The analysis of spatial data: individual movements and species and community models

Prof. Otso Ovaskainen
University of Helsinki, Finland
Department of Biosciences

group leader

Otso Ovaskainen

Project cordinator

Juri Kurhinen

IT-designer

Evgeniy Meyke

technician

Heini Ali-Kovero

Mathematical Biology Group

post docs

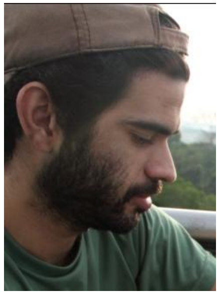
Maria Delgado

Dmitry Schigel

Henjo de Knegt

Guillaume Blanchet

PhD students



Ulisses Cameron

Jussi Jousimo

Sonja Koskela

Henna Fabritius

Veera Norros

Markku Karhunen

Tanjona Ramiadantsoa

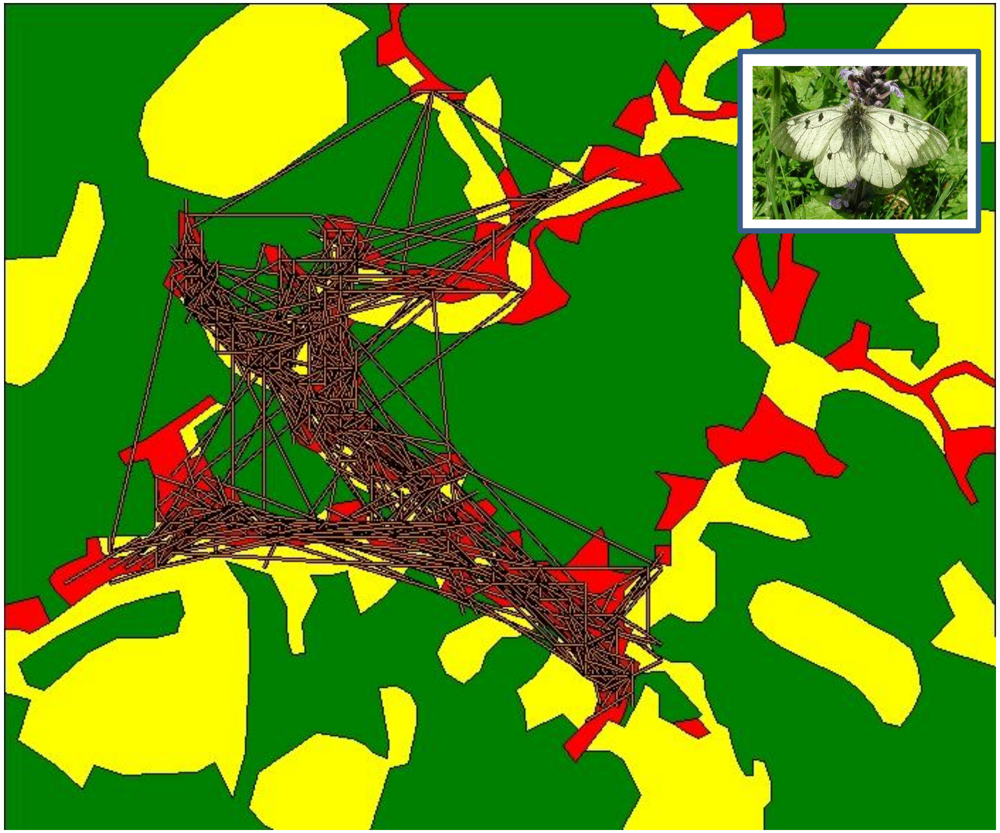
movements, populations, communities, genetics, evolution, bioinformatics

Movement plays a central role in ecology

- All organisms move!
- Understanding movement is central to all questions in spatial ecology
- Applications (monitoring, managing and conserving populations) often require an understanding of movement.
- Habitats are fragmenting – can the organisms move between the fragments?
- Climate is changing – can the organisms move to the areas where climate will be suitable in the future?

Examples of movement data:

Mark-recapture on butterfly movement



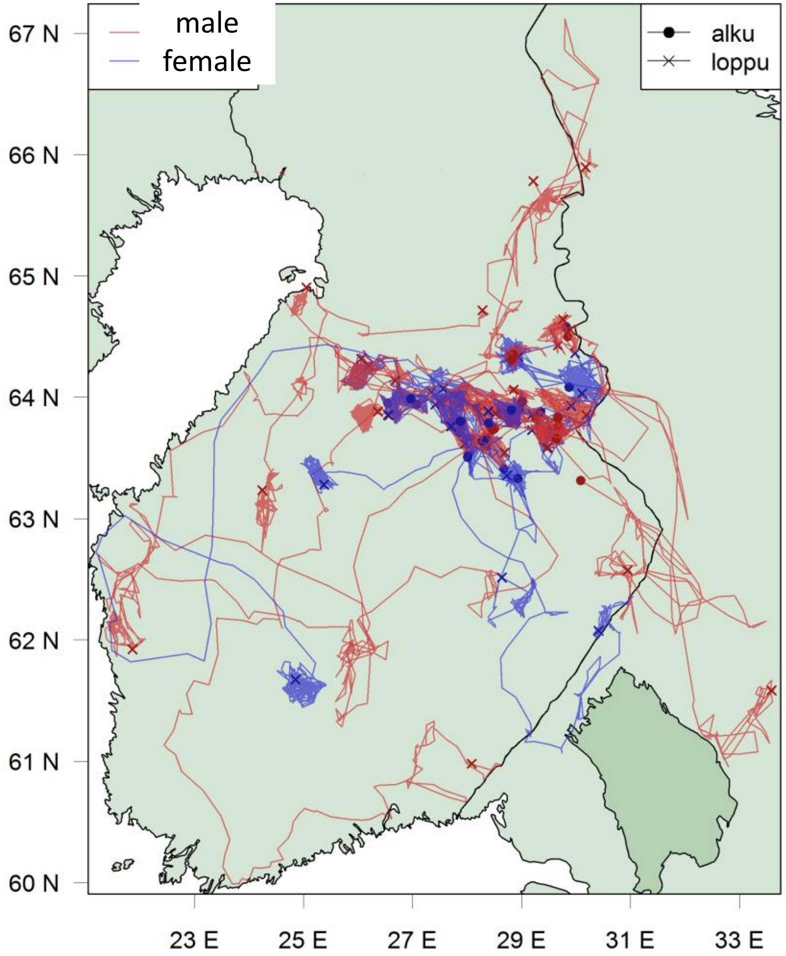
Ovaskainen, O. et al. 2008. An empirical test of a diffusion model: predicting clouded apollo movements in a novel environment. *American Naturalist* **171,** 610-619.

Examples of movement data:

GPS data on wolf, bear, lynx, moose, forest reindeer, ...

S. Ronkainen

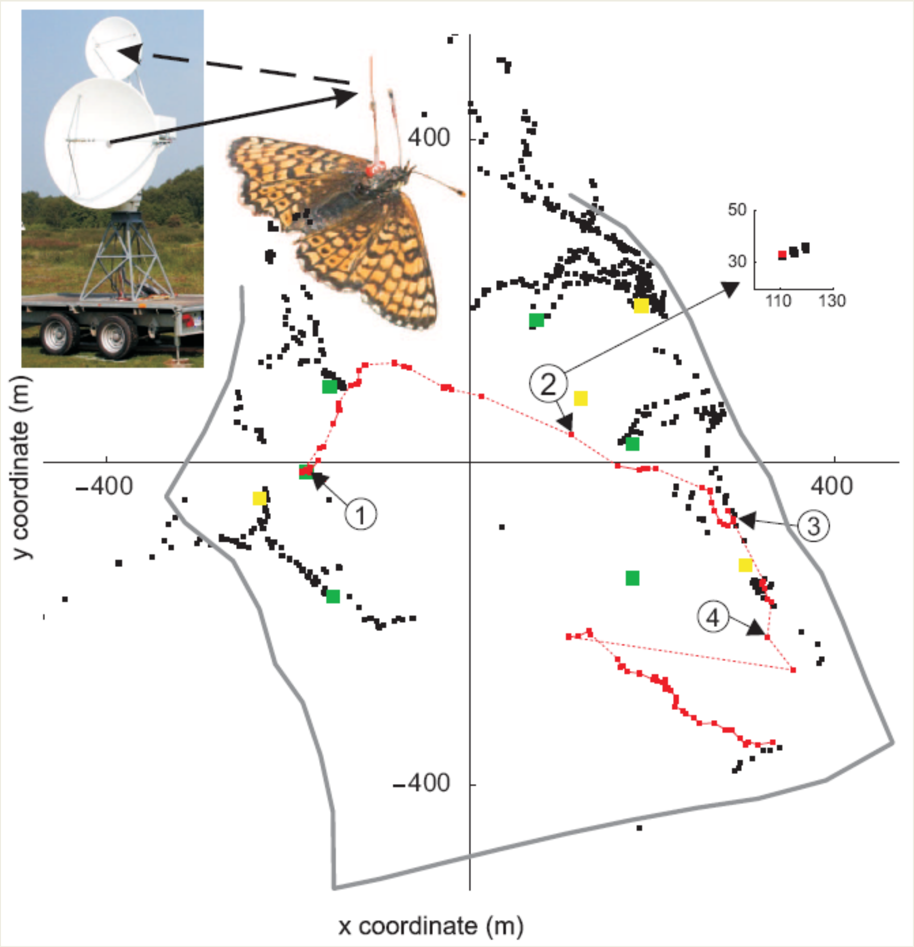
Movements by GPS collared wolves (2002-2008)



Source: Ilpo Kojola / GFR

Examples of movement data:

Harmonic radar butterfly movement



Ovaskainen, O. et al. 2008. Tracking butterfly movements with harmonic radar reveals an effect of population age on movement distance. *PNAS* **105**, 19090-19095.

Examples of movement data:

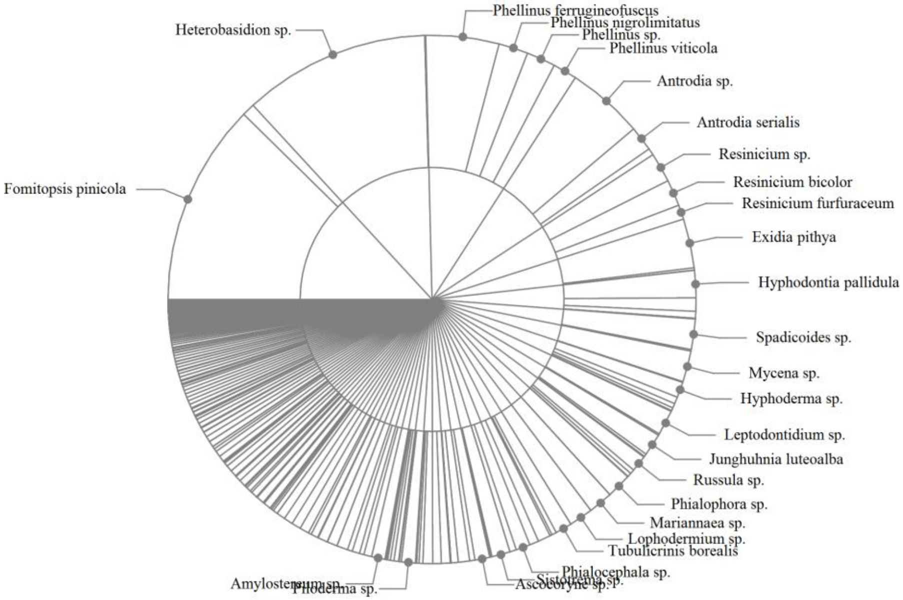
DNA data on dispersing fungal spores

Veera Norros collecting spore samples with a cyclone sampler

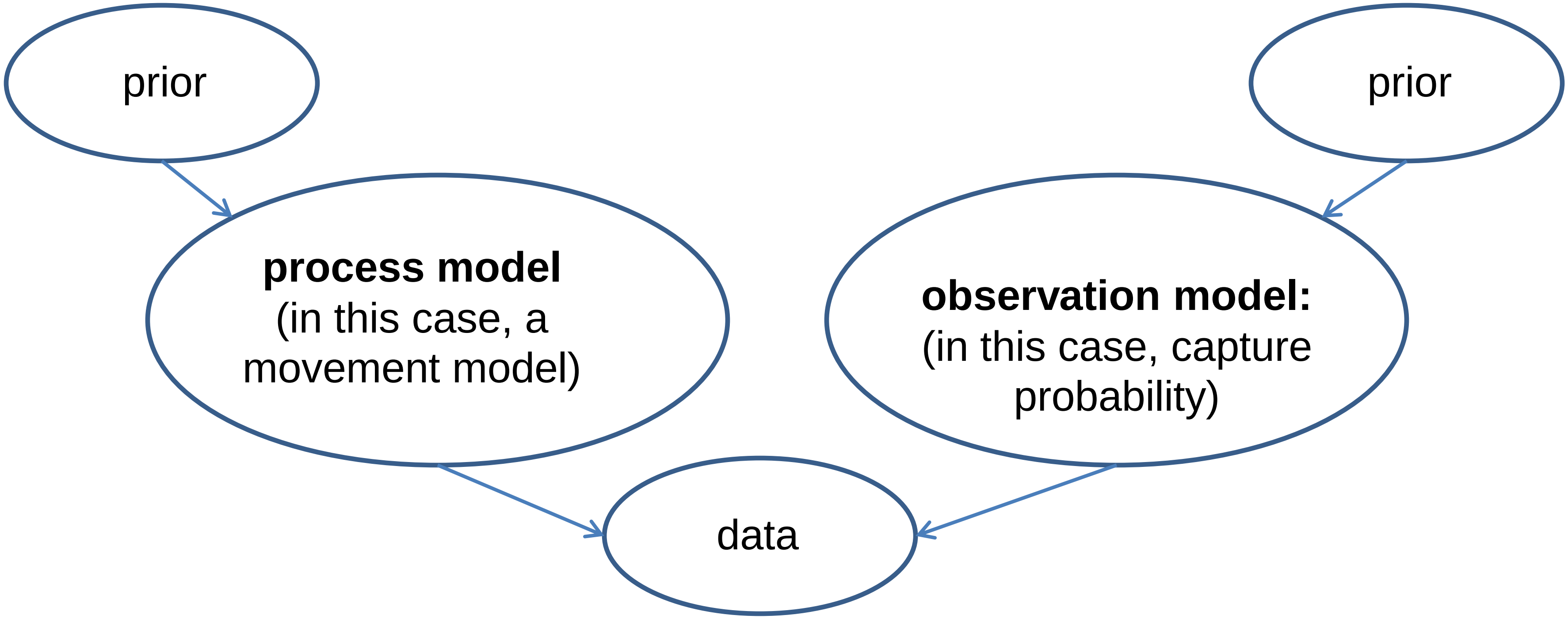
454-sequencing

```
>sample1
TTTCCGTAGGTGAACCTGCGGAAGGATCATTAATGAATACAATTCGGTCGGCGGGAAGGAGGGGGAG
CTGTCGCTGGCCTTGTGGCATGTGCACGCTCTCTTTGGAACGTCGGTCGTCTTTCATATTTTCACCAGTG
CACCCAATGTAGGATGCCTCTCCTCCGGGAGGGGGGACCTATGTCTTTTTCAGACGCCCCCACAGTTTA

>sample2
GAAAGTCTCAGAATGTTTACTATCGTCGAACCATGACTTCCAGGAGACGTGGGTCGGCGAGATAAAAG
TTATCACAACTTTCAGCAACGGATCTCTTGGCTCTCGCATCGATGAAGAACGCAGCGAATTGCGATATG
TAATGTGAATTGCAGATCTACAGTGAATCATCGAATCTTTGAACGCACATTGCGCTCCTCGGTGTTCCG

>sample3
```

identification



Bayesian state-space approach



prior

prior

process model
(in this case, a
movement model)

observation model:
(in this case, capture
probability)

data

Technical details on computation of likelihood and MCMC sampling:

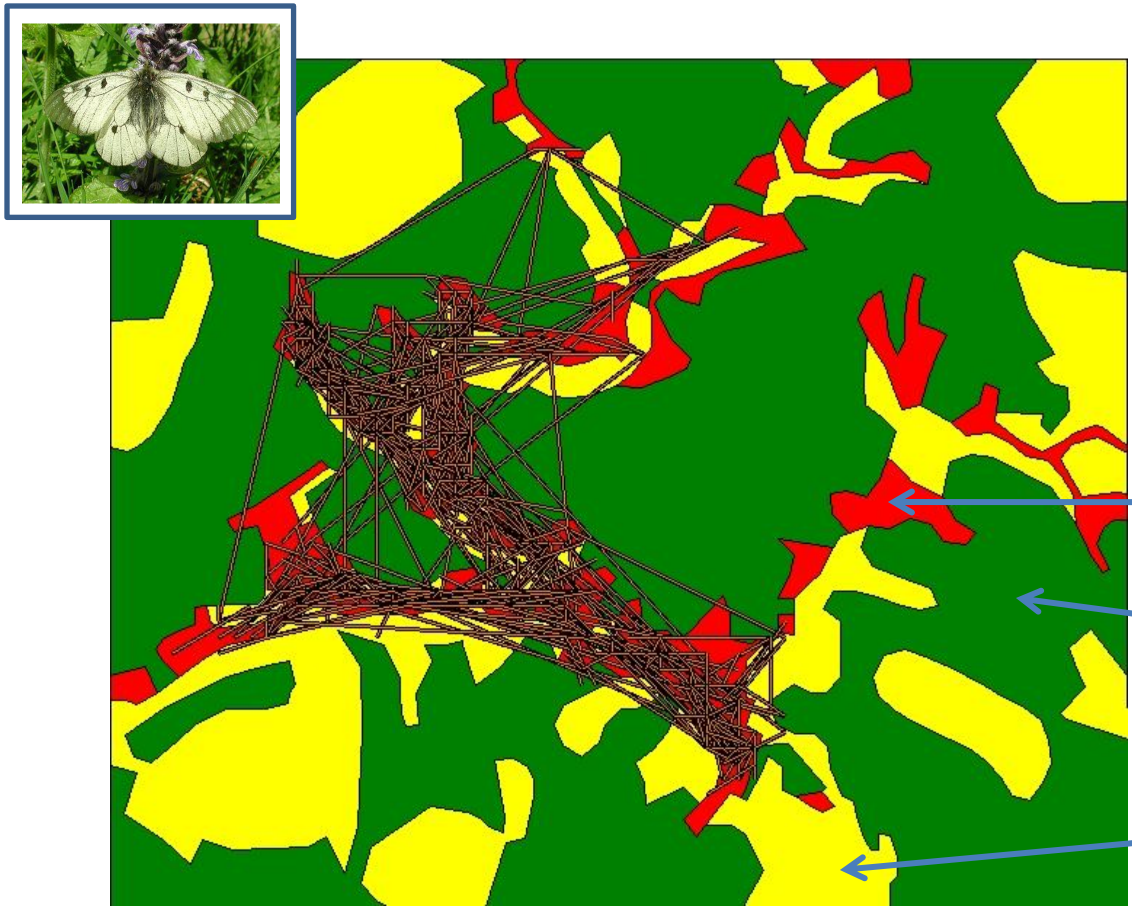
Ovaskainen, O. 2004. Habitat-specific movement parameters estimated using mark–recapture data and a diffusion model. *Ecology* **85**, 242-257.

