# A Tale of Two Parasites

## Statistical modelling to support NTD control in Africa

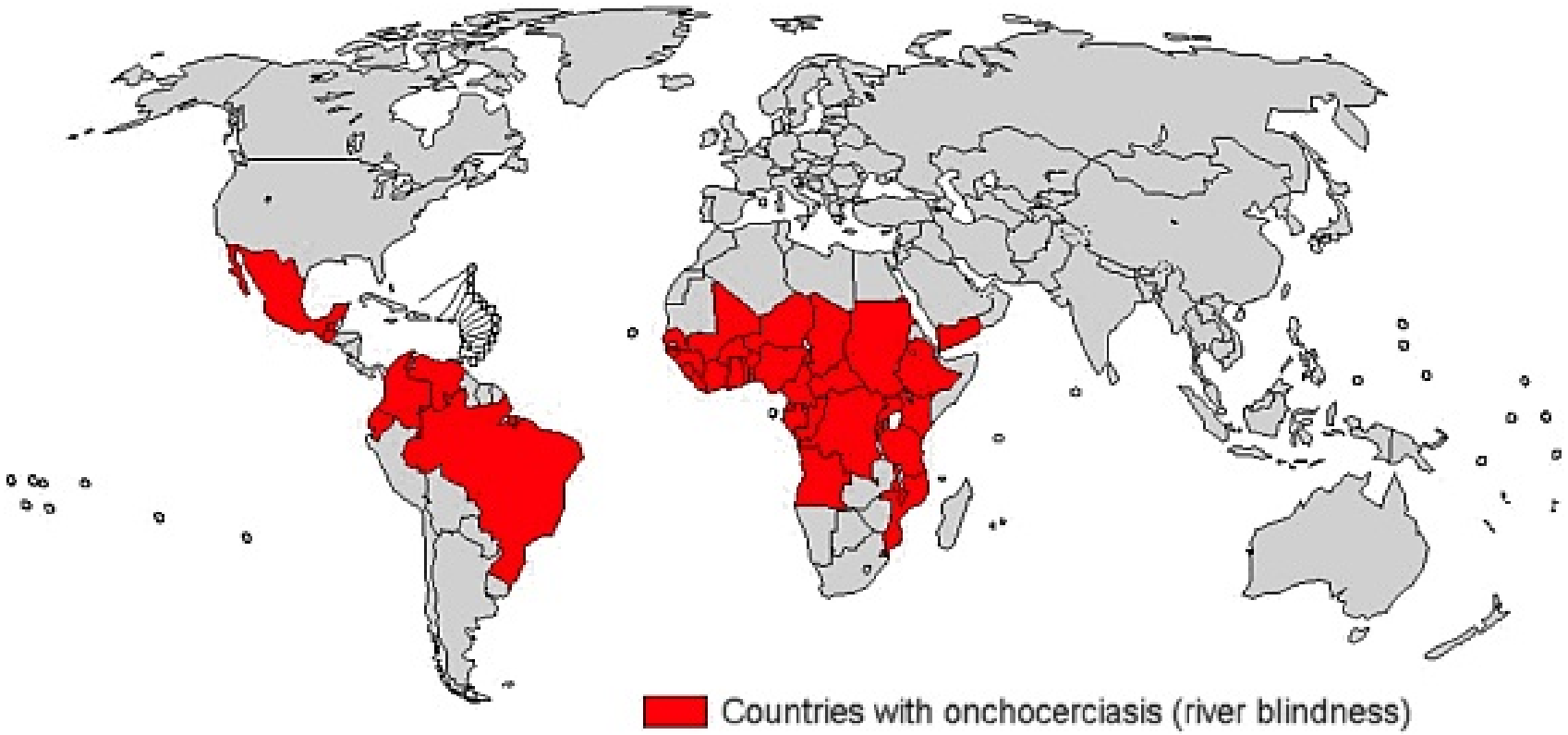
Peter J Diggle

Lancaster University


# MDA: a tool for control of vector-borne filarial disease

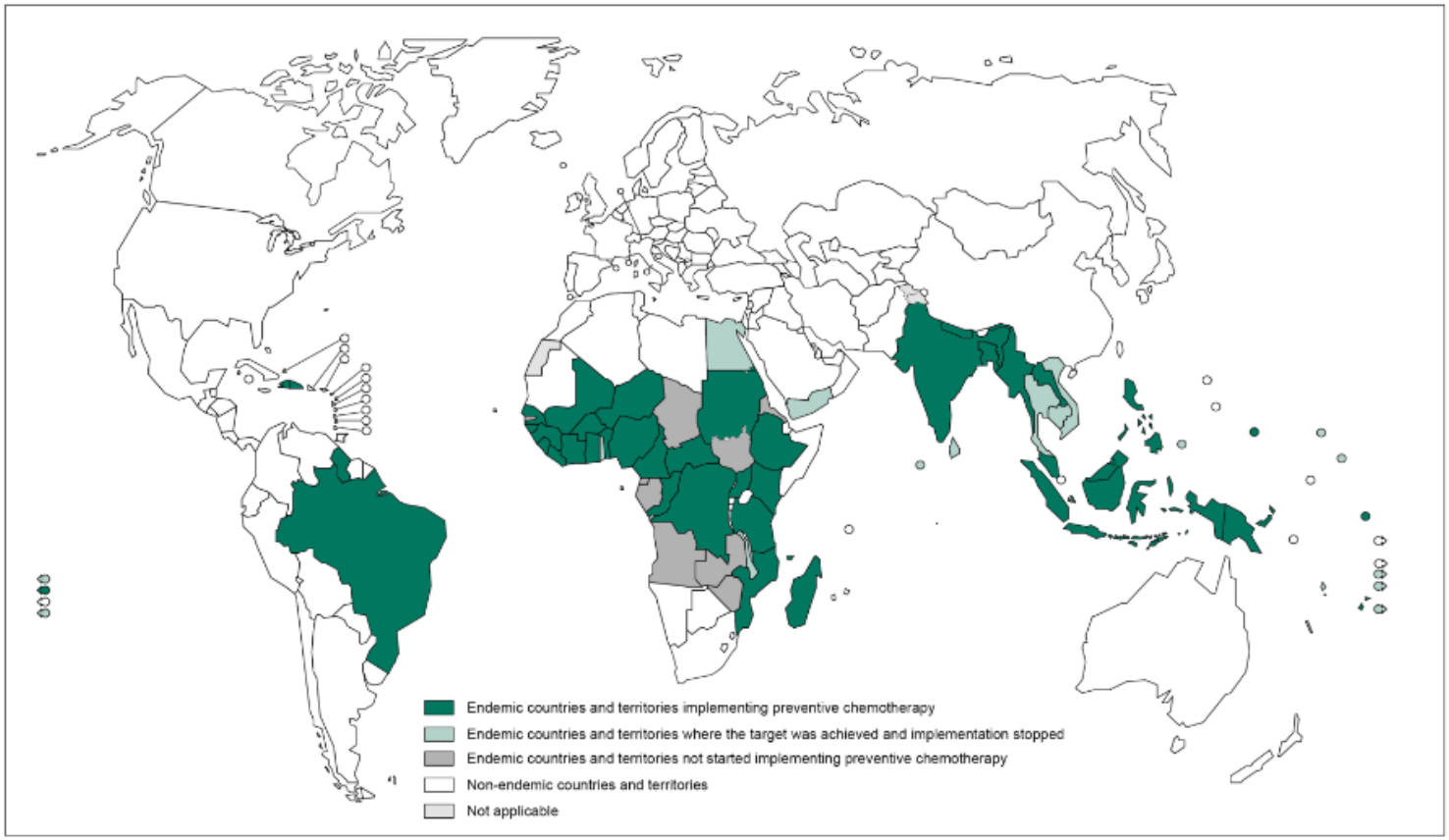
- **Ivermectin (Mectizan): annual dose clears microfilarial infections of the blood**
- **generally considered safe, with no serious side-effects**
- **mass distribution made possible by donation programme (Merck)**
- **used in multi-national programmes to combat onchocerciasis and lymphatic flariasis**

# Onchocerciasis: distribution



# Lymphatic filariasis: distribution

Distribution and status of preventive chemotherapy for lymphatic filariasis, worldwide, 2014

