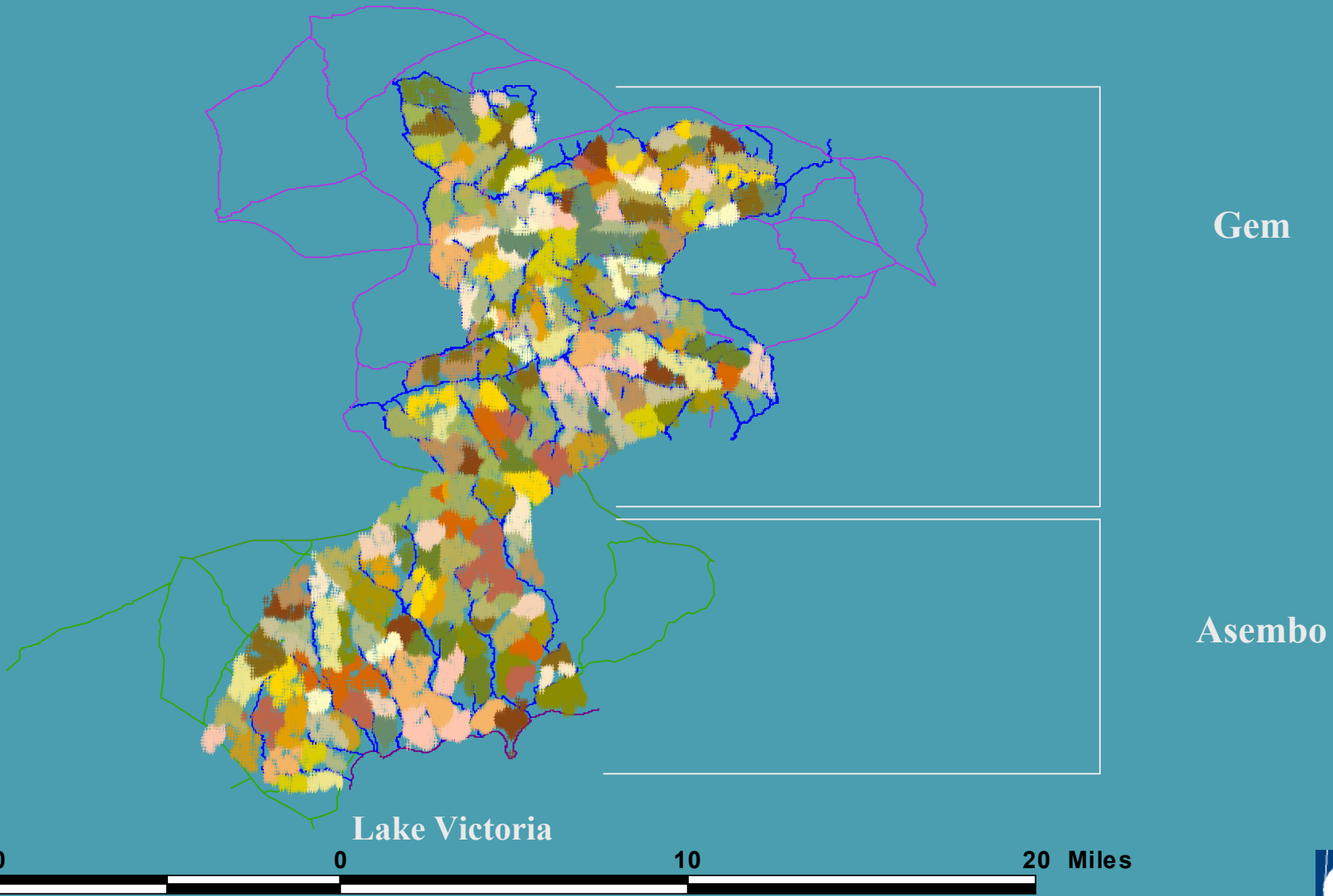- Solar battery recharging system works very well.