Ovaskainen, O., Rekola, H., Meyke, E. and Arjas, E 2008. Bayesian methods for analyzing movements in heterogeneous landscapes from mark-recapture data. *Ecology* **89**, 542-554.

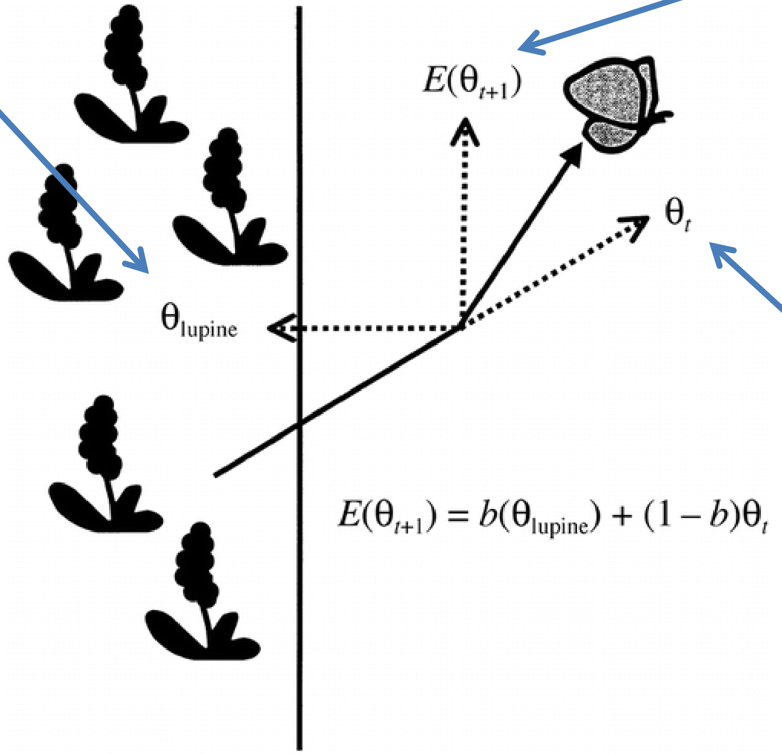
Ovaskainen, O. 2008. Analytical and numerical tools for diffusion based movement models. *Theoretical Population Biology* **73**, 198-211.

A model of animal movement in heterogeneous space

Diffusion parameter *D*
= movement rate within
a given habitat type



Edge-mediated behavior (habitat selection at edges)



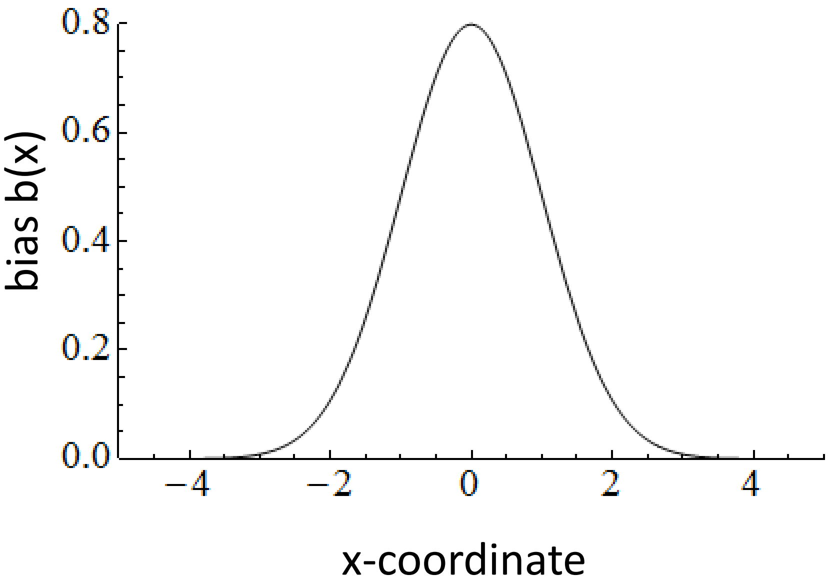
tendency to move toward habitat patch

expected direction for the next move

tendency to continue in the direction already moving

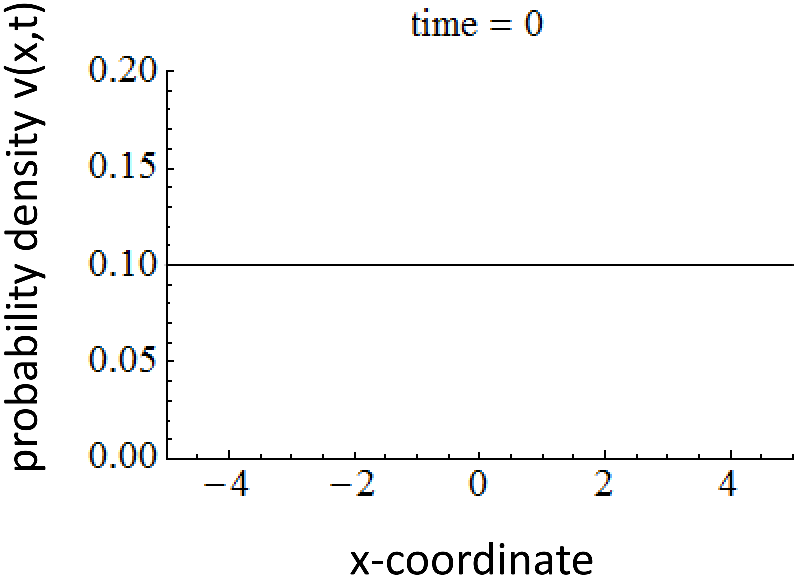
Schultz, C. B., and E. E. Crone. 2001. Edge-mediated dispersal behavior in a prairie butterfly. *Ecology* **82**, 1879-1892.

Edge-mediated behaviour pushes the individual towards the preferred habitat



← EDGE →

PATCH MATRIX

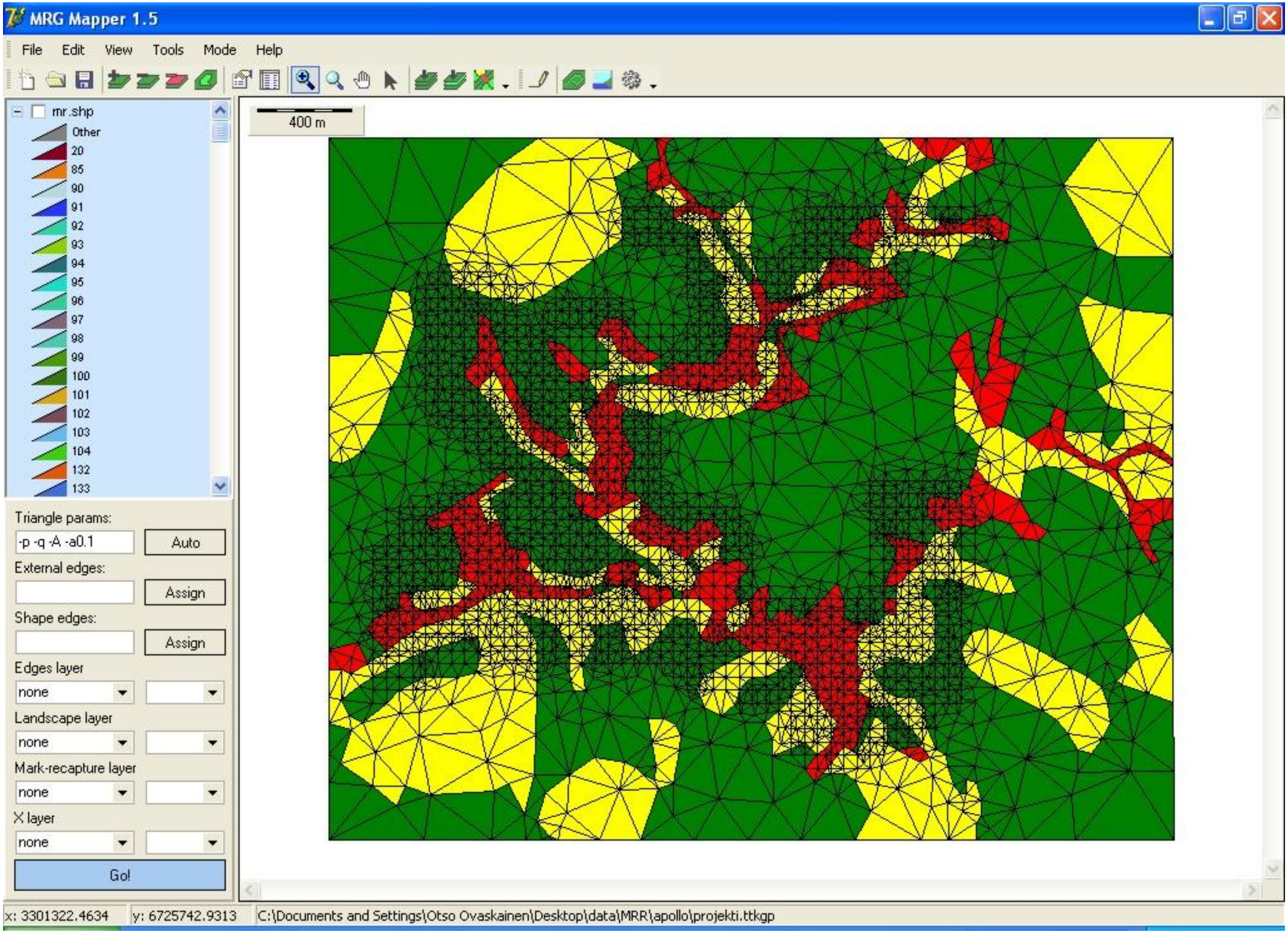


← EDGE →

PATCH MATRIX

Ovaskainen, O. and Cornell, S. J. 2003. Biased movement at a boundary and conditional occupancy times for diffusion processes. *Journal of Applied Probability* **40**, 557-580.

Solving the diffusion model numerically



Triangle params: -p -q -A -a0.1

External edges:

Shape edges:

Edges layer: none

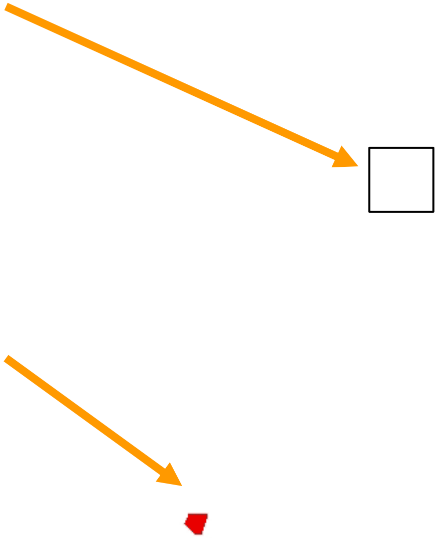
Landscape layer: none

Mark-recapture layer: none

X layer: none

x: 3301322.4634 y: 6725742.9313 C:\Documents and Settings\Otso Ovaskainen\Desktop\data\MRR\apollo\projekti.ttkgp

Simulating the time-evolution of the probability density



location of a
site that is
searched for

initial location

Searching but not finding gives information

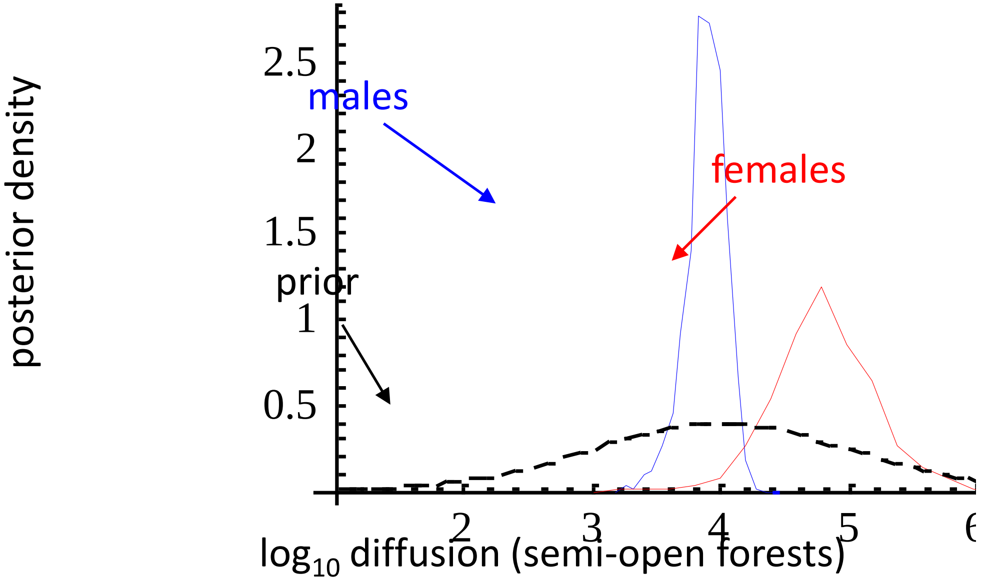
The capture probability *p* is the probability of observing an individual given that it actually is at the site

	site	somewhere else
Probability that an individual is in a site before the search	x	$1-x$
Probability that an individual is in the site after the site is searched for (without finding the individual)	$\frac{x(1-p)}{1-px}$	$\frac{1-x}{1-px}$

Ovaskainen, O. 2004. Habitat-specific movement parameters estimated using mark–recapture data and a diffusion model. *Ecology* **85**, 242-257.

Example of biological inference

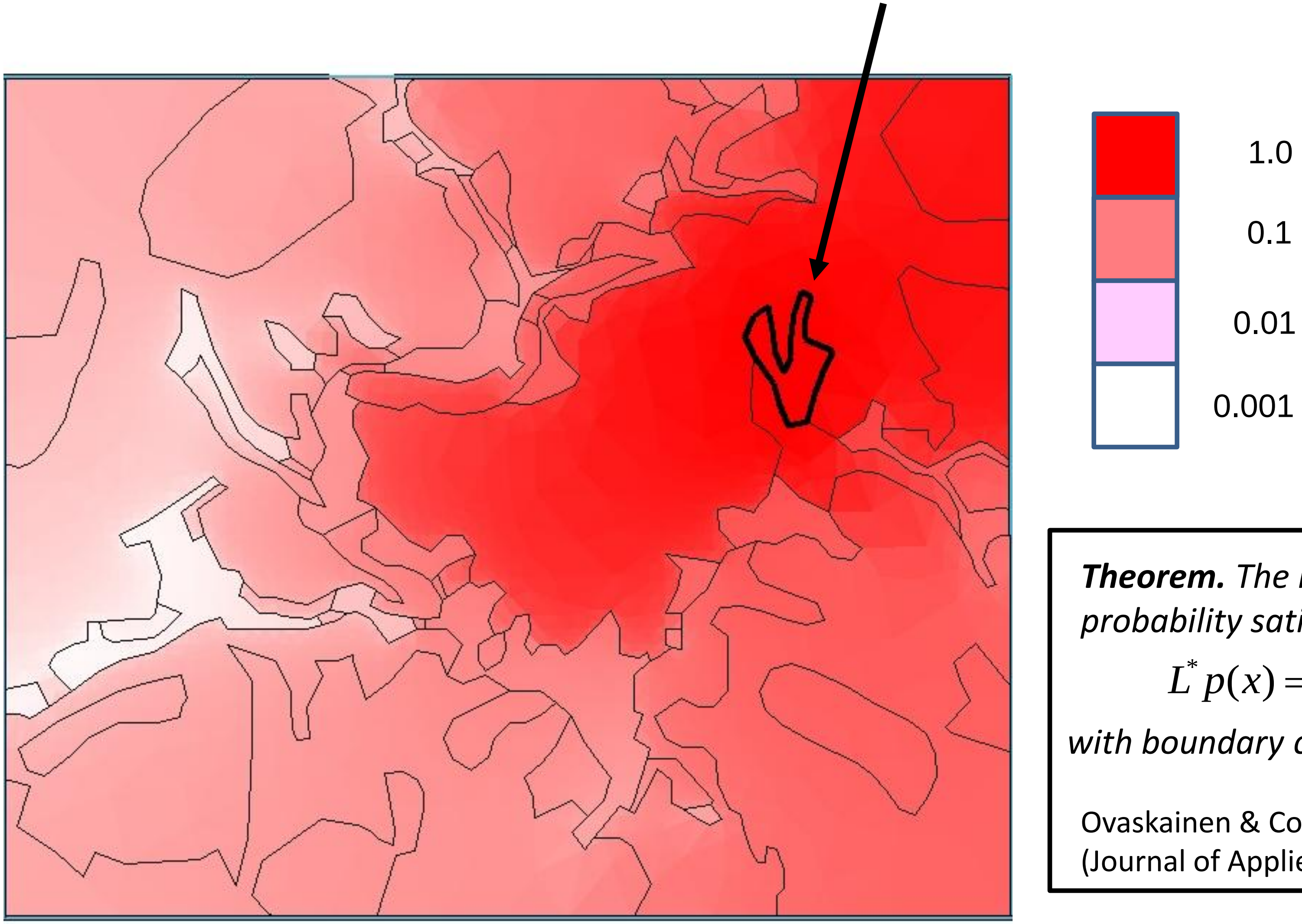
Females move faster than males outside the breeding habitat



Ovaskainen, O. et al. 2008. An empirical test of a diffusion model: predicting clouded apollo movements in a novel environment. *American Naturalist* **171,** 610-619.

