HANDLING UNCERTAIN SPATIAL DATA IN MONITORING AND CONTROL PROBLEMS

John Ellis, **Natalia Petrovskaya**, Wenxin Zhang

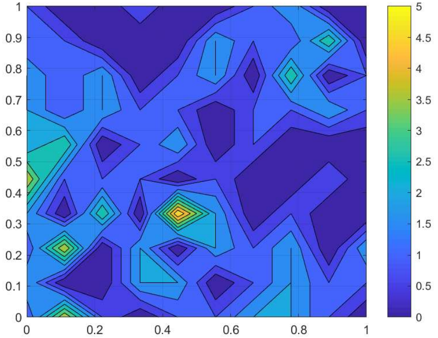
School of Mathematics, University of Birmingham, UK

Outline

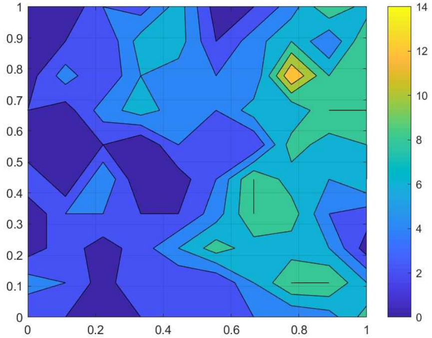
- Introduction: processing spatial data in ecological applications.
  - Monitoring and control (M&C) procedure as a data filter.
- Spatial data with measurement error.
  - Can data filters deal with measurement error?
  - Case study: M&C in the problem of biological invasion.
  - Inherent uncertainty in the model.
- Conclusions

SPATIAL DATA PROCESSING

Examples of spatial distributions



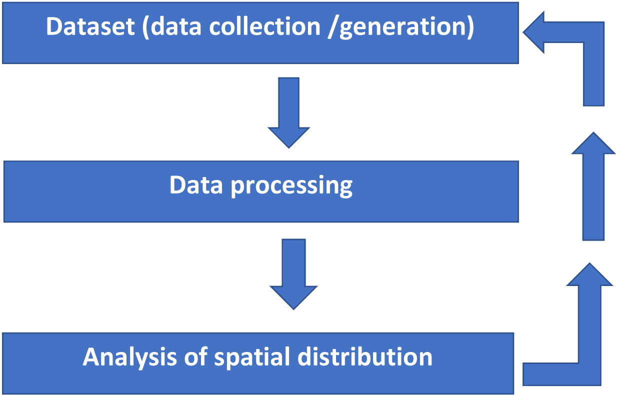
(a) Grey field slug
Deroceras reticulatum
(data collection: E.Forbes)



(b) New Zealand flatworm
Arthurdendyus triangulatus
(data collection: A.Murchie & A.Harrison)

N.B.Petrovskaya et al. (2018) Towards the development of a more accurate monitoring procedure for invertebrate populations, in the presence of an unknown spatial pattern of population distribution in the field. Insects, 9:29

Reconstruction of spatial distributions



Example of spatial data: grey field slug

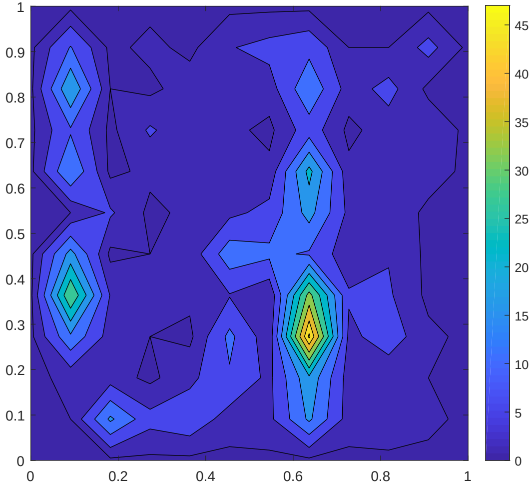
Data on the grey field slug *Deroceras reticulatum*:

11	0	2	0	5	7	9	1	1
21	1	0	3	1	3	15	3	7
10	0	6	2	2	0	8	0	3
16	0	2	2	1	2	23	2	5
1	6	0	2	4	7	18	4	4
18	0	1	2	14	12	10	2	5
32	5	3	2	5	1	34	6	6
15	3	1	0	12	2	48	4	8
2	3	0	4	10	4	20	3	2
1	17	7	9	3	4	17	3	4

Trap counts collected from a sampling grid in a farm field at Shropshire, UK.