Con's for PDA-based data collection

- Expensive if there are a large number of data collectors .
  - If there are more than 15 data collectors, consider scanning systems.
- Hardware and battery failures
  - Have an extra PDA or two on hand. Use PDA's with replaceable batteries if possible.
  - Not a major concern. Only one project of 10 has experienced loss of field data.
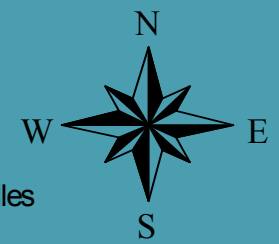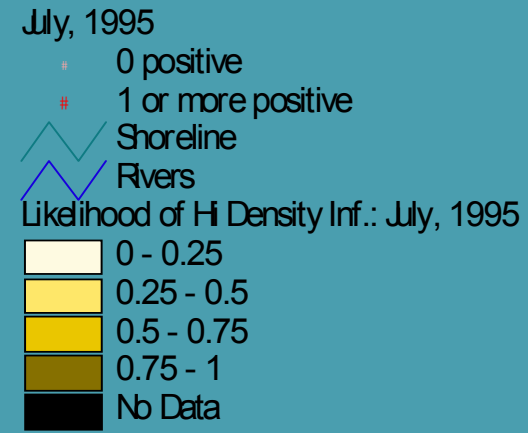
Staffing and Equipment

- Three GPS technicians.

- One base station ($10K) for differential GPS.

- Three laser units ($3K each).   A compass and tape measure will work, too ($50, but is low-tech!).

- The GPS units are $500 each, the PDA's are $500 each (with cables), software is $800, the solar battery recharger is $60.

- GIS software is $1K to $3K (additional modules).

We've mapped approximately 20,000 compounds in Asembo and Gem covering over 450 sq km and 125,000 people .

Gem

Asembo

Lake Victoria

10   0   10   20 Miles

CDC
SAFER · HEALTHIER · PEOPLE

High density malaria infections
July 1995

July, 1995
- 0 positive
- 1 or more positive
- Shoreline
- Rivers

Likelihood of Hi Density Inf.: July, 1995
- 0 - 0.25
- 0.25 - 0.5
- 0.5 - 0.75
- 0.75 - 1
- No Data

3   0   3   6 Miles

N
W   E
S
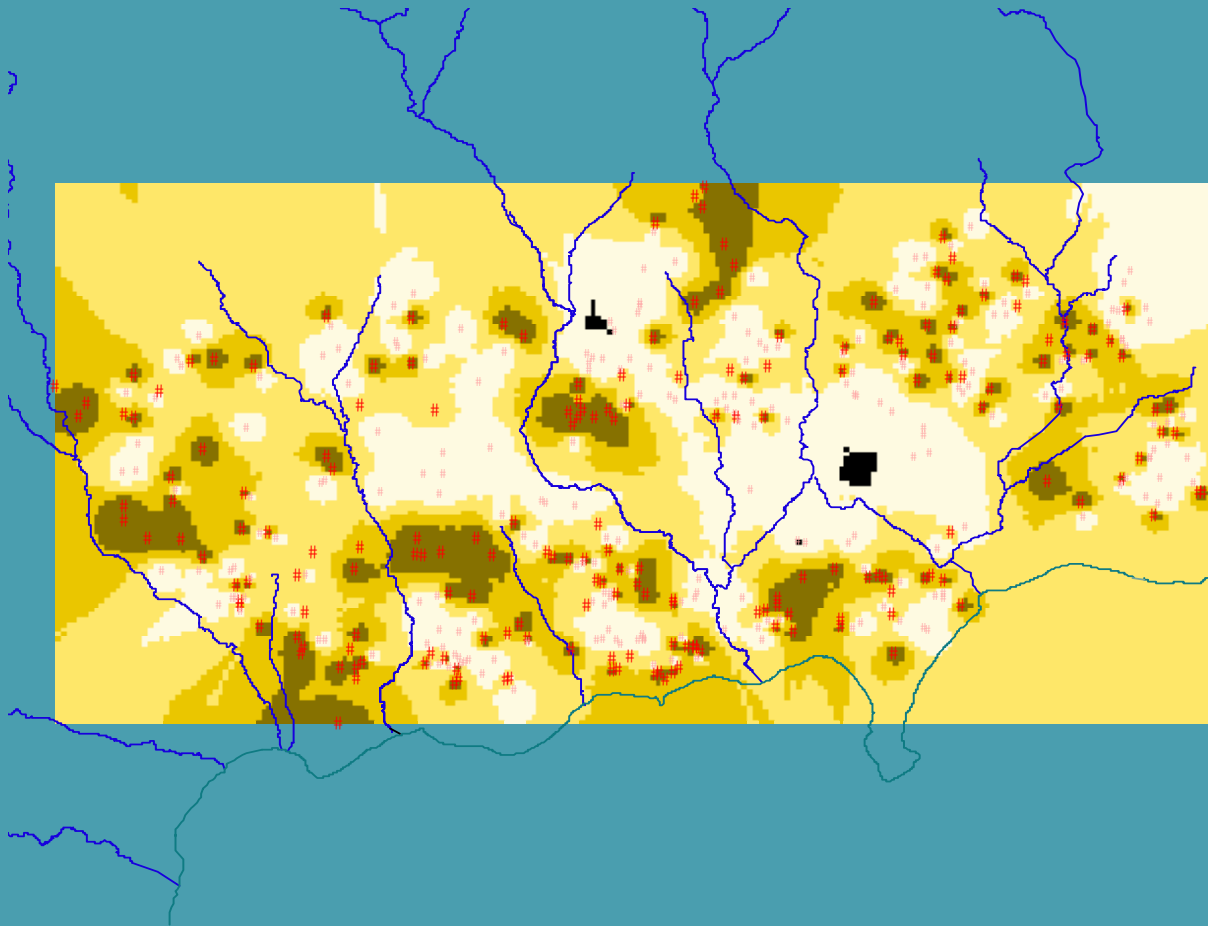
Abundance of *A. gambiae,* June 1995