Example of model prediction

What is the probability that the butterfly ever visits this meadow?



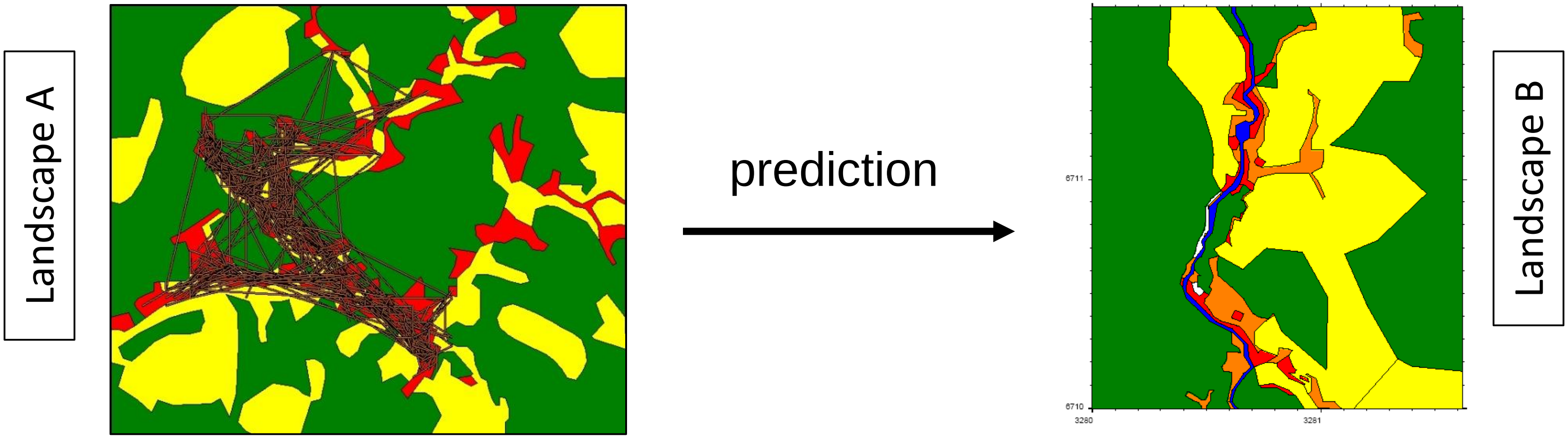
Theorem. *The hitting probability satisfies*

$$L^{*} p(x) = 0$$

with boundary condition $B^{C_{1*}}$

Ovaskainen & Cornell 2003
(Journal of Applied Probability)

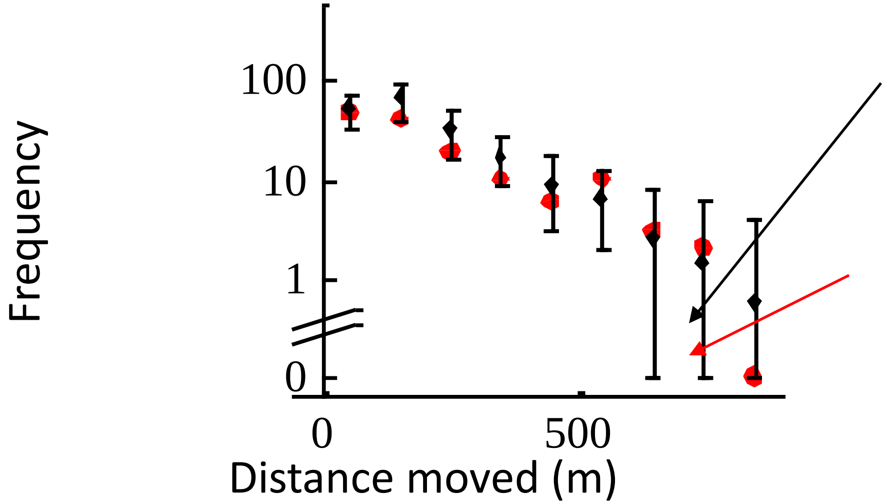
Example of model validation



Landscape A

prediction

Landscape B

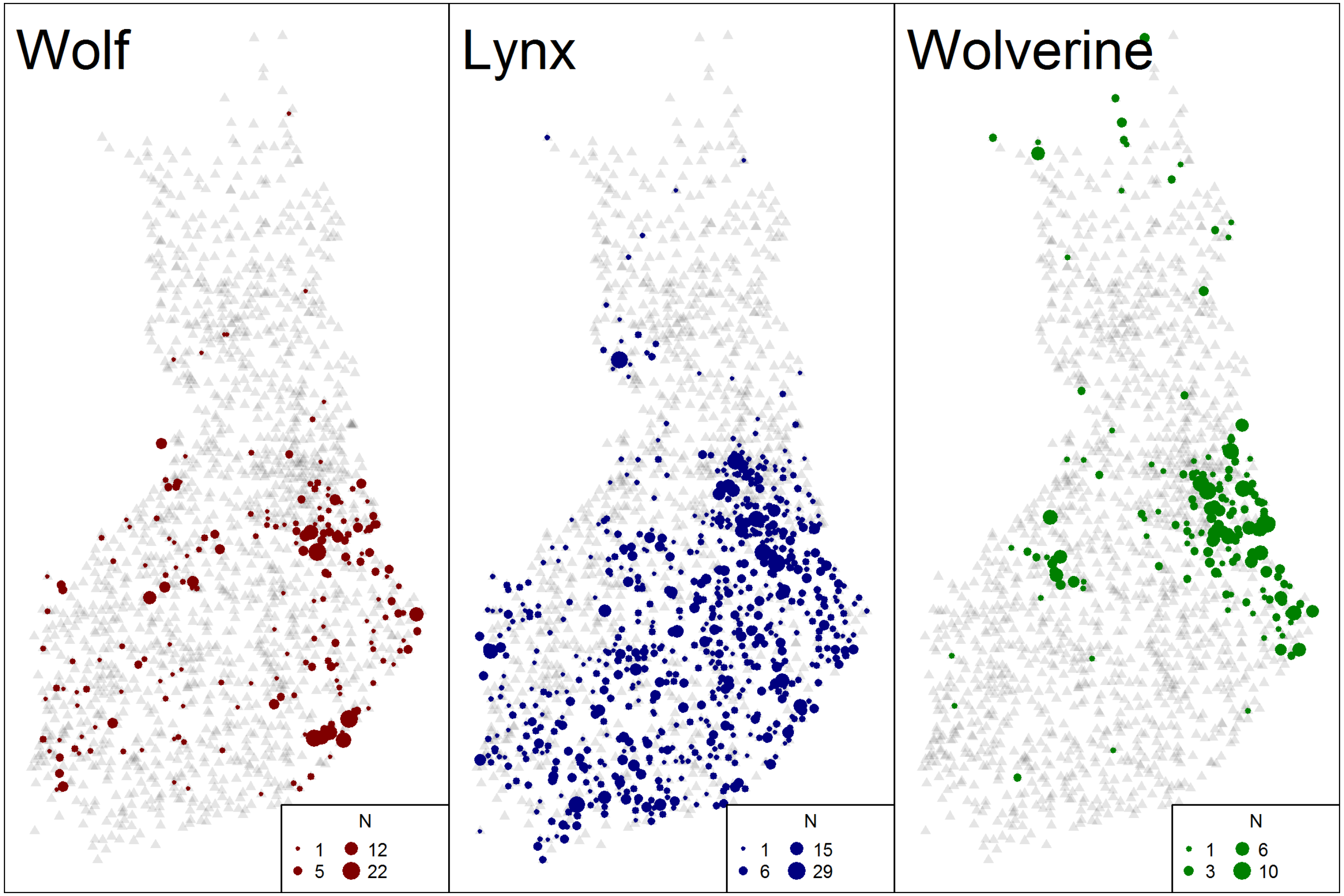


Model parameterized with data from Landscape A, prediction for Landscape B

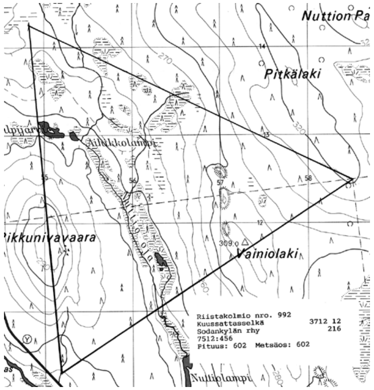
Empirical data on Landscape B

Ovaskainen, O., Luoto, M., Ikonen, I., Rekola, H., Meyke, E. and Kuussaari, M. 2008. An empirical test of a diffusion model: predicting clouded apollo movements in a novel environment. *American Naturalist* **171,** 610-619.

Example of population-level data: winter track data have been collected in Finland for ca. 30 game animal species since 1989



Graphics: Eliezer Gurarie

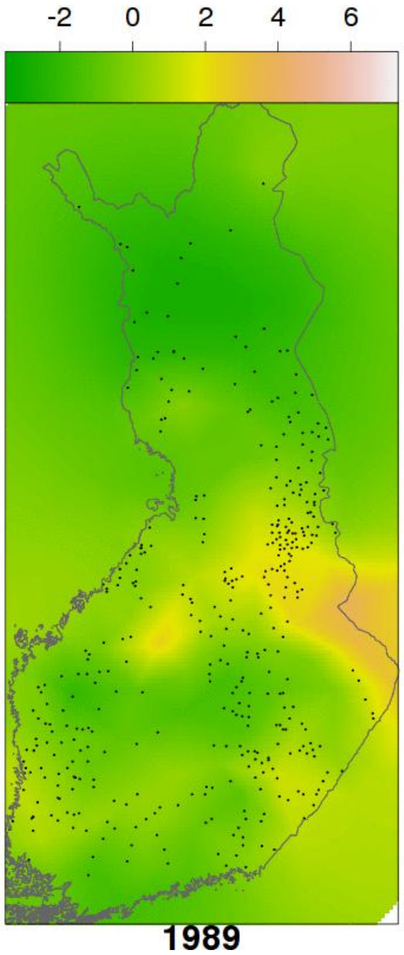


12 km long skiing route

Photo: I.Kojola

Spatio-temporal statistics (with INLA)

$$
\begin{aligned}
y(\boldsymbol{s}_i, t) &\sim \text{Poisson}(\eta(\boldsymbol{s}_i, t)) \\
\log(\eta_t) &= \mu + \beta t + x_t + \epsilon \\
x_t &= \phi x_{t-1} + \omega_t \\
\epsilon &\sim \text{MVN}(0, \sigma_\epsilon^2 I_n) \\
\omega_t &\sim \text{MVN}(0, \sigma_\omega^2 \Sigma(\kappa)).
\end{aligned}
$$



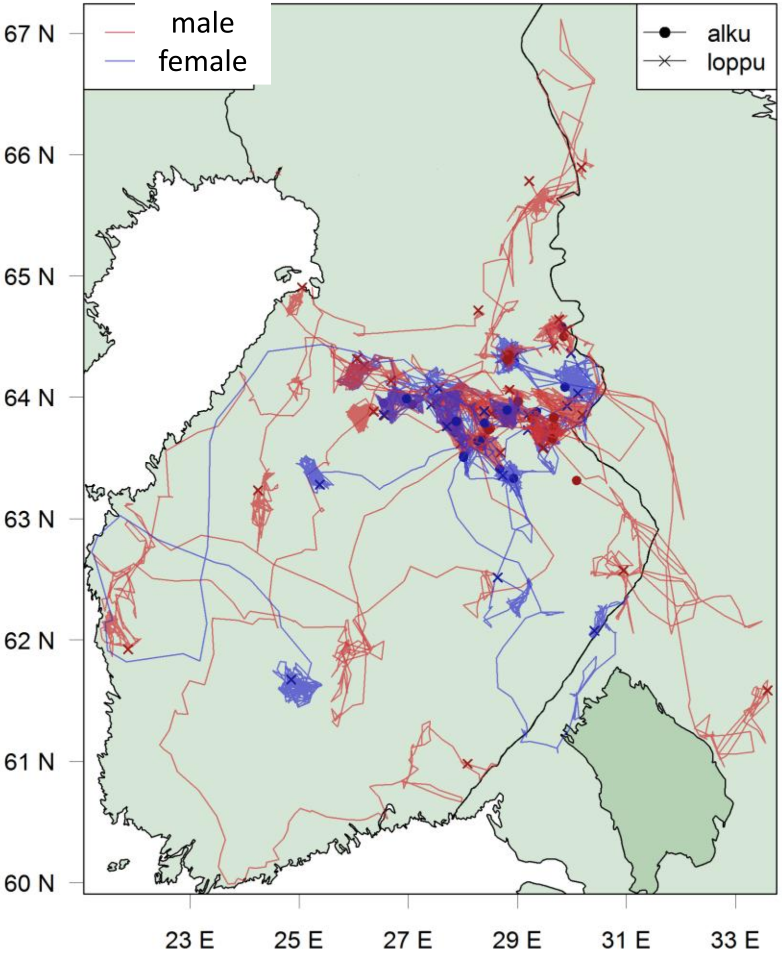
Jussi Jousimo

Number of tracks / km of searching effort

GPS data on wolf movement

S. Ronkainen

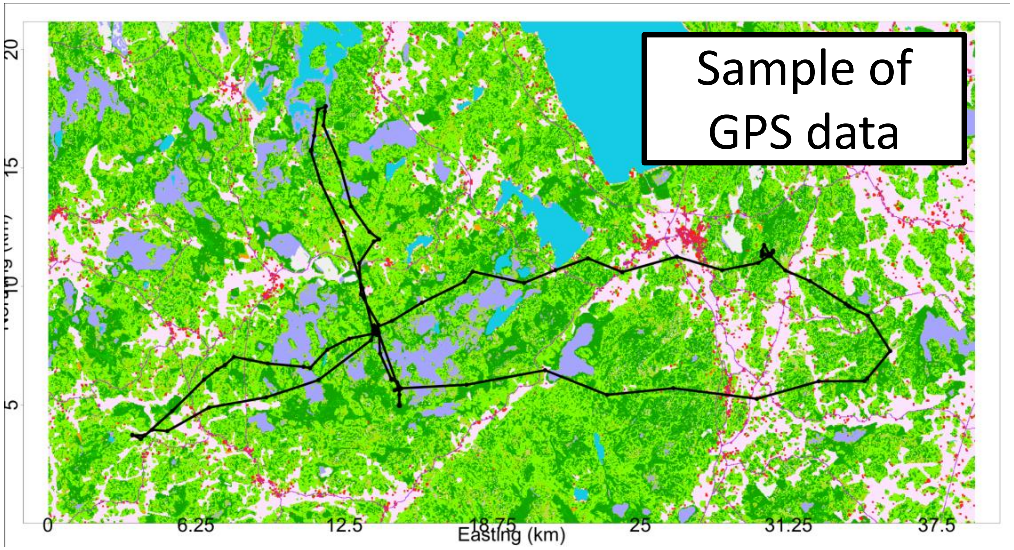
Movements by GPS collared wolves (2002-2008)



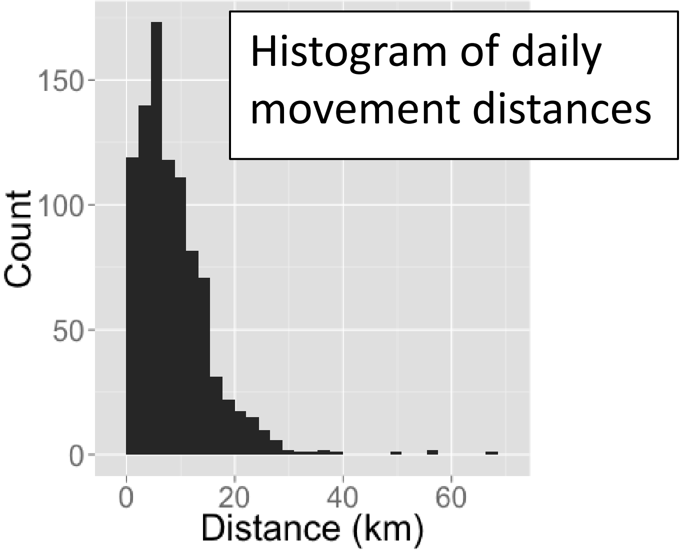
Gurarie, E., Suutarinen, J., Kojola, I. and Ovaskainen, O. (2011)

Source: Ilpo Kojola / GFR

GPS data tells about movement distances and habitat use



Sample of GPS data

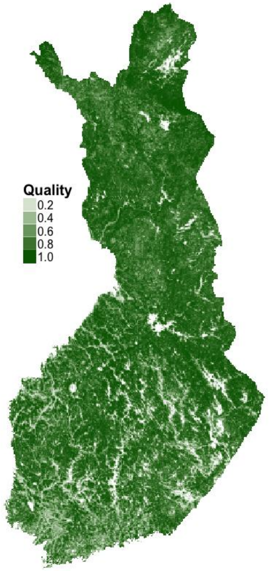


Histogram of daily movement distances

	Urban	Field	Forest	Bog	Water
Expected use	0.04	0.08	0.69	0.08	0.11
Realized use	0.01	0.00	0.94	0.04	0.01
Relative preference	0.13	0.02	1.00	0.39	0.06