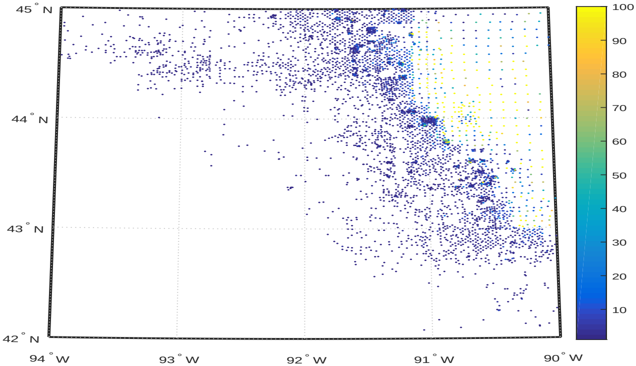
by courtesy of Emily Forbes

Spatial distribution of the grey field slug



Continuous spatial distribution reconstructed from discrete data

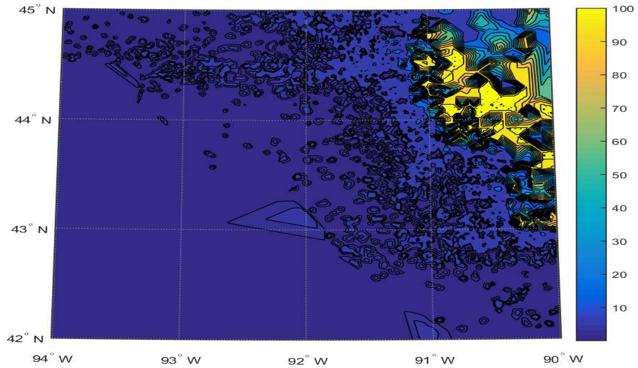
Example of spatial data: gypsy moth *Lymantria dispar*



(*by courtesy of Andrew Liebhold*)

- Data collection: $\sim$ 80000 traps across the USA territory
- More sophisticated data processing algorithm is required: data have been fragmented

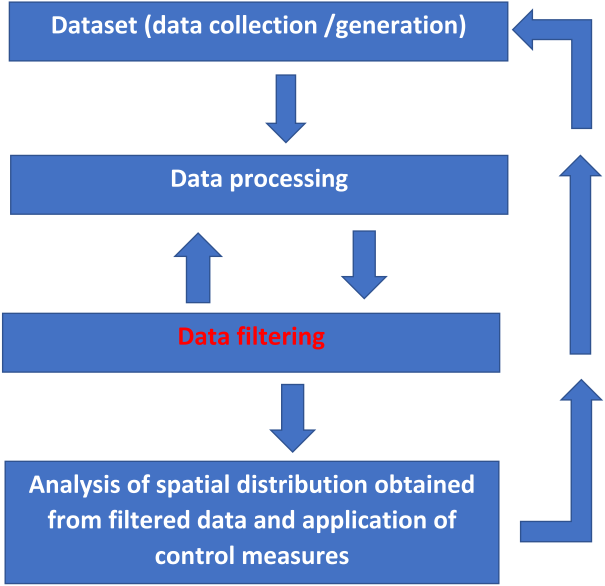
Handling spatial data



Spatial distribution of the gypsy moth in a selected geographic region

P.C. Tobin, L.M. Blackburn, eds. (2007) Slow the Spread: a national program to manage the gypsy moth. Gen. Tech. Rep. NRS-6. PA: U.S.D.A.