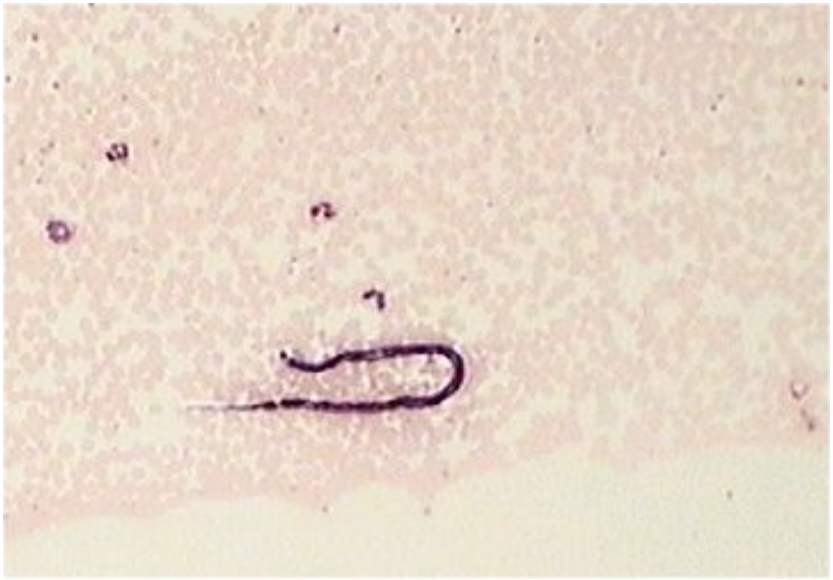


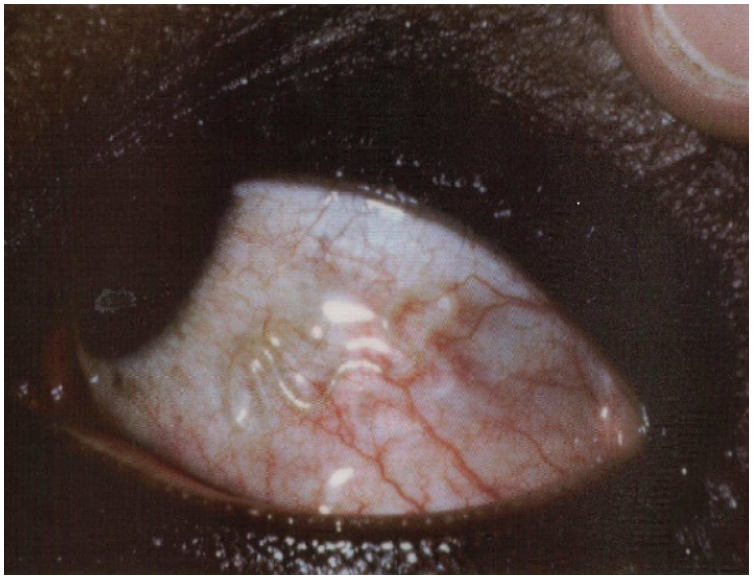



Data Source: World Health Organization
Map Production: Control of Neglected Tropical Diseases (NTD)
World Health Organization

World Health Organization

# The Loa loa problem

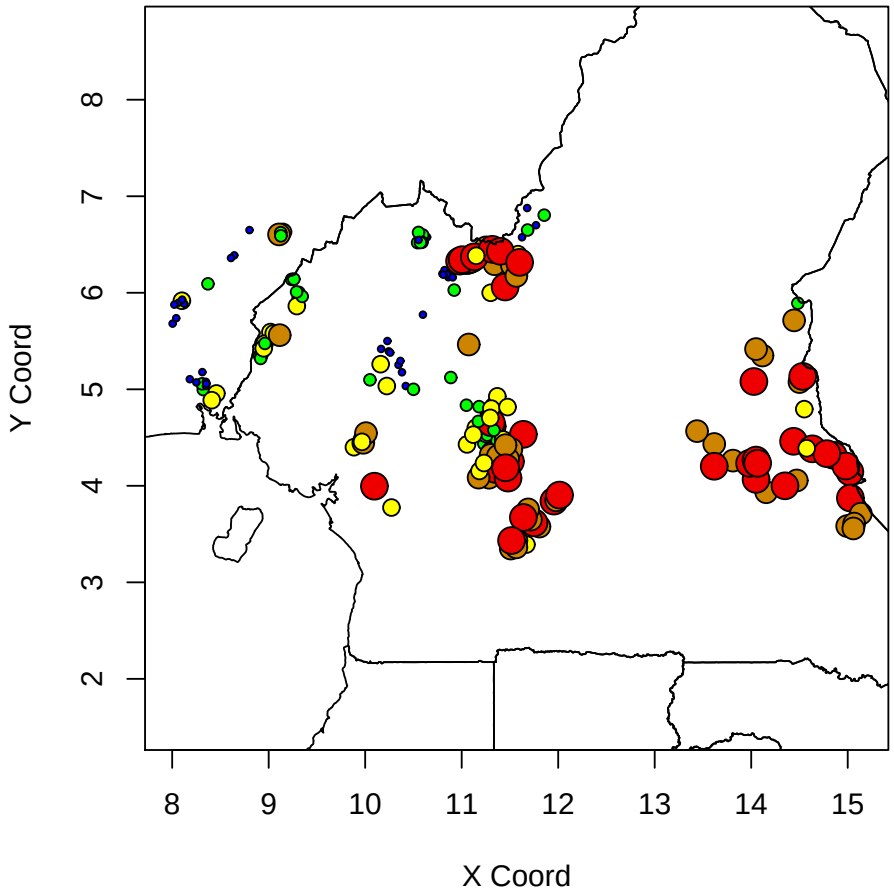




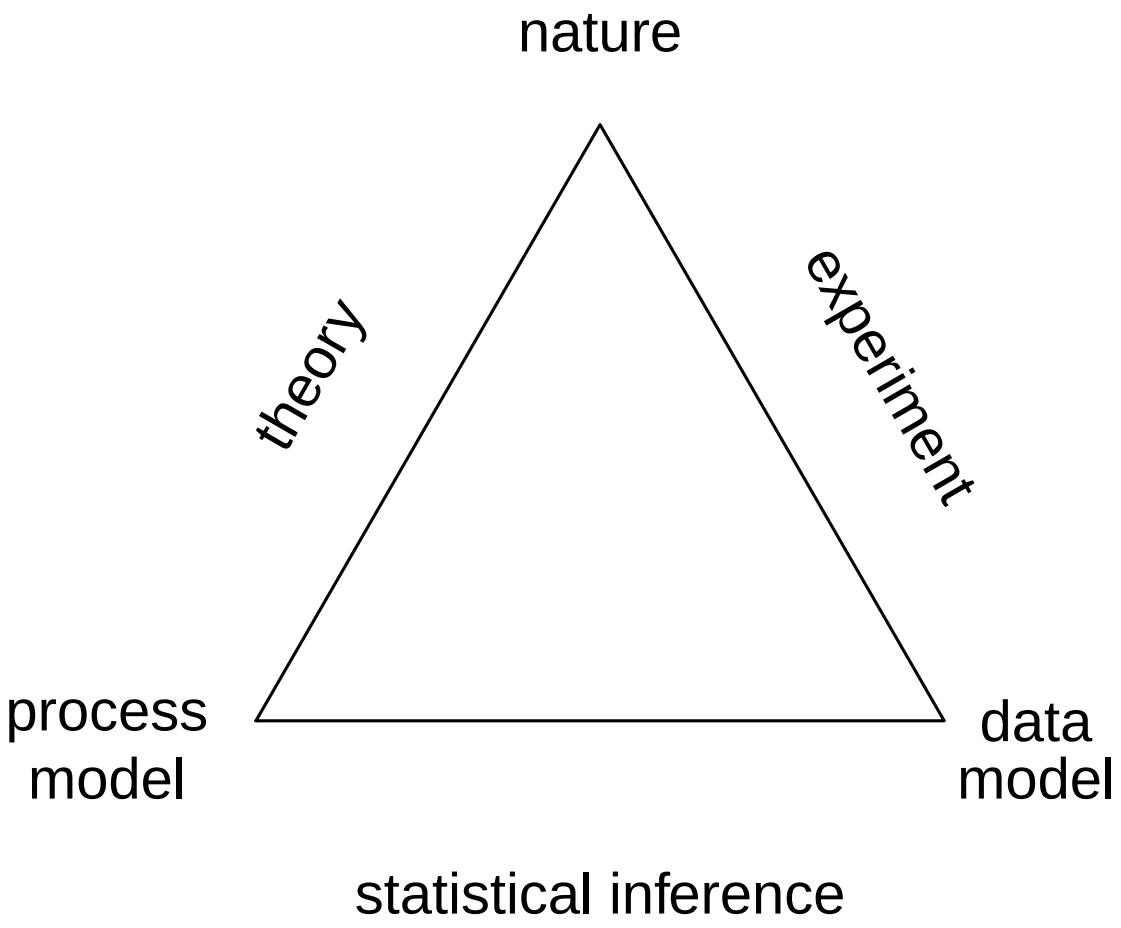
**People who are heavily co-infected with *Loa loa* parasites can experience serious (occasionally fatal) adverse reactions to ivermectin**

# First solution to the Loa loa problem: prevalence mapping

- **Individuals at risk of experiencing an SAE are those with high microfilarial loads**
- **Difficult/expensive to measure individual MF load**
- **Individuals with high MF loads are likely to be found in high-prevalence areas**




**High-risk area $\Leftrightarrow$ Prevalence $> 20\%$**


nature
theory
experiment
process
model
data
model
statistical inference

"The answer to any prediction problem is a probability distribution"

Peter McCullagh

---

$S$ = state of nature
$Y$ = all relevant data
$T$ = $\mathcal{F}(S)$ = target for prediction