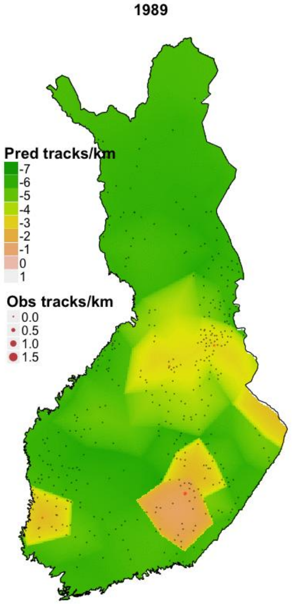
Correcting population density by habitat preferences

Habitat preference

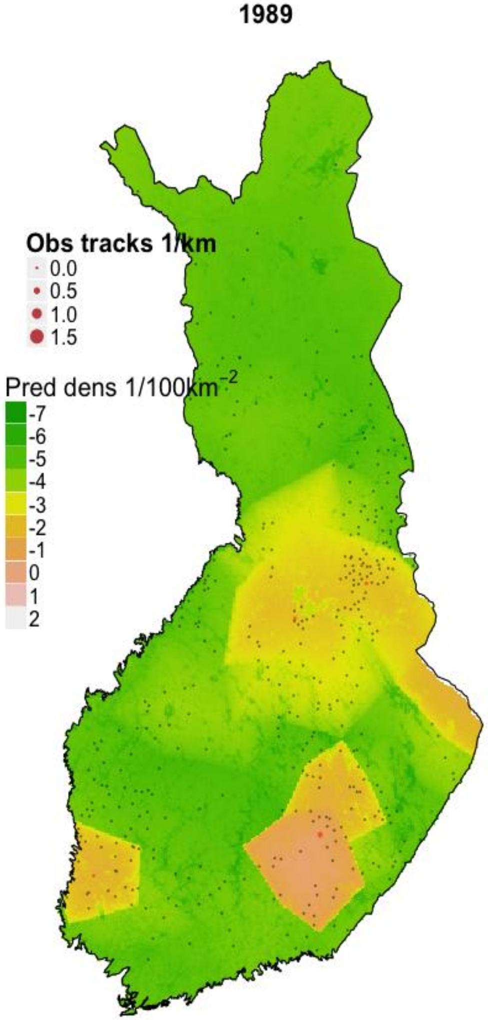


X

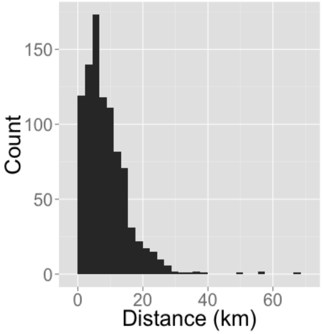
Number of tracks / km of searching effort



=



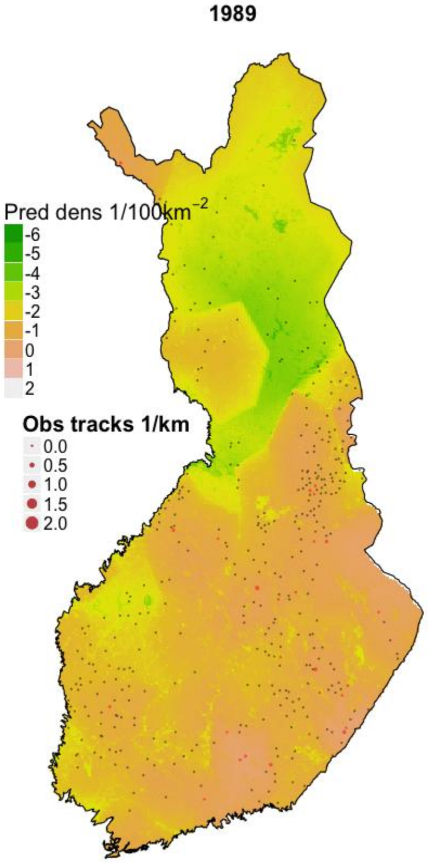
Daily movement distance



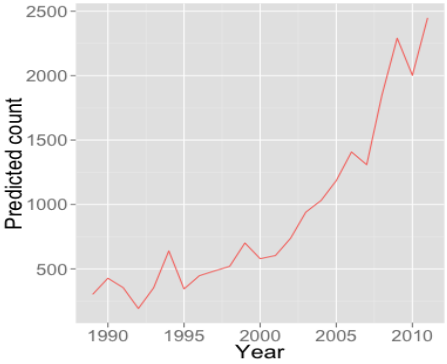
Individuals per 100 km^2

Comparison to independent data

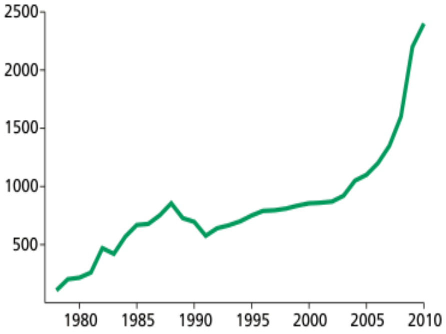
Lynx



Individuals per 100 km^2



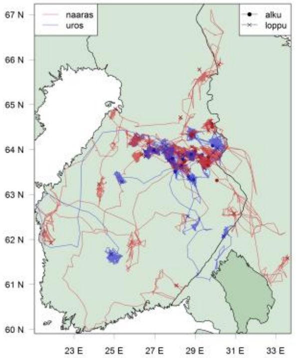
Predicted number of lynxes from winter track data

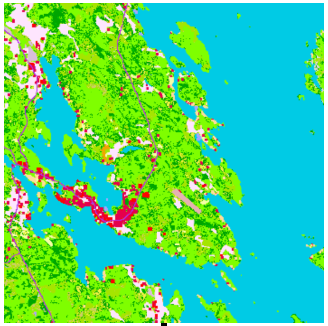


Source: GFR

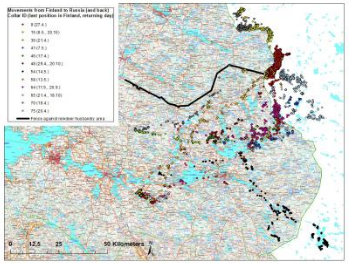
Official statistics based on independent data

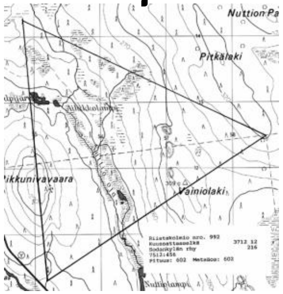
Long-term aim: integrate information from various data sources



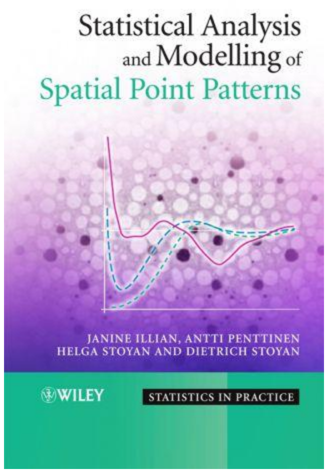


$$
\begin{aligned}
y(\boldsymbol{s}_i, t) &\sim \text{Poisson}(\eta(\boldsymbol{s}_i, t)) \\
\log(\eta(\boldsymbol{s}_i, t)) &= \mu + \beta \log(d(\boldsymbol{s}_i, t)) + x(\boldsymbol{s}_i, t) + \epsilon \\
x(\boldsymbol{s}_i, t) &= \phi x(\boldsymbol{s}_i, t-1) + \omega(\boldsymbol{s}_i, t) \quad |\phi| < 1 \\
\text{Cov}(\omega(\boldsymbol{s}_i, t), \omega(\boldsymbol{s}_j, t')) &= \begin{cases} 0 & \text{if } t \neq t' \\ \sigma_\omega^2 C_{\text{Matérn}}(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|, \kappa, \nu = 2) & \text{if } t = t' \end{cases} \\
\epsilon &\sim \text{N}(0, \sigma_\epsilon^2)
\end{aligned}
$$





Mathematical methods for spatio-temporal point processes

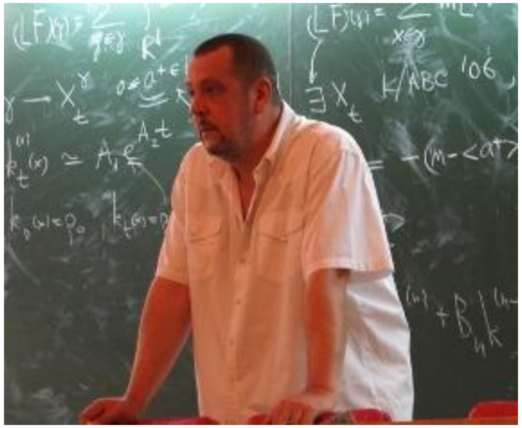

Statistical Analysis and Modelling of Spatial Point Patterns

JANINE ILLIAN, ANTTI PENTTINEN
HELGA STOYAN AND DIETRICH STOYAN

WILEY

STATISTICS IN PRACTICE

We talk about marked point processes...


Yuri Kondratiev

...or Markov evolutions in the space of locally finite configurations...


Ben Bolker

...and these models can be written down as a spatial moment equation

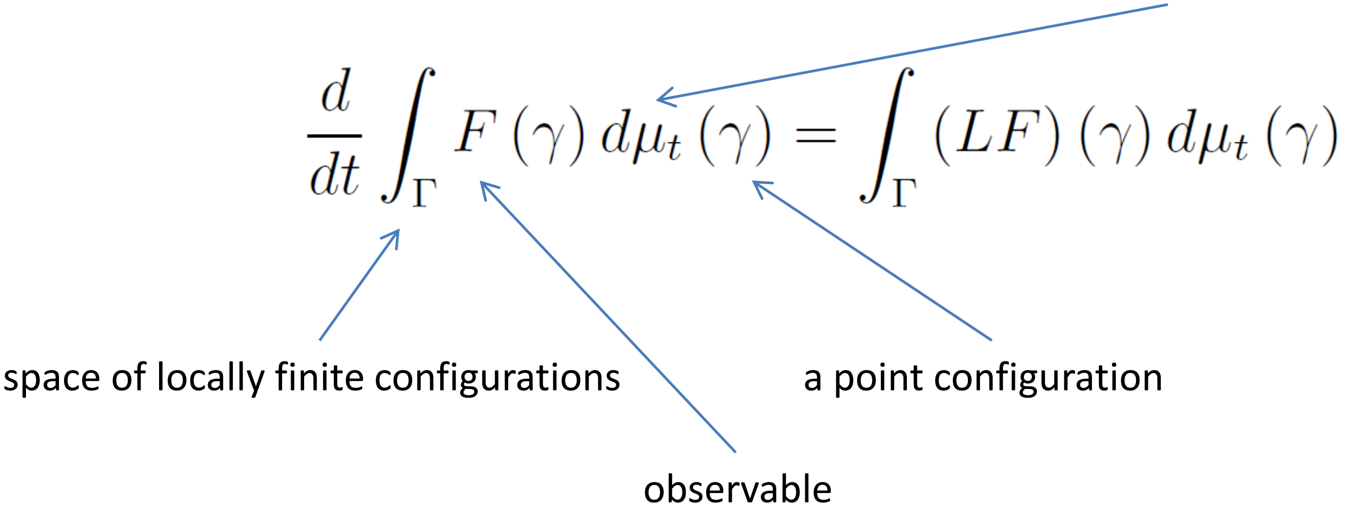
Markov evolutions in the space of finite configurations

Configuration space $\Gamma = \left\{ \gamma \subset \mathbb{R}^d : |\gamma \cap \Lambda| < \infty \right\}$

Particles may

- follow birth-death dynamics
- move by jumps
- have marks (e.g. resource-consumer, predator-prey, genotypes)
- interact with other particles (or groups of particles)

General notation: evolution of measures



$$\frac{d}{dt}\int_{\Gamma} F(\gamma)\, d\mu_t(\gamma) = \int_{\Gamma} (LF)(\gamma)\, d\mu_t(\gamma)$$

space of locally finite configurations

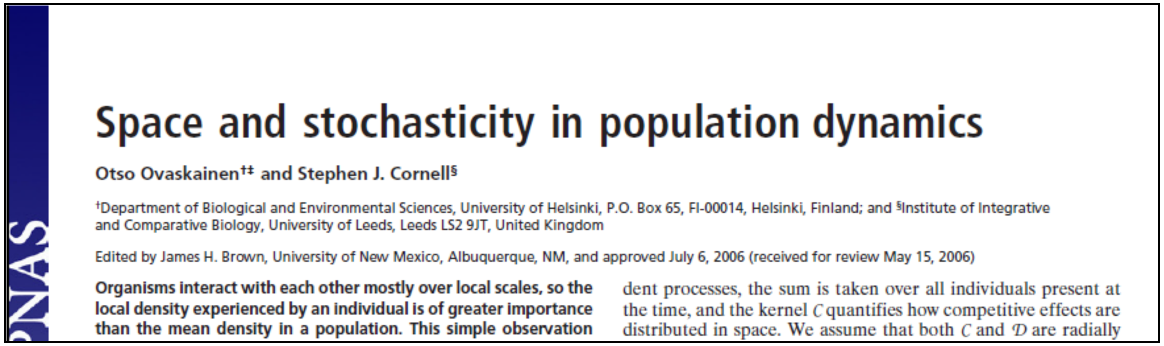
a point configuration

observable

Example model: spatial logistic model

$$\begin{aligned}(LF)(\gamma) \;&=\; \sum_{x\in\gamma}\left(m+\sum_{y\in\eta\setminus x} a^{-}(x-y)\right)\left[F(\gamma\setminus x)-F(\gamma)\right] \\ &\quad + \sum_{y\in\gamma}\int_{\mathbb{R}^d} a^{+}(x-y)\left[F(\gamma\cup x)-F(\gamma)\right]dx.\end{aligned}$$

Mathematical methods of predicting such how models behave



Space and stochasticity in population dynamics

Otso Ovaskainen†‡ and Stephen J. Cornell§

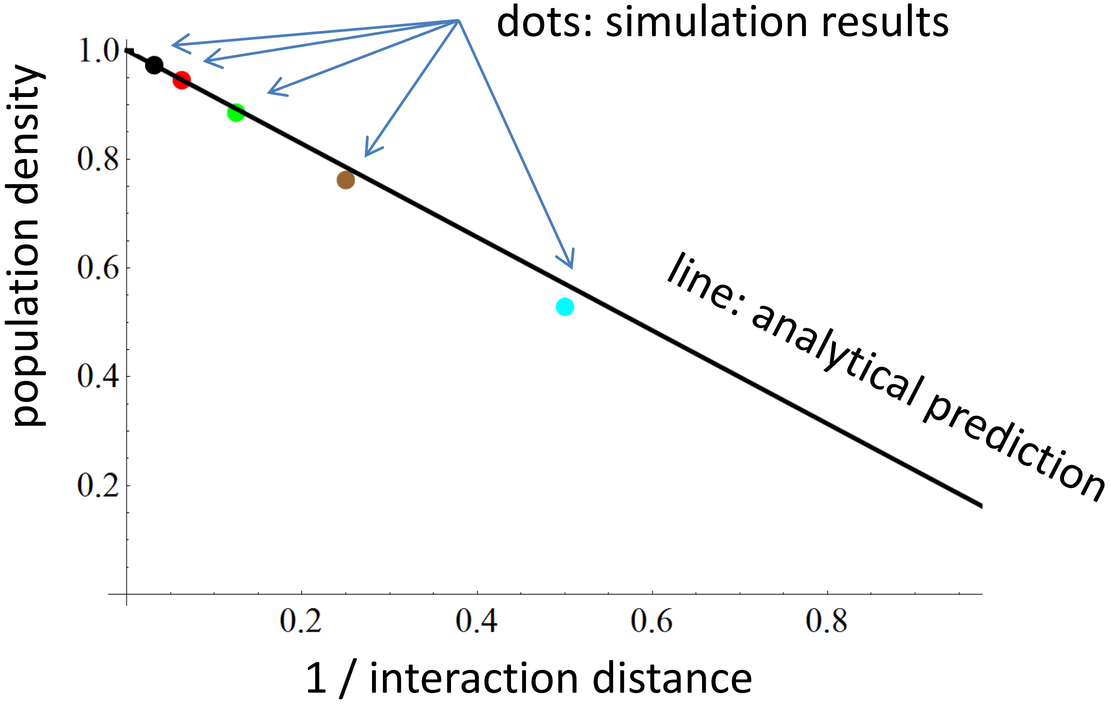
†Department of Biological and Environmental Sciences, University of Helsinki, P.O. Box 65, FI-00014, Helsinki, Finland; and §Institute of Integrative and Comparative Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

Edited by James H. Brown, University of New Mexico, Albuquerque, NM, and approved July 6, 2006 (received for review May 15, 2006)

Organisms interact with each other mostly over local scales, so the local density experienced by an individual is of greater importance than the mean density in a population. This simple observation

dent processes, the sum is taken over all individuals present at the time, and the kernel $\mathcal{C}$ quantifies how competitive effects are distributed in space. We assume that both $\mathcal{C}$ and $\mathcal{D}$ are radially

