# *CATCHING GHOSTS*

# *WITH A COARSE NET:*

# REAL AND IMAGINARY EFFECTS

# IN ECOLOGICAL MONITORING ROUTINE

# BASED ON SPARSE SAMPLING

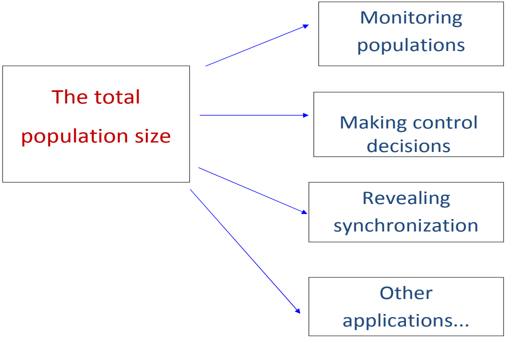
**Natalia Petrovskaya**

School of Mathematics,University of Birmingham, UK

# *Introduction*

# Introduction

- In many ecological applications knowledge of the total population size (population abundance) is essential
- The total population size is considered as important input information for a number of ecological problems



The total population size

Monitoring populations

Making control decisions

Revealing synchronization

Other applications...

# Sheep in the field

The total population size can be obtained by *direct counting*

# Aphids on leaves in the agricultural field

The total population size can be obtained by *evaluation* only

Since the population abundance is often used as input information *its inaccurate evaluation may result in wrong conclusions about the ecological problem!*

# Outline

- Evaluation of the total population size from sampling
  - Sampling procedure for a spatial population distribution
  - Accuracy of sampling results on coarse grids: deterministic vs. probabilistic approach
  - Dealing with a single realisation of a random variable

- Consequences of inaccurate evaluation: synchronization of population fluctuations
  - Absence of synchronization on coarse sampling grids
  - False synchronization on coarse sampling grids

- Conclusions

# *Sampling procedure for a spatial distribution*

# Sampling procedure

- At fixed time $t$ the population is confined within a certain physical domain.

- The population can be monitored and sampled with help of 'detectors' (e.g. traps, soil cores, nets,...) installed over the domain.

- We assume that sampling brings us reliable information about the population density $u(x, y)$.

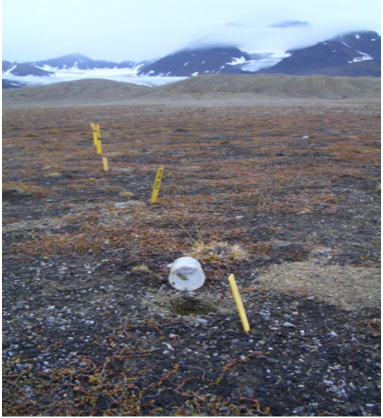
- *Given the population density values* $u(x_i, y_i)$ *at sampling locations* $(x_i, y_i)$*,* $i = 1 \ldots N$*, the total population size must be evaluated.*

# ‘Detectors’ used for sampling

*A slug trap in the wheat field*

*Bat detectors, Fruska Gora National Park.*
©The Rufford Foundation

# Examples of sampling grids



*A line of pitfall traps, Kongsfjord.* ©S.J. Coulson

*Experimental layout with flatworm traps.*

# Evaluation of the population size from discrete data

- The population size $I$ is given by $I = \iint\limits_D u(x,y)dxdy$.

- Values $u_i \equiv u(x_i, y_i)$, $i = 1, \ldots, N$ are given at the locations $(x_i, y_i)$ (*grid points*) only. Hence we have

$$I \;\approx\; I_a(N) = A\bar{u} = \frac{A}{N}\sum_{i=1}^{N} u_i,$$

  where $A$ is the area of the domain.

- The *evaluation error* is

$$e(N) = \frac{|I - I_a(N)|}{I}$$

- The *accuracy* requirement is $e \leq \tau$, where $\tau$ is the specified tolerance

# The coarse grid problem

- *Financial and labor resources available for sampling are often limited* ⇒ we consider a coarse sampling grid.

- A coarse sampling grid means the total number of samples taken in a given area is not large enough (or, similarly, the distance between the neighboring sampling locations is too large).

- Insufficient information about the population density $u(x, y)$ may result in inaccurate evaluation of the total population size.

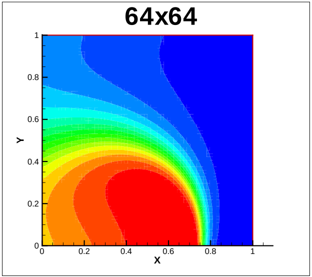
- That may, in turn, result in incorrect conclusion about the ecological problem.