---

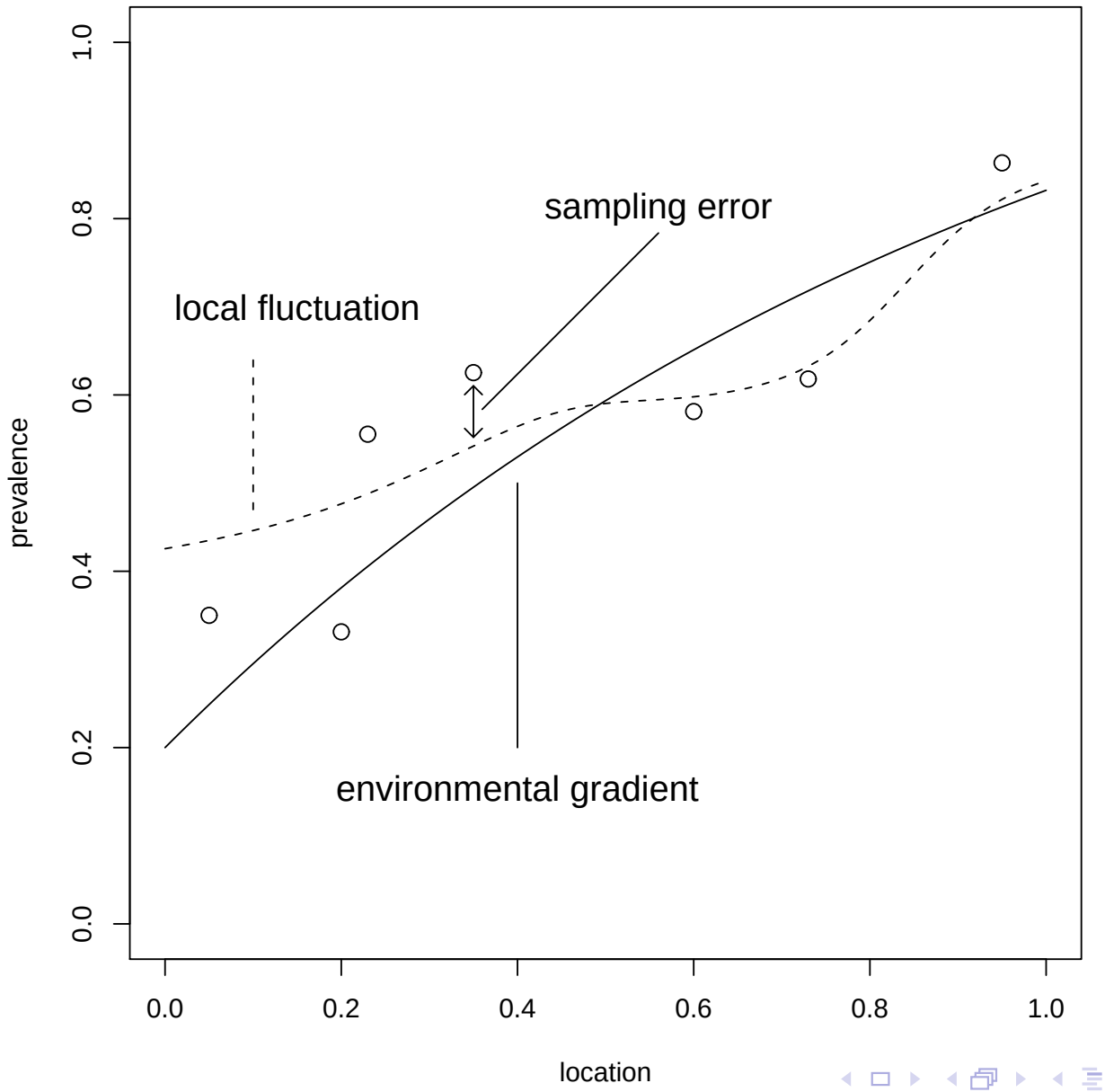
---

**Model:** $[S, Y] = [S][Y|S]$
**Prediction:** $[S, Y] \Rightarrow [S|Y] \Rightarrow [T|Y]$

---

$T(x) =$ prevalence $> 20\%$ (yes/no)

# Geostatistical model schematic



# The pink map

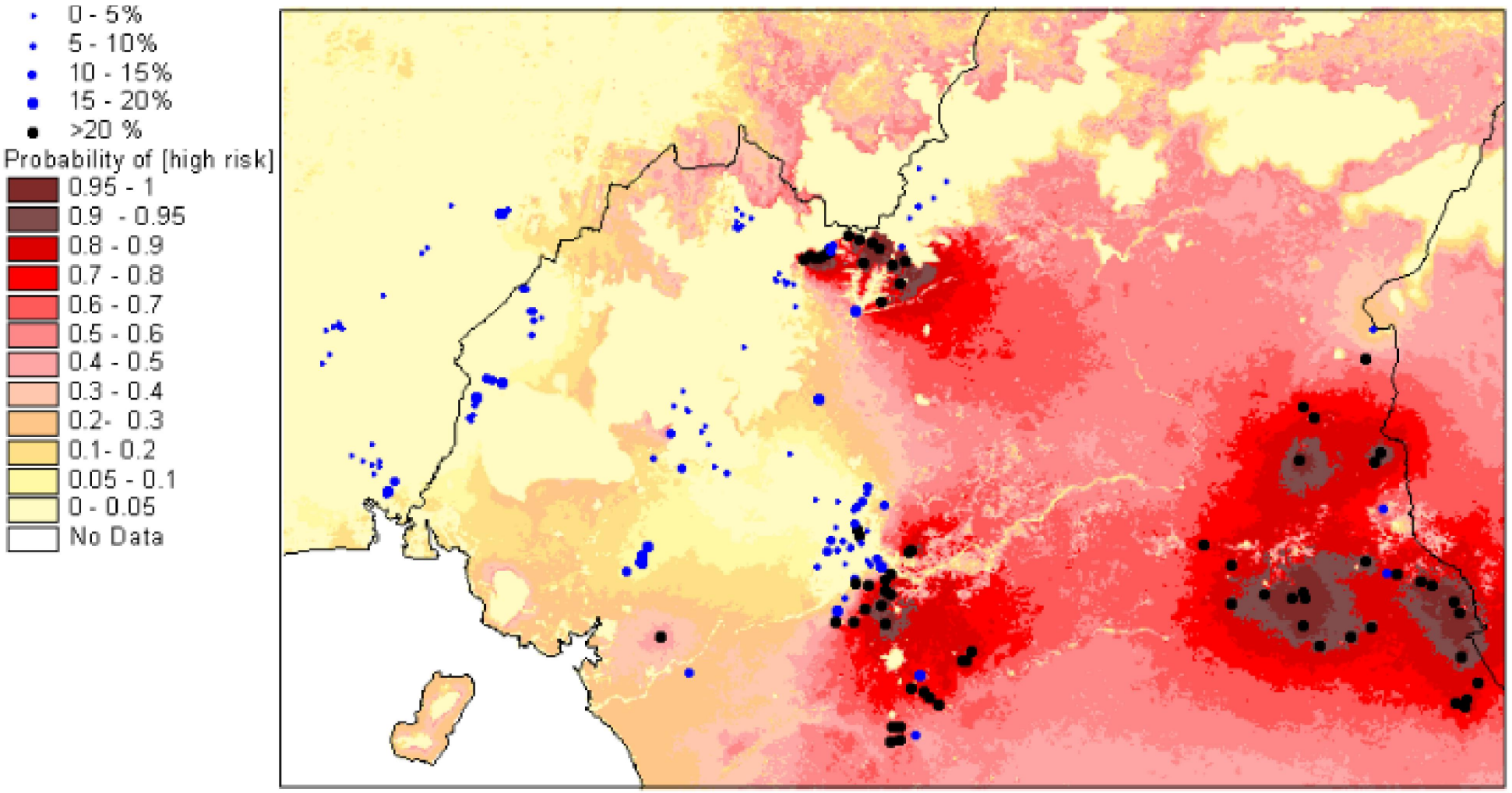
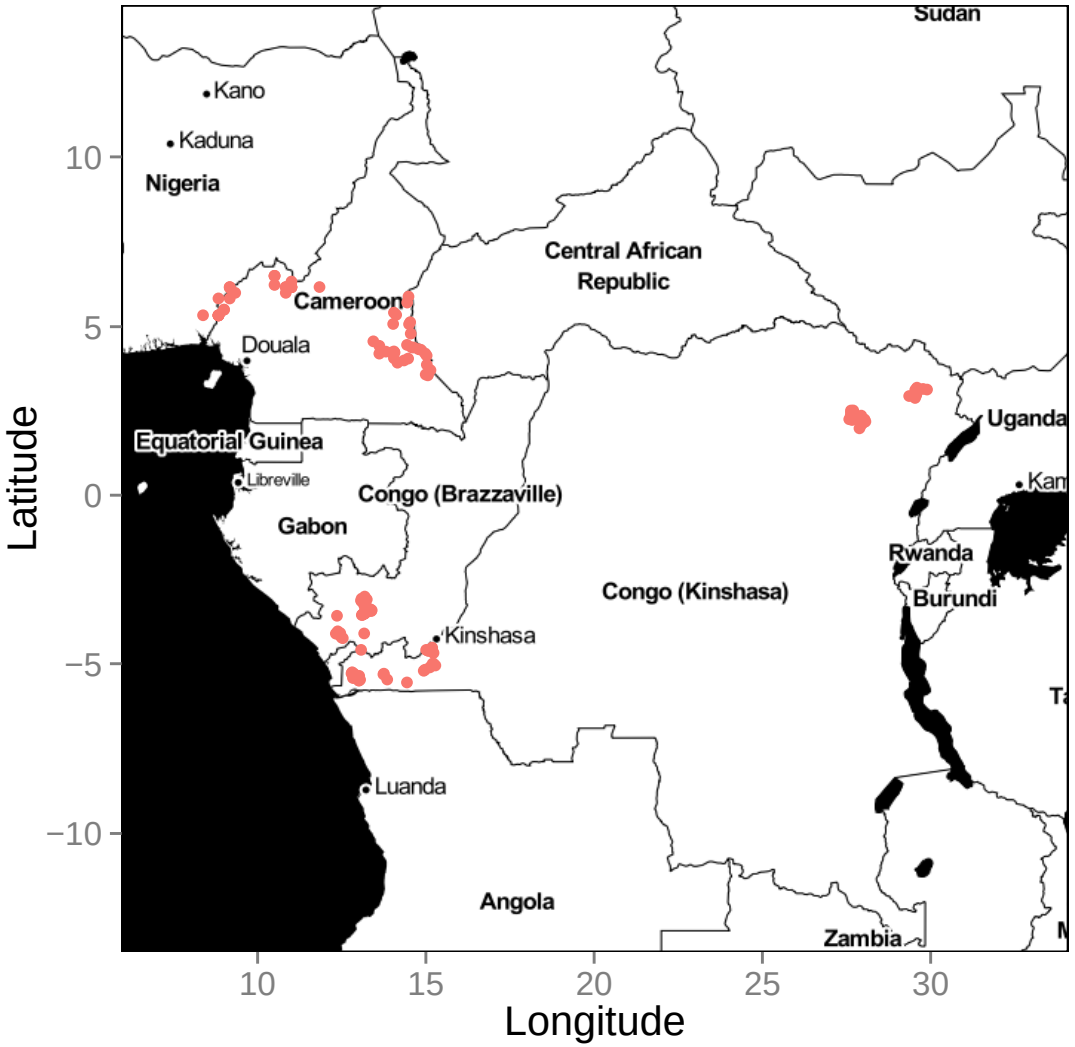