Stephen Cornell



Example model:
spatial logistic model

Summer 2012: research program in stochastic dynamics in Bielefeld / Germany

Leonid Bogachev

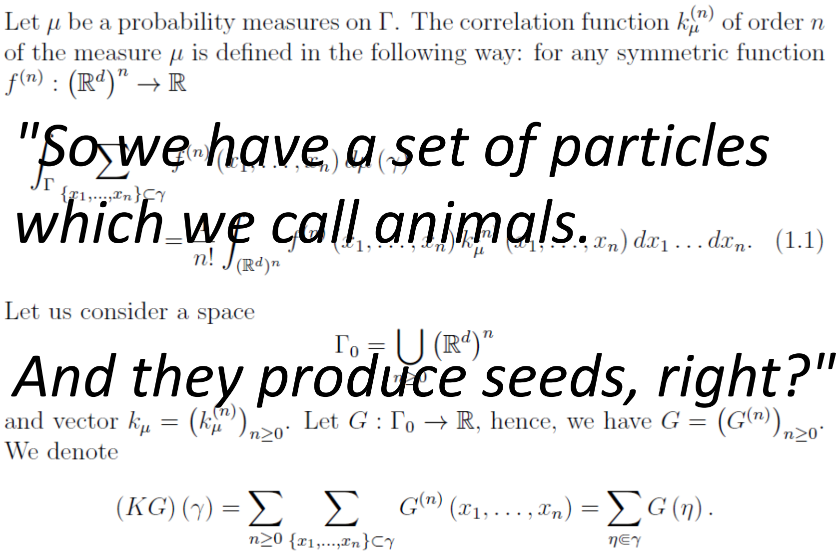
Yuri Kondratiev

Ben Bolker

Otso Ovaskainen

Four months, ca. 100 participants, with expertise in mathematics, physics and spatial ecology

Translating mathematics to biology and vice versa



Let μ be a probability measures on Γ. The correlation function $k_\mu^{(n)}$ of order n of the measure μ is defined in the following way: for any symmetric function $f^{(n)}:\left(\mathbb{R}^d\right)^n \to \mathbb{R}$

$$\int_\Gamma \sum_{\{x_1,\ldots,x_n\}\subset\gamma} f^{(n)}(x_1,\ldots,x_n)\,d\mu(\gamma) = \frac{1}{n!}\int_{(\mathbb{R}^d)^n} f^{(n)}(x_1,\ldots,x_n)\,k_\mu^{(n)}(x_1,\ldots,x_n)\,dx_1\ldots dx_n. \quad (1.1)$$

Let us consider a space

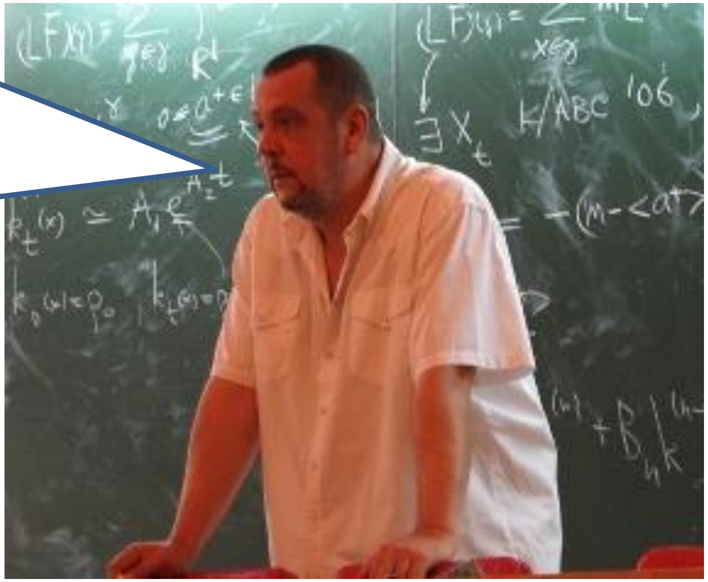
$$\Gamma_0 = \bigcup_{n\geq 0}\left(\mathbb{R}^d\right)^n$$

and vector $k_\mu = \left(k_\mu^{(n)}\right)_{n\geq 0}$. Let $G:\Gamma_0 \to \mathbb{R}$, hence, we have $G = \left(G^{(n)}\right)_{n\geq 0}$. We denote

$$(KG)(\gamma) = \sum_{n\geq 0}\sum_{\{x_1,\ldots,x_n\}\subset\gamma} G^{(n)}(x_1,\ldots,x_n) = \sum_{\eta\Subset\gamma} G(\eta).$$

"So we have a set of particles which we call animals.

And they produce seeds, right?"



Prof. Yuri Kondratiev (Bielefeld, Germany)

Evolution of measures
(model definition, what the individuals do?)

$$\frac{d}{dt}\int_{\Gamma} F(\gamma)\, d\mu_t(\gamma) = \int_{\Gamma} (LF)(\gamma)\, d\mu_t(\gamma)$$

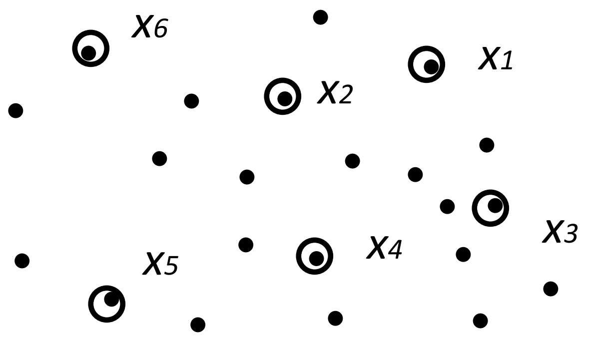
Lesson from mathematicians

$$\frac{d}{dt} k_t(\eta) = \left(L^{\Delta} k_t\right)(\eta)$$

Evolution of correlation functions

(how the model behaves, what the population does)?

Correlation functions



Let μ be a probability measures on Γ. The correlation function $k_\mu^{(n)}$ of order n of the measure μ is defined in the following way: for any symmetric function $f^{(n)} : \left(\mathbb{R}^d\right)^n \to \mathbb{R}$

$$\int_\Gamma \sum_{\{x_1,\ldots,x_n\}\subset\gamma} f^{(n)}\left(x_1,\ldots,x_n\right) d\mu\left(\gamma\right)$$
$$= \frac{1}{n!} \int_{(\mathbb{R}^d)^n} f^{(n)}\left(x_1,\ldots,x_n\right) k_\mu^{(n)}\left(x_1,\ldots,x_n\right) dx_1 \ldots dx_n. \quad (1.1)$$

Vector of all correlation functions: $k_\mu = \left(k_\mu^{(n)}\right)_{n\geq 0}$

Example: spatial logistic model

"Model definition"

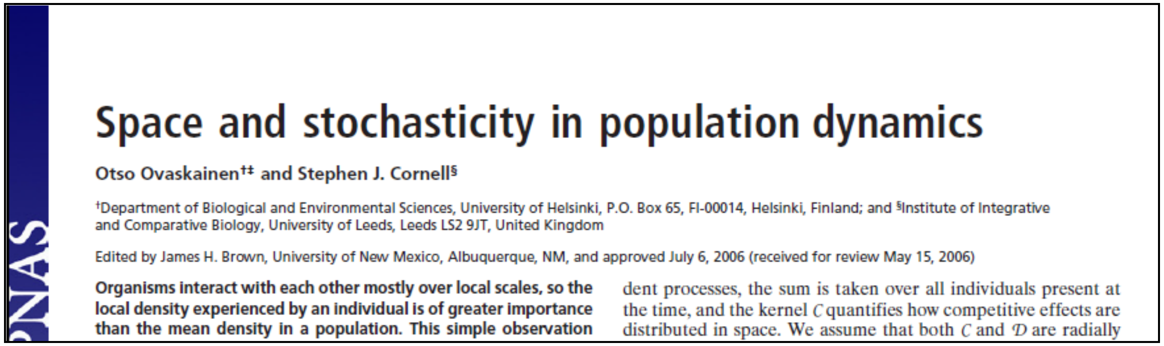
$$
\begin{aligned}
(LF)(\gamma) \quad &= \quad \sum_{x\in\gamma}\Bigg(m+\sum_{y\in\eta\setminus x} a^{-}(x-y)\Bigg)\left[F(\gamma\setminus x)-F(\gamma)\right] \\
&\quad +\sum_{y\in\gamma}\int_{\mathbb{R}^d} a^{+}(x-y)\left[F(\gamma\cup x)-F(\gamma)\right]dx.
\end{aligned}
$$

⇒

"How the model behaves?"

$$
\begin{aligned}
(L^{\Delta}k)(\eta) \quad &= \quad -\Bigg(m|\eta|+\sum_{x\in\eta}\sum_{y\in\eta\setminus x} a^{-}(x-y)\Bigg)k(\eta) \\
&\quad -\sum_{y\in\eta}\int_{\mathbb{R}^d} a^{-}(x-y)k(\eta\cup x)dx \\
&\quad +\sum_{y\in\eta}\Bigg(\sum_{x\in\eta\setminus y} a^{+}(x-y)\Bigg)k(\eta\setminus y) \\
&\quad +\sum_{y\in\eta}\int_{\mathbb{R}^d} a^{+}(x-y)k((\eta\setminus y)\cup x)dx
\end{aligned}
$$

Mathematical methods of predicting such how models behave



PNAS

Space and stochasticity in population dynamics

Otso Ovaskainen†‡ and Stephen J. Cornell§

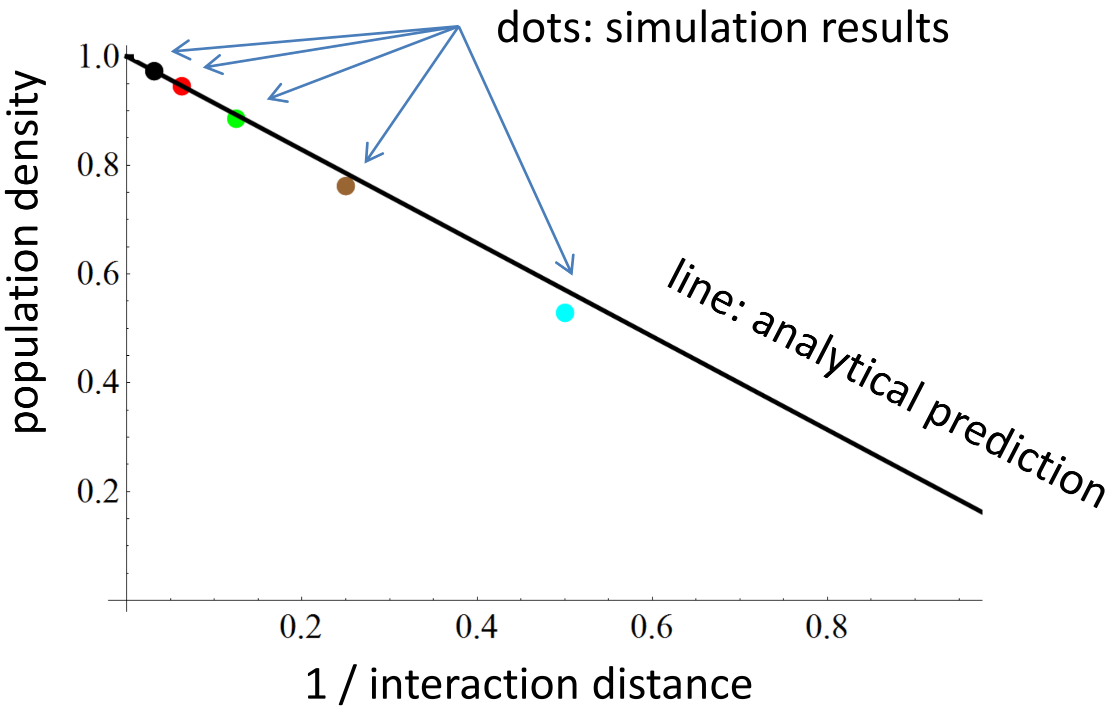
†Department of Biological and Environmental Sciences, University of Helsinki, P.O. Box 65, FI-00014, Helsinki, Finland; and §Institute of Integrative and Comparative Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

Edited by James H. Brown, University of New Mexico, Albuquerque, NM, and approved July 6, 2006 (received for review May 15, 2006)

Organisms interact with each other mostly over local scales, so the local density experienced by an individual is of greater importance than the mean density in a population. This simple observation

dent processes, the sum is taken over all individuals present at the time, and the kernel C quantifies how competitive effects are distributed in space. We assume that both C and $\mathcal{D}$ are radially

Stephen Cornell

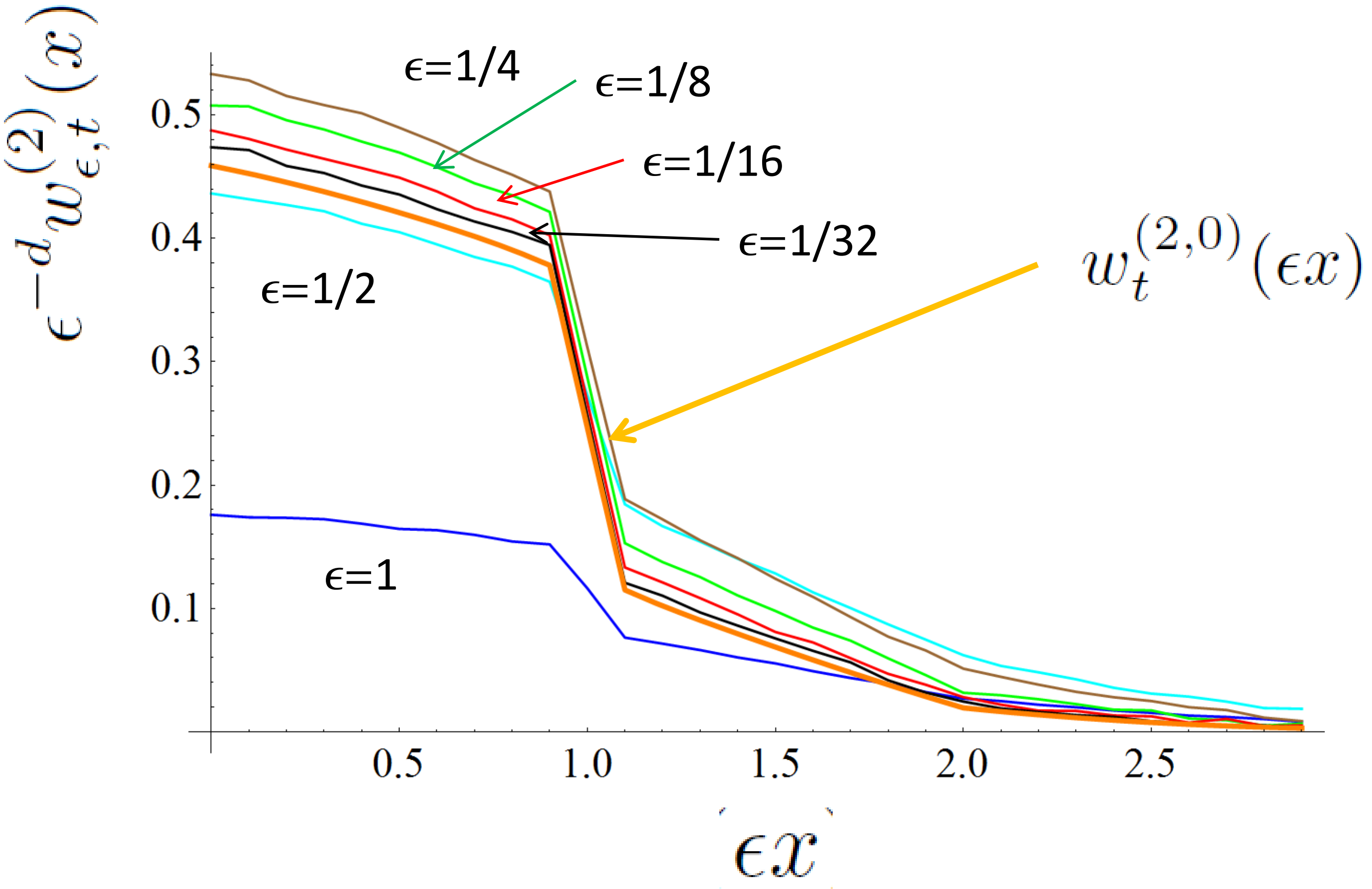


dots: simulation results

line: analytical prediction

Example model:
spatial logistic model

Example: second truncated correlation function



Ingredients for cooking your model

(Ovaskainen, Cornell, Bolker, Kutovyi, Finkelshtein and Kontratiev, in prep.)