June 95 Mosquitoes
- 0 - 0.99
- 1 - 9.99
- 10 - 29.99
- 30 - 2999.99

June 95 Surface
- -15.751 - 0.999
- 1 - 9.999
- 10 - 29.999
- 30 - 59.999
- 60 - 813.756

N
W   E
S

3   0   3   6 Mles
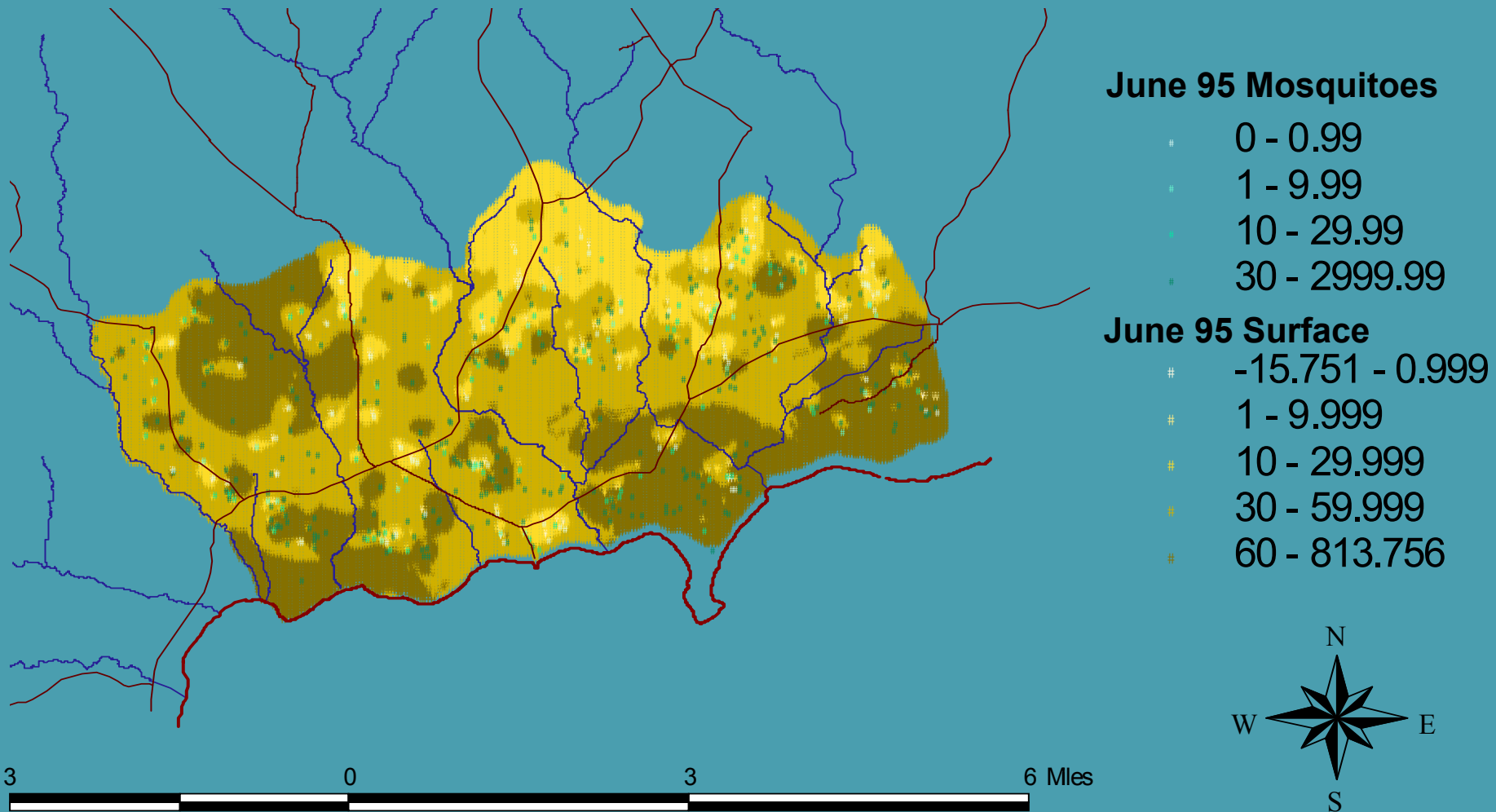
Contours of Probability of Complete Vaccination
by Household Location, Asembo Bay, Western Kenya