*Figure 6: PCM for [high risk] in Cameroon based on ERMr with ground truth data.*

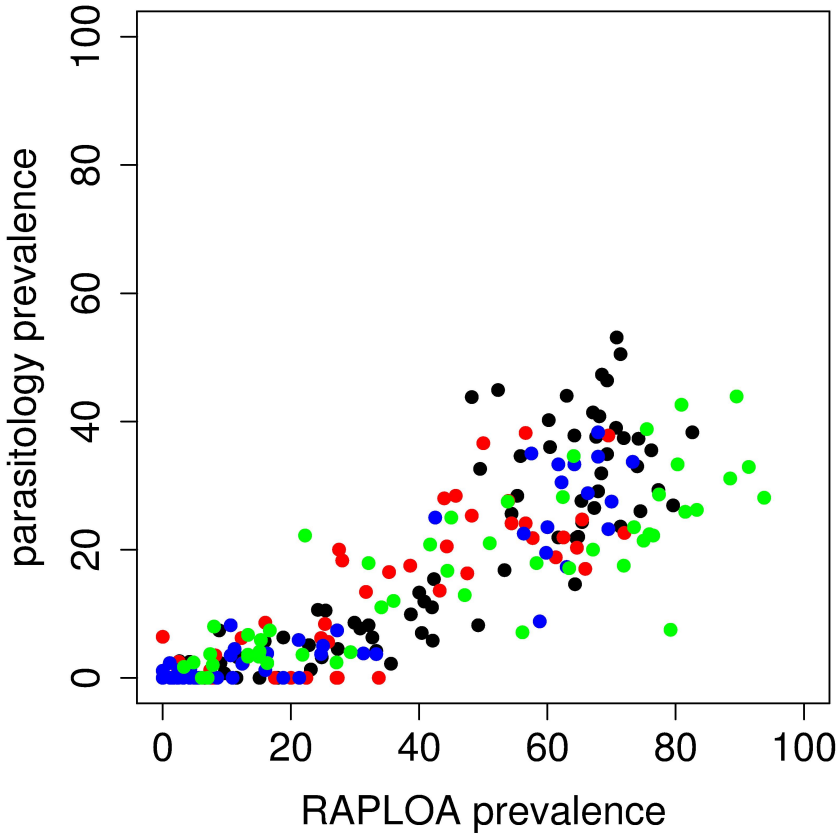
# Second solution... RAPLOA

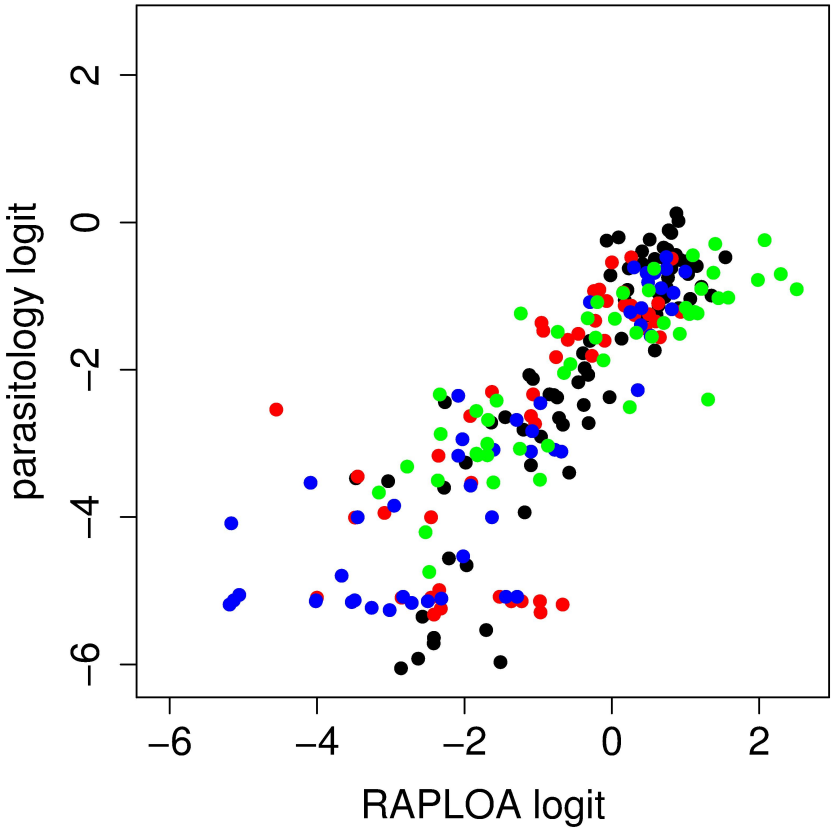
# Calibration surveys



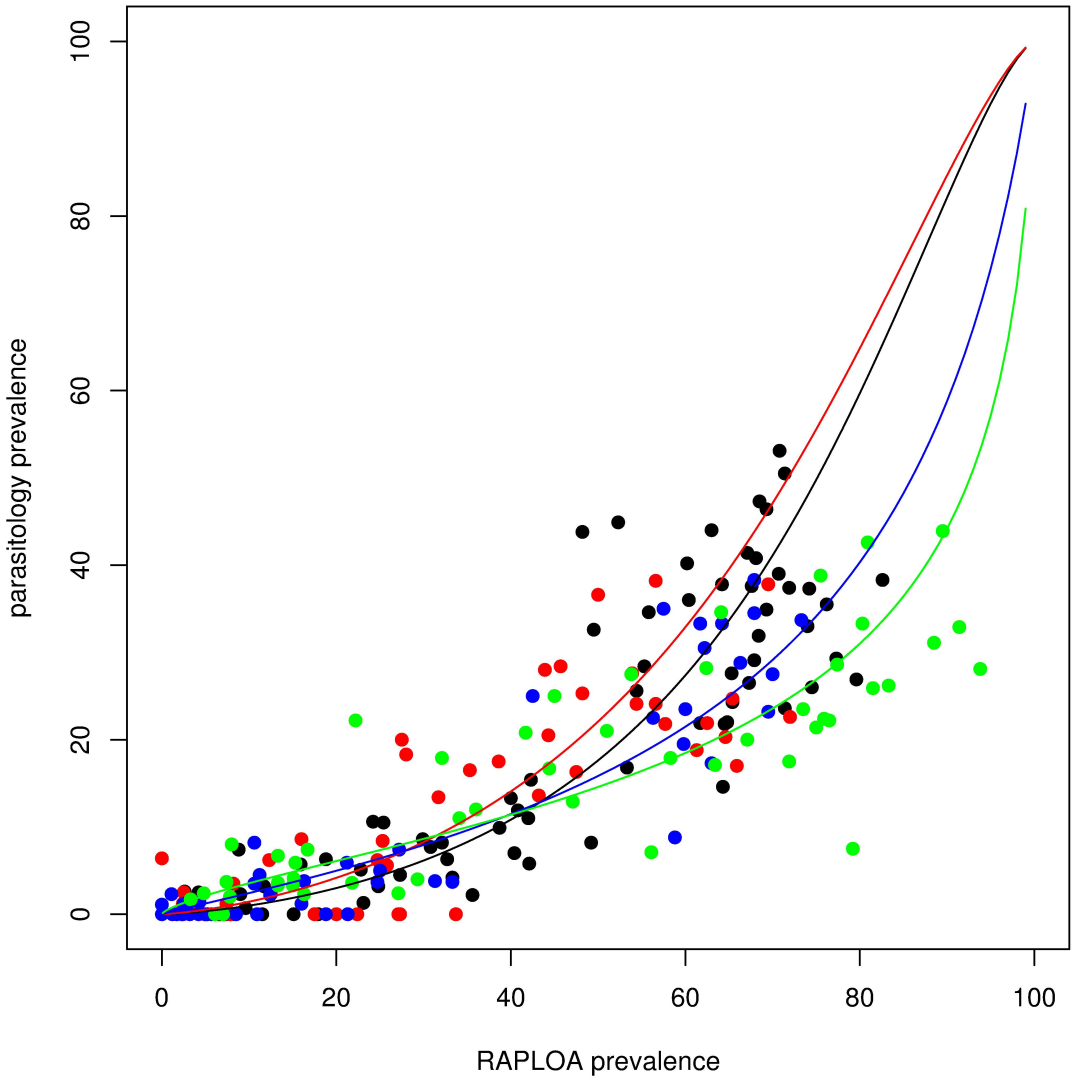

- **223 villages**
- **24 to 229 individuals per village, total 19,128**

# Calibration data

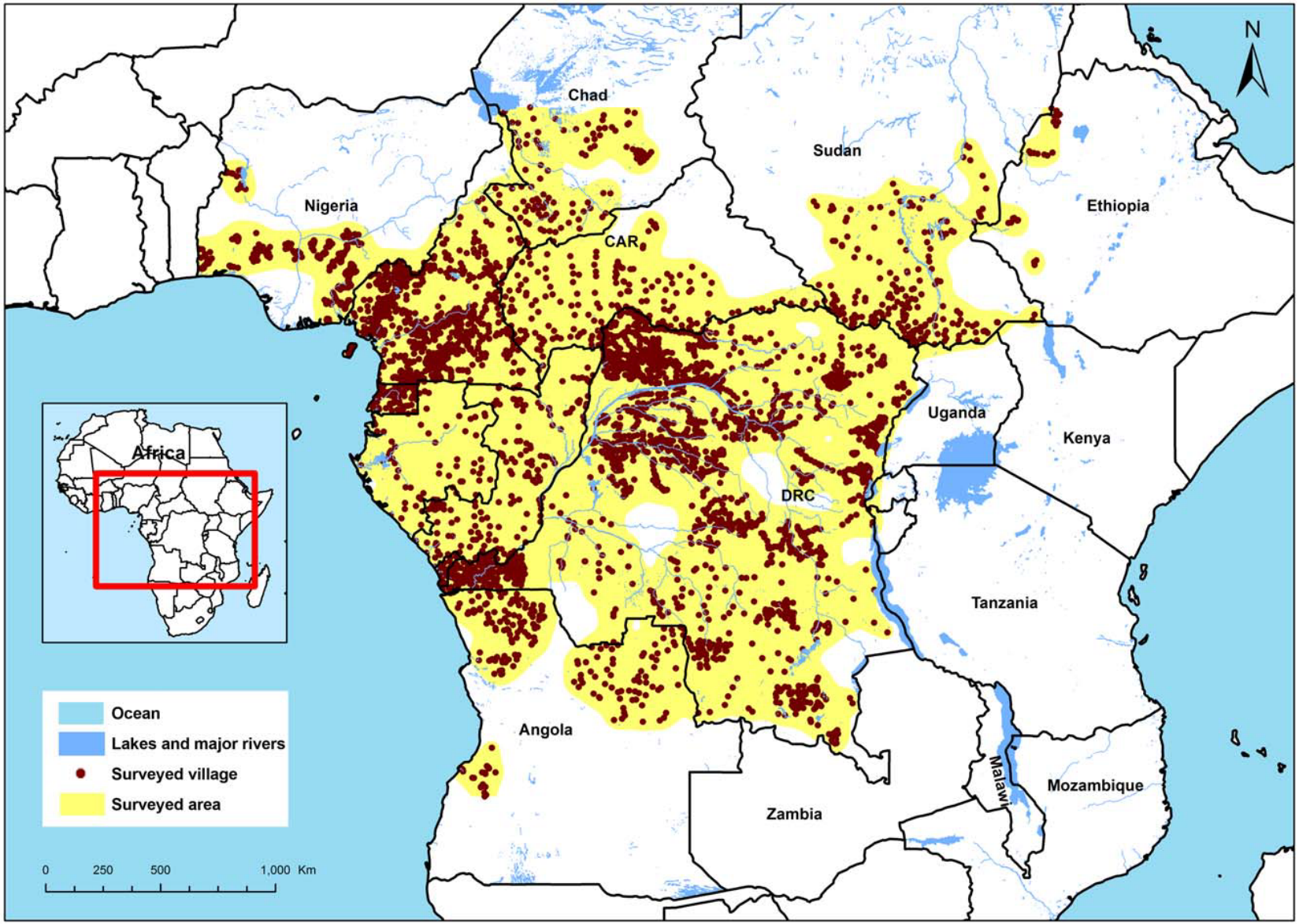


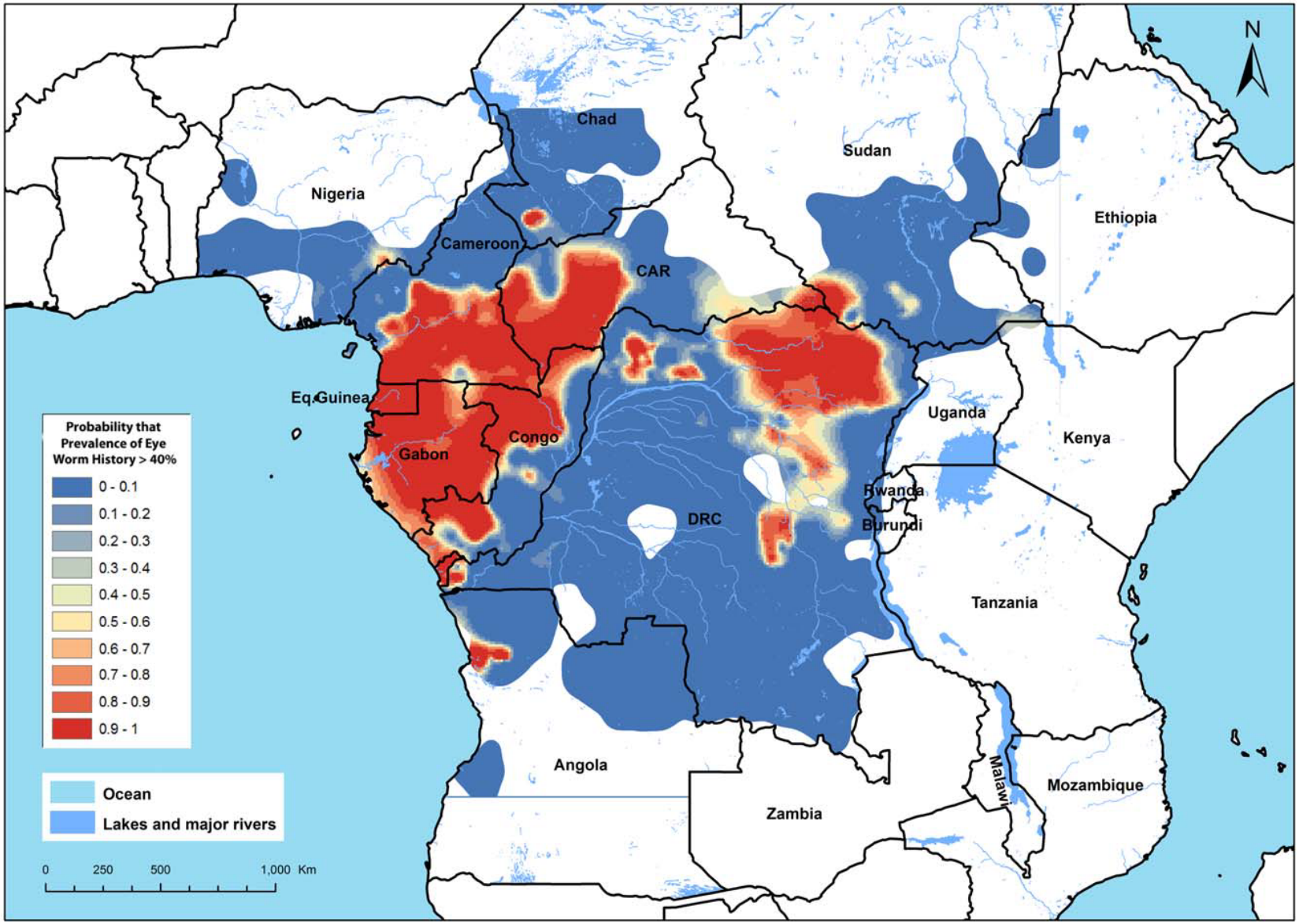
