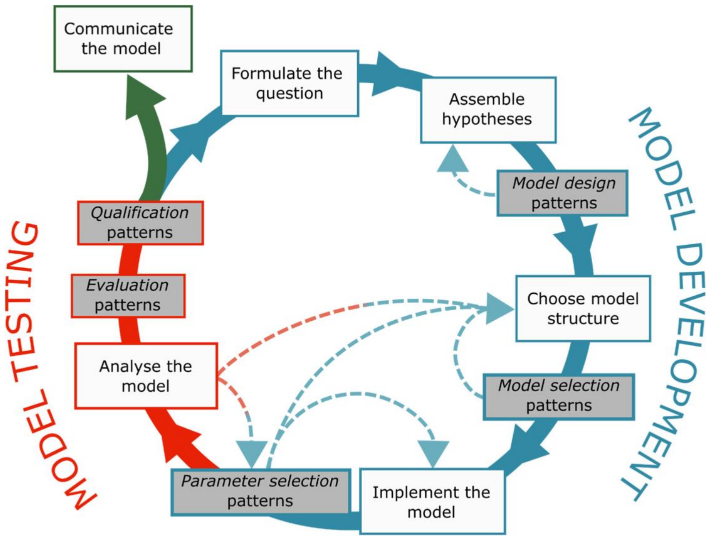
More design concepts, uncertainty, model analysis and parameterisation

THE UNIVERSITY of EDINBURGH
informatics
Modelling of Systems for Sustainability
INFR10088

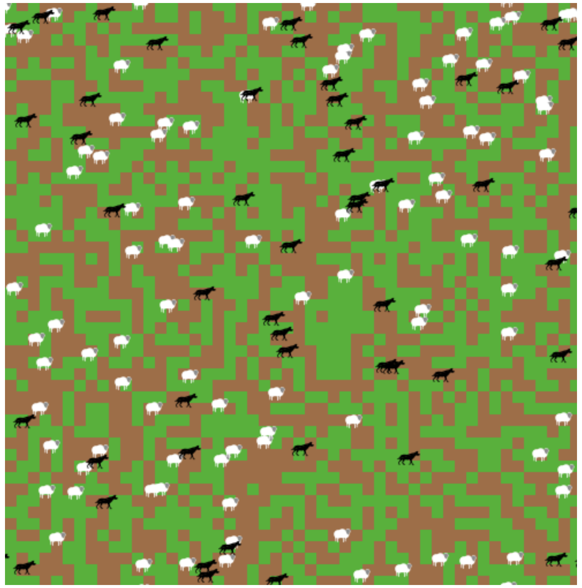
Group formation and project selection

- **Week 4, Monday-Thursday 9am:** If you have a clear idea of questions or a system for the group project (CW2):
  - Use the form Nigel shared last week to propose a project you might want to work on with others
  - Nigel and David will check the proposals to make sure they are feasible
- **Week 5, Monday-Thursday:** Register your interest in one or more project areas (e.g. social, ecological, economic) and/or proposed projects
- **Week 5, Friday:** Nigel and David will form groups that are interdisciplinary and bring together (as much as possible) students with similar interests
- **Week 6, Monday:** Groups announced

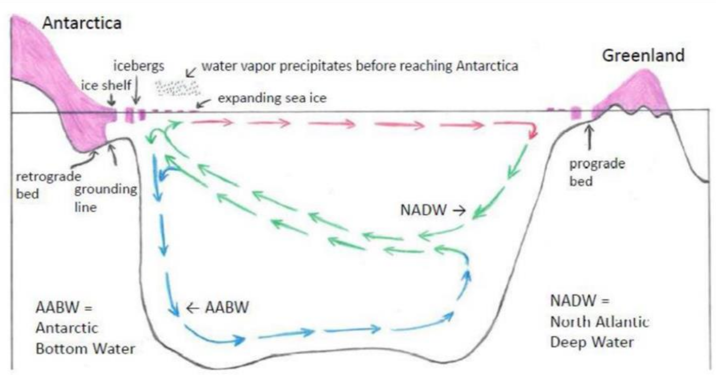
So far...



The modelling cycle

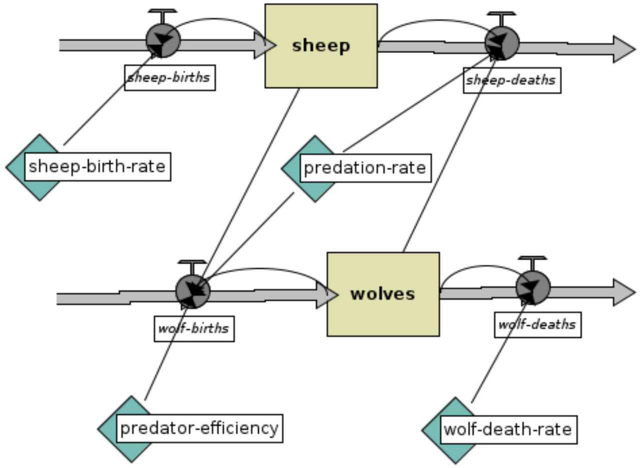


Agent-based modelling...



