Unité BIA Toulouse
Biométrie et Intelligence Artificielle

# Model-based adaptive spatial sampling for occurrence map construction

N. Peyrard and R. Sabbadin

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

mia
Département de mathématiques et informatique appliquées

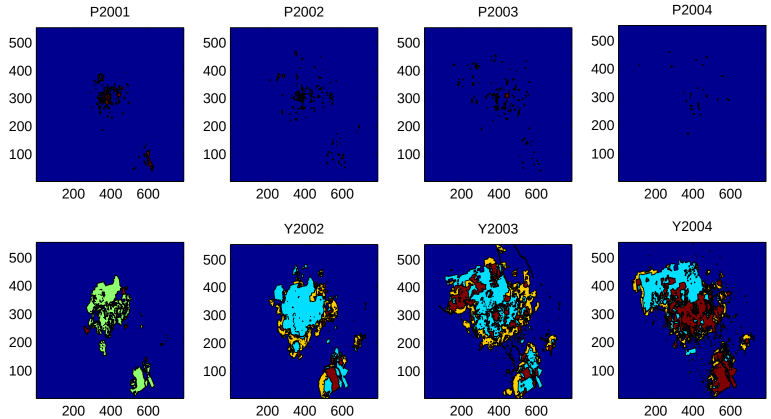
INRA

Unité BIA Toulouse
Biométrie et Intelligence Artificielle

# Mapping spatial processes in environmental management

## Mapping pest occurrence

- Building pest occurrence map in order to eradicate
- Observations costly
- Errors in mapping also costly



# Mapping spatial processes in environmental management

**Different problems depending on observations nature**

- Data visualization
  - Complete observations (everywhere)
  - Perfect observations (No errors/missing data)

  $\Rightarrow$ How to visualize data?
- Map reconstruction
  - Complete observations
  - Noisy observations

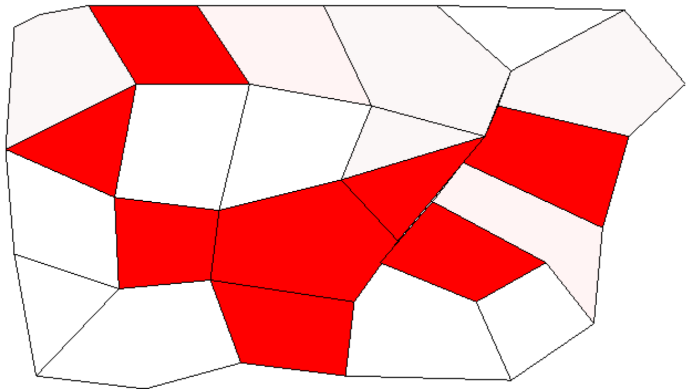
  $\Rightarrow$ How to reconstruct the “true” map?
- **Sampling and map construction**
  - Incomplete observations (not everywhere)
  - Noisy observations

  $\Rightarrow$ Where to observe? / How to reconstruct?

# Mapping spatial processes in environmental management

**How to design an efficient spatial sampling method to estimate an occurrence (0/1) map when**

- ✓ process to map has spatial structure
- ✓ observations are imperfect/incomplete
- ✓ sampling is costly
- ✓ process does not evolve during the sampling period



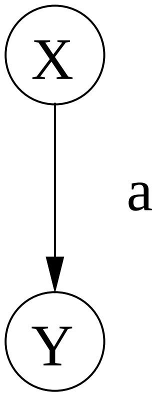
# Overview of the proposed approach

**Optimization approach for designing spatial sampling policies**

The Hidden Markov Random Field model is used for:

- Representing current uncertain knowledge about map to reconstruct
- Updating knowledge after observations
- Defining a unique criterion for
  - map reconstruction from observed data
  - spatial sampling actions selection

# Optimal sampling problem




Hidden variable $X$

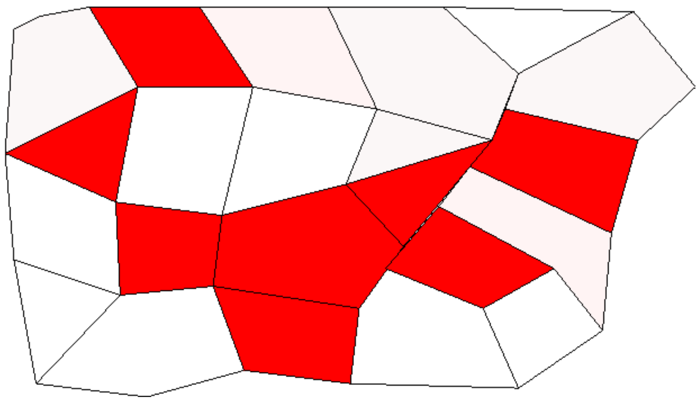
Sampling action $a$

Observation model
$p(Y = o|x, a)$

**Question**: How to reconstruct hidden variable $X$ using sampling actions?

1. Hidden variable model
2. Updated model after sampling result
3. Hidden variable reconstruction
4. Sampling action optimization

# Spatial sampling optimization



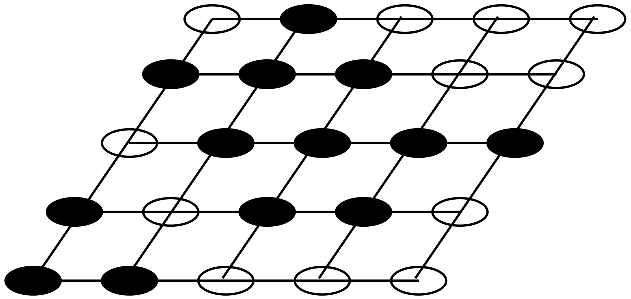
The hidden variable $x$ is a map

$\Rightarrow$ The sampling optimization problem has to be revisited

**Question**: How to reconstruct hidden map $x$ using sampling actions?

1. Hidden map model
2. Updated model after sampling result
3. Hidden map reconstruction
4. Sampling action optimization

# Pairwise Markov random field (1)



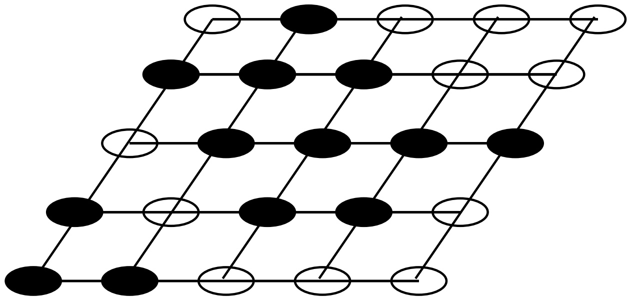
- Multiple interacting variables
- Independence given neighborhood