# Fitted model



# RAPLOA survey locations



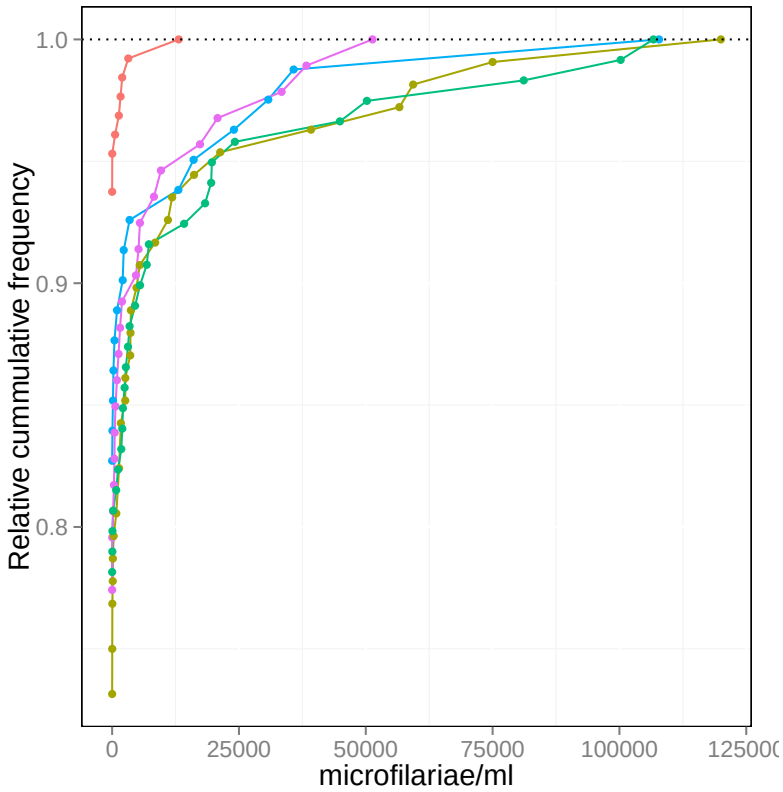
# RAPLOA exceedance map



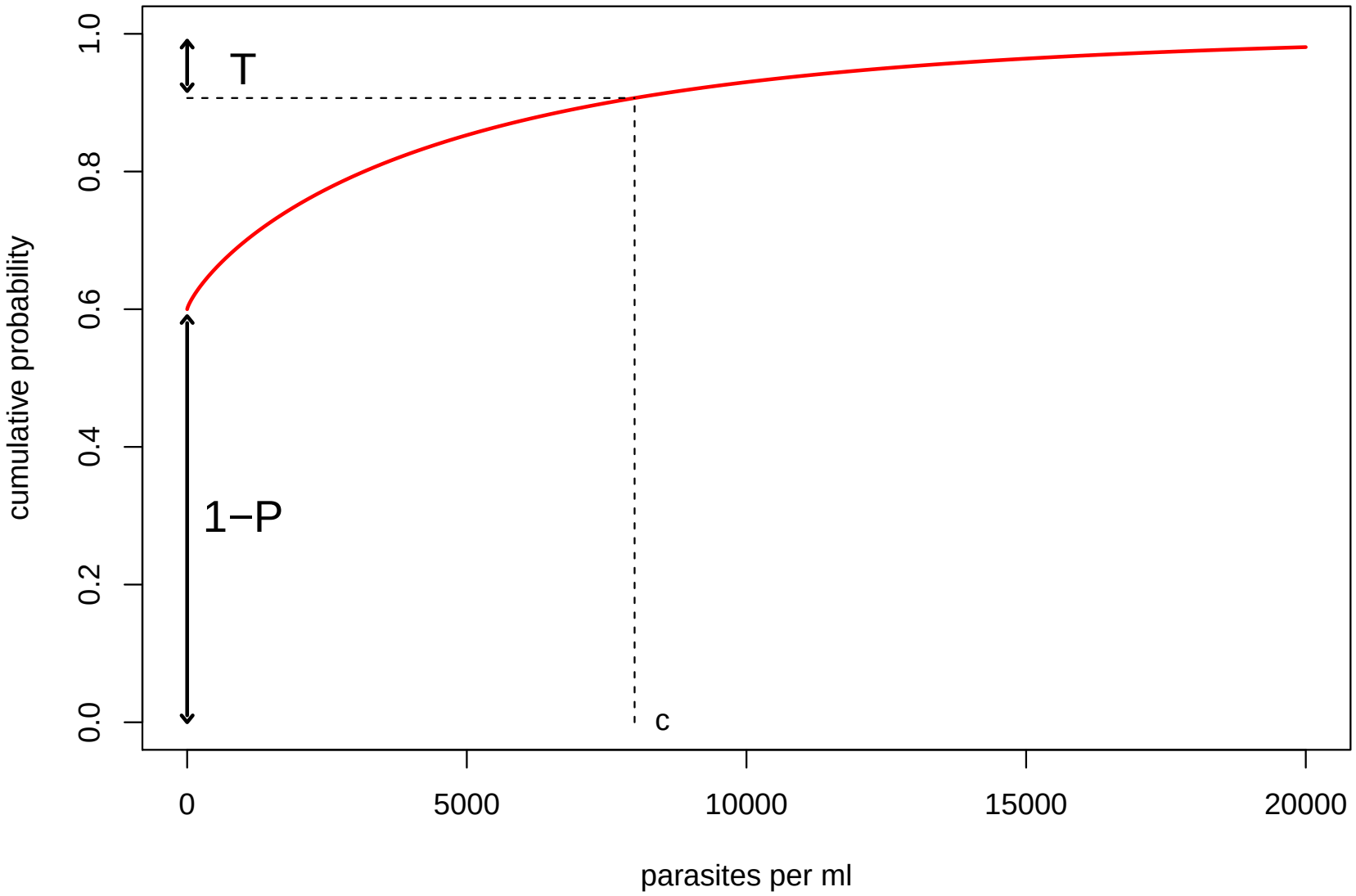
# Third solution... use prevalence to predict numbers of high-risk individuals

- **calibration data include individual MF loads**
- **wide variation in prevalence and MF load**
- **but distributional shape broadly consistent across all villages**