Real systems, e.g. Atlantic Meridional Overturning Circulation

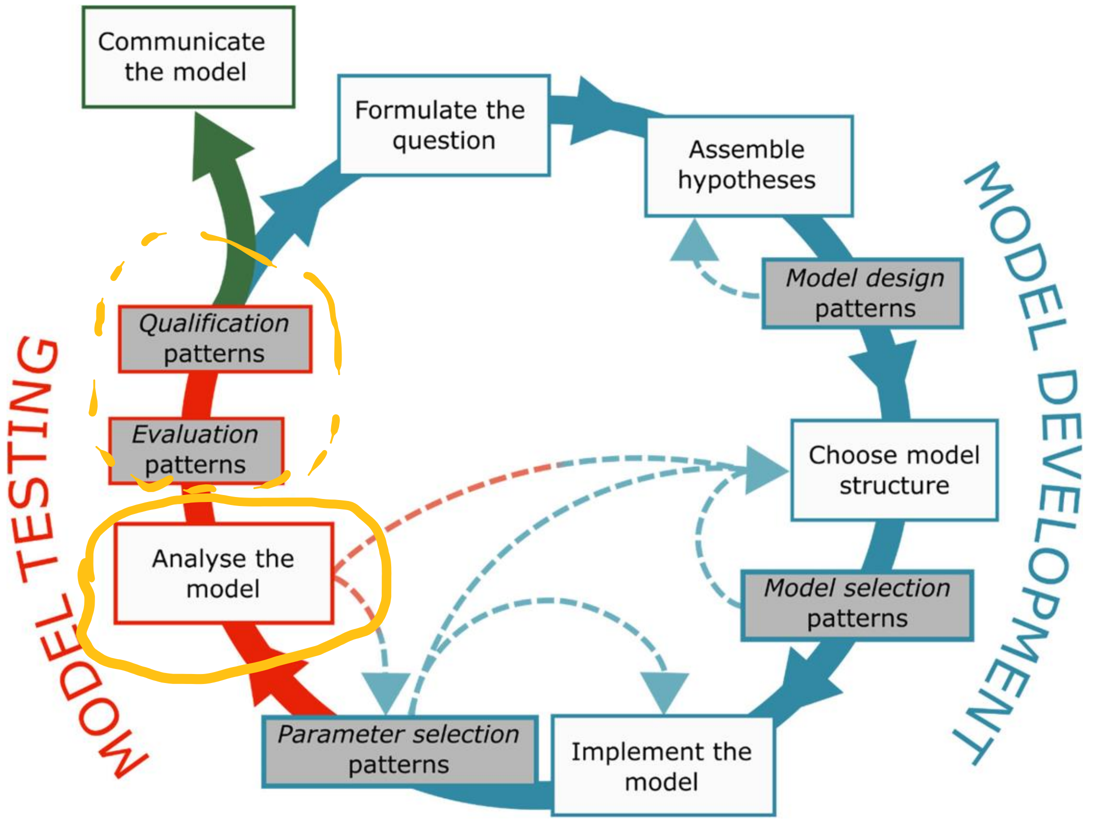
Overview
Design concepts
Details



...and system dynamics

Today...

- **D**esign concepts
- Model analysis
- Parameter selection, including patterns



Overview - aims

- Overall aim (Course Learning Outcome 2):
  investigate a sustainability system question, identify system elements and their interactions, and codify a system model using an appropriate model description framework
- This lecture:
  - Design concepts in the context of an example
    - Collectives
    - Sensing
    - Stochasticity
    - Interaction
  - Model analysis: visualisation and sensitivity analysis
  - Parameterisation and calibration

Example model: African Wild Dogs

Railsback and Grimm, Chapter 10

African wild dogs in Hluhluwe-iMfolozi Park, South Africa

- Sub-Saharan Africa's most endangered carnivore, <6000 in wild
- Can small populations exist in small dispersed habitats?
- What is the optimal reintroduction strategy?
- Gusset et al (2009, *Biological Conservation*) investigated these questions with an agent-based model

Dogs on the catwalk: Modelling re-introduction and translocation of endangered wild dogs in South Africa

Markus Gusset [a b c], Oliver Jakoby [c], Michael S. Müller [c], Michael J. Somers [b d], Rob Slotow [a], Volker Grimm [c]