Complete vaccination
- # No
- # Yes

Prob. of vaccination
- 0 - 0.2
- 0.2 - 0.4
- 0.4 - 0.6
- 0.6 - 0.8
- 0.8 - 1

Contours for Prob. of vaccination

Ñ Health centers
Roads
Streams
Lake Shore
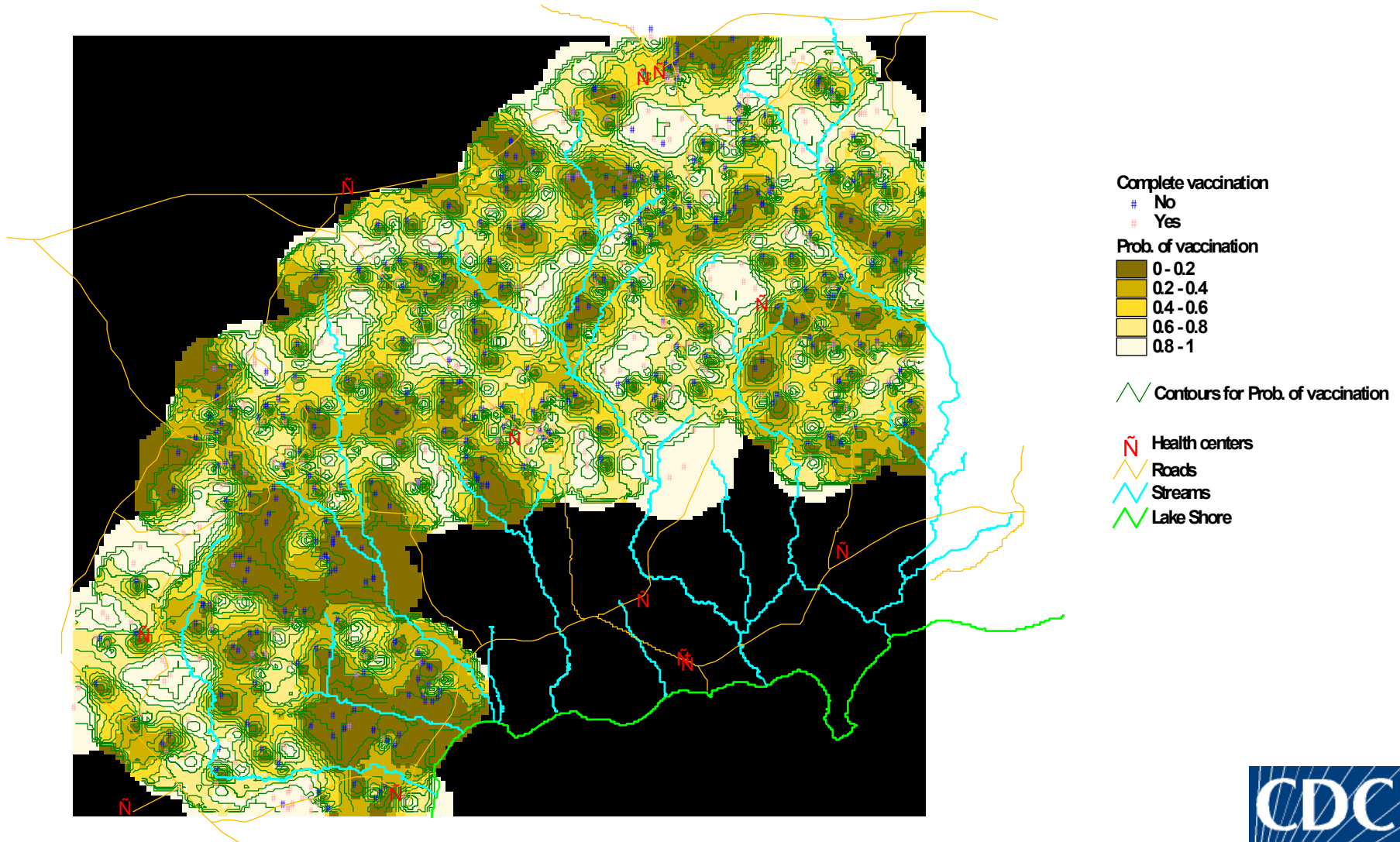
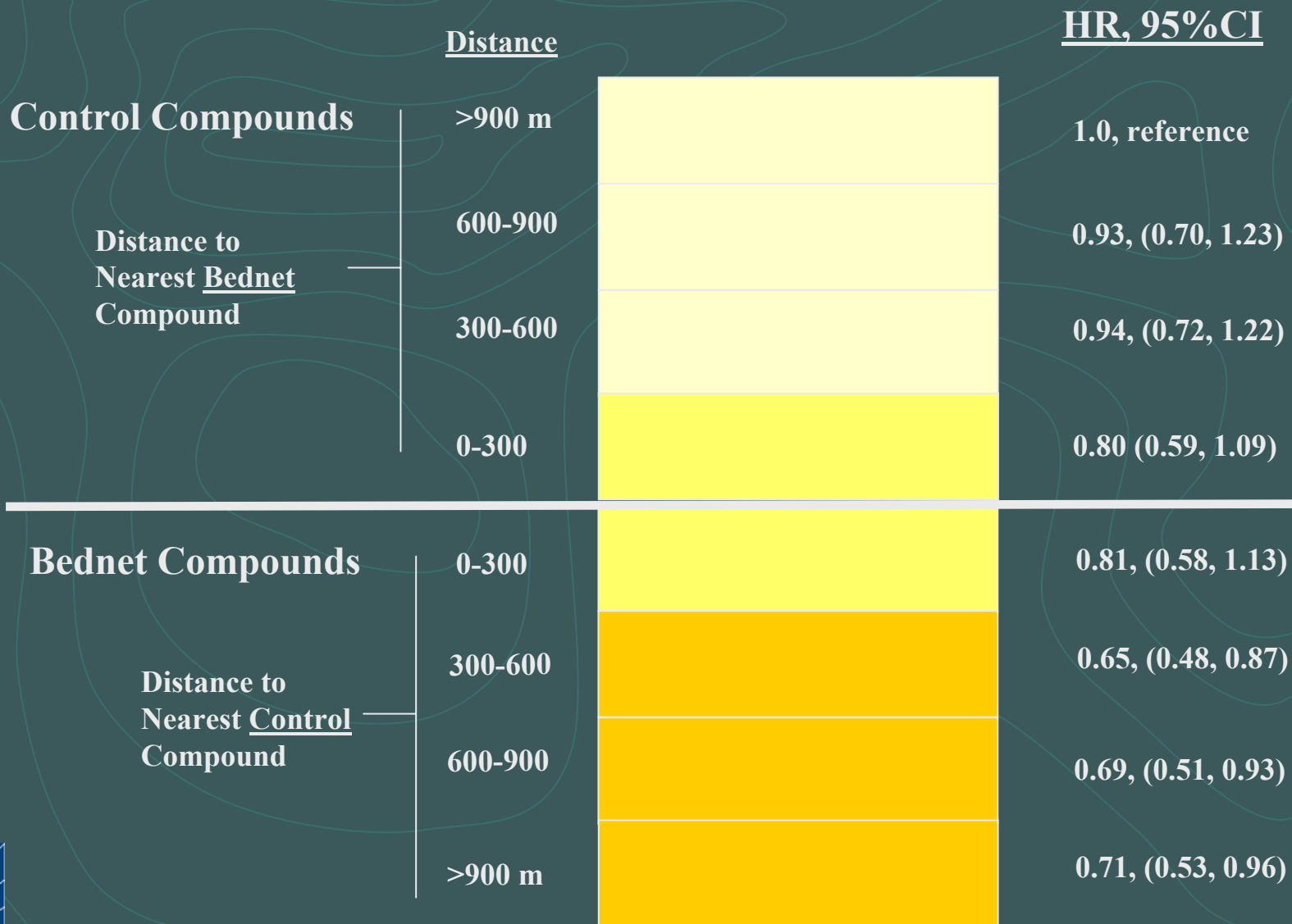Cholera cases by week of onset. Note the east (darker) to west (lighter) movement of the epidemic.