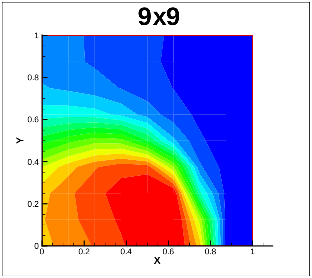
# *Evaluating the total population size on coarse sampling grids*

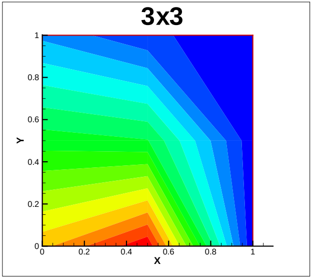
# Sampling on coarse grids

**Distribution 1 (D1)**:

regular sampling grids of 64*x*64, 9*x*9 and 3*x*3 nodes


64x64

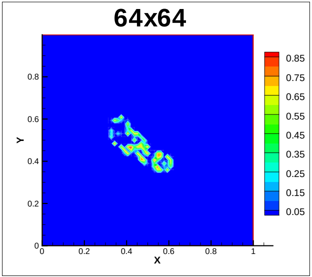

9x9

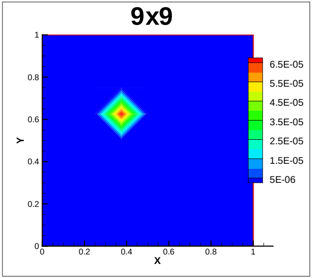

3x3

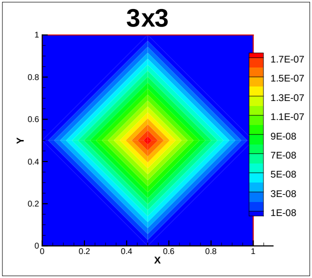
# Sampling on coarse grids

**Distribution 2 (D2)**:

regular sampling grids of 64*x*64, 9*x*9 and 3*x*3 nodes


64x64

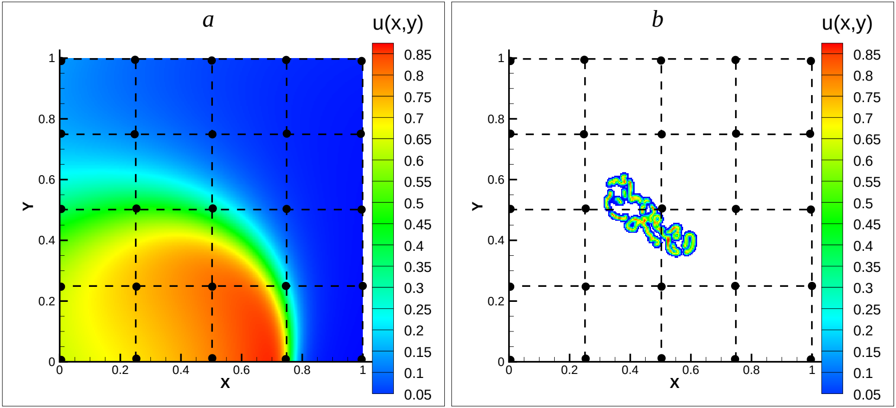

9x9


3x3

# Accuracy of evaluation on coarse grids

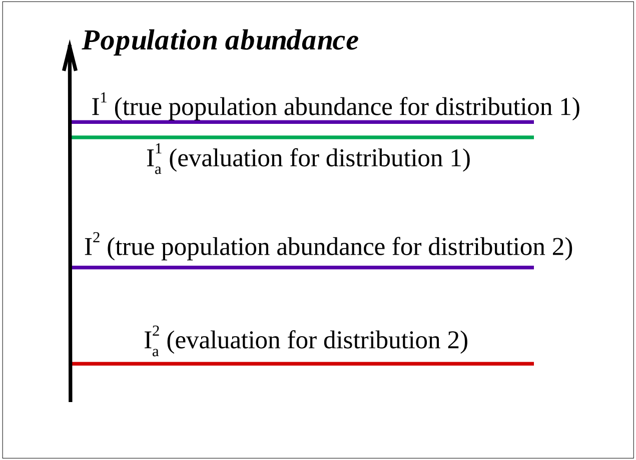
Sampling grid of $N = 5x5$ nodes:

**D1**: $e(N) = \dfrac{|I - I_a|}{I} \sim 10^{-2}$ **D2**: $e(N) \sim 1.0$



# Evaluation of the total population size

Comparison of the evaluation results for distributions D1 and D2



# Accuracy of evaluation on coarse grids

- On coarse grids we may have low/high accuracy of population abundance evaluation
- The whole approach seems to be deterministic (go to the field, install the grid, collect data, obtain evaluation $I_a$...)
- However...

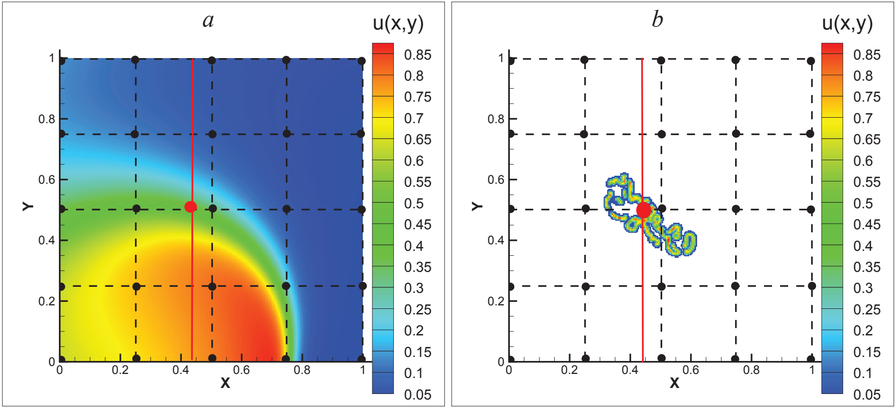
# Accuracy of evaluation on coarse grids