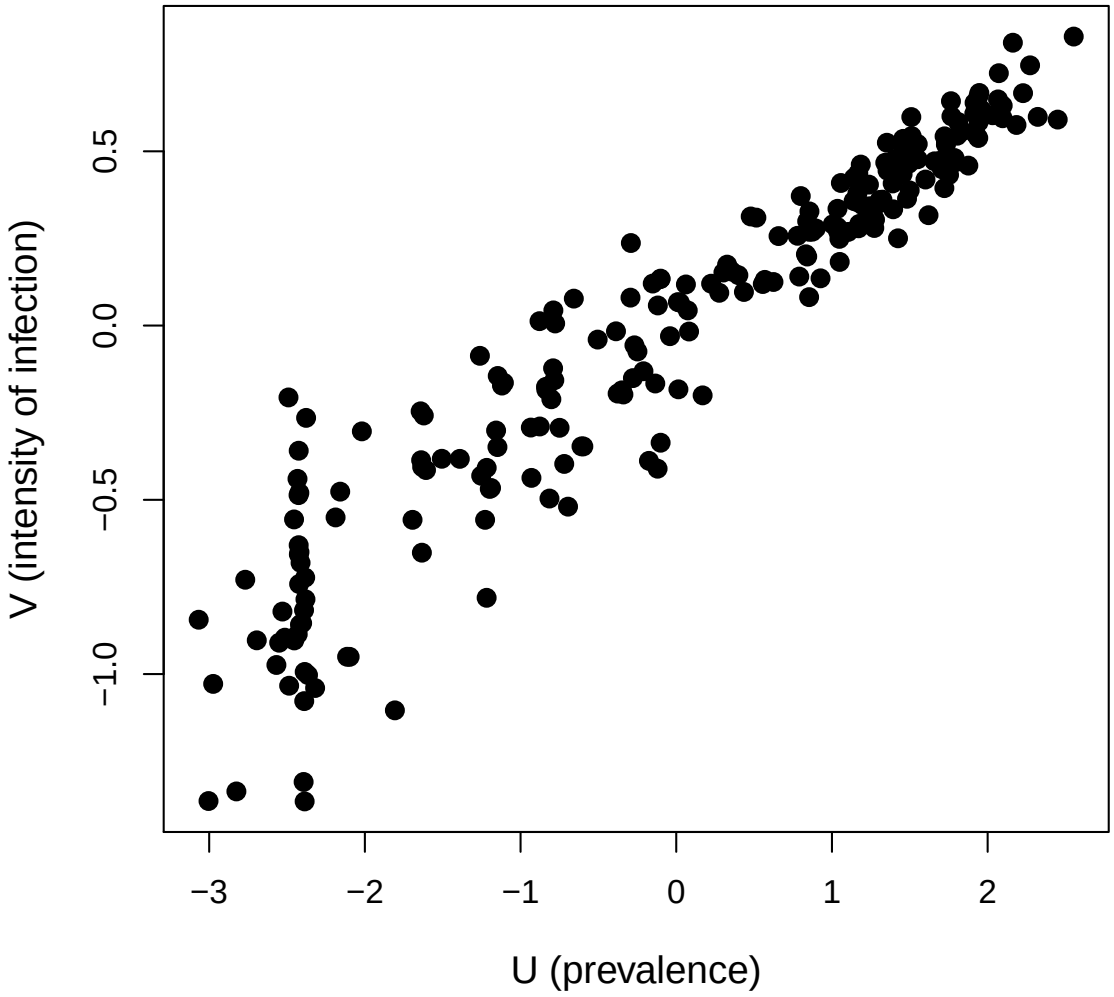
**Cumulative distributions of MF load**



# Schematic: P=prevalence; T=proportion highly infected

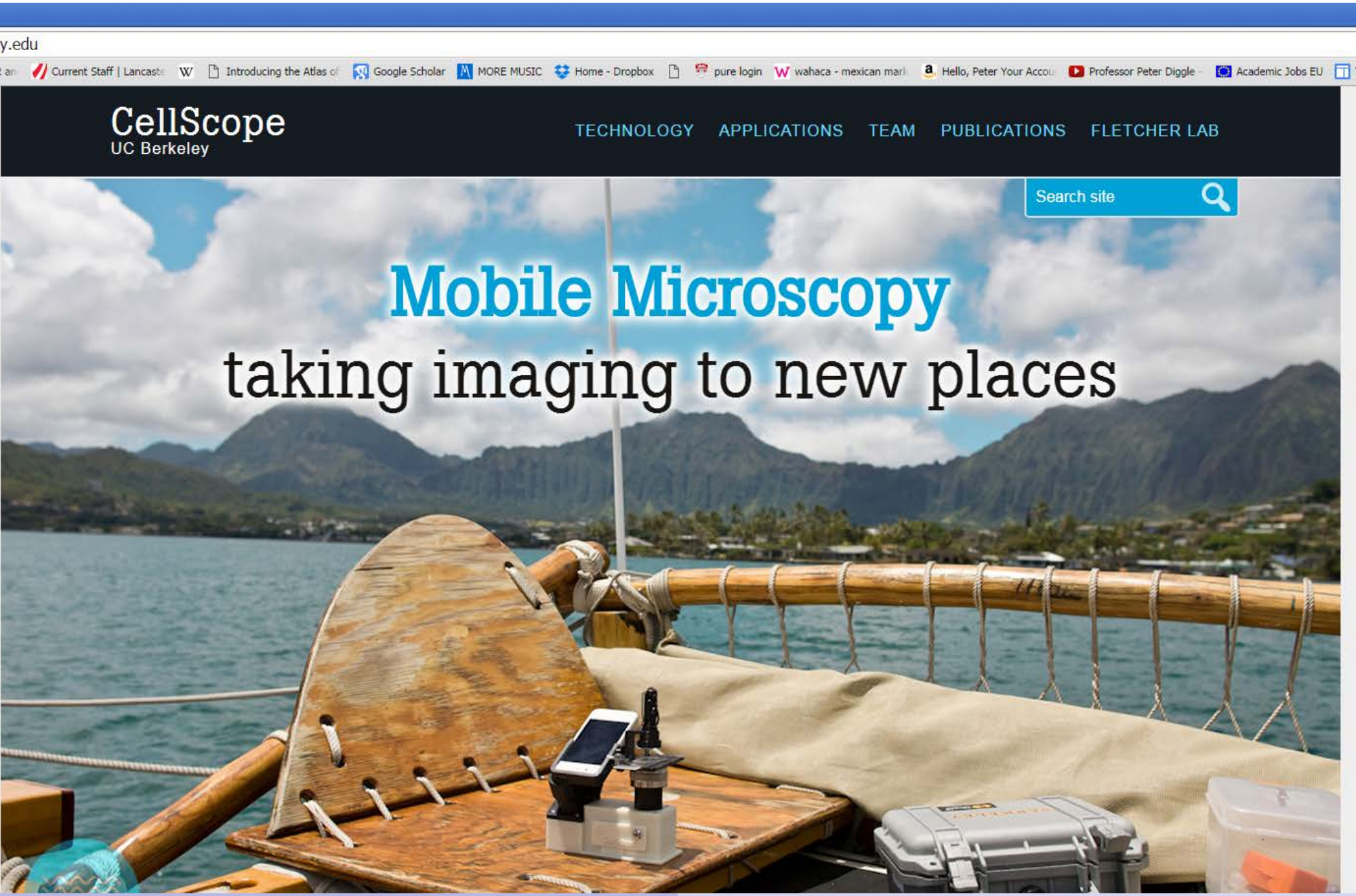


# Predicted random effects




**Target for prediction:** $T = \rho(U) \times \{1 - G(20000; \lambda(V), \kappa)\}$

# Fourth solution... the Loascope



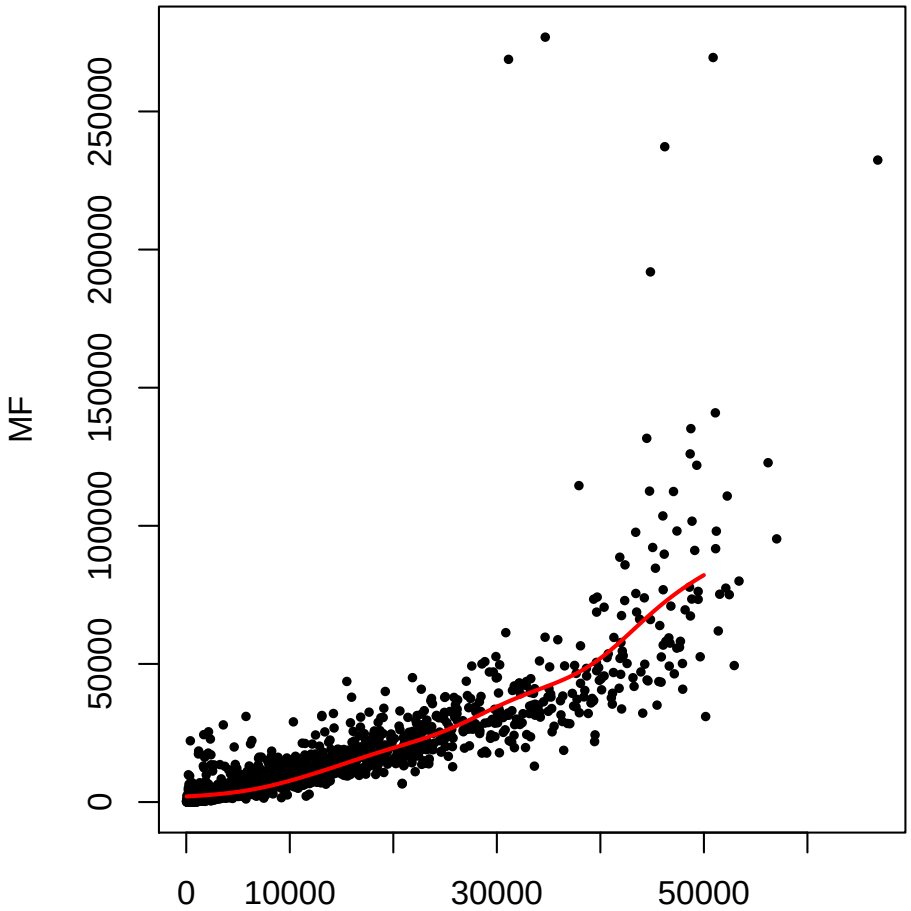
# Comparing traditional microscopy and Loascope

**15,181 people from 88 villages in Cameroon, each tested by traditional miroscopy (MF) and Loascope (LS).**

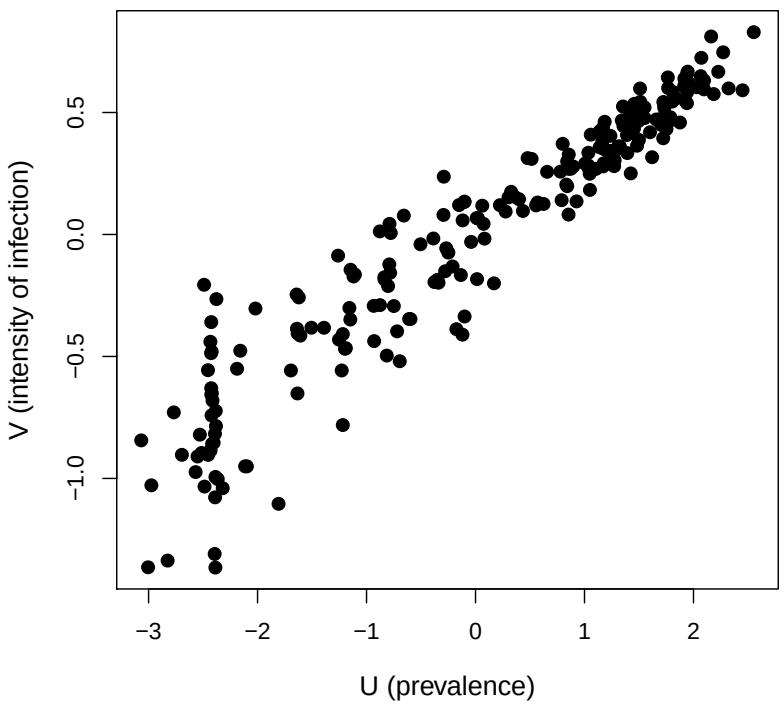
- **High concordance wrt infection status ($\kappa = 0.858$)**

| | CS | | |
|---|---|---|---|
| | 0 | 1 | Total |
| MF 0 | 12135 | 357 | 12492 |
| 1 | 268 | 2421 | 2689 |
| Total | 12413 | 2779 | 15181 |

- **Ranges of discordant values:**
  - $MF = 0 \Rightarrow 31 \leq LS \leq 7261$
  - $LS = 0 \Rightarrow 20 \leq MF \leq 19940$
- **Parasite-loads well-calibrated except at very high MF loads**



# Variation between villages: exploiting correlation




- **The proportion of a rare condition (here, $MF > 20K$) in a sample is an imprecise estimate of the proportion in the population**
- **Correlation between community-level prevalence and intensity of infection $\Rightarrow$ information gain (narrower prediction intervals)**