Birth: $L = L^B(a)$

$$(LF)(\gamma) = \sum_{y\in\gamma} \int_{\mathbb{R}^d} a(x-y)(L^{x+}F)(\gamma)dx$$

Death: $L = L^D(r)$

$$(LF)(\gamma) = r \sum_{x\in\gamma} (L^{x-}F)(\gamma)$$

Death by competition: $L = L^C(a)$

$$(LF)(\gamma) = \sum_{x\in\gamma} \left(\sum_{y\in\gamma\setminus x} a(x-y) \right) (L^{x-}F)(\gamma)$$

Death by external factor: $L = L_{12}^{DE}(a), \gamma = \{\gamma_1, \gamma_2\}$

$$(LF)(\gamma) = \sum_{x\in\gamma_2} \left(\sum_{y\in\gamma_1} a(x-y) \right) (L_2^{x-}F)(\gamma)$$

Infection: $L = L_{12}^I(a), \gamma = \{\gamma_1, \gamma_2\}$

$$(LF)(\gamma) = \sum_{y\in\gamma_2} \sum_{x\in\gamma_1} a(x-y)(L_2^{x+}L_1^{x-}F)(\gamma)$$

Birth by facilitation: $L = L_{12}^{BF}(a,b), \gamma = \{\gamma_1, \gamma_2\}$

$$(LF)(\gamma) = \sum_{y\in\gamma_2} \sum_{z\in\gamma_1} \int_{\mathbb{R}^d} a(z-y)b(x-y)(L_2^{x+}F)(\gamma)dx$$

Change to another type: $L = L_{12}^{CT}(r), \gamma = \{\gamma_1, \gamma_2\}$

$$(LF)(\gamma) = r \sum_{x\in\gamma_1} (L_2^{x+}L_1^{x-}F)(\gamma)$$

Immigration: $L = L^{IM}(r)$

$$(LF)(\gamma) = r \int_{\mathbb{R}^d} (L^{x+}F)(\gamma)dx$$

Jump: $L = L^J(a)$

$$(LF)(\gamma) = \sum_{x\in\gamma} a(x-y)(L^{y+}L^{x-}F)(\gamma)$$

Birth by consumption: $L = L_{12}^{BC}(a,b), \gamma = \{\gamma_1, \gamma_2\}$

$$(LF)(\gamma) = \sum_{y\in\gamma_2} \sum_{z\in\gamma_1} \int_{\mathbb{R}^d} a(z-y)b(x-y)(L_2^{x+}L_1^{z-}F)(\gamma)dx$$

Birth to another type: $L = L_{12}^B(a), \gamma = \{\gamma_1, \gamma_2\}$

$$(LF)(\gamma) = \sum_{y\in\gamma_1} \int_{\mathbb{R}^d} a(x-y)(L_x^{2+}F)(\gamma)dx$$

Conclusions

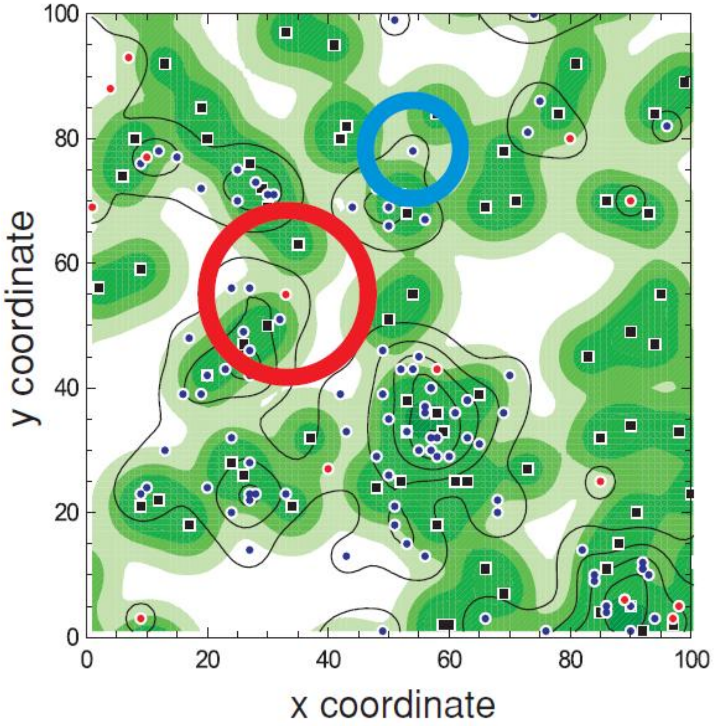
- State-space models combine a process model with an observation model. They allow one to bring biological knowledge into statistical inference, combine different data sources and use data with missing observations.

- Movement models can be integrated into models of demographic, genetic and evolutionary dynamics. Bringing different kinds of information together can help to get a more full picture.

- Hierarchical modeling approaches make it possible to build community-level models from species-specific considerations. Such approaches allow one to assess the influences of different kinds of factors (environmental covariates, species interactions, phylogenetic effects, ...) on community structure.

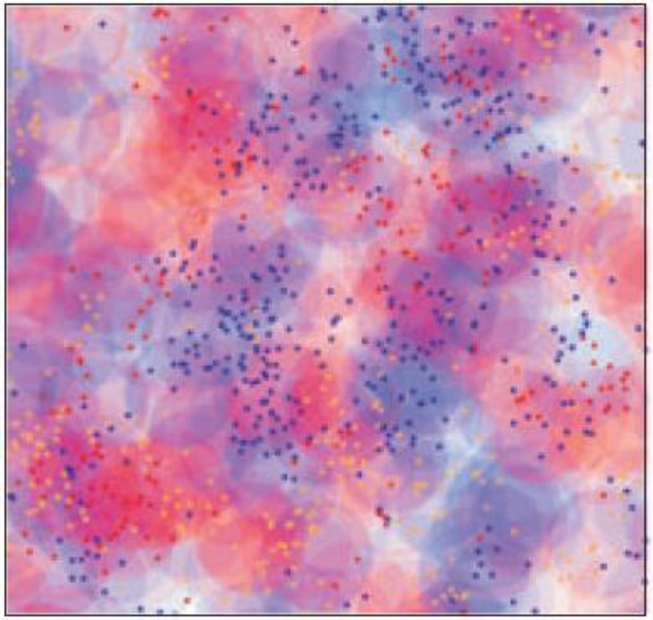
The method applies to a wide range of models

Case study 1:
evolution of dispersal distance



Ace North

Case study 2:
local adaptation

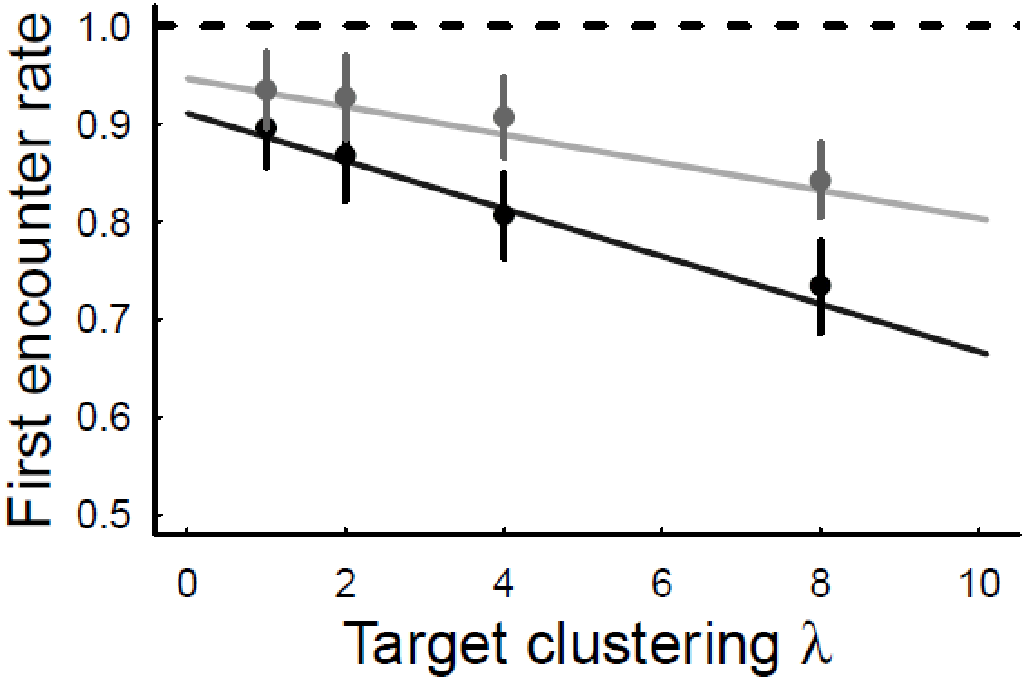


North, Cornell and Ovaskainen.
Evolution 2011

North, Pennanen, Ovaskainen and Laine.
Evolution 2010

The method applies to a wide range of models

Case study 3: encounter rates between searcher (e.g. predators) and targets (e.g. prey)

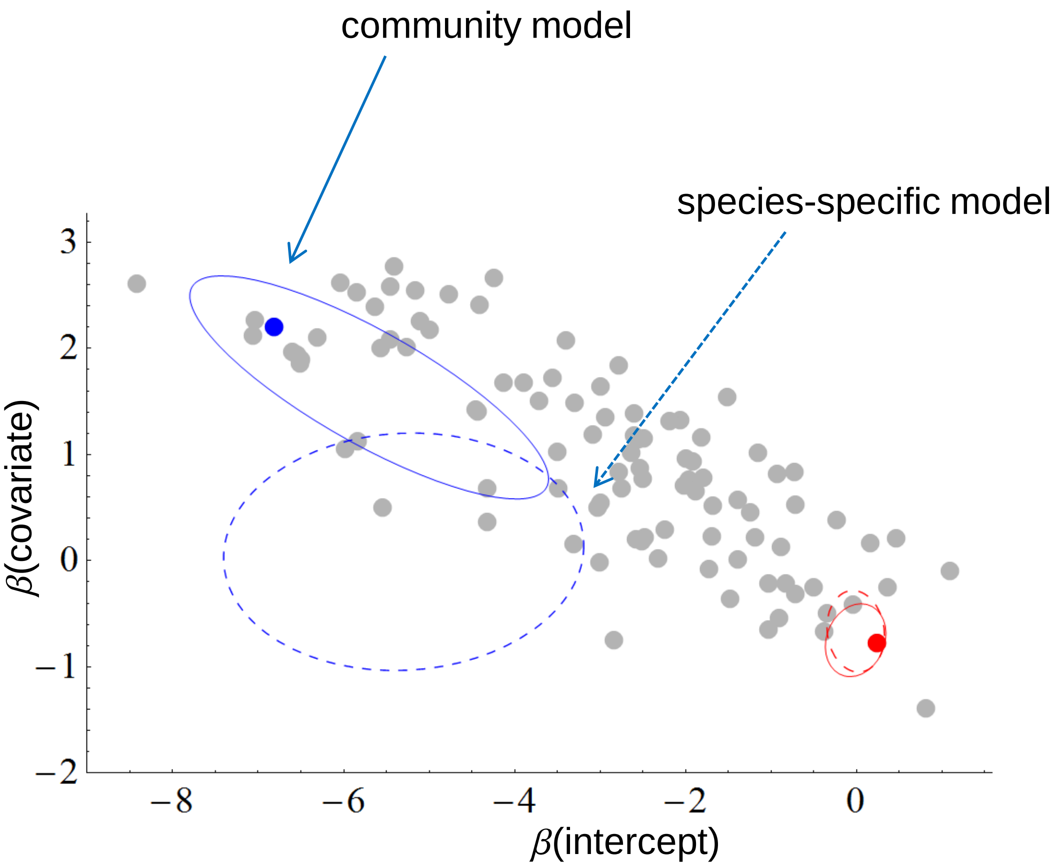


Eli Gurarie

Gurarie and Ovaskainen. Theoretical Ecology 2012

The community model gives improved estimates for species-specific parameters

The ellipses show the 75% quantiles of the parameter estimates



community model

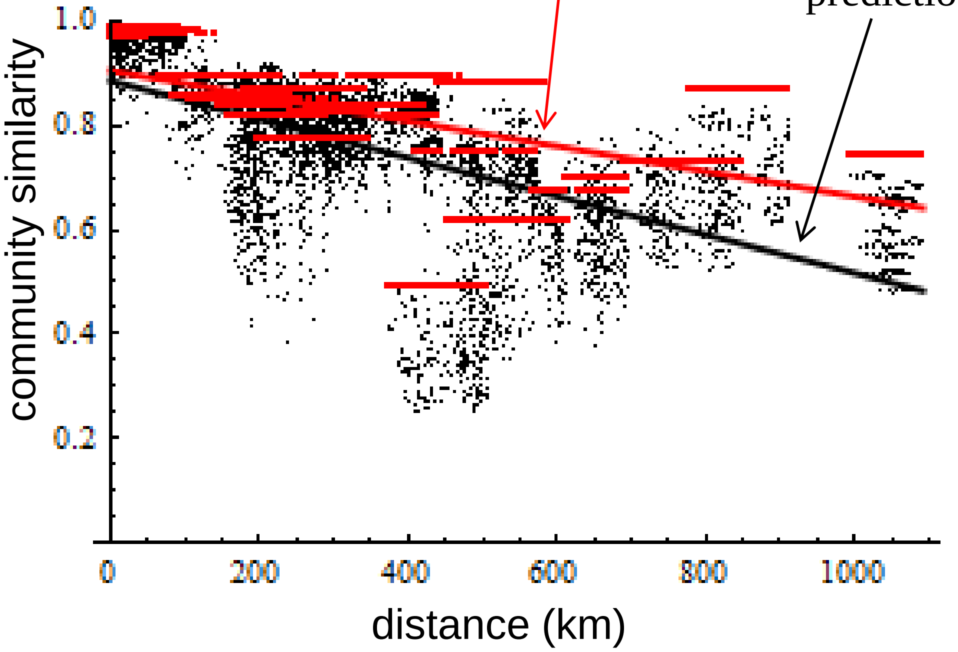
species-specific model

Inference based on the community-level parameters

$$(\beta_0, \beta_1) \sim N(\boldsymbol{\mu}, \mathbf{V})$$

Mean response (over species) to the covariates

Variation (among species) in response to the covariates, and correlation between pairs of covariates

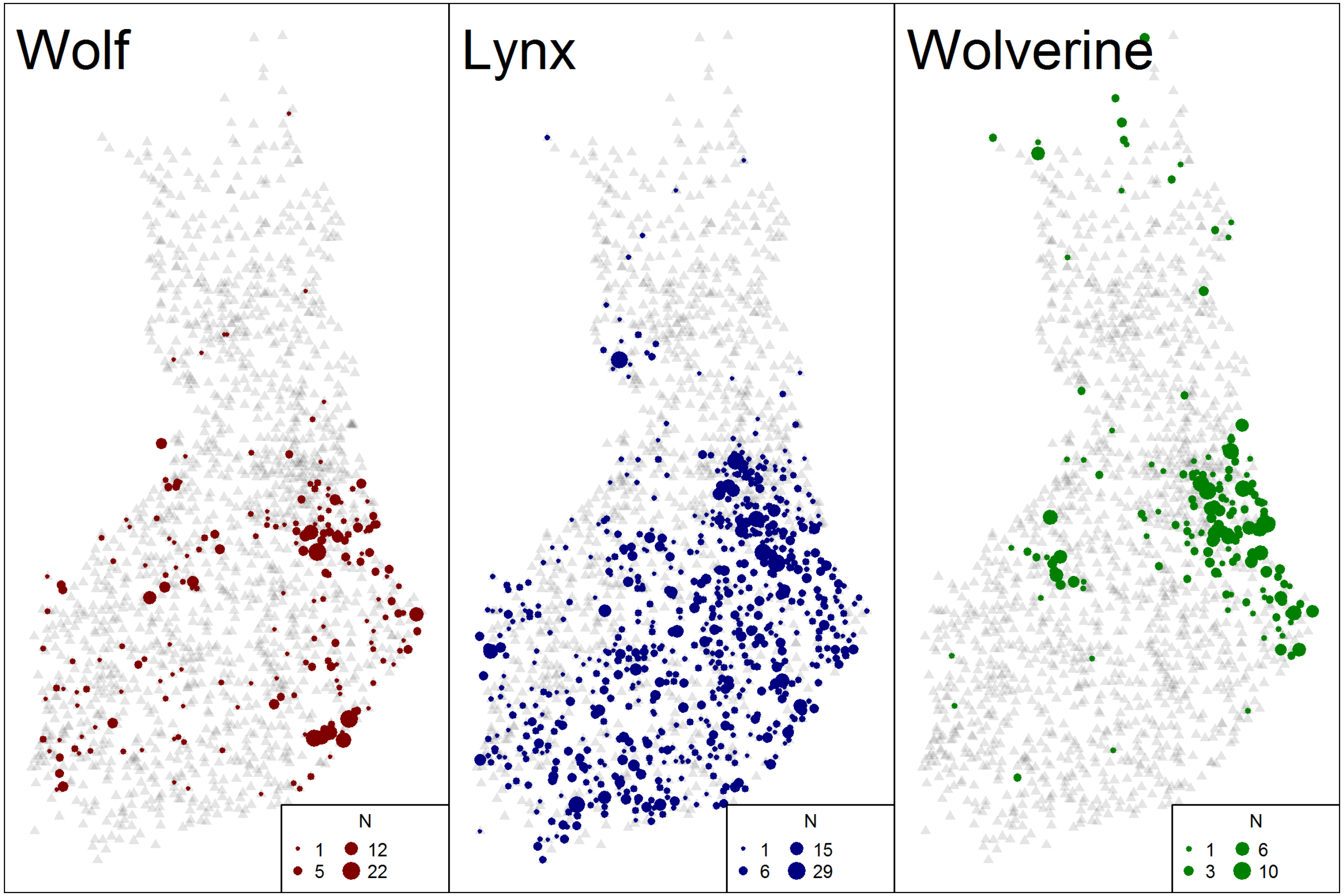

prediction with spatial covariates only

prediction with spatial and environmental covariates

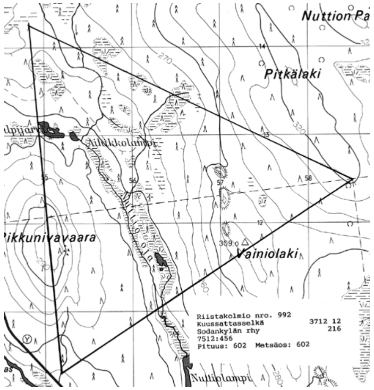
35% of the distance decay can be attributed environmental covariates parameters, 65% to spatial covariates parameters

Rare species are specialised to nutrient poor and high pH waters, whereas common species are generalists

Winter track data have been collected in Finland for ca. 30 game animal species since 1989