- On coarse grids we may have low/high accuracy of population abundance evaluation
- The whole approach seems to be deterministic (go to the field, install the grid, collect data, obtain evaluation $I_a$...)
- On a fixed grid of $N$ points the accuracy depends on a spatial pattern of the density distribution and on the location of each grid point

# Accuracy of evaluation on coarse grids

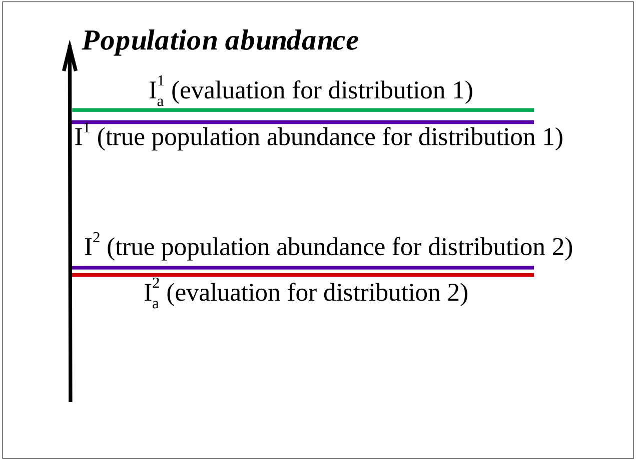
Moving grid nodes on the same sampling grid:

**D1**: $e(N) = \dfrac{|I - I_a|}{I} \sim 10^{-2}$ **D2**: $e(N) \sim 10^{-3}$



# Evaluation of the total population size

Comparison of the evaluation results for distributions D1 and D2 on a shifted grid



# Accuracy of evaluation on coarse grids

- On coarse grids we may have low/high accuracy of population abundance evaluation

- On a fixed grid of $N$ points the accuracy depends on a spatial pattern of the density distribution and on the location of each grid point

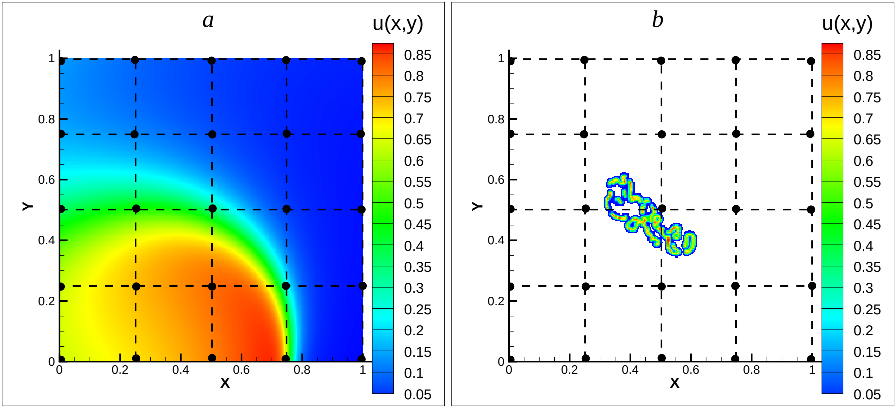
- Hence the evaluation result $I_a$ can be considered as a single realisation of random variable $\tilde{I}_a \in [I_{min}, I_{max}]$

- *This uncertainty in evaluation appears because we do not have sufficient information about the spatial density distribution* $u(x, y)$ *(this is NOT because of measurement errors in the problem!)*

# Accuracy of evaluation on coarse grids

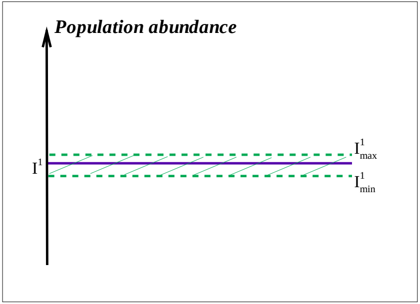
Sampling grid of $N = 5x5$ nodes:

**D1**: $e(N) = \dfrac{|I - I_a|}{I} \sim 10^{-2}$ **D2**: $e(N) \sim 1.0$



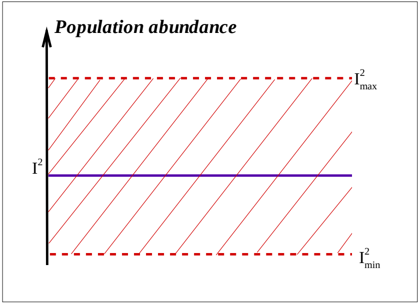
# The evaluation range for density distribution D1