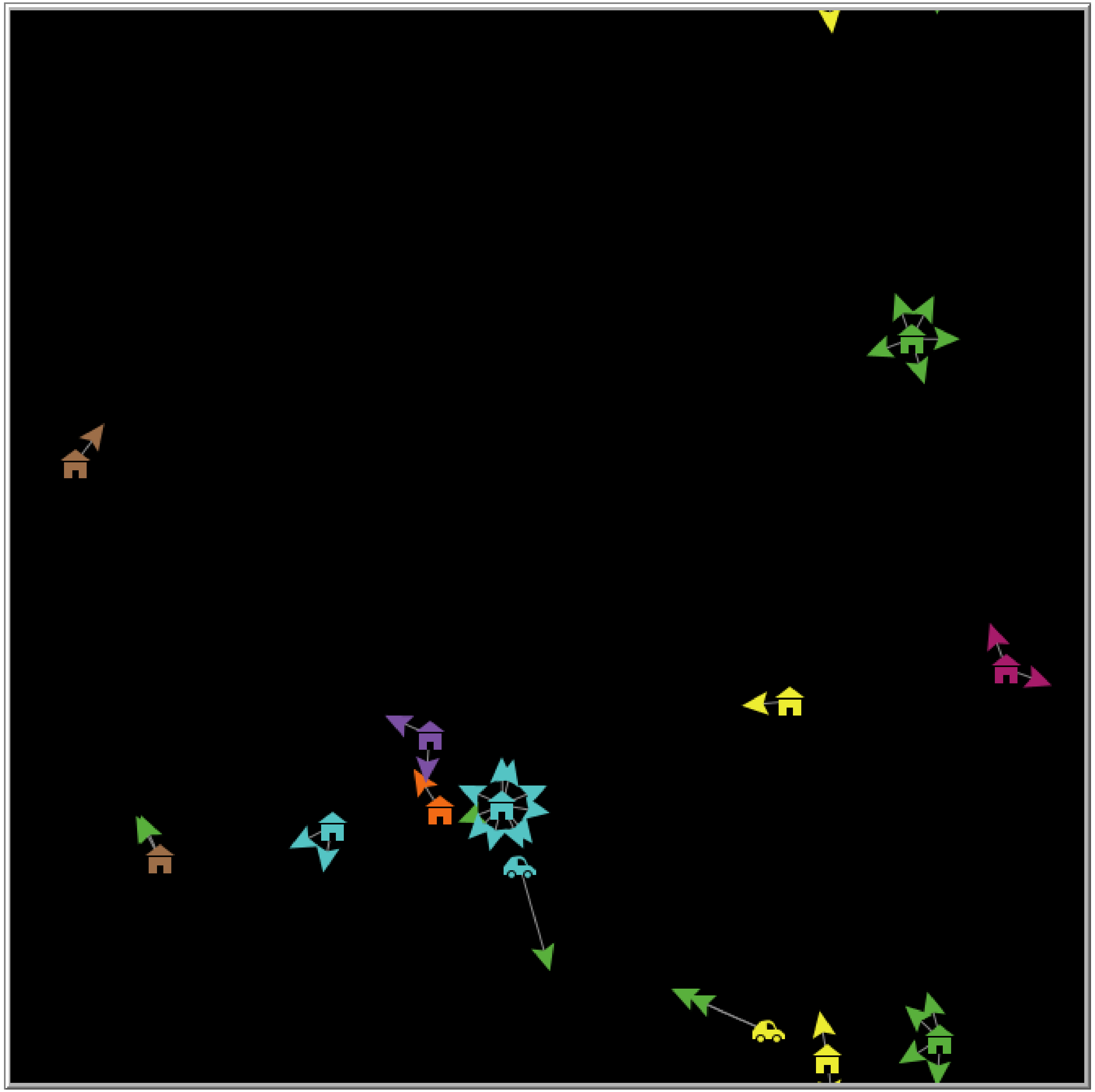
Hluhlwe-iMfolozi Park, Bjørn Christian Tørrissen, CC BY-SA-4.0. *Lycaon pictus* Charles J Sharp CC BY-SA 4.0, Wikipedia

African wild dog behaviour

- **Dogs** live in **packs** with one **alpha female** and one **alpha male**, the **only individuals that reproduce**
- **Non-alpha siblings** (subordinates) of the same sex sometimes form **disperser groups**, which search for other disperser groups
- If **disperser groups** meet, they may form a new **pack**
- Dogs in **disperser groups** are **more likely to die** than if in a pack
- Landscape has limited carrying capacity, so reproduction rate goes down as population increases

Overview of model

- Space doesn't matter (just for visualisation)
- Dogs (status: pups, yearlings, subordinates or alphas)
- Packs
- Disperser groups



Design concept: collectives

- **Collectives** are collections of co-operating individuals or entities
- Two types:
  - **Emerging** from simple rules – e.g. flocking behaviour
  - **Explicitly coded** as collective agent
- Explicitly coded type
  - Own state variables, including list of individuals belonging to the collective
  - Submodels
  - NetLOGO implementation: **breeds** (think wolves & sheep)
- Confusingly `dogs` are also `turtles`

Collectives in African Wild dog example

```
breed [dogs dog] ; agent
breed [packs pack] ; collective
breed [disperser-groups disperser-group] ; collective

dogs-own

[

  age

  sex

  status

]

hatch-dogs 1 [ … ]
```