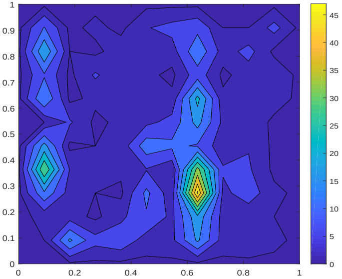
Handling spatial data in M&C routine



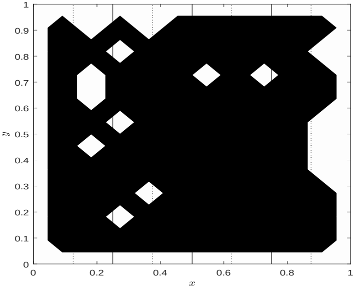
Example of data filtering: targeted use of pesticide

- Original dataset: trap counts C_i collected in monitoring of the grey field slug population
- Filtered dataset: presence/absence data $\hat{C}_i$ generated from the dataset C_i

$$\hat{C}_i = 1 \quad \text{for} \quad C_i > 0, \qquad \hat{C}_i = 0 \quad \text{for} \quad C_i = 0, \; i = 1, \ldots, N$$



Original dataset

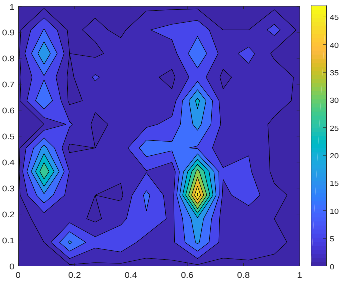


Filtered dataset

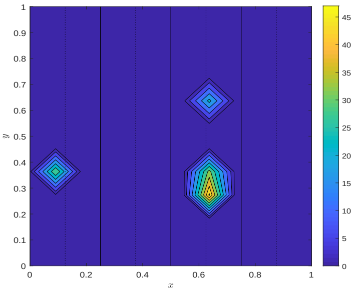
Example of data filtering: targeted use of pesticide

An alternative filter definition: identification of slug patches with the high population density within the patch

$$\hat{C}_j = C_j \quad \text{if} \quad S(N_j) = \frac{1}{N_j}\sum_{j=1}^{N} C_j > S_{th},$$
$$\hat{C}_j = 0 \quad \text{otherwise}$$



Original dataset



Filtered dataset

SPATIAL DATA WITH MEASUREMENT ERRORS

Random measurement errors in spatial data

Corrupted data on the grey field slug *Deroceras reticulatum*:

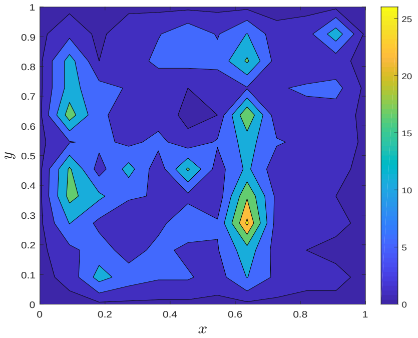
0	3	1	1	4	0	0	1	1
4	1	0	2	1	0	0	0	0
2	1	0	2	1	0	1	0	2
1	3	0	3	2	4	1	1	1
0	2	1	0	3	0	3	0	0
2	3	3	1	3	1	1	0	2
0	2	3	0	1	1	0	1	1
0	3	2	3	3	1	2	0	1
2	2	4		0	2	0	0	2
3	0	1	0	2	2	1	0	0

(*by courtesy of Emily Forbes*)

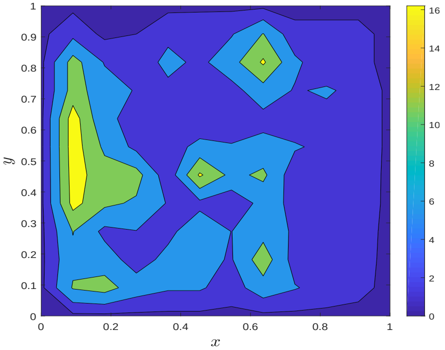
Joint Committee for Guides in Metrology (2008) Evaluation of measurement data. Guide to the expression of uncertainty in measurement, JCGM 100:2008, Paris