**Both prevalence and intensity give important information and both are necessary for efficient prediction**

# Current debate: MDA or TNT (test-and-not-treat)

## Model-validation study in Cameroon

1. **Loascope data from approx 100 individuals in each of 30 villages**
2. **Use model-based decision rule:**
   - **$P$ = probability that $< 1\%$ of individuals have $> 20000$ parasites/ml**
   - **MDA indicated if $P > 0.95$**
3. **Additional sample/census in each village to confirm (or not) model-based decision**

# Summary Results

**Phase 1: Results based on algorithm predications**

| **Phase 2: Validation** | MDA | TNT | |
|---|---|---|---|
| MDA | 14 | 5 | **19** |
| TNT | 0 | 5 | **5** |
| | **14** | **10** | |

Algorithm-based predictions are conservative, resulting in 5 villages receiving TNT that *could* have received MDA, based on validation results

# Summary Results

**Phase 1: Results based on algorithm predications**

| **Phase 2: Validation** | MDA | TNT | |
|---|---|---|---|
| MDA | 14 | 5 | **19** |
| TNT | 0 | 5 | **5** |
| | **14** | **10** | |

No villages deemed unsafe to treat with MDA in validation phase were targeted for MDA based on algorithm predictions

# Relationship of model-based to test-and-not-treat

- **Model-based strategy converges to test-and-not-treat as appetite for risk decreases**
- **Zero risk is unattainable**
  - **20000/ml a uniquely correct threshold?**
  - **measurement error in estimates of parasites/ml?**
- **Sequential algorithm:**