Design concept: sensing

- Sensing is about what information agents have, including what they can know about other agents
  - For example, how does a dog in a pack know who the alphas are
- Sensing can also relate to how reliable the information is
  - For example, in a business situation, how reliable is a salesperson's estimate of the profit from a particular investment
- NetLOGO concept of `links:` exist between pairs of agents at any locations in the environment, and allow information sharing
- Links can be directional (from/to) or bidirectional

Sensing: connection between dogs and pack

```
create-packs initial-num-packs
  [ ; now in pack context
    let num-dogs random-poisson initial-mean-pack-size
    hatch-dogs num-dogs
    [ ; now in dog context
      ...
      ; create a link between the dog and its pack
      create-link-with myself ; "myself" is the pack
      ...
    ] ; end of hatch dogs
```

Design concept: stochasticity

- **Stochastic** describes processes that depend at least party on random numbers and events; cf **deterministic**
- Choices to make and consequences of stochastic models
  - What probability distribution?
  - What parameters for the distribution, or how do they depend on other simulation quantities?
  - Need to run replications to understand how much of variability is due to stochastic processes
- Random number generation is only *pseudo-random*
- Set **seed** to replicate behaviour of particular model, set seed
  - But **do not** to get replicates and **be careful** in BehaviourSpace!

Stochasticity: African dog distributions

```
let num-dogs random-poisson mean-birth-rate
```

- Why Poisson?
- Other distributions are built in, e.g.

```
random ; uniform distribution, already seen
random-normal
random-gamma
random-exponential
```

Design concept: Interaction

- **Interaction**: how agents communicate with or affect each other, such as by exchanging information or competing for resources
- Indirect interaction
  - for example, competition for a limited resource
- Direct interaction
  - For example, dispersing packs meeting one another
- Interactions can be **global** or **local**
  - The space need not be geographic – for example connections between relatives in different countries

Direct interaction in African Dogs example (highlights)

```
to do-pack-formation ; disperser-group context
  let other-groups other disperser-groups ; agentset
  let source-group self
  hatch-packs 1 [ ; new pack context
    if (([sex] of other-group) != sex and
      ([natal-pack-ID] of other-group != natal-pack-ID)) [
        let all-dogs (turtle-set [link-neighbors] of source-group
                          [link-neighbors] of other-group)
        ask all-dogs [ create-link-with myself ]
    ask other-group [die] ; get rid of disperser groups
    die ; get rid of myself (this disperser group)
  ]
end
```

Insights from building the model

- Test subroutines: code up one-by-one
- Print helpful output
- Click forward step-by step with "Go"
- Create plots to assess behaviour