Systematic measurement errors in spatial data

Spatial distribution of the grey field slug reconstructed from trap counts collected on various sampling grids



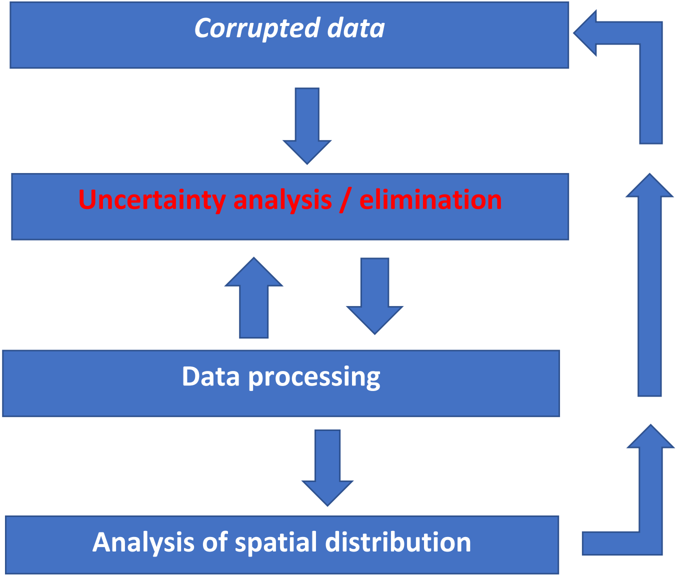
Trap counts from 100 traps



Trap counts from 50 traps

Similar results in: D. Bohan et al. (2000) Parametric intensity and spatial arrangement of the terrestrial mollusc herbivores Deroceras reticulatum and Arion intermedius. Journal of Animal Ecology 69:1031-1046

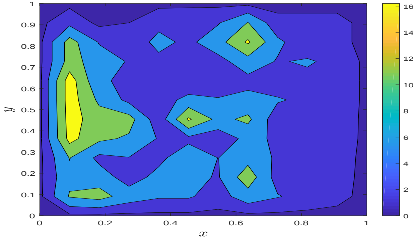
Handling measurement errors in spatial data



Handling measurement errors in the M&C routine

0	3	1	1	4	0	0	1	1	0
4	1	0	2	1	0	0	0	0	2
2	1	0	2	1	0	1	0	2	0
1	3	0	3	2	4	1	1	1	1
0	2	1	0	3	0	3	0	0	0
2	3	3	1	3	1	1	0	2	0
0	2	3	0	1	1	0	1	1	1
0	3	2	3	3	1	2	0	1	2
2	2	4		0	2	0	0	2	0
3	0	1	0	2	2	1	0	0	0

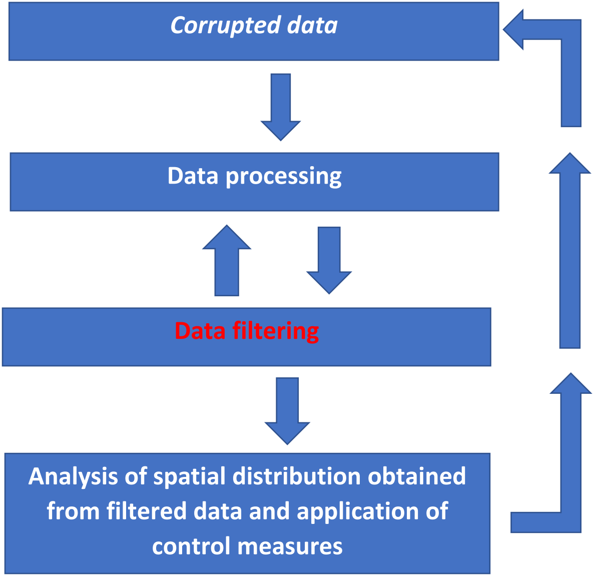
Random measurement error



Systematic measurement error

- Original data are affected by measurement errors yet filtered data may or may not be affected depending on the filter definition.
- Hence, using a filter in the M&C routine can in some cases replace the uncertainty analysis/elimination procedure.