- all realisations of random variable $\tilde{I}_a$ are close to the true population size $I$ (i.e. the mean is close to $I$, the variance is small)
- any single realisation $I_a$ of random variable $\tilde{I}_a$ can be considered as a good approximation to the true population abundance
- all realisations of random variable $\tilde{I}_a$ can be found with almost equal probability



# The evaluation range for density distribution D2

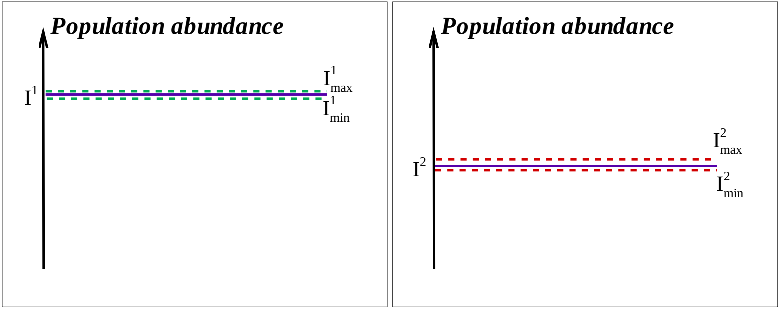
- the range $[I_{min}, I_{max}]$ of random variable $\tilde{I}_a$ is big (i.e. the mean is not close to $I$, the variance is big)
- only some realisations $I_a$ of random variable $\tilde{I}_a$ can be considered as a good approximation to the true population abundance
- the p.d.f. function of random variable $\tilde{I}_a$ is not constant (i.e. 'inaccurate' realisations $I_a$ are more likely to be seen)



# "Keep calm and refine the grid!"

In our evaluation procedure

- we have to deal with a single realisation of random variable $\tilde{I}_a$
- in many cases we do not have any information about the mean $< \tilde{I}_a >$ and variance $\sigma$ of $\tilde{I}_a$
- however, we do know that $< \tilde{I}_a > \rightarrow I$ and $\sigma \rightarrow 0$ on fine grids



## *Dealing with a single realisation of the total population size*

# The evaluation error on coarse grids

The evaluation error $e(N) = |I - I_a|/I$ is a random variable

Any single realisation of $e(N)$ depends on the number $N$ of grid points, a spatial pattern of the density distribution, and a location of each grid point

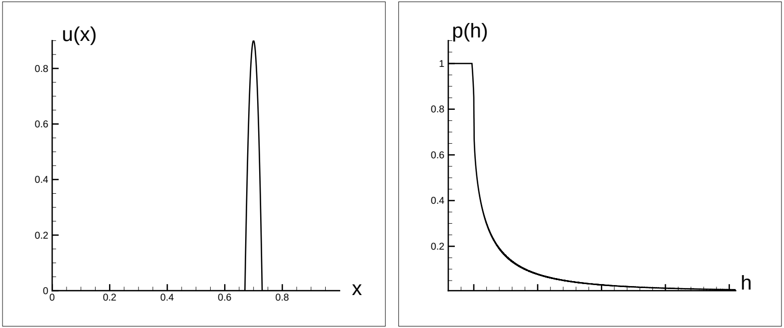
Instead of considering the mean and the variance it is more convenient to formulate the following problem:

- What is the probability $p$ of the event $e \leq \tau$?
- What is is the number $N^*$ of grid points to provide the probability $p(e \leq \tau) = 1$?

# Highly aggregated density: Example I

$$u(x) = -A(x - x^*)^2 + B$$

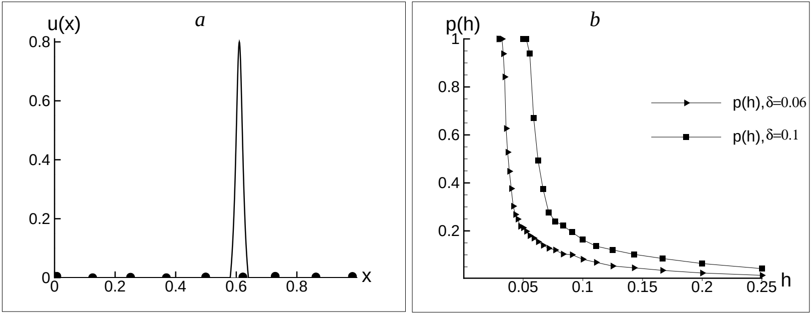
For fixed $A$, $B$ and $\tau$, find the threshold grid step size $h^*$ (and hence the number of nodes $N^*$) required to evaluate the total population size with the error $e \le \tau$ wherever the peak is located.



# Highly aggregated density: Example II

$$u(x) = \begin{cases} \dfrac{\delta^2}{4}\dfrac{1}{4(x-x^*)^2+\delta^2/4}, & x \in [x^* - \delta/2, x^* + \delta/2], \\ 0, & \textit{otherwise}. \end{cases}$$

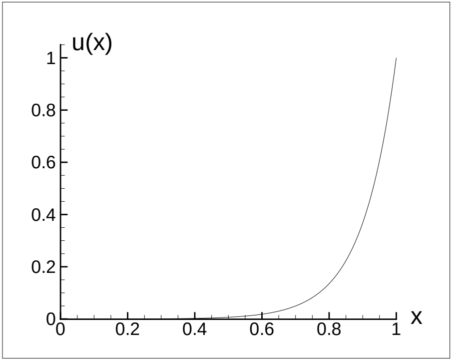
For fixed tolerance $\tau = 0.25$, find the probability of $e \le \tau$ for peak width $\delta = 0.06$ and $\delta = 0.1$.