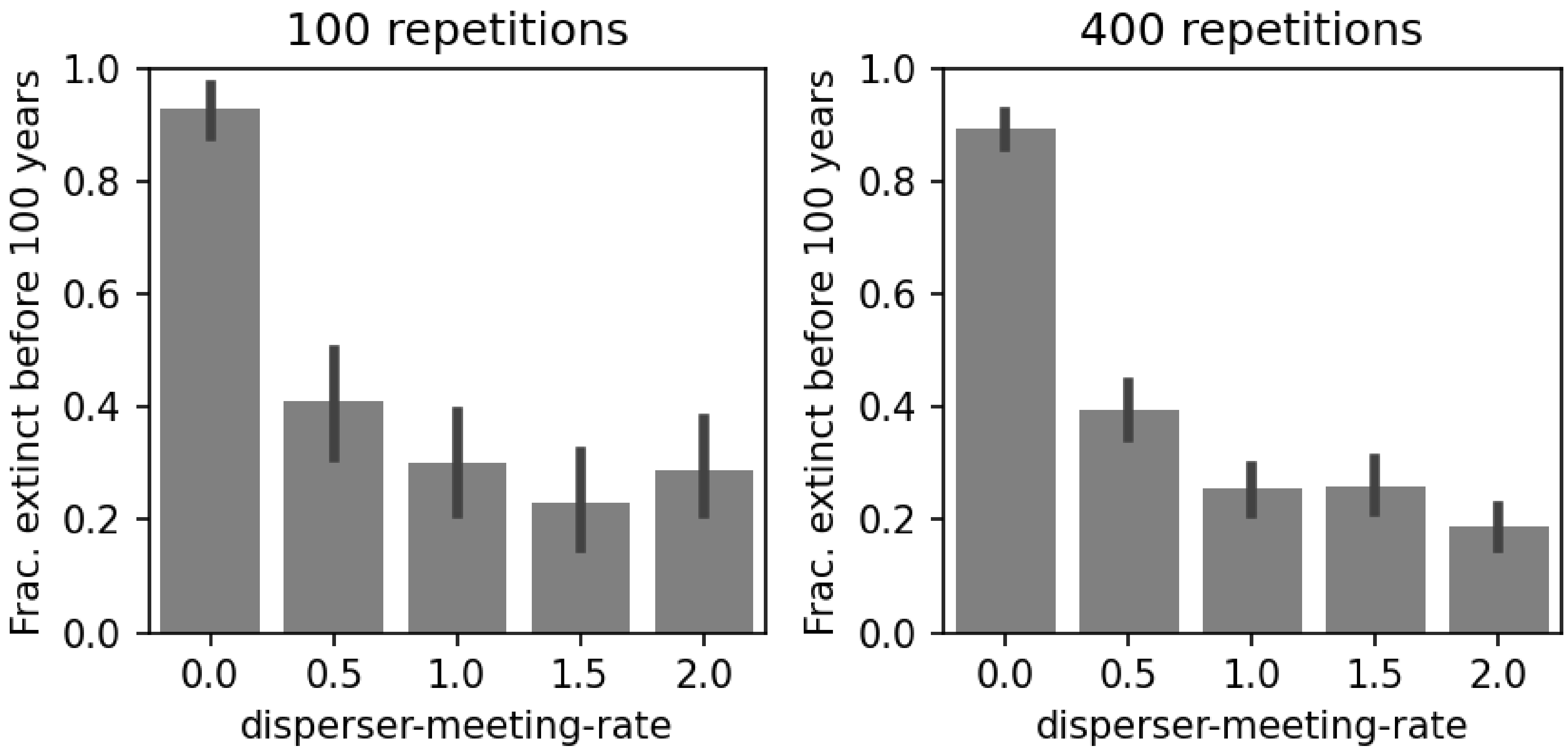
Expressing uncertainty

Characterising results

- **Define** and **observe** a **statistic** of the simulation
  - E.g. Corridor width for butterfly hill-topping
  - Wild dogs: are the dogs extinct by 100 years? (True/false)
- If simulation is **stochastic,** results will vary from run to run
  - Hence, summarise statistics from multiple runs, e.g. using mean
- The **"true" mean** would derive from an infinite number of runs
  - **Estimate of mean** based on a practical number of runs is **uncertain**
  - If possible, visualise/state uncertainty as **confidence interval** inversely proportional to the square root of the number of runs
- Best practice: indicate confidence intervals graphically and in tables

Uncertainty with 1 parameter: Disperser Meeting Rate

- Could control rate, e.g. using fencing
- Error bars are 95% confidence intervals
  - Always state what measure of uncertainty you are using!
- Interpretation:
  - no meetings definitely bad
  - Things improve quickly up to a rate of 1.0