Handling measurement errors in the M&C routine



‘Strongly’ vs. ‘slightly’ corrupted datasets

WHAT IS THE DATA RANGE WHERE A FILTER

CAN BE USED WITHOUT UNCERTAINTY ANALYSIS

IN THE M&C ROUTINE?

Research methodology

I. Selection of a baseline problem and generation of spatial data from a reliable mathematical/computational model

II. Simulation of M&C routine: definition of data filter

III. Simulation of corrupted dataset

IV. Application of data filter to the corrupted dataset to determine the parameter range where the filter works

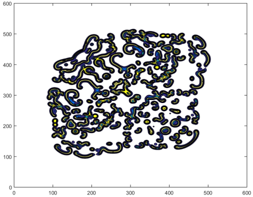
The case study

M&C of biological invasion

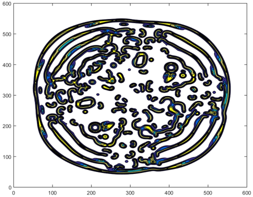
Advantages

- well developed models of biological invasion (more than one model have been investigated)
- the wealth of spatial distributions can be generated
- a simple definition of the filter can be employed in the problem
- easy and straightforward simulation of measurement errors in the M&C routine

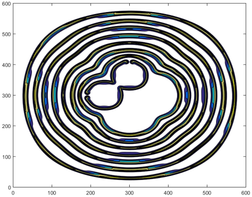
Generation of spatial distributions



(a)



(b)



(c)

(a) Patchy spatial distribution of the density of invasive species obtained as a result of numerical solution of the mathematical model of biological invasion.

(b) Continuous front density distribution obtained from the same model. The spatial pattern has patchy structure behind the front.

(c) Continuous front density distribution with quasi-regular oscillations behind the front.

Definition of a data filter in the M&C routine

The ‘n’-filter: data are filtered to separate continuous front spatial distributions from patchy spatial distributions

Control measures are different if a patchy spatial distribution is detected

The filter definition is based on the image analysis:

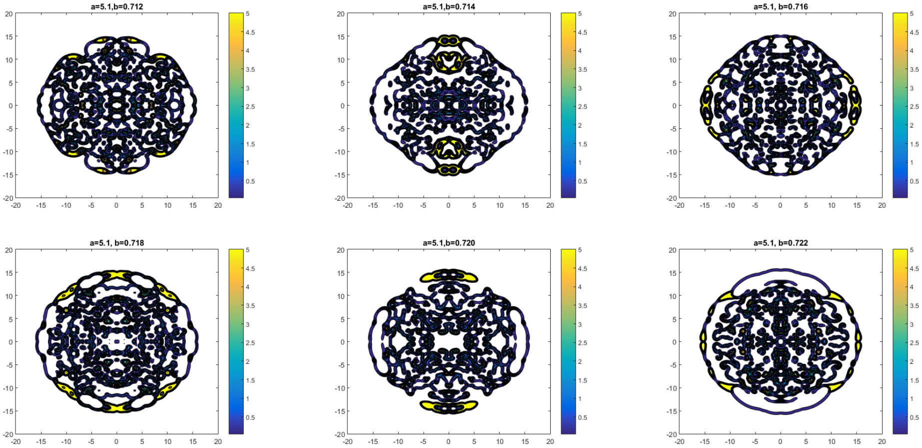
$n = 1$ – continuous front spatial distribution,

$n > 1$ – patchy spatial distribution,

where n is the number of objects in the image of spatial distribution

N.B. Petrovskaya et al. (2017) Patchy, not patchy, or how much patchy? Classification of spatial patterns appearing in a model of biological invasion. Math. Model. Nat. Phenom., 12:208-225

Spatial distributions of invasive species



Spatial distributions of the density of invasive species where continuous front and patchy distributions look similar to each other.