$\Rightarrow$ Pairwise Markov random field

**Question**: How to reconstruct hidden map $x$ using sampling actions?

1. Hidden map model
2. Updated model after sampling result
3. Hidden map reconstruction
4. Sampling action optimization

# Pairwise Markov random field (2)



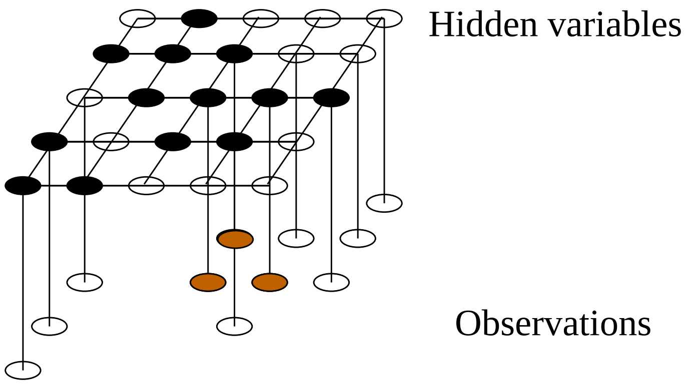
- Multiple interacting variables
- Independence given neighborhood

$\Rightarrow$ Pairwise Markov random field

- Interaction graph $G = (V, E)$
- $\psi_i$: “weights” on states of vertex $i$
- $\psi_{ij}$: correlations “strength” between neighbor vertices
- $Z$: normalizing constant / partition function

$$P(x) = \frac{1}{Z}\Big(\prod_{i \in V} \psi_i(x_i)\Big)\Big(\prod_{(i,j)\in E} \psi_{ij}(x_i, x_j)\Big)$$

# Hidden Markov random field (1)



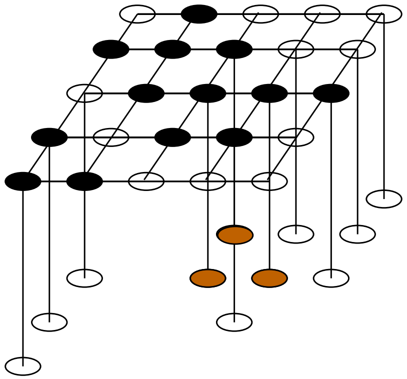

- $a \in \{0,1\}^{|V|}$: subset of $V$ selected for sampling
- Independent observations:

$$P(o|x,a) = \prod_{i \in V} P_i(o_i|x_i,a_i)$$

**Question**: How to reconstruct hidden map $x$ using sampling actions?

1. Hidden map model
2. Updated model after sampling result
3. Hidden map reconstruction
4. Sampling action optimization

# Hidden Markov random field (2)




- $a \in \{0,1\}^{|V|}$: subset of $V$ selected for sampling
- Independent observations:

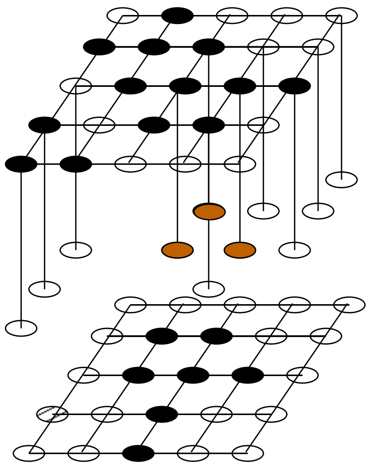
$$P(o|x,a) = \prod_{i \in V} P_i(o_i|x_i,a_i)$$

## Updated Markov random field (Bayes' theorem)

$$P(x|o,a) = \frac{1}{Z}\Big(\prod_{i \in V} \psi'_i(x_i,o_i,a_i)\Big)\Big(\prod_{(i,j)\in E} \psi_{ij}(x_i,x_j)\Big) \text{ where}$$

$$\psi'_i(x_i,o_i,a_i) = \psi_i(x_i)P_i(o_i|x_i,a_i)$$

# Hidden map reconstruction (1)



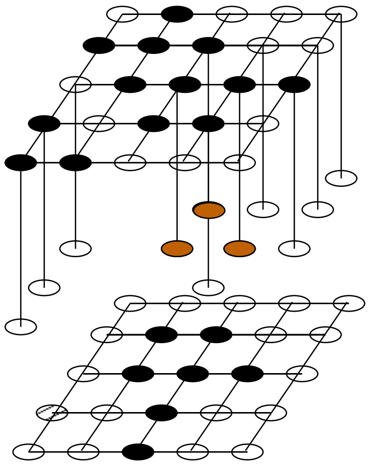

Local (MPM):

$$x_i^* = \arg\max_{x_i} P_i(x_i|o,a), \forall i \in V$$

**Question**: How to reconstruct hidden map $x$ using sampling actions?

1. Hidden map model
2. Updated model after sampling result
3. Hidden map reconstruction
4. Sampling action optimization

# Hidden map reconstruction (2)




Local (MPM):