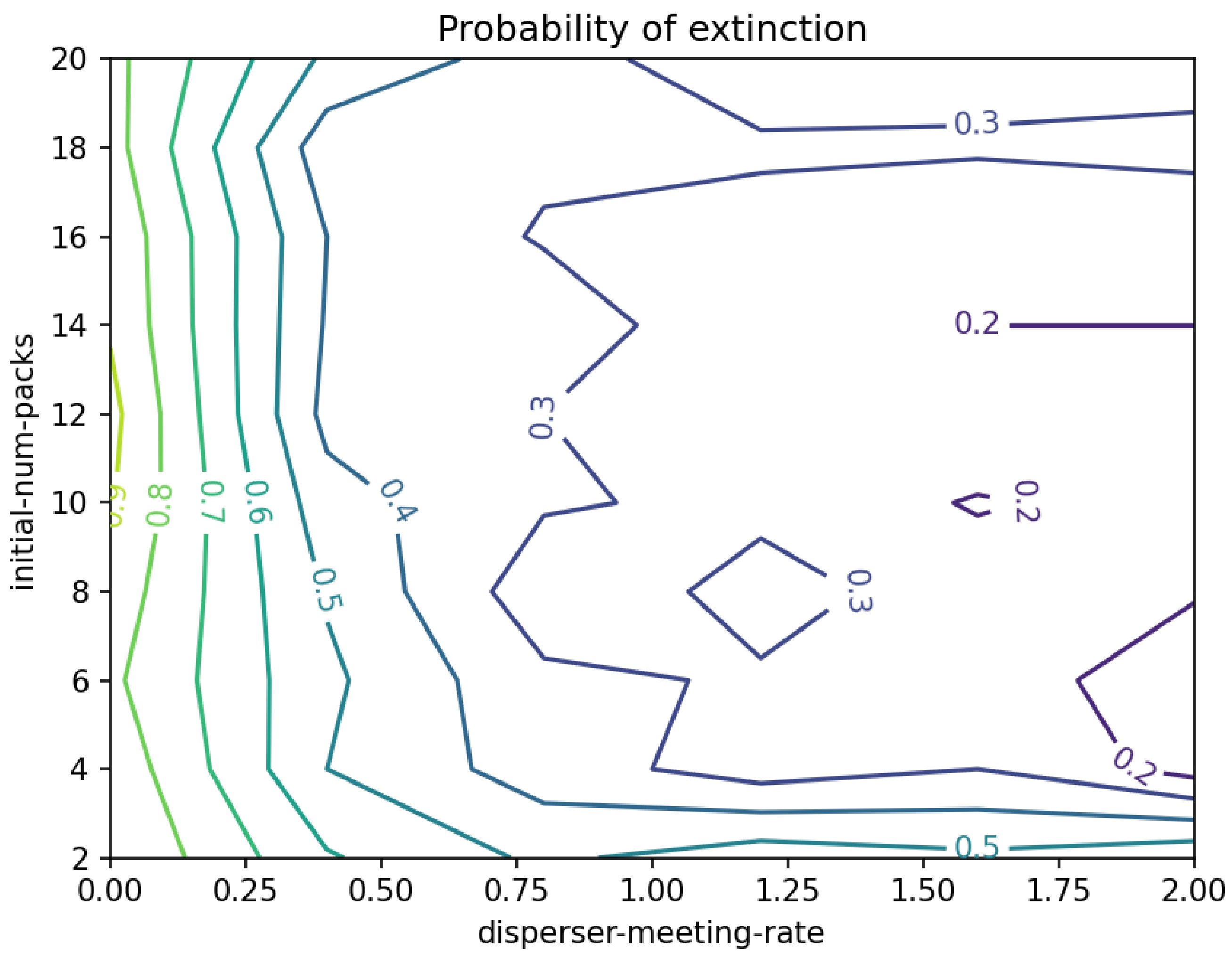


95% C.I. = sqrt(p*(1-p)/n)*1.96

Model analysis

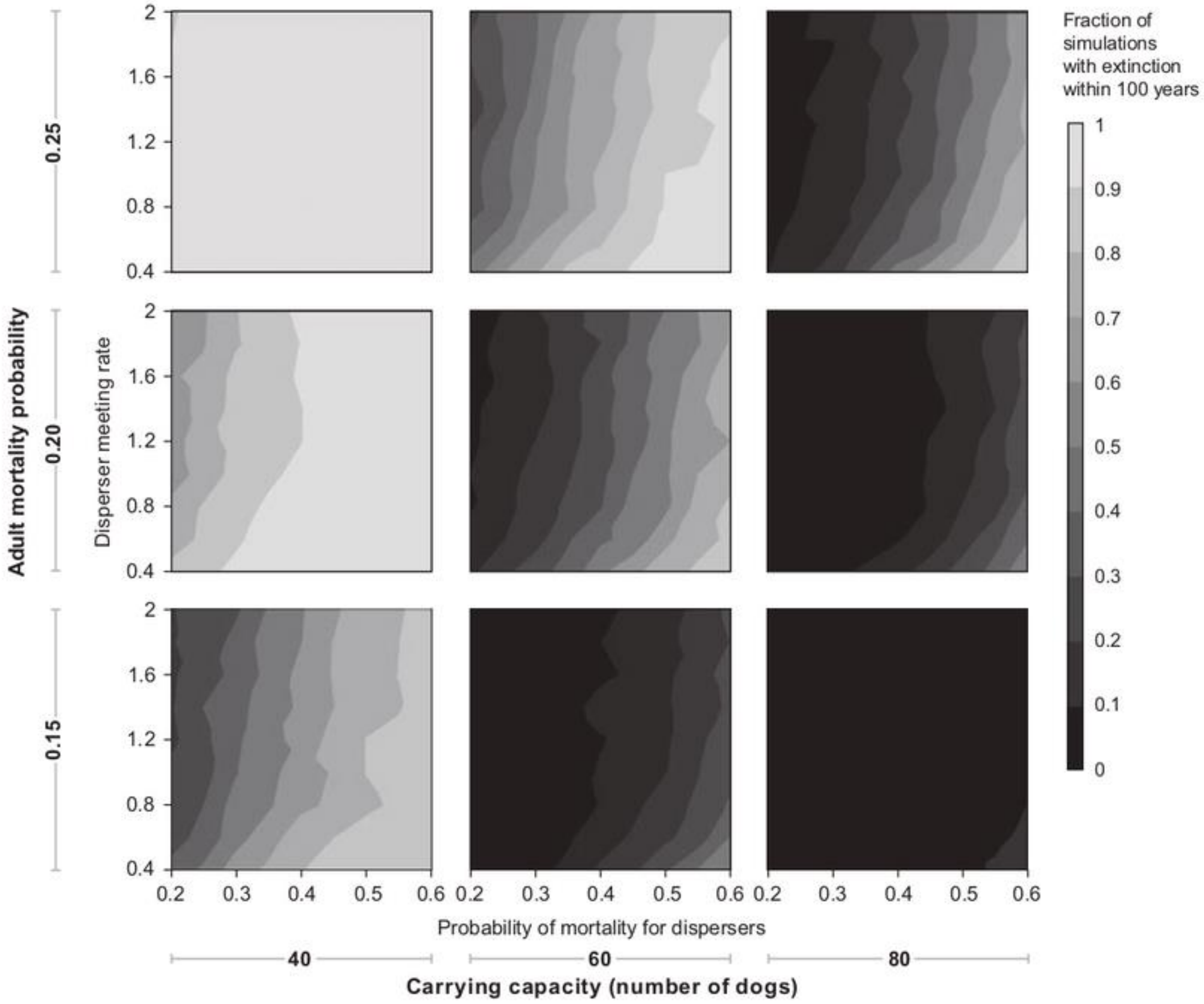
Effect of 2 parameters: Disperser meeting rate and initial number of packs

- Contour plots for visualisation
- Confidence intervals can't be shown easily
- Interactions visible
- Interpretation?