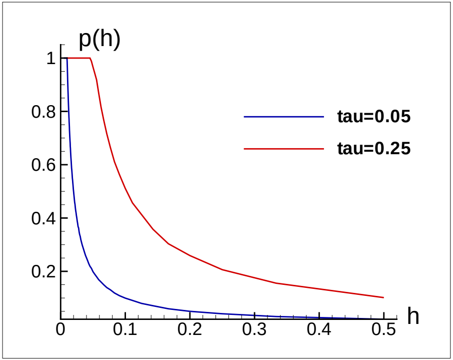


# Highly aggregated density: Example III

$$u(x) = \frac{e^{x/\mu} - 1}{e^{1/\mu} - 1}, \qquad x \in [0, 1]$$

For fixed $\mu = 0.1$, find the probability of $e \leq \tau$ for the tolerance $\tau = 0.25$ and $\tau = 0.05$.





# Conclusions about sampling on coarse grids

- Standard methods of evaluation are not always reliable on coarse grids where the number of samples is small. The results of evaluation should be explained from a probabilistic viewpoint.
- In some cases (but not always!) we can think of a single realisation of the random variable obtained in the sampling procedure as of a deterministic variable. Otherwise the probability of accurate evaluation has to be computed.
- The definition of ‘grid coarseness’ depends on a spatial pattern of the density function.

# *Revealing synchronization from sparse data*

# The phenomenon of synchronization

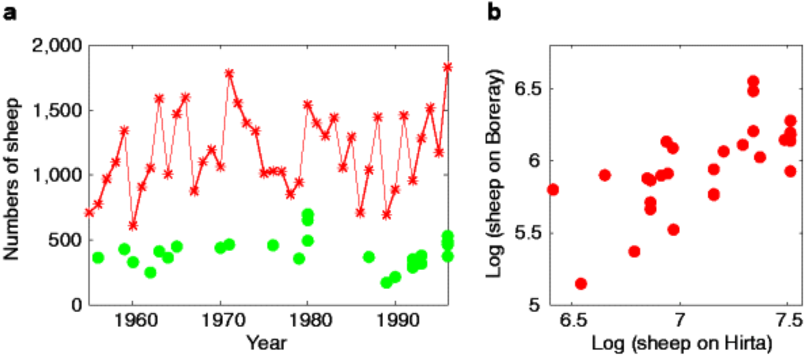
- In many cases, the population fluctuations in different habitats appear to be correlated.

- This phenomenon is known as synchronization.

# Example of synchronization

Sheep populations on two islands in the St Kilda archipelago, Scotland

B.T. Grenfell et al. Nature 394, 674–677.



# Biological studies of synchronization

- Synchronization was observed between densities of white-footed mouse *Peromyscus leucopus* collected at eight sites in western Massachusetts, USA.

  J.S. Elkinton, W.M. Healy, J.P. Buonaccorsi, G.H. Boettner, A.M. Hazzard, and H.R. Smith 1996 Ecology 77:2332–2342.

- Populations of caribou and musk oxen on opposite coasts of Greenland showed the high degree of cross-species synchronization (being separated by a minimum of $1,000$ km of inland ice).

  E. Post and M.C. Forchhammer 2002 Nature 420: 168–171

- More examples of synchronization along with discussion of the mechanisms of synchronisation can be found in the literature...

  A. Liebhold, W.D. Koenig, and O.N. Bjornstad. 2004 Annu. Rev. Ecol. Evolut. Sys., 35:467–490.

  T.S. Rosenstock, A. Hastings, W.D. Koenig, D.J. Lyles, and P.H. Brown. 2011 Oikos, 120:1434–1440.

  E. Ranta, V. Kaitala, K. Lindstrom, E. Helle. 1997 Oikos, 78:136–42.

# Revealing synchronization between two habitats

- Consider two different habitats labeled as field $A$ and field $B$.
- Introduce scalar characteristics $I^A$ and $I^B$ of field $A$ and field $B$ respectively (in most cases $I$ is the total population size).
- Evaluate $I^A$ and $I^B$ at times $t_1, t_2, \ldots, t_M$ to generate two time series $S_A = \{I_1^A, I_2^A, \ldots, I_M^A\}$ and $S_B = \{I_1^B, I_2^B, \ldots, I_M^B\}$.
- The degree of synchronization between the two populations is usually quantified by the Pearson correlation coefficient between the time series $S_A$ and $S_B$.