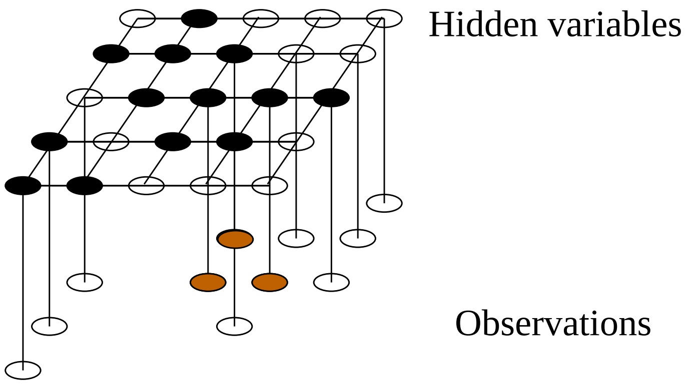
$x_i^* = \arg\max_{x_i} P_i(x_i|o,a)$

**Value of reconstructed map**

Expected number of well classified sites in $x^*$

$$V^{MPM}(o,a) \quad = \quad f\Big(\sum_{i\in V} \max_{x_i} P_i(x_i|o,a)\Big)$$

# Sampling action optimization (1)



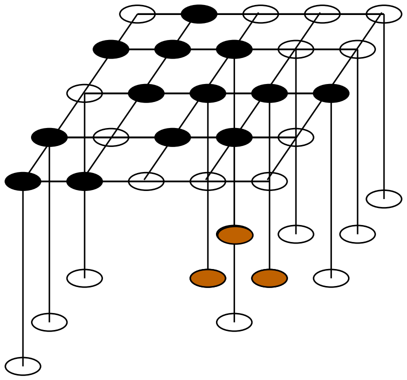

- $a \in \{0,1\}^{|V|}$ selected for sampling
- Independent observations $o \in \{0,1\}^{|V|}$

$\Rightarrow$ How to optimize the choice of $a$?

**Question**: How to reconstruct hidden map $x$ using sampling actions?

1. Hidden map model
2. Updated model after sampling result
3. Hidden map reconstruction
4. Sampling action optimization

# Sampling action optimization (2)




- $a \subseteq V$ selected for sampling
- Independent observations $o$ result

$\Rightarrow$ How to optimize the choice of $a$?

$$U(a) = -c(a) + \sum_o P(o|a)V(o,a)$$
$$a^* = \arg\max_a U(a)$$

- The computation of $a^*$ is hard! (NP-hard)
- Only feasible for small problems or needs approximation!

Unité BIA Toulouse
Biométrie et Intelligence Artificielle

# Approximate spatial sampling (1)

Approximate the computation of

$$a^* = \arg\max_a -c(a) + \sum_o P(o|a) V^{MPM}(o,a)$$

- Explore cells where initial knowledge is the most uncertain: marginal $P_i(x_i|o,a)$ closest to $\frac{1}{2}$

$$\tilde{a} = \arg\max_a -c(a) + f\left(\sum_{i,a_i=1} \min\left\{P_i(X_i=1), P_i(X_i=0)\right\}\right)$$

- Marginals computation is itself *NP-hard*
$\Rightarrow$ approximation using belief propagation (sum prod) algorithm

# Approximate spatial sampling (2)

The approximation results from simplifying assumptions:

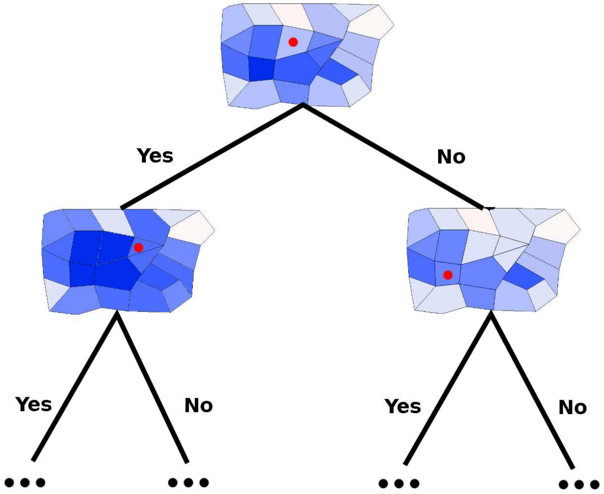
- Sampling actions are reliable
- No passive observations
- Joint probability approximated by one with idependent factors

# Adaptive spatial sampling (1)

- Idea:
  - Sampling locations not chosen once for all before the sampling campaign
  - Intermediate observations are taken into account to design next sampling step
  - Possibility to visit a cell more than once

# Adaptive spatial sampling (2)

- a sampling strategy $\delta$ is a tree
- a trajectory in $\delta$: $\tau = (a^1, o^1, \dots, a^K, o^K)$




**Value of a leaf**

$$U(\tau) = -\sum_{k=1}^{K} c(a^k) + V^{MPM}(o^0, o^1, \dots, o^K, a^0, a^1, \dots, a^K)$$

**Value of a strategy** $V(\delta) = \sum_\tau U(\tau) P(\tau \mid \delta)$

# Heuristic adaptive spatial sampling

- Exact computation is *PSPACE-hard* !

⇒ Heuristic algorithm

- on line computation
- approximate method for static sampling at each step

# Concluding remarks

- A framework for spatial sampling optimization:
  - based on Hidden Markov random fields
  - different map quality criteria
  - extended to "adaptive" sampling
- Problems too complex for exact resolution

  $\Rightarrow$ Heuristic solution based on approximate marginals computation
- Empirical validation on simulated problems:
  - Comparison of SSS, ASS and classical sampling methods (random sampling, ACS)
  - Markov random fields parameters learned from real data
  - ASS $>$ SSS $>$ classical methods

# Ongoing work

- Exact algorithms for small problems (Usman Farrokh): combining variable elimination and tree search
- “Random sets + kriging” approach (Mathieu Bonneau): development of a dedicated approximate method and comparison to the HMRF approach
- PhD thesis on *adaptive spatial sampling for weeds mapping at the scale of an agricultural area* (Sabrina Gaba, INRA-Dijon).
- Future?