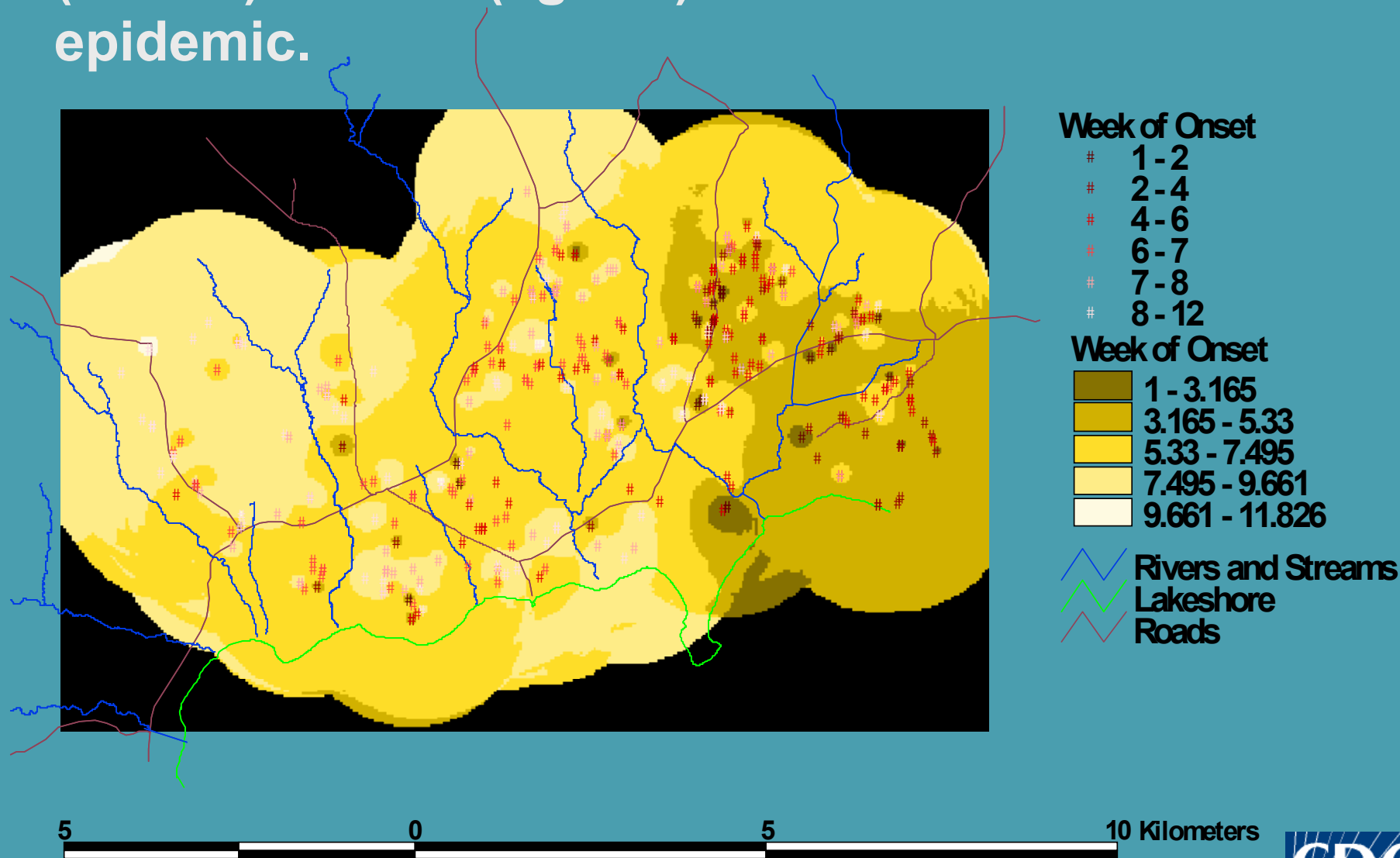
GPS data has been combined with Satellite Image Data to study vector abundance as a function of distance to water/altitude measures (GPS) and vegetation/ground cover measures (remote sensing).
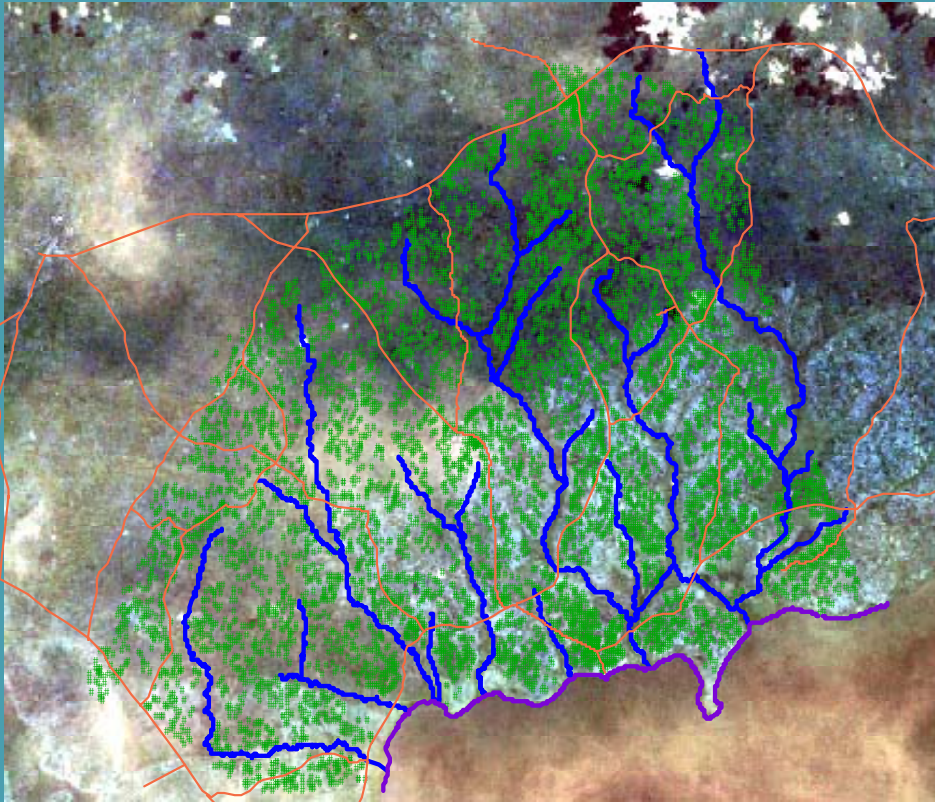
Odds Ratios for Childhood Mortality by Distance
Darker Colors Mean More Protection

	Distance		HR, 95%CI
Control Compounds	>900 m		1.0, reference
Distance to Nearest Bednet Compound	600-900		0.93, (0.70, 1.23)
	300-600		0.94, (0.72, 1.22)
	0-300		0.80 (0.59, 1.09)
Bednet Compounds	0-300		0.81, (0.58, 1.13)
Distance to Nearest Control Compound	300-600		0.65, (0.48, 0.87)
	600-900		0.69, (0.51, 0.93)
	>900 m		0.71, (0.53, 0.96)

Summary

- Paperless data collection systems have worked well for relatively simple data collection projects.
  - We plan to use them in more sophisticated systems soon.
  - They also can have integrated GIS data collection capability.
- These systems allow rapid collection of high quality data.
- GIS systems can provide data that provide unique information of public health importance.