# Correlation coefficient as a measure of synchronization

- The total population size in the domain of ‘size’ $L$ is

$$I \approx \frac{L}{N} \sum_{n=1}^{N} u(\mathbf{r}_n, t).$$

- The degree of synchronization between the two populations is quantified by the Pearson correlation coefficient:

$$\rho_{AB} = \frac{\sum_{i=1}^{M} \left(I_A^{(i)} - \mu_A\right)\left(I_B^{(i)} - \mu_B\right)}{\sqrt{\left(\sum_{i=1}^{M} \left(I_A^{(i)} - \mu_A\right)^2\right)\left(\sum_{i=1}^{M} \left(I_B^{(i)} - \mu_B\right)^2\right)}} ,$$

  where $\mu_A$ and $\mu_B$ are the sample means:

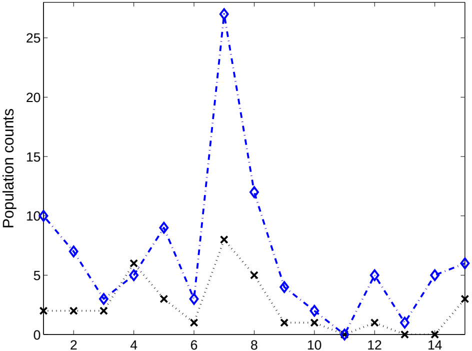
$$\mu_A = \frac{1}{M} \sum_{i=1}^{M} I_A^{(i)} , \qquad \mu_B = \frac{1}{M} \sum_{i=1}^{M} I_B^{(i)} .$$

# Example: synchronization of *Tipula paludosa* in agricultural landscape

The metapopulation of crane fly (*Tipula paludosa*) was monitored for 15 years in 38 agricultural fields across south-west Scotland.

D. Bearup, S.V.Petrovskii, R. Blackshaw, A.Hastings 2013 Am. Nat. 182:393–409.

The correlation coefficient is $\rho = 0.88$



# Implications of synchronization

- Synchronized fluctuations in abundance among spatially segregated populations can be used to quantify and understand the distribution of abundance in space and time.

- Synchronization has many implications across the whole range of ecological sciences,in particular, for agro-ecology and pest control.

- Inaccurate evaluation of $I^A$ and $I^B$ may affect the conclusion about synchronization between habitats $A$ and $B$.

- Our aim is to investigate how conclusions about synchronization work when populations with complex spatial distributions are monitored on realistically coarse grids.

# *True and false synchronization between sampling data on coarse grids*

# Numerical simulation of the population dynamics

- We simulate two sequences of spatial distributions where the total population size is a certain known function of time and space,

$$u_j(x,t) = \omega_j(t) F_j(x), \quad j = A, B.$$

- We consider $\omega_A(t)$ and $\omega_B(t)$ for populations in field $A$ and field $B$ respectively and require $\omega_A(t) = \omega_B(t) \equiv \omega(t)$.
- The population dynamics in $A$ and $B$ is perfectly synchronized and $\rho_{AB} = 1$.

# Numerical simulation of the population dynamics

- We consider the population density being a superposition of normal distributions:

$$u_j(x,t) = \frac{\omega(t)}{p} \sum_{m=1}^{p} f_j^{(m)}(x), \quad j = A, B,$$

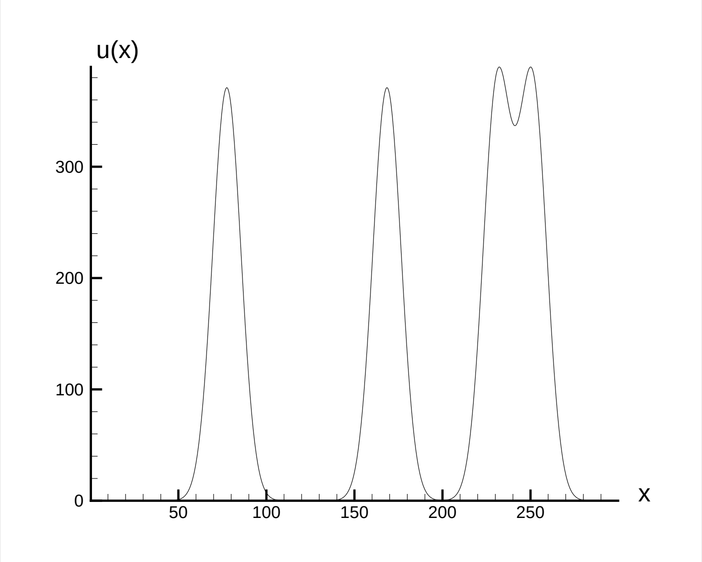
where

$$f_j^{(m)}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[ -\frac{\left(x - \hat{x}_j^{(m)}\right)^2}{2\sigma^2} \right],$$

and the location $\hat{x}_j^{(m)}$ of every peak is a random variable.