$\Rightarrow$ Spatial partially observed Markov decision processes

## Questions?

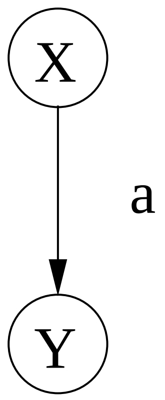
# Thanks for listening

# Contents

1- Optimal sampling of a hidden random variable
2- Defining optimal spatial sampling problems
3- Approximate computation of an optimal strategy
4- Evaluation of proposed method on simulated data

# Optimal sampling problem

## Hidden variable model



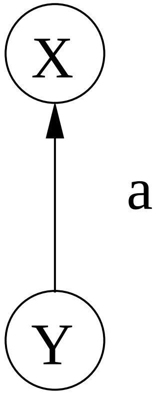

Prior model $P(x)$

**Question**: How to reconstruct hidden variable $X$ using sampling actions?

1. Hidden variable model
2. Updated model after sampling result
3. Hidden variable reconstruction
4. Sampling action optimization

# Optimal sampling problem

## Updated model



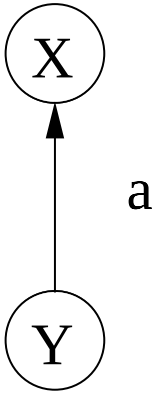

Posterior: $P(x|o,a) = \dfrac{P(o|x,a)P(x)}{P(o|a)}$

**Question**: How to reconstruct hidden variable $X$ using sampling actions?

1. Hidden variable model
2. Updated model after sampling result
3. Hidden variable reconstruction
4. Sampling action optimization

# Optimal sampling problem

## Hidden variable reconstruction



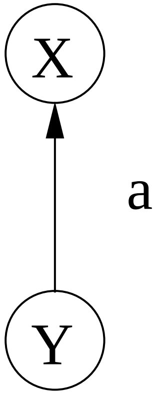

$$x^*(o,a) = \arg\max_x P(x|o,a)$$
$$V(o,a) = f(P(x^*|o,a))$$

**Question**: How to reconstruct hidden variable $X$ using sampling actions?

1. Hidden variable model
2. Updated model after sampling result
3. Hidden variable reconstruction
4. Sampling action optimization

# Optimal sampling problem

## Hidden variable reconstruction



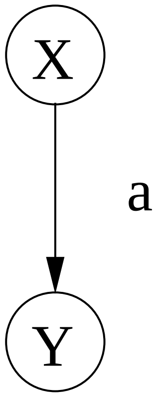

$$x^*(o,a) = \arg\max_x P(x|o,a)$$
$$V(o,a) = f(P(x^*|o,a))$$

**Question**: How to reconstruct hidden variable $X$ using sampling actions?

- $x^*(o,a)$ is the best reconstruction given sampling result $(o,a)$
- $V(o,a)$ is the value of reconstructed variable after sampling result $(o,a)$

# Optimal sampling problem

## Sampling action optimization




$$U(a) = -c(a) + \sum_o P(o|a)V(o,a)$$

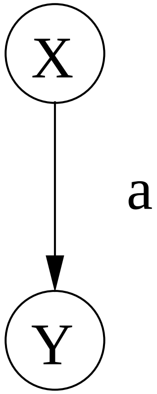
$$a^* = \arg\max_a U(a)$$

**Question**: How to reconstruct hidden variable $X$ using sampling actions?

1. Hidden variable model
2. Updated model after sampling result
3. Hidden variable reconstruction
4. Sampling action optimization

# Optimal sampling problem

## Sampling action optimization




$$U(a) = -c(a) + \sum_o P(o|a)V(o,a)$$
$$a^* = \arg\max_a U(a)$$

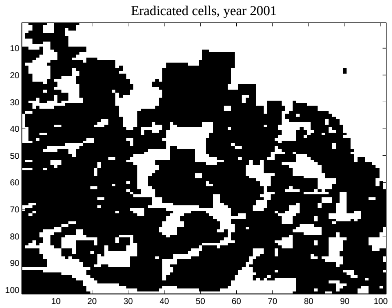
**Question**: How to reconstruct hidden variable $X$ using sampling actions?
The value of an action is a tradeoff between

- The cost $c(a)$ of the action and
- The expected quality of the reconstructed variable (over all possible sample results)

# Contents

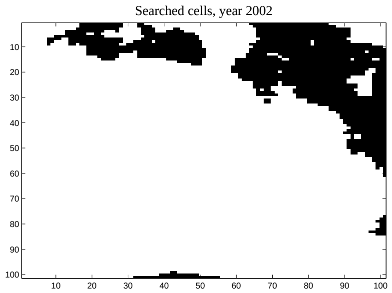
# Contents

# HMRF model for fire ants problem (1)



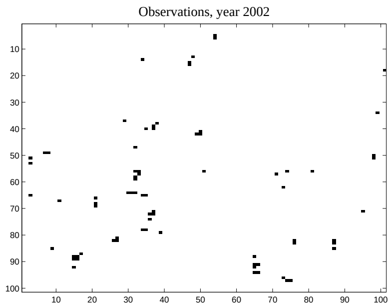

Eradication

$(e)$




Search actions

$(a)$




Observations

$(o)$

- eradication (at previous year): $e_i \in \{0, 1\},\ i = 1, \dots n$
- search actions: passive search or active search, $a_i \in \{0, 1\},\ i = 1, \dots n$
- observations: no nest detected / at least one nest detected, $o_i \in \{0, 1\},\ i = 1, \dots n$

# HMRF model for fire ants problem (2)

- Distribution on maps = Potts model

$$P_e(x \mid \alpha, \beta) = \frac{1}{Z} \exp\Big( \sum_{i \in V} \alpha_{e_i} \operatorname{eq}(x_i, 1) + \beta \sum_{(i,j) \in E} \operatorname{eq}(x_i, x_j) \Big)$$

- Distribution of observation given map, $P_{a_i}(o_i \mid x_i, \theta)$

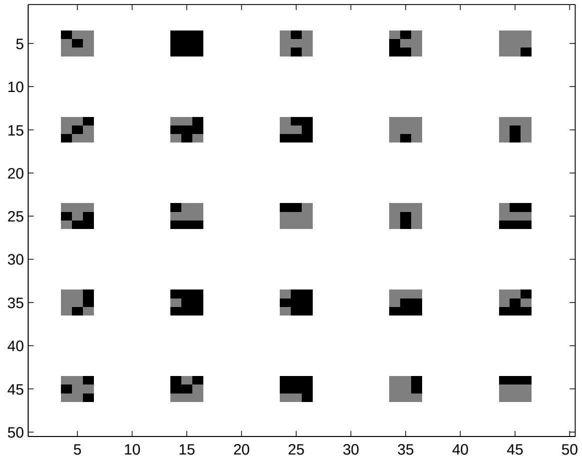
| $o_i \setminus x_i$ | 0 | 1 |
|---|---|---|
| 0 | 1 | $1 - \theta_{a_i}$ |
| 1 | 0 | $\theta_{a_i}$ |

with $\theta_0 < \theta_1$

# HMRF model for fire ants problem (3)

An initial arbitrary sampling $(a^0, o^0)$ is used for:

- Parameters estimation: $\lambda = (\alpha, \beta, \theta)$
  approximate version of EM for HMRF (Simul field EM)
  - identification problem between $\alpha$ and $\theta$
  - OK if $\theta$ known: use of expert values
- Marginals computation: $P_i(x_i | o_i^0, a_i^0)$



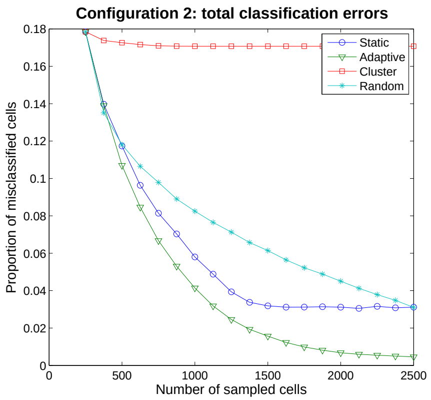
# Heuristic sampling methods evaluation (1)

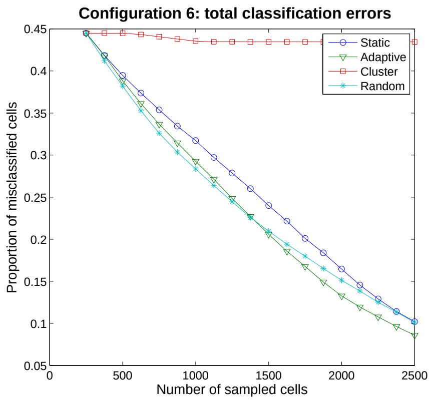
- Evaluation on simulated data
- Comparison of behavior of
  - random sampling (RS)
  - adaptive cluster sampling (ACS)
  - static heuristic sampling (SHS)
  - adaptive heuristic sampling (AHS)

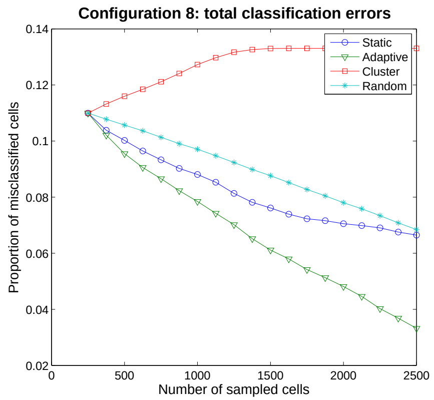
# Heuristic sampling methods evaluation (2)

- Procedure: repeat 10 times
  - simulate hidden map $x$ from $P(x \mid \alpha, \beta)$ ($50 \times 50$ cells)
  - apply regular sampling (about $10\%$ of area): $a^0$
  - simulate $o^0$ from $P_{a_i}(o_i \mid x_i, \theta)$ (regular sampling plus passive search)
  - estimate initial knowledge
  - apply RS, ACS, SHS, AHS, 10 times

# Rate of misclassified cells




$\alpha = (0, -2), \beta = 0.8 \quad \alpha = (0, 0), \beta = 0.5 \quad \alpha = (1 - 1), \beta = 0.4$

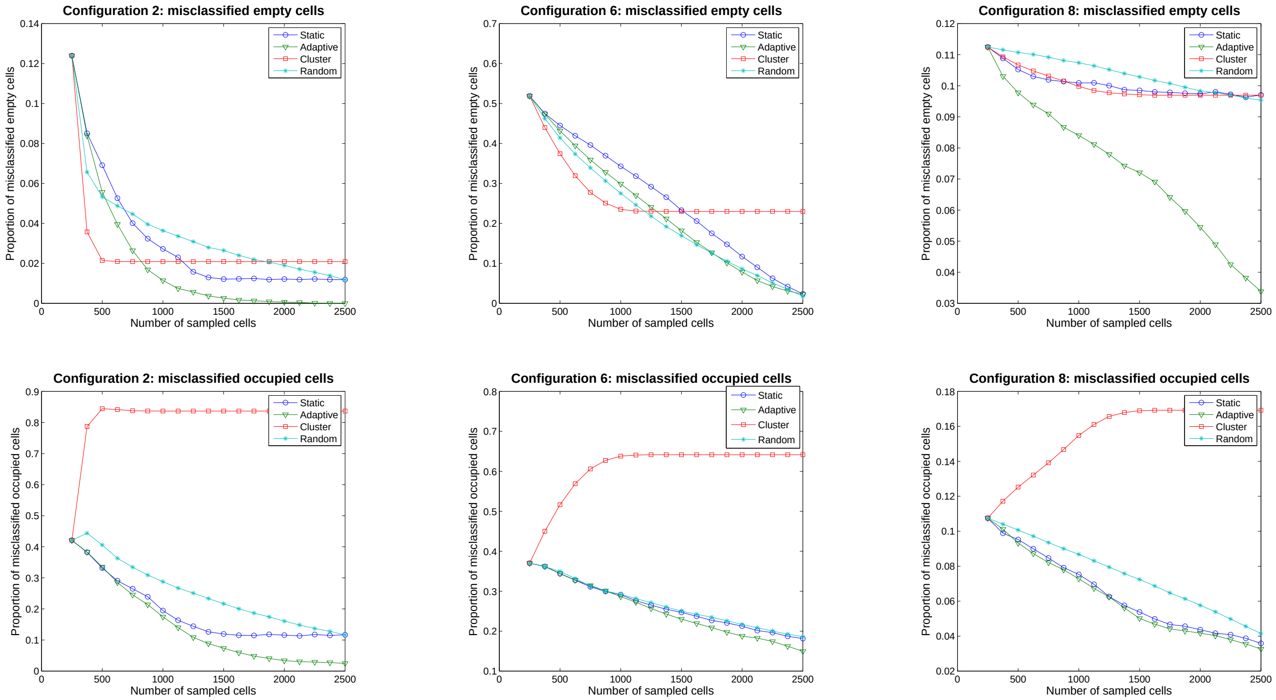
$$\theta = (0, 0.8)$$

legend: SHS AHA ACS RS

# Per color error rate

misclassified
empty cells

misclassified
occupied
cells




$$\alpha = (0, -2) \quad \alpha = (0, 0) \quad \alpha = (1 - 1)$$

$$\beta = 0.8 \quad \beta = 0.5 \quad \beta = 0.4$$

egend:

SHS AHA ACS

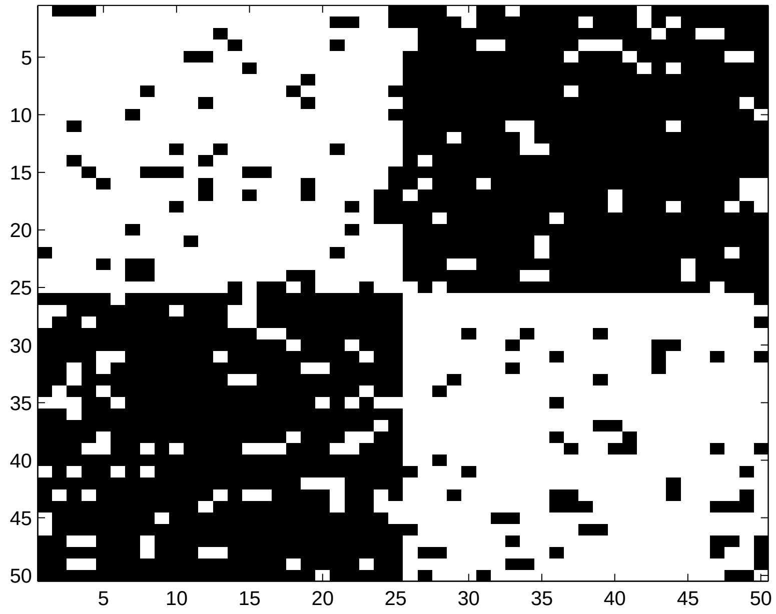
RS

# General behavior

- ACS is not adapted (as expected): poor results
- Adaptive HS $\geq$ Static HS $\geq$ Random S
- Discrepancy between Adaptive HS and Static HS increases with
  - sampling ressources
  - hidden map structure

# Where do we sample?

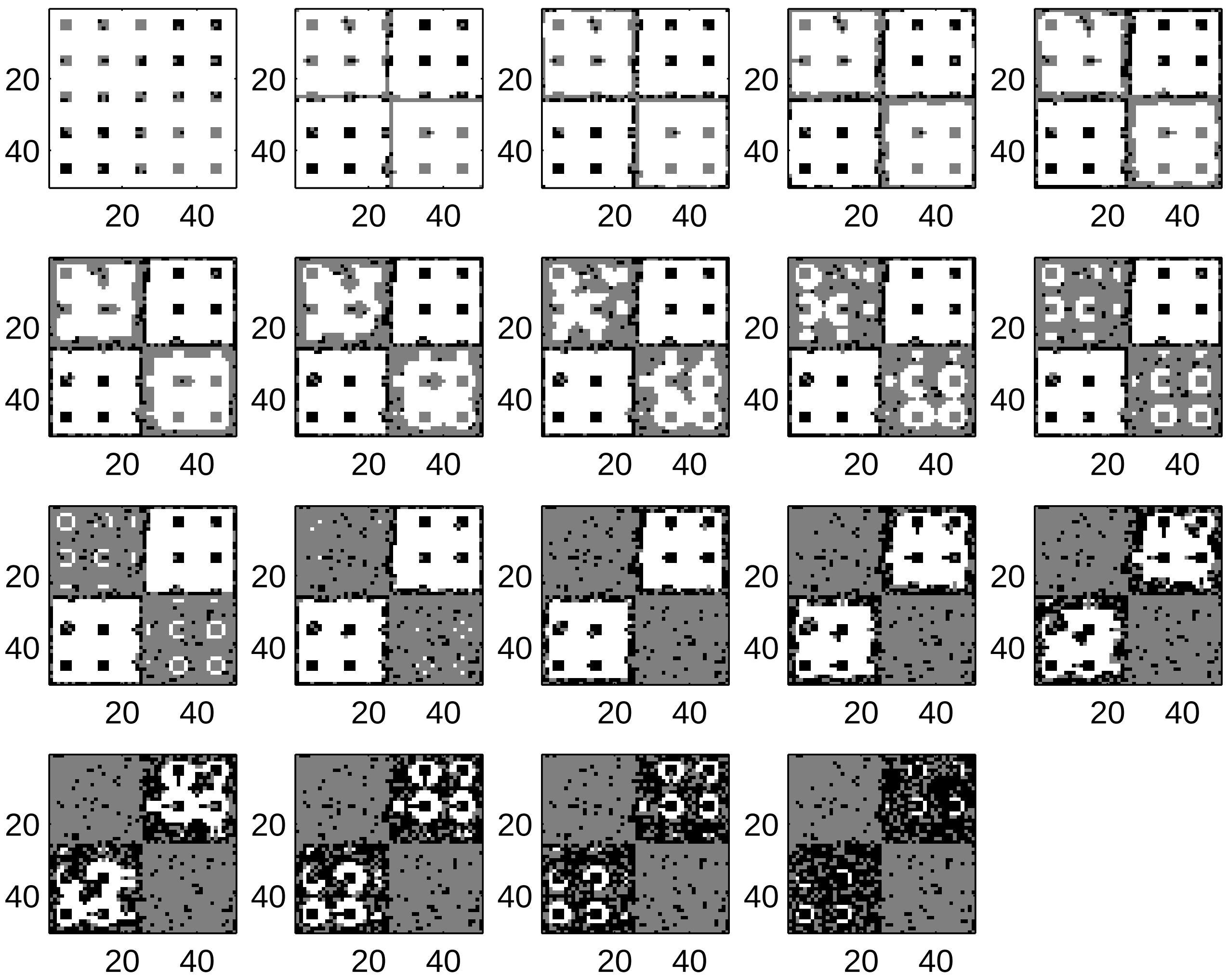
## Hidden map




$$\alpha = (1, -1), \beta = 0.4, \theta = (0, 0.8)$$

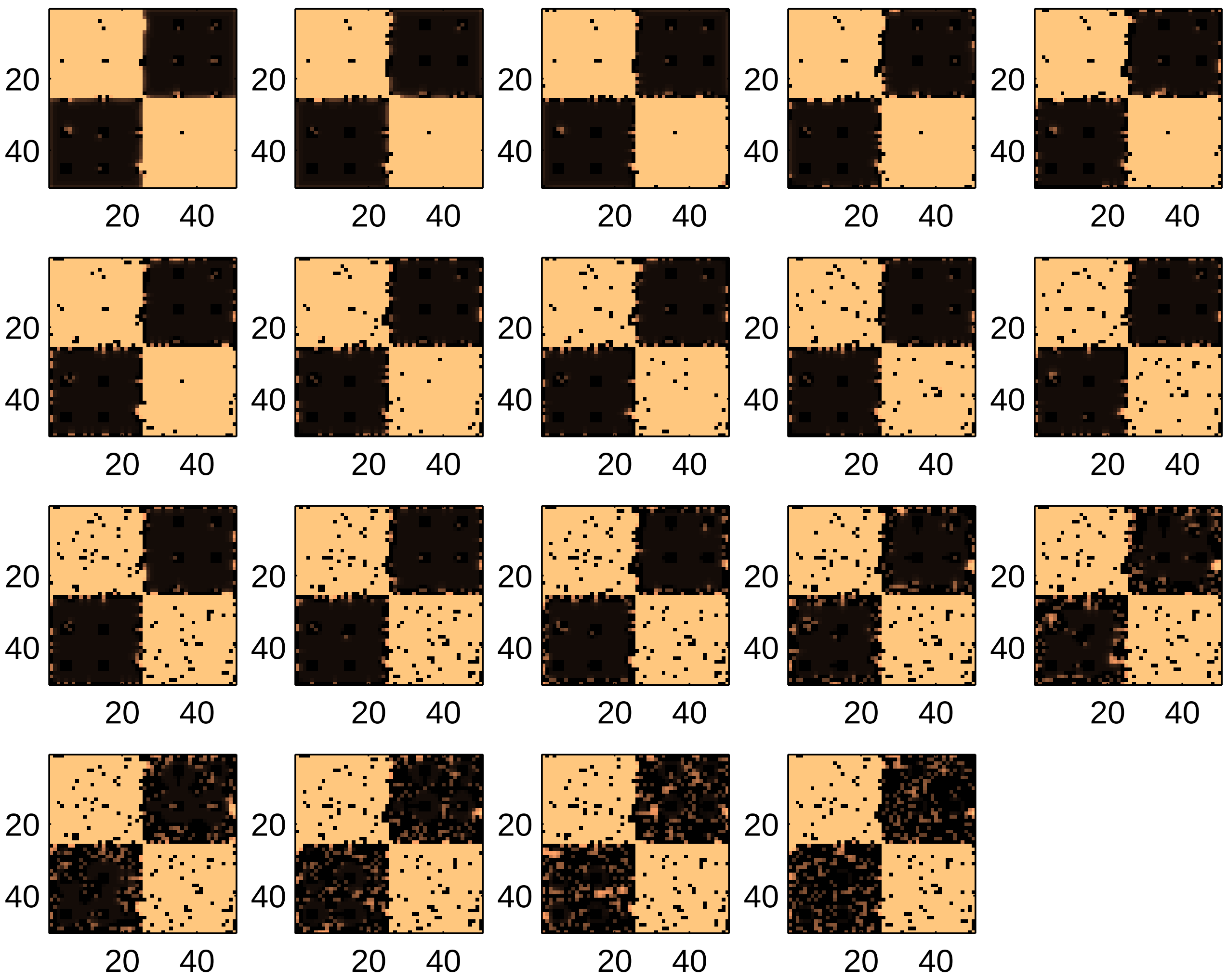
# Where do we sample?

Static sampling: A and O



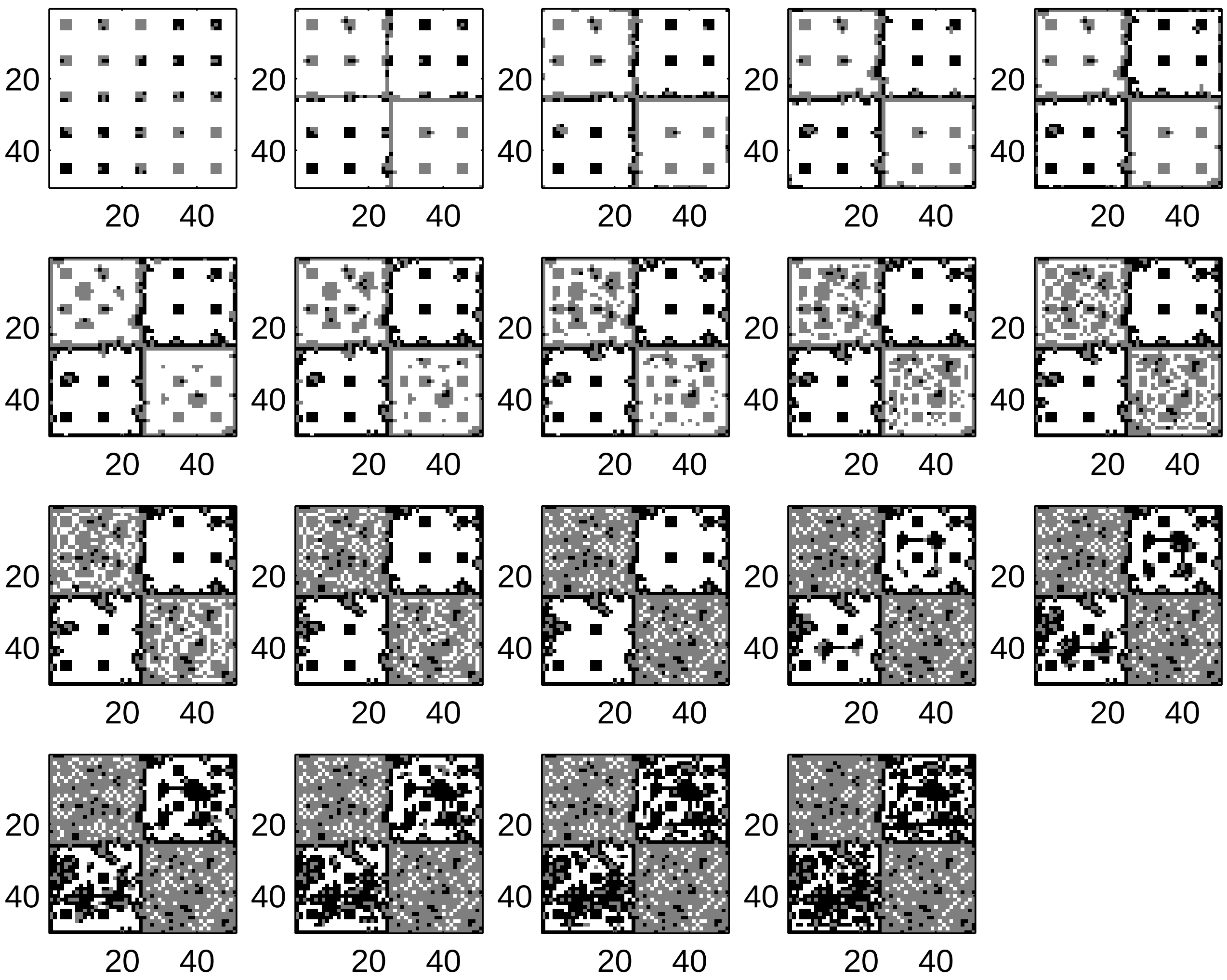
# Where do we sample?

Static sampling:marginals



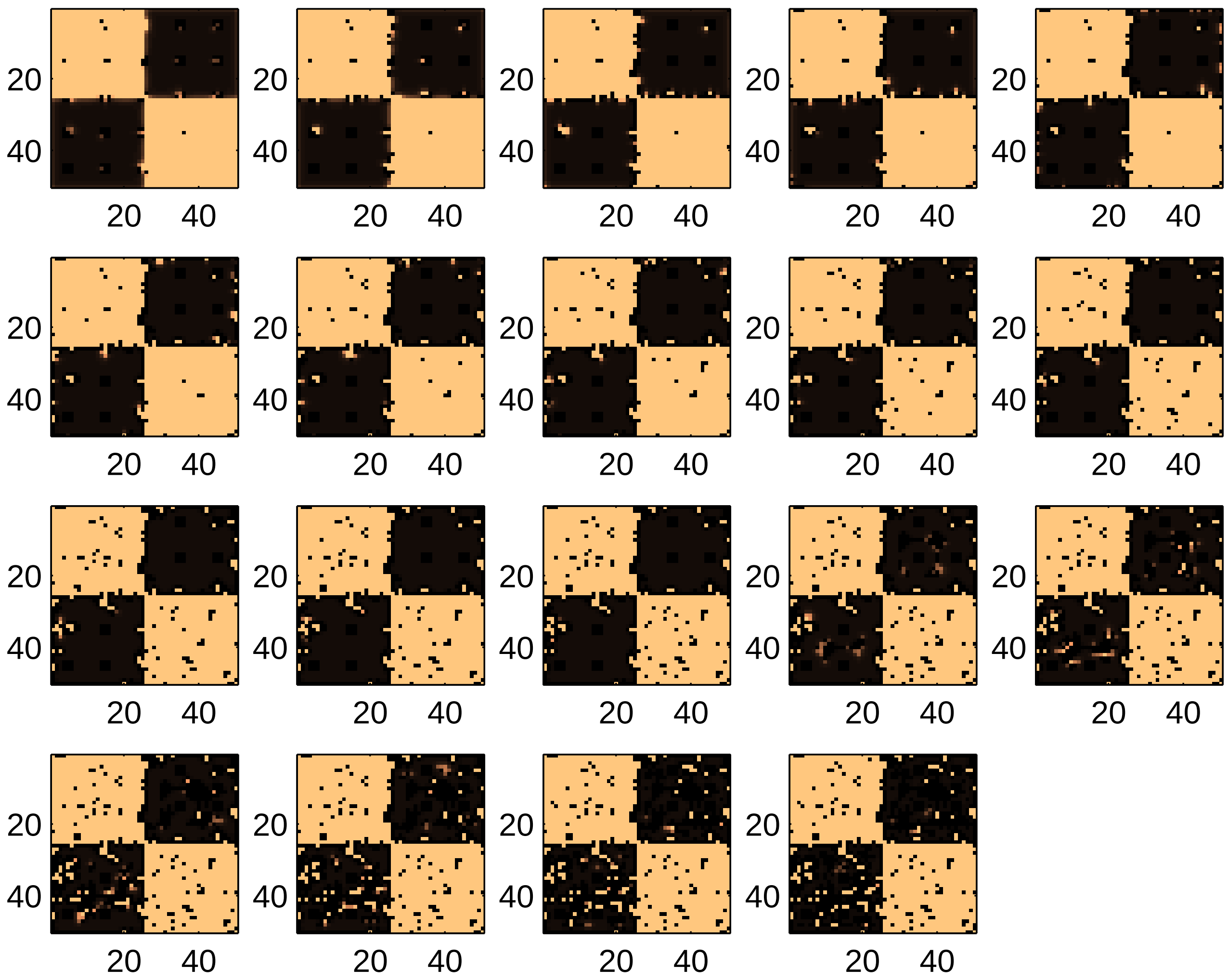
# Where do we sample?

Adaptive sampling: A and O (cumul)



# Where do we sample?

Adaptive sampling: marginals



# Where do we sample?

- No sampling in large empty areas
- Sampling preferably near detected occupied sites within low density areas
- If sampling ressources increase
  - SHS complete exploration until the whole area is covered
  - AHA can visit several times a site before extending exploration to another area