Simulation of corrupted datasets

Assumptions:

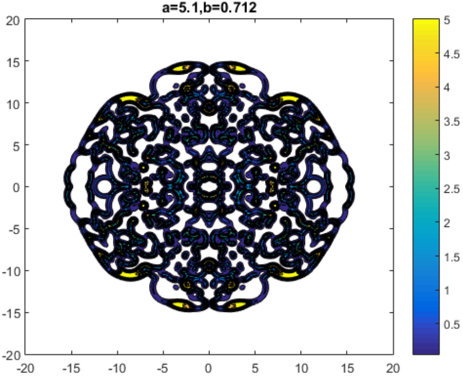
- The dataset generated as a result of numerical solution to a mathematical model of biological invasion is accurate.
- Nodes of computational grid where the solution is available are thought of as trap locations.

I. Low trap counts are replaced by zero trap count (trap exposition time is shorter than required)

II. Every nth trap is removed from a sampling grid (the sampling grid is coarser than required)

I. Low trap counts are replaced by zero trap count

$$\hat{u}(x,y) = u(x,y) \quad \text{for} \quad u(x,y) > C, \qquad \hat{u}(x,y) = 0 \quad \text{for} \quad u(x,y) \leq C$$

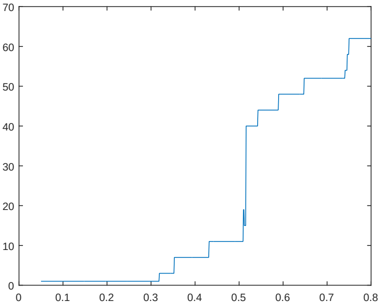


C	0.05	$\cdots$	0.826	$\cdots$	0.858	$\cdots$	0.879
n	1	1	5	5	6	6	10

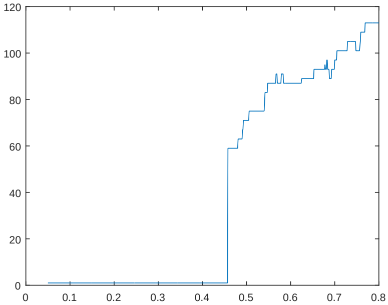
I. Low trap counts are replaced by zero trap count

The number of objects as a function of threshold density C for two continuous front distributions:

The ‘n’-filter works on corrupted datasets till (a) $C^* = 0.32$, and (b) $C^* = 0.46$ (the biologically meaningful maximum threshold value is $C \sim 0.1$)

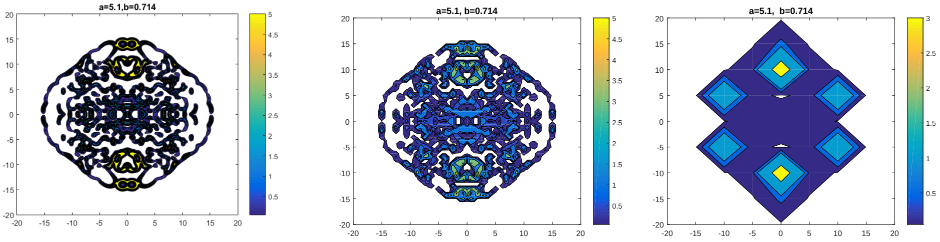


(a)



(b)

Every nth trap is removed from a sampling grid



(a) (b) (c)