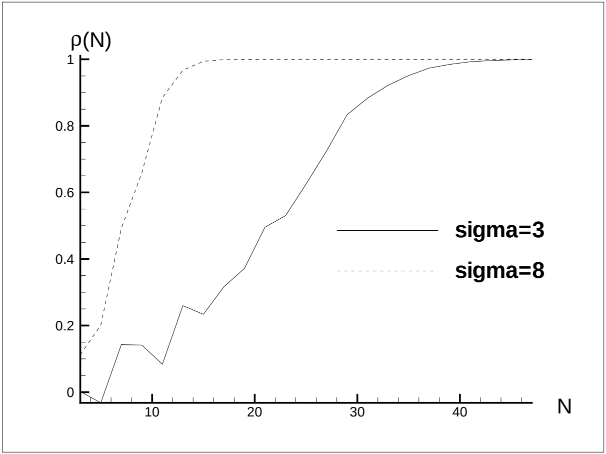
- The function $\omega(t)$ is $\omega(t) = Ct$, where $C = const$.

# Superposition of several peaks at fixed time *t*



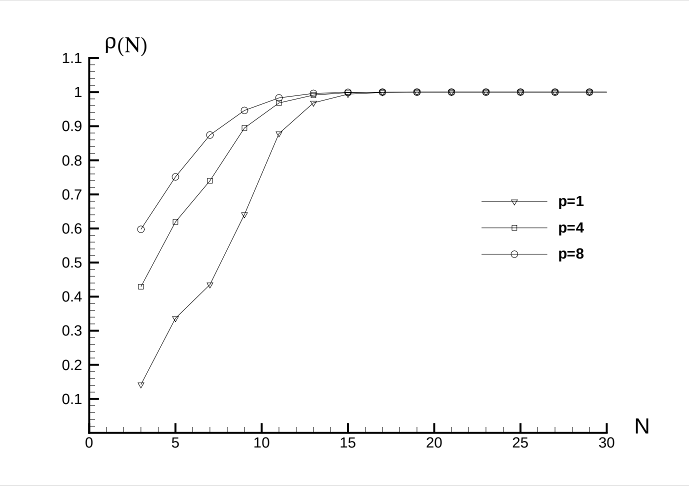
# Synchronization on coarse grids

The study of a single peak ($p = 1$). The correlation coefficient for various values of $\sigma$.



# Synchronization on coarse grids

The correlation coefficient $\rho(N)$ for different number of peaks *p* in the population density distribution ($\sigma = 8$).

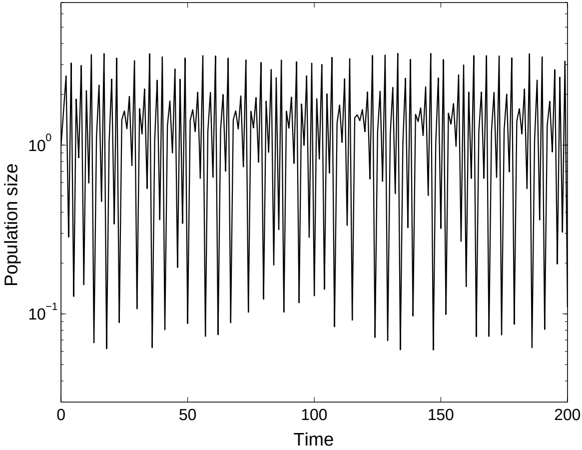


# Synchronization on coarse grids

The Ricker map:

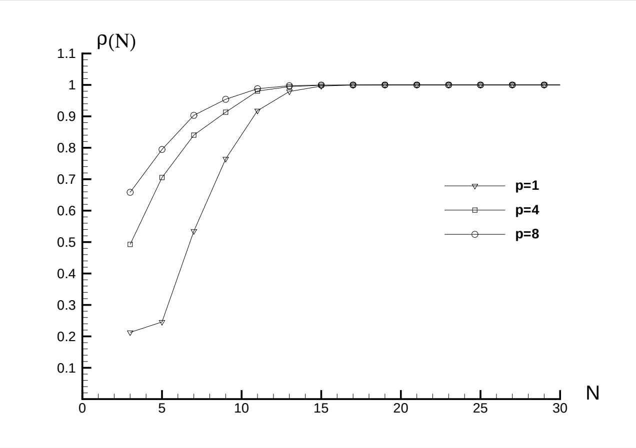
$$\omega(t+1) = \omega(t) r e^{-\alpha\omega(t)},$$

where $r$ and $\alpha$ are parameters.



# Synchronization on coarse grids

The correlation coefficient for the Ricker model.



# False synchronization on coarse grids

Sampling on coarse grids may also result in false synchronization:

– the correlation coefficient is $\rho \approx 1$ whilst the dynamics is actually not correlated.

# False synchronization on coarse grids

- Consider a hypothetical system consisting of two $1-D$ domains $A$ and $B$ where the population of a certain species has a unimodal spatial distribution with the maximum at the domain boundary:

$$u_j(x,t) = \frac{\omega_j(t)}{\sigma_j(t)\sqrt{2\pi}} \exp\left[-\frac{x^2}{2\sigma_j(t)^2}\right], \quad 0 \le x \le L_j, \qquad j = A, B.$$

- Consider the case where the populations are affected by factors different in the two domains, so that the population size in domain $A$ and domain $B$ evolves differently with time but preserving the shape of the spatial distribution.