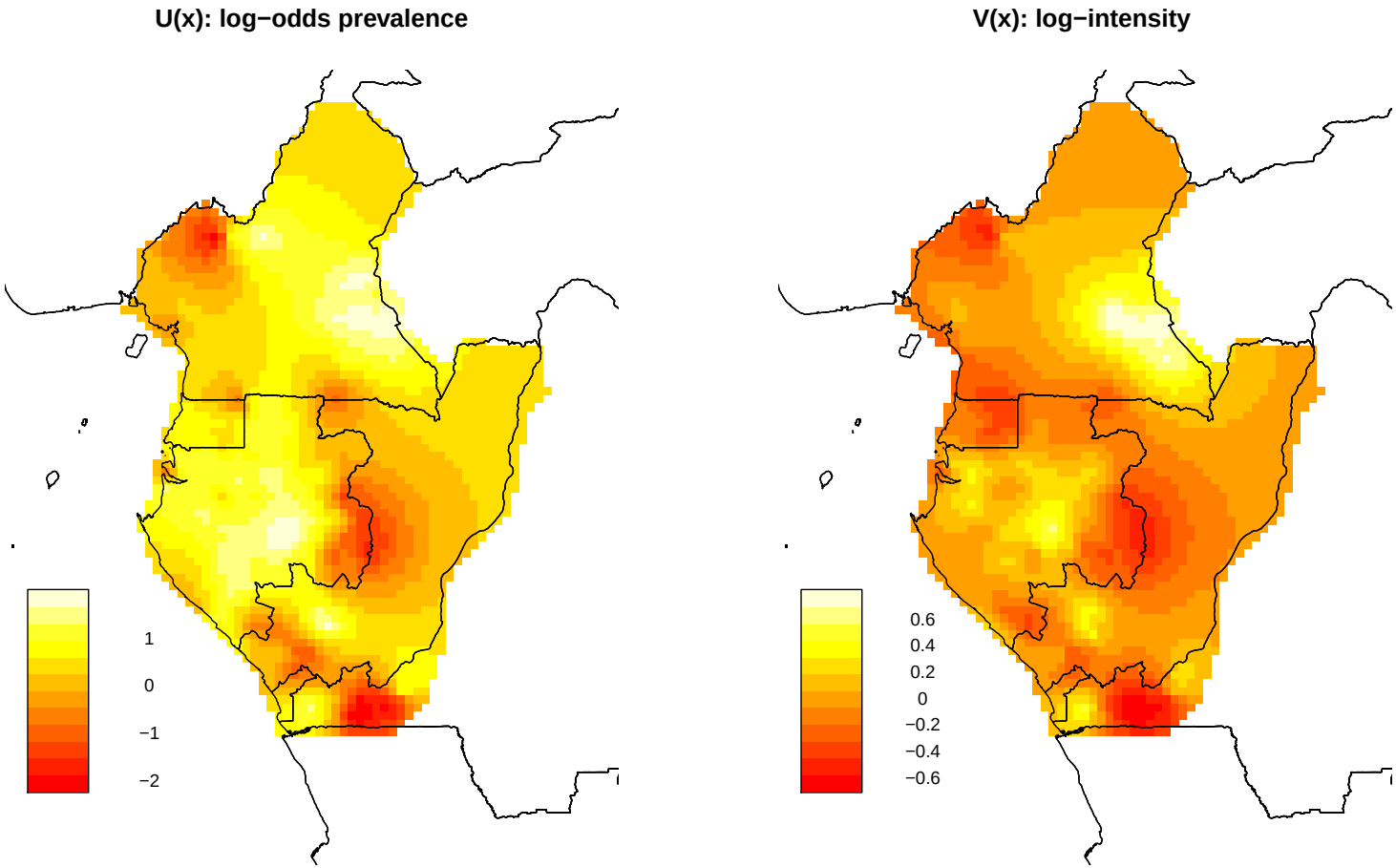
**Initial sample** $\Rightarrow \mathrm{P}(\text{“safe” for MDA}) = p$

$$p > 0.95 \quad \text{“safe”}$$

$$p < 0.05 \quad \text{“not safe”}$$

$$0.05 < p < 0.95 \quad \text{increase sample size}$$

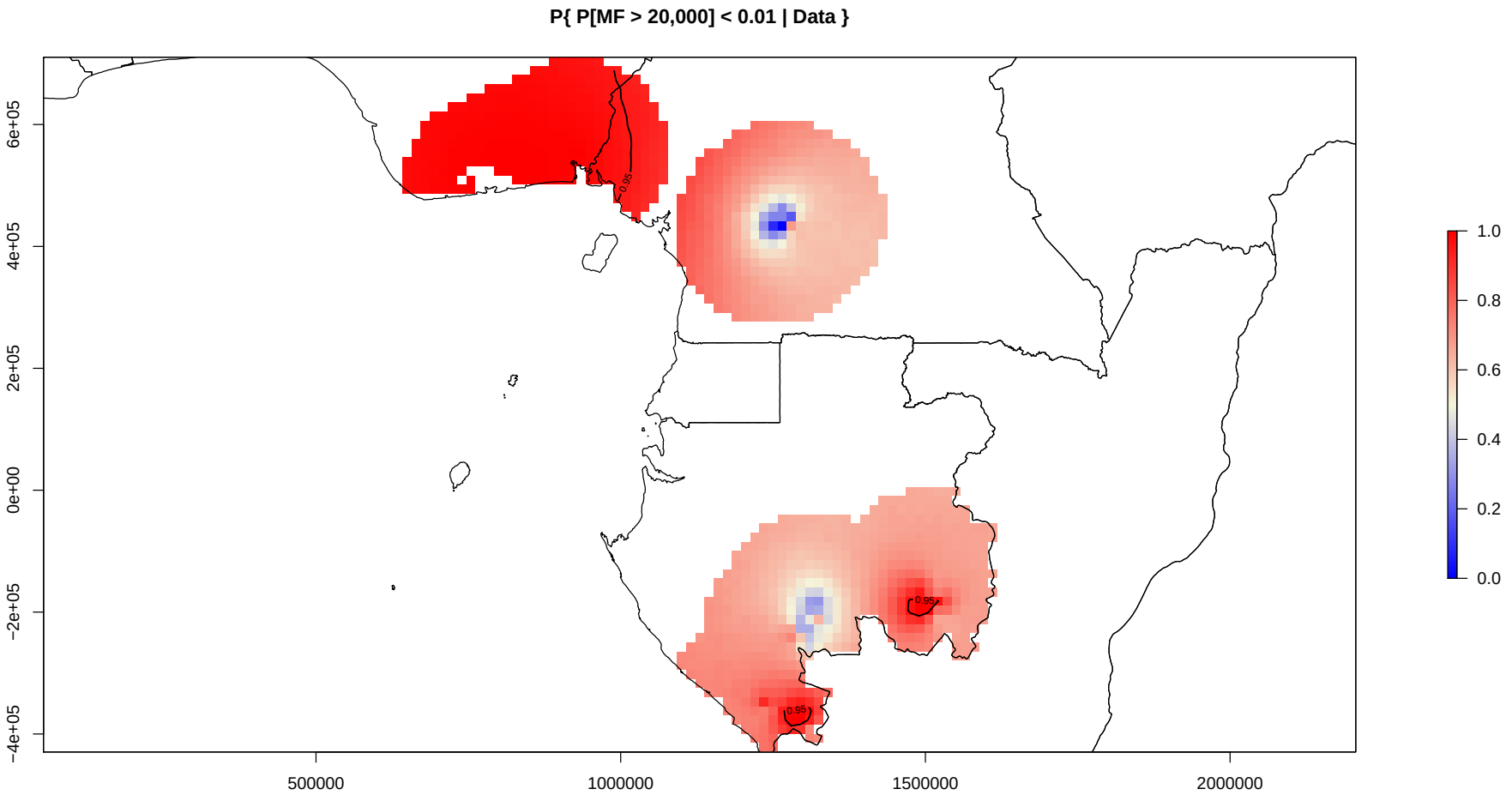
# The geospatial extension




$\Rightarrow$ **narrower prediction intervals (borrowing strength)**

# Mapping areas “safe” for MDA

**“safe” ⇔ “at least 95% certain that at most 1% of individuals have more than 20,000 parasites/ml**



# Fifth solution...what's the real question?

- **Vaccine safety analogue** $\mathrm{P}(SAE) = ?$
- **Loa loa...what do we know?**
  - **Predictive distribution** of $Y = MF$ load at location $x$: $f_x(y|\text{data})$
  - **Risk of SAE** given $MF = y$

$$p(y) = \Phi\{-1.05 + (1.87[\log_e(y) - 11.06])\} \pm ?$$

Boussinesq et al (2003, corrected)

- **Predictive target**

$$T(x) = \mathrm{Prob}(\text{SAE at } x) = \int p(y) f_x(y|\text{data}) dy$$

- **Draw samples** from predictive distribution of $T(x)$ over required area

# Conclusions

- **Principled statistical methods**
  - force assumptions into the open
  - deliver optimal solutions within the declared assumptions
- **But there is no free lunch**

  **"We buy information with assumptions"**

  **C H Coombs**
- **analyse problems, not data**

  **"Better an approximate answer to the right question than a precise answer to the wrong question"**

  **John Tukey**

# References


Boussinesq, M., Gardon, J., Gardon-Wendel, N. and Chippaux, J. (2003). Clinical picture, epidemiology and outcome of Loa-associated serious adverse events related to mass ivermectin treatment of onchocerciasis in Cameroon. *Filaria Journal*, **2**, 1–13.

Diggle, P.J., Moyeed, R.A. and Tawn, J.A. (1998). Model-based Geostatistics (with Discussion). *Applied Statistics* **47** 299–350.

Diggle, P.J., Thomson, M.C., Christensen, O.F., Rowlingson, B., Obsomer, V., Gardon, J., Wanji, S., Takougang, I., Enyong, P., Kamgno, J., Remme, H., Boussinesq, M. and Molyneux, D.H. (2007). Spatial modelling and prediction of Loa loa risk: decision making under uncertainty. *Annals of Tropical Medicine and Parasitology*, **101**, 499–509.

Schlüter, D.K., Ndeffo-Mbah, M.L., Takougang, I., Ukety, T., Wanji, S., Galvani, A.P. and Diggle, P.J. (2016). Using community-level prevalence of Loa loa infection to predict the proportion of highly-infected individuals: statistical modelling to support lymphatic filariasis elimination programs. *PLoS Neglected Tropical Diseases*, **10**, 12, e0005157. doi:10.1371/journal.pntd.0005157