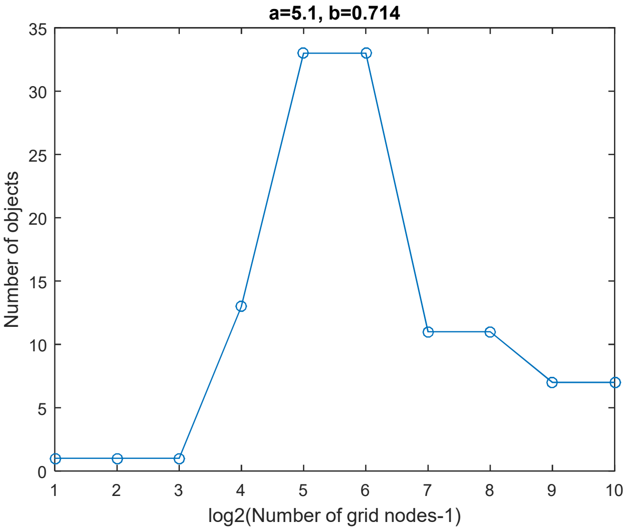
(a) The original patchy spatial distribution obtained on a fine grid, $N = 1025$;

(b) Reconstruction of the spatial distribution on a coarse 'sampling' grid, $N = 65$;

(c) Reconstruction of the spatial distribution on a very coarse grid, $N = 9$;

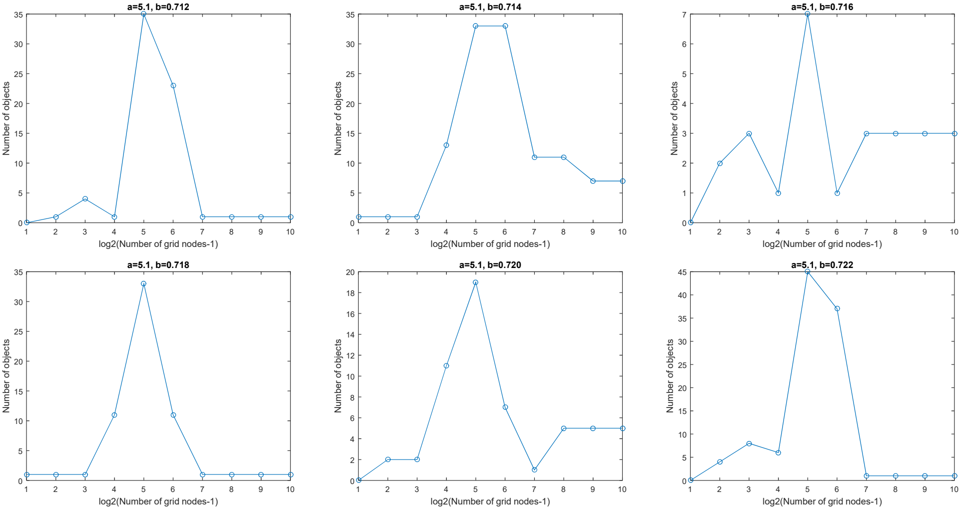
Petrovskaya, N.B. & Zhang, W. (2019) Accurate recognition of spatial patterns arising in spatio-temporal dynamics of invasive species. in: L. Formaggia (ed.), Dynamical Systems Applied to Biology and Natural Sciences, SEMA-SIMAI Springer Series, Springer, Berlin (accepted for publication)

II. Every *n*th trap is removed from a sampling grid



$N^* = 9$ - the ‘n’-filter works for data ‘collected’ when very coarse ‘sampling’ grids are used

II. Every *n*th trap is removed from a sampling grid



The number of objects as a function of the number of grid points in a sampling grid used to reconstruct the visual image of spatial pattern.

INHERENT UNCERTAINTY IN THE MODEL

Mathematical model of the gypsy moth invasion

$$\frac{\partial u(x,y,t)}{\partial t} = \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + \gamma u(u-\beta)(1-u) - uv \; , \qquad (1)$$

$$\frac{\partial v(x,y,t)}{\partial t} = \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) + uv - mv \; , \qquad (2)$$

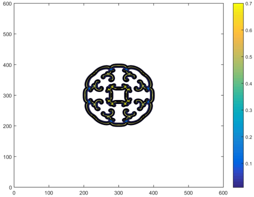
– The initial conditions are given by ‘continuous front’ ($n = 1$) spatial distributions of $u(x, y)$ and $v(x, y)$

– The model is solved numerically in a large spatial domain with no-flux boundary conditions