Graphics: Eliezer Gurarie



12 km long skiing route

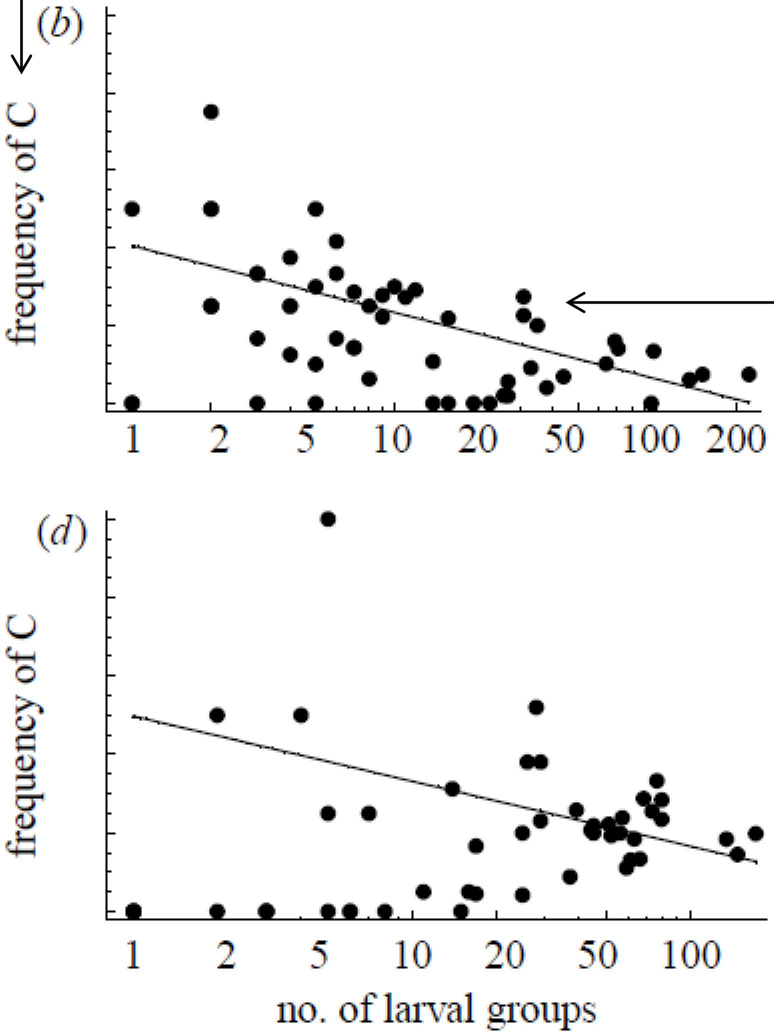
Photo: I.Kojola

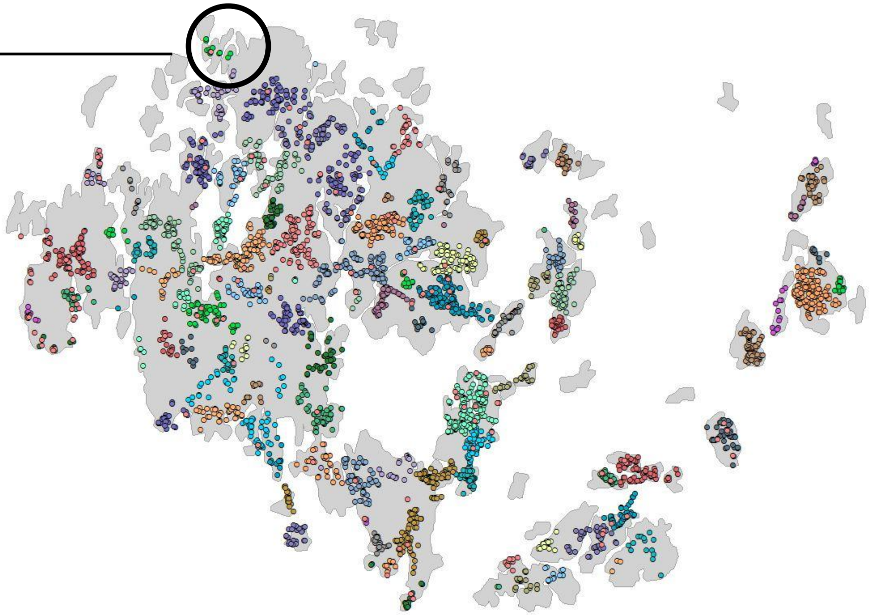
Evolutionary dynamics (evolution of dispersal): model vs. data

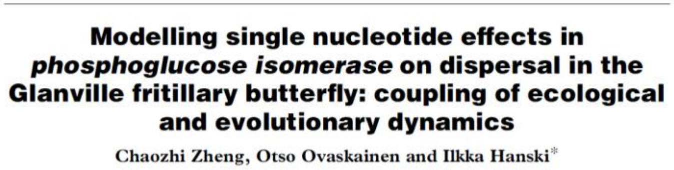
C-allele for PGI makes the individual especially dispersive

model prediction

empirical data





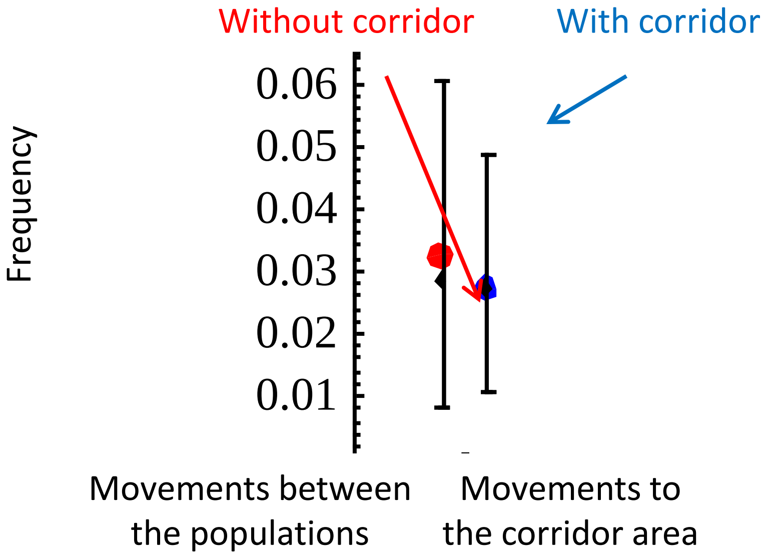


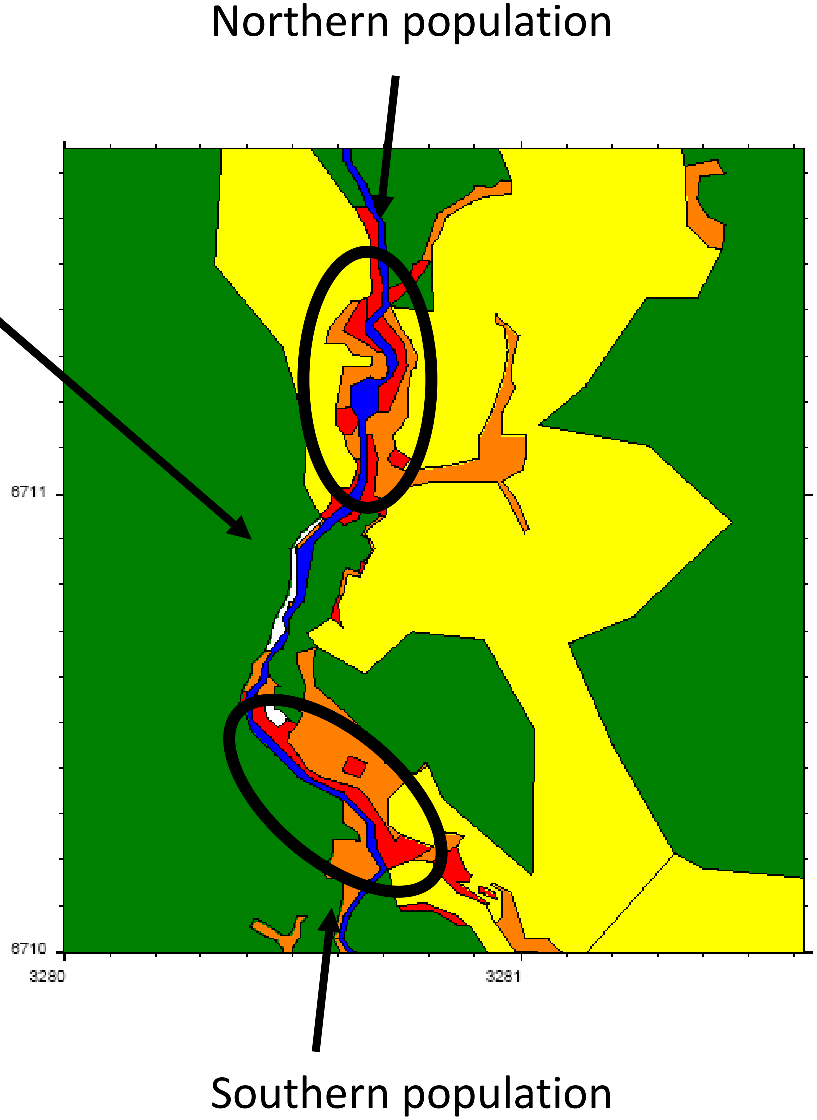
Modelling single nucleotide effects in *phosphoglucose isomerase* on dispersal in the Glanville fritillary butterfly: coupling of ecological and evolutionary dynamics

Chaozhi Zheng, Otso Ovaskainen and Ilkka Hanski*

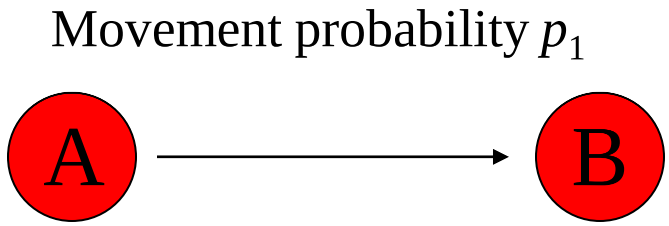
The effect of a movement corridor

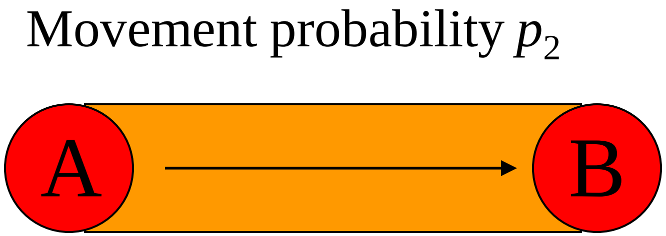
A movement corridor was cut through the forest

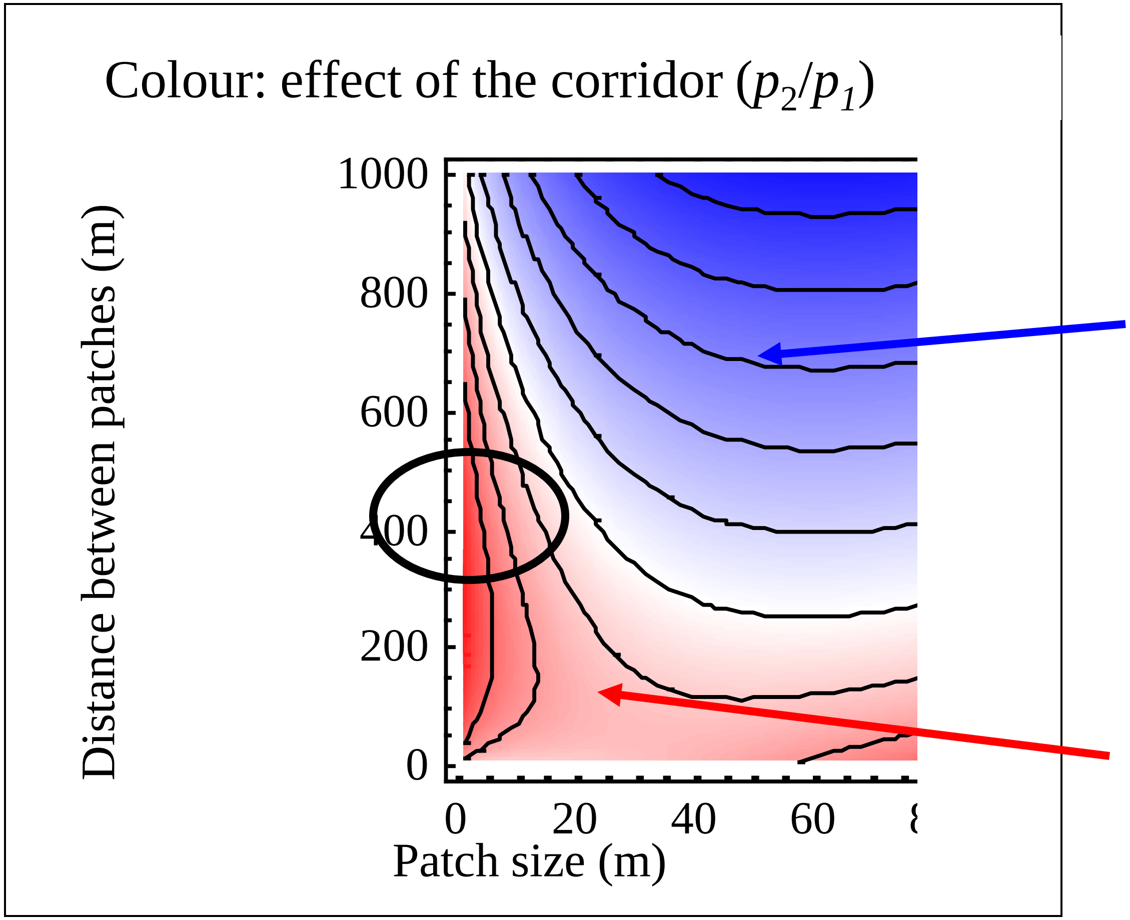




What kind of a corridor would increase movements?


Movement probability p_1


Movement probability p_2


Colour: effect of the corridor (p_2/p_1)

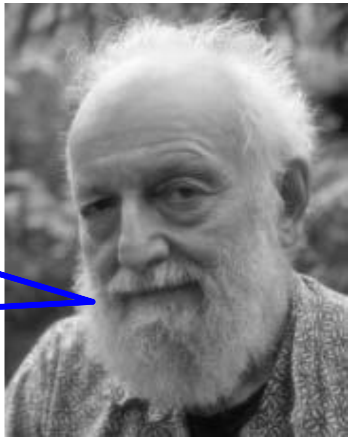
Corridor decreases movements ($p_2<p_1$)

Corridor increases movements ($p_2>p_1$)

The Levins metapopulation model

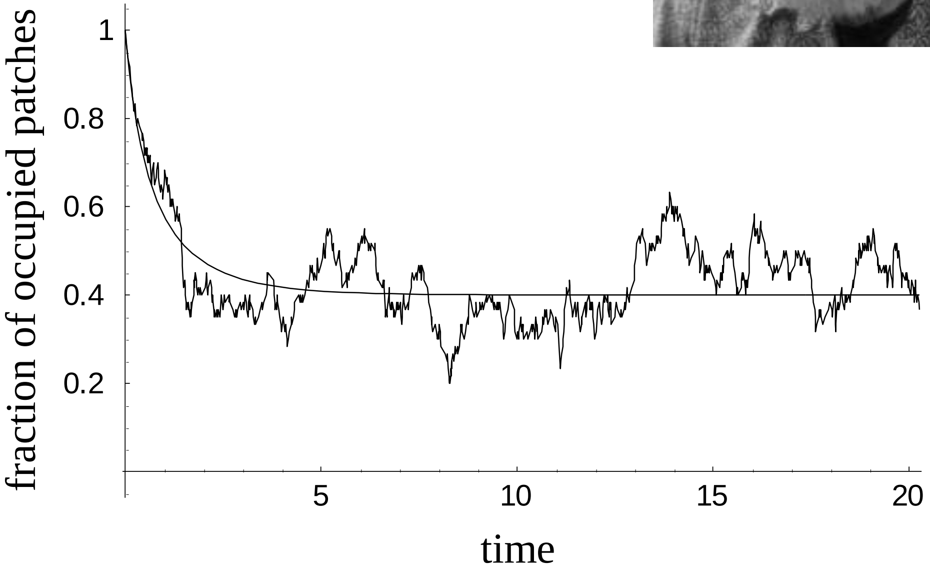
- Dynamic variable: *p*, fraction of occupied patches
- Extinction rate *e*
- Amount of suitable habitat: *h*
- Colonization rate parameter *c*

$$\frac{dp}{dt} = cph(1-p) - ep$$



Threshold condition for persistence:

$$h > e/c$$



Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Am.* **15**, 237-240.

Lande, R. 1987. Extinction thresholds in demographic models for territorial populations. *Am. Nat.* **130**, 624-635.