Visualising effect of 3+ parameters

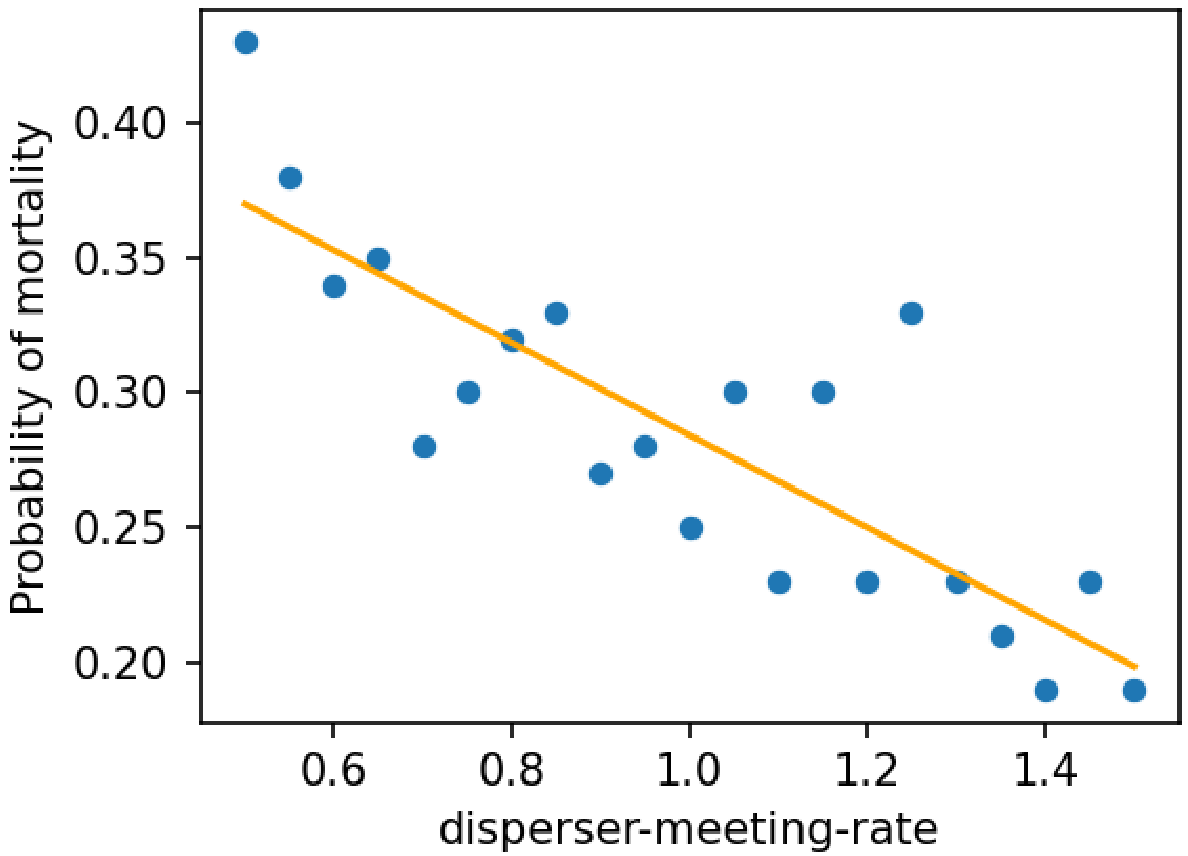
- More challenging that two parameters...
- … but possible to an extent by using "small multiples" (aligned plots)



Railsback & Grimm (2019)

Single-parameter local sensitivity analysis

- Assume we have determined a default set of parameters
- Vary **one** parameter around this standard value
    - For deterministic or minimally stochastic simulations, could be ±5% of parameter range
    - Here ±50% due to high level of stochasticity
- Examine gradient, which is the **local sensitivity**