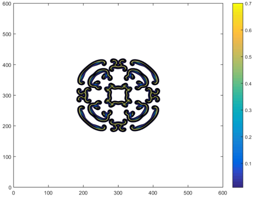
– The topology of spatial distribution $u(x, y)$ depends on parameter m (mortality)

Jankovic, M., & Petrovskii, S.V. (2013) Gypsy moth invasion in North America: a simulation study of the spatial pattern and the rate of spread. Ecol. Compl. 14:132-144

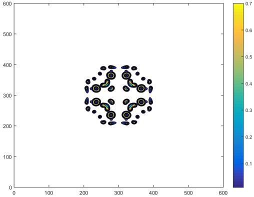
Occasional switches over transition time



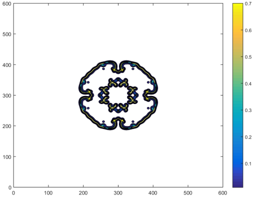
(a) continuous front pattern
time $t_1 = 200$ ($t^* = 310$)



(b) patchy pattern
time $t_2 = 250 > t_1$

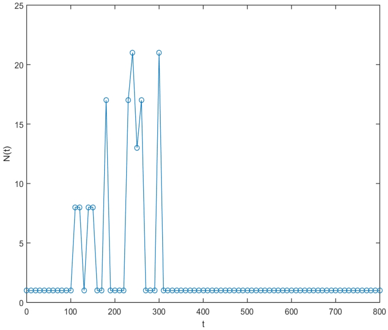


(a) patchy pattern
time $t_1 = 230$ ($t^* = 290$)

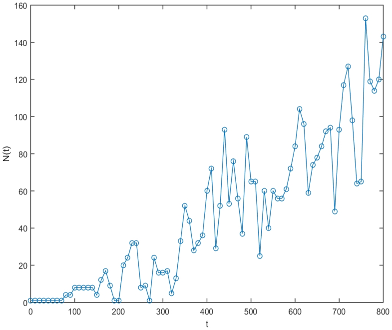


(b) continuous front pattern
time $t_2 = 270 > t_1$

The number of objects n as a function of time t



(a)

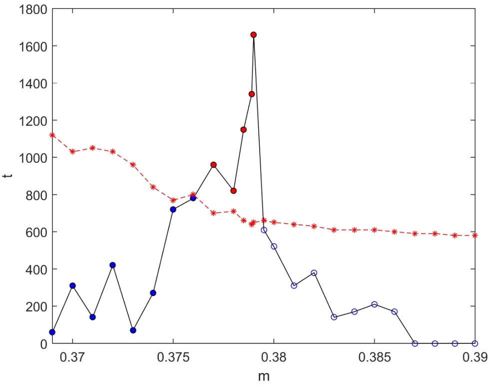


(b)

(a) $m = 0.381$: convergence to continuous front spatio-temporal dynamics, $n(t) = 1$ at any time $t > t^* = 310$

(b) $m = 0.374$: convergence to patchy spatio-temporal dynamics; $n(t) > 1$ at any time $t > t^* = 290$.

The transient dynamics



Time t^D required for the invasive species to spread over domain D (dashed red curve, red stars) and transition time t^* (solid blue curve, open and closed circles) as functions of parameter m.

Petrovskaya, N.B. & Zhang, W. (2019) When seeing is not believing: comparative study of various spatial distributions of invasive species (submitted to JTB)

Conclusions

- The M&C protocol can be thought of as a data filter as its application transforms the original dataset.
- Data filtering in the M&C procedure may alleviate negative impact of uncertainty on the accuracy of results when spatial distributions are reconstructed from data with measurement errors. In some cases, there is no need to ask for more accurate data collection as measurement errors will be 'eliminated' by application of the M&C protocol.
- The filter's applicability bounds have to be further investigated to determine the data range where uncertainty can be neglected.
- If inherent uncertainty presents in the model, the M&C data filter may become useless, no matter how accurate the data are.

THANK YOU!