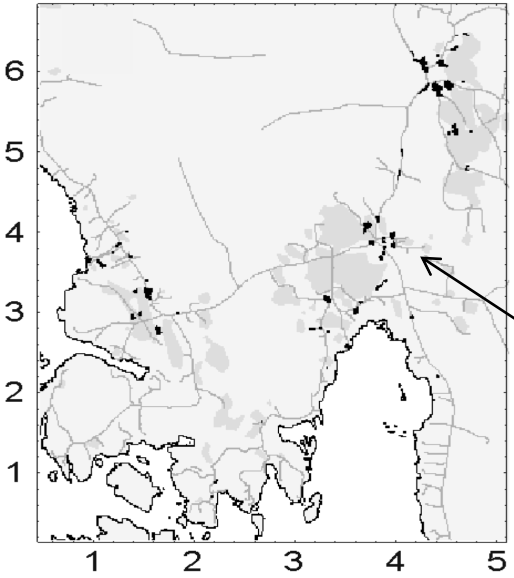
The Hanski metapopulation model

$$\frac{dp_i}{dt} = (\text{Colonization rate})_i (1 - p_i) - (\text{Extinction rate})_i \, p_i$$

- Dynamic variable: *pi*, the probability that patch *i* is occupied

- Extinction rate decreases with patch area

- Colonization rate increases with connectivity and patch area

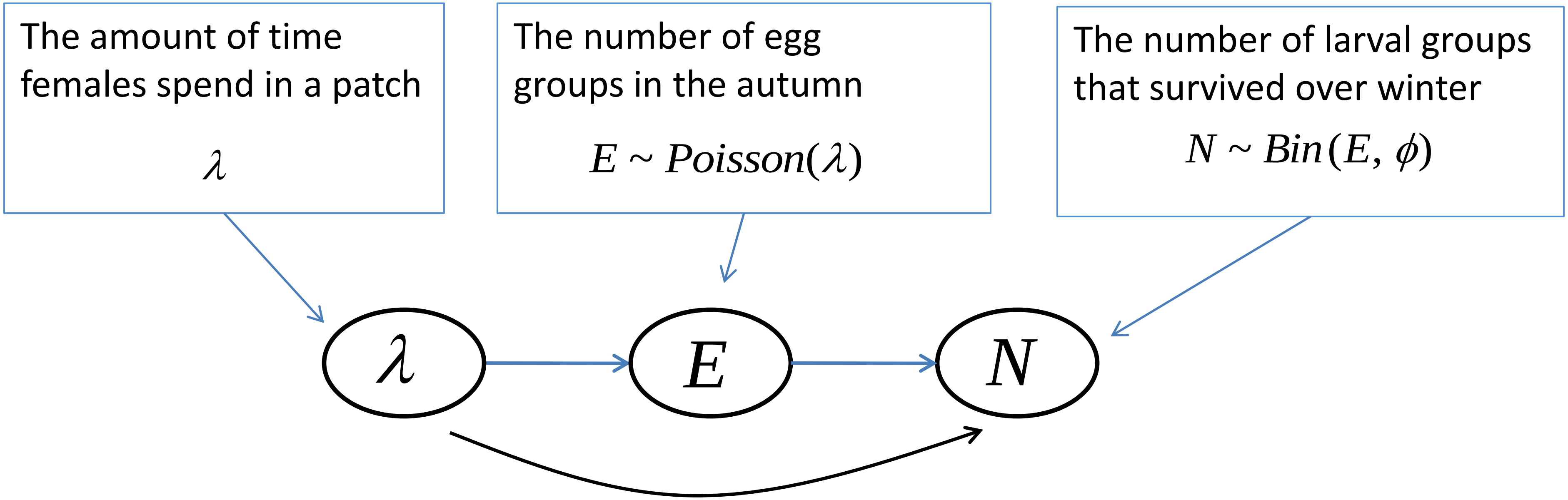


black dots: habitat patches

Threshold condition for persistence: $\lambda_M > e/c$

Hanski, I. and Ovaskainen, O. 2000. The metapopulation capacity of a fragmented landscape. *Nature* **404**, 755-758.

Deriving a stochastic patch occupancy model (SPOM) from the individual-based model (IBM)



The amount of time females spend in a patch

λ

The number of egg groups in the autumn

$E \sim Poisson(\lambda)$

The number of larval groups that survived over winter

$N \sim Bin(E, \phi)$

Can we integrate *E* out?

The probability that no egg groups successfully spin a winter nest is $P\left[N=0\right]=(1-\phi)^{E}$

Summing over the Poisson distribution gives $P\left[N=0\right]=\exp(-\phi\lambda)$

Non-random co-occurrence among species

Environmentally constrained null-models

1. Fit species-specific models independently
2. Do a randomization test for the residuals

Peres-Neto. P. R., Olden, J. D. and Jackson, D. A. 2001. Environmentally constrained null models: site suitability as occupancy criterion. *Oikos* **93**, 110-120.

Multivariate species community models

1. Fit one model for the species community

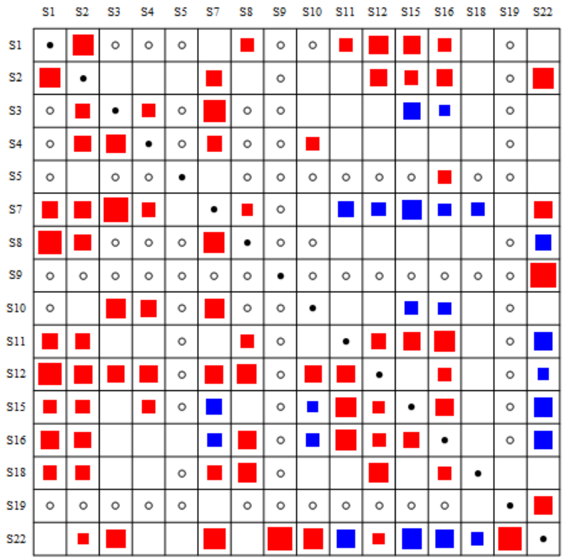
$$y_{ij} \sim x_j^T \beta_i + \varepsilon_{ij}$$

$$R_{ii'} = \mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i'j})$$

Ovaskainen, O., Hottola, J. and Siitonen, J. 2010. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology* **91**, 2514-2521.

Sebastián-Conzález, E., Sánchez-Zapata, J. A., Botella, F. and Ovaskainen, O. 2010. Testing the heterospecific attraction hypothesis with time-series data on species co-occurrence. *Proceedings of the Royal Society B: Biological Sciences* **277**, 2983-2990.

Matrix $\mathbf{R}$

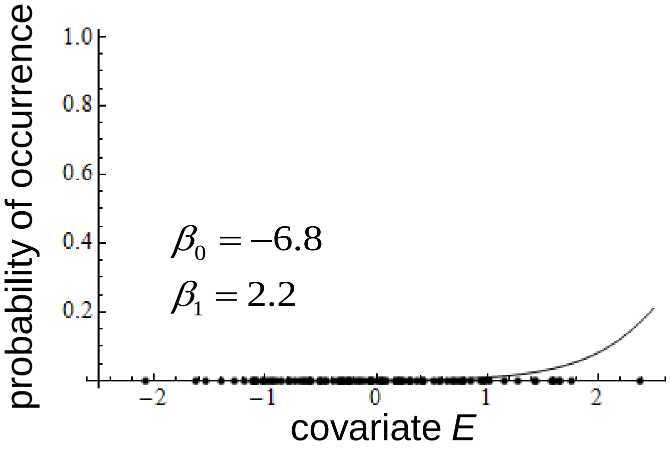


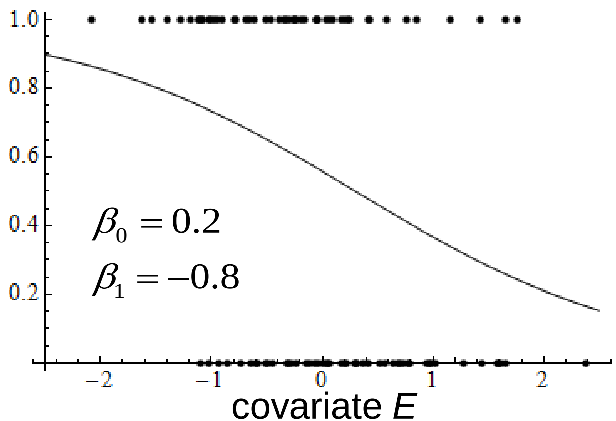
In wood decaying fungi, some species pairs co-occur more often and others less often than expected from independent occurrences

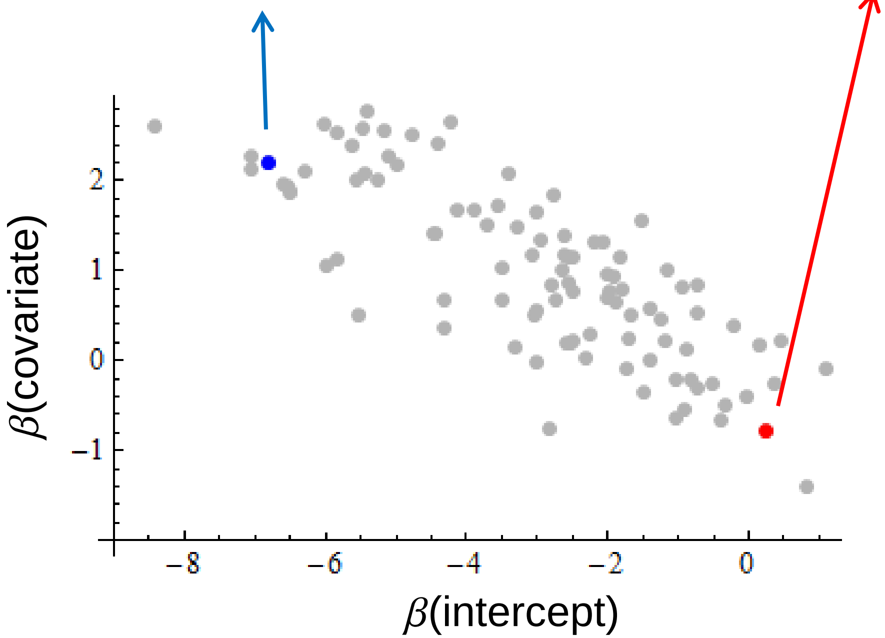
(after accounting for the covariates)

Shared responses to environmental covariates

Species-level model: $\mathrm{P}[y=1]=\mathrm{logit}^{-1}(\beta_0+\beta_1 E)$







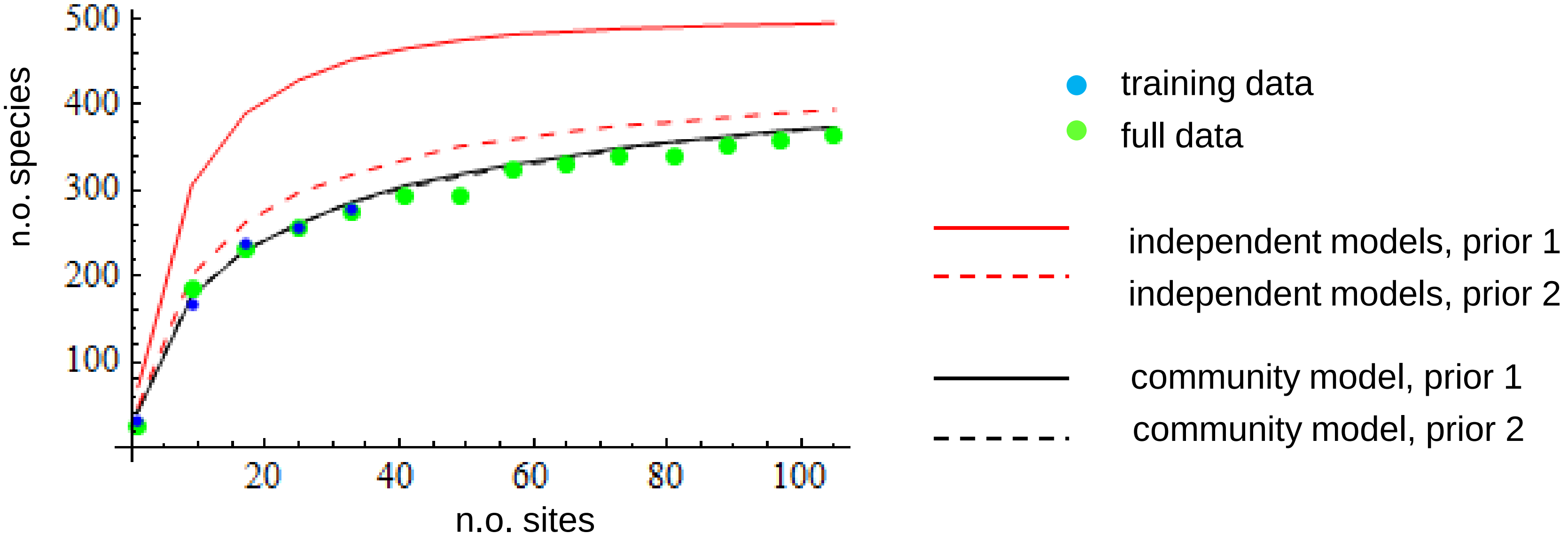
Community-level model:

$$(\beta_0,\beta_1)\sim N(\mu,V)$$

Testing the predictive power with real data (500 diatom species on 105 streams)

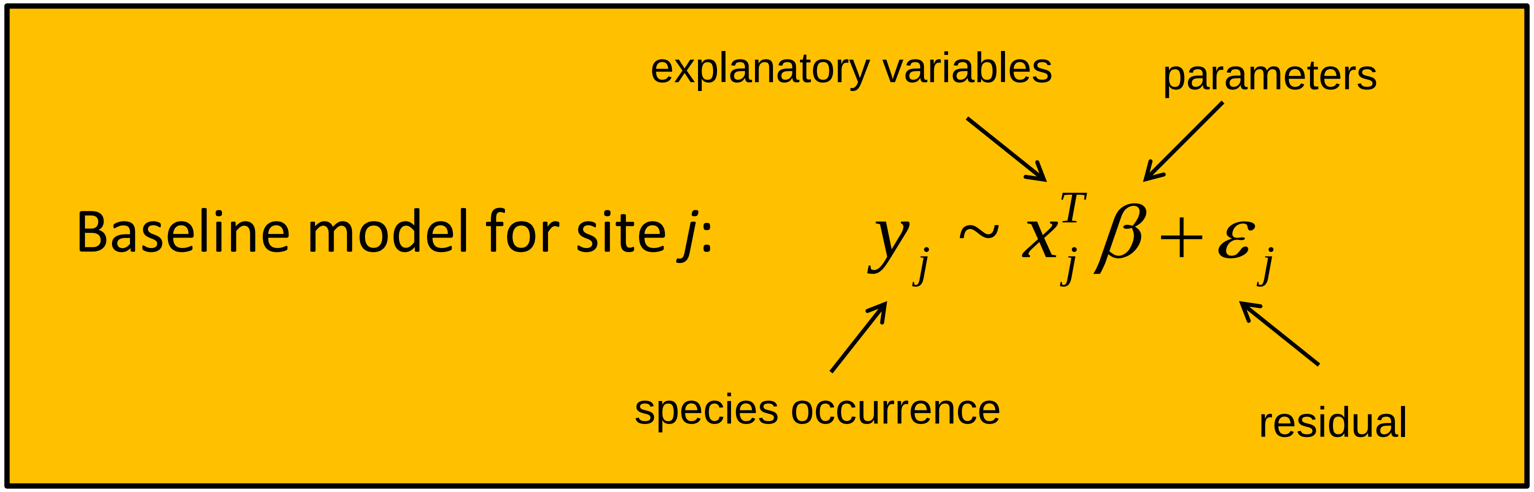
Species-level models: $$y_{ij} \sim x_j^T \beta_i + \varepsilon_{ij}$$

Community-level model: $$\beta_i \sim N(\mu,\Sigma)$$



Ovaskainen, O. and Soininen, J. 2011. Making more out of sparse data: hierarchical modeling of species communities. *Ecology* **92**, 289-295.

Hierarchical modelling approaches



explanatory variables

parameters

Baseline model for site *j*: $$y_j \sim x_j^T \beta + \varepsilon_j$$

species occurrence

residual

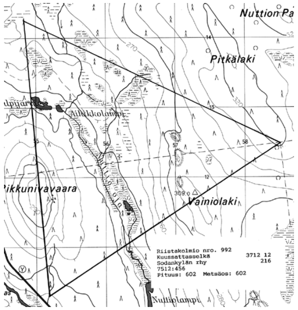
Spatial & spatio-temporal models: $$\mathrm{Cov}(\varepsilon_j, \varepsilon_{j'}) \neq 0$$

Multispecies models (*i*=species): $$y_{ij} \sim x_j^T \beta_i + \varepsilon_{ij}$$

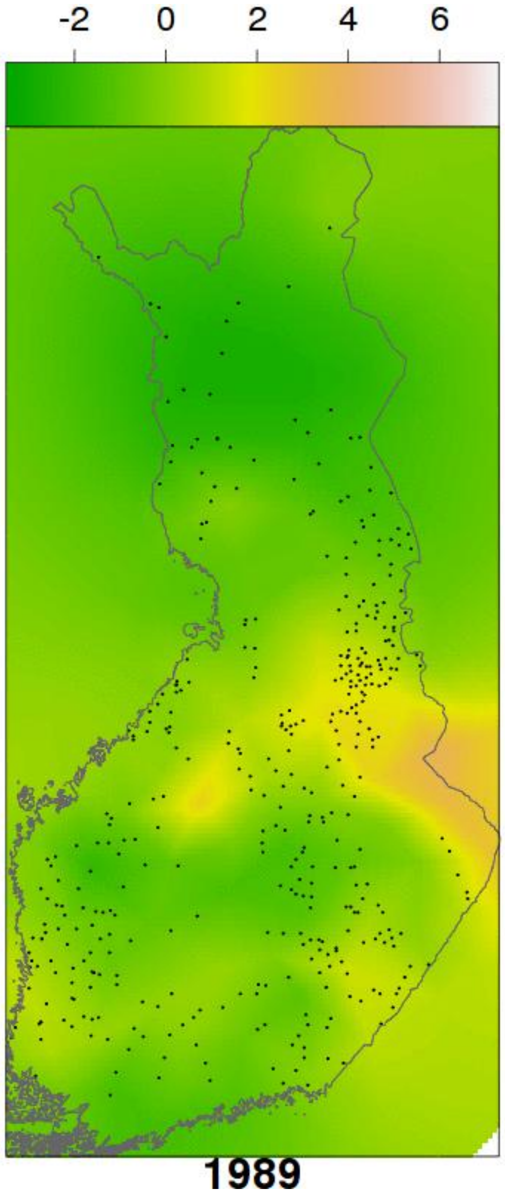
Co-occurrence patterns: $$\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i'j}) \neq 0$$

Shared responses to covariates $$\beta_i \sim \ldots$$

Spatio-temporal models (with INLA)



$$\begin{aligned}\log(\eta_t) &= \mu + \beta t + x_t + \epsilon \\ x_t &= \phi x_{t-1} + \omega_t \\ \epsilon &\sim \mathrm{MVN}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I}_n) \\ \omega_t &\sim \mathrm{MVN}(\mathbf{0}, \sigma_\omega^2 \boldsymbol{\Sigma}(\kappa)).\end{aligned}$$



Jussi Jousimo et al., *in prep.*