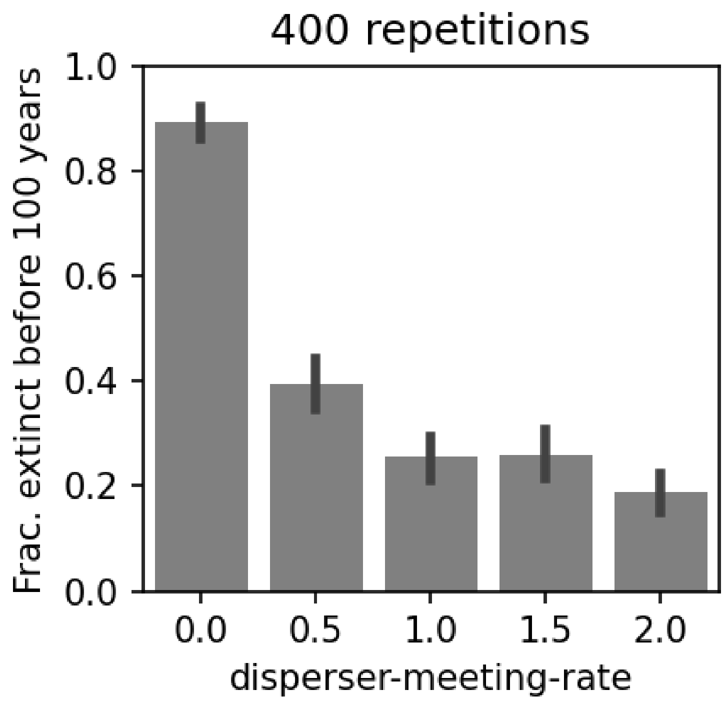
Single-parameter local sensitivity analysis: regression method

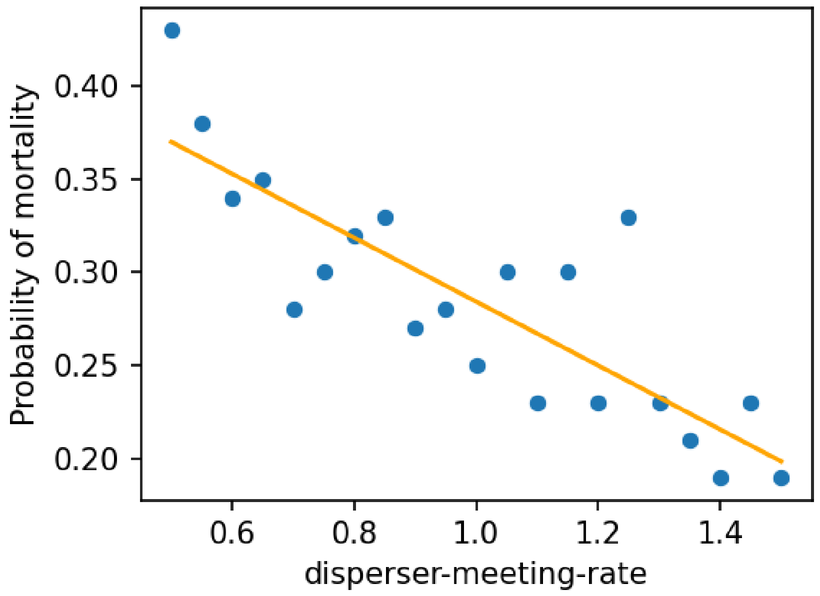
- Here there the survival fraction decreases by 17±3% for every unit increase with the disperser meeting rate.
- **Question:** does this interpretation have problems?

```
                                 OLS Regression Results
==========================================================================================
Dep. Variable:     Frac. extinct before 100 years   R-squared:                       0.013
Model:                                        OLS   Adj. R-squared:                  0.013
Method:                             Least Squares   F-statistic:                     28.06
Date:                            Mon, 09 Oct 2023   Prob (F-statistic):           1.30e-07
Time:                                    13:51:31   Log-Likelihood:                -1294.0
No. Observations:                            2100   AIC:                             2592.
Df Residuals:                                2098   BIC:                             2603.
Df Model:                                       1
Covariance Type:                        nonrobust
==========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                     0.4555      0.034     13.491      0.000       0.389       0.522
disperser-meeting-rate   -0.1712      0.032     -5.298      0.000      -0.235      -0.108
==============================================================================
Omnibus:                      861.038   Durbin-Watson:                   2.116
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              405.314
Skew:                           0.938   Prob(JB):                     9.71e-89
Kurtosis:                       1.944   Cond. No.                         6.76
==============================================================================
```

Global sensitivity analysis

- The **local** sensitivity analysis can be misleading when there are nonlinear dependencies of statistics on parameters
- **=> Global sensitivity analysis:** various methods for large parameter spaces
    - But visualisation possible for up to 4 parameters





Parameterisation and Calibration

Parameters

- **Parameters**: the constants used in equations (system dynamics & ABMs) and algorithms (ABMs only)
  - E.g. randomness q in the Butterfly hill-topping model
  - The efficiency in the Fisheries model
- **Parameterisation:** the general process of setting parameters:
  - Some parameters may be well known, measurable or estimatable: e.g. average number of pups in a litter
  - Some parameters may be harder to estimate, e.g. carrying capacity of environment
- **Calibration (**aka **parameter fitting**) is the process of adjusting parameters to match observed **patterns**

Pattern-oriented modelling

- Idea: find as many possible observed **patterns** that the model should produce
- => Greater likelihood of falsification, hence "Strong inference" (Platt, 1964) if patterns are produced
- **Qualitative patterns are OK**: "Many – perhaps most – of the great issues of science are qualitative, not quantitative, even in physics and chemistry" (