# False synchronization on coarse grids

- The population of a certain species has a unimodal spatial distribution with the maximum at the domain boundary:

$$u_j(x,t) = \frac{\omega_j(t)}{\sigma_j(t)\sqrt{2\pi}} \exp\left[-\frac{x^2}{2\sigma_j(t)^2}\right], \quad 0 \le x \le L_j, \qquad j = A, B.$$

- We consider the following hypothetical situation:

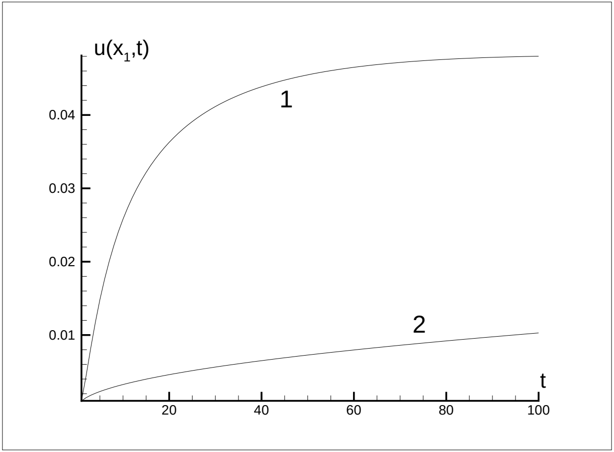
$$\omega_A(t) = \omega_0 t^a, \quad \sigma_A(t) = \sigma_0, \qquad \omega_B(t) = \omega_0 - \gamma t, \quad \sigma_B(t) = \sigma_0 t^b,$$

  where $a, b, \gamma, \omega_0$ and $\sigma_0$ are parameters.

- We take $L_A = L_B = 300$, $a = 0.5$, $b = 0.25$, $\gamma = 0.0001$, $\omega_0 = 2$ and $\sigma_0 = 3$.

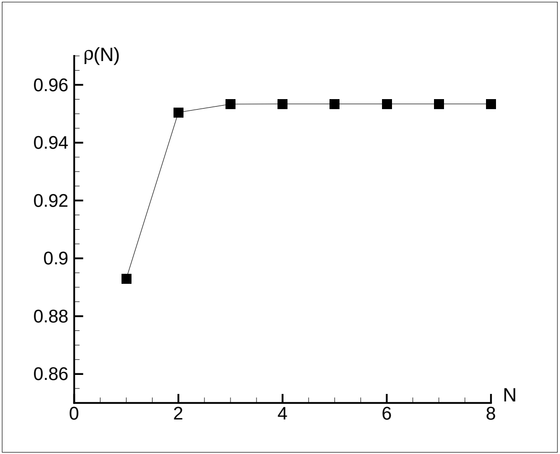
# False synchronization on coarse grids

Population density at the fixed sampling location $x_1$ in domain *A* (curve 1) and domain *B* (curve 2).



# False synchronization on coarse grids

The correlation coefficient on a sequence of coarse grids



# Conclusions

- Identification of synchronization relies heavily on the accuracy of the population size estimation. That, in its turn, strongly depends on the number of samples in the census.

- Standard methods of evaluation are not reliable on coarse grids where the number of samples is small. The results of evaluation should be explained from a probabilistic viewpoint when data used for evaluation are sparse. This must be taken into account when synchronization is investigated.

- Conclusions about synchronization are likely to be wrong if a patchy population distribution is monitored on coarse grids.

- The knowledge of spatial pattern is crucial and any information about it must be used to its fullest extent.

# References

N.B.Petrovskaya, S.V.Petrovskii. *Catching Ghosts with a Coarse Net: Use and Abuse of Spatial Sampling Data in Detecting Synchronization* submitted, 2016

N.B.Petrovskaya, N.L.Embleton. *Computational Methods for Accurate Evaluation of Pest Insect Population Size.* In: Ecological Modelling Applied to Entomology, W.A.C. Godoy and C.P. Ferreira (eds.), Springer-Verlag Berlin Heidelberg, 2014, pp.171-218

S.V.Petrovskii, N.B.Petrovskaya, D.Bearup. *Multiscale Approach to Pest Insect Monitoring: Random Walks, Pattern Formation, Synchronization, and Networks.* Physics of Life Reviews, 2014, vol.11, pp.536-539

N.L.Embleton, N.B.Petrovskaya. *A Novel Approach to Evaluation of Pest Insect Abundance in the Presence of Noise.* Bulletin of Mathematical Biology, 2014, vol.76, pp. 718-743

N.B.Petrovskaya, N.L.Embleton. *Evaluation of Peak Functions on Ultra-Coarse Grids.* Proc. R. Soc. A, 2013, vol.469, doi: 10.1098/rspa.2012.0665

N.B.Petrovskaya, E.Venturino. *Numerical Integration of Sparsely Sampled Data.* Simulat. Modell. Pract. Theory, 2011, vol.19, pp.1860–1872

N.B.Petrovskaya, S. V. Petrovskii. *The Coarse-Grid Problem in Ecological Monitoring.* Proc. R. Soc. A, 2010, vol.466, pp. 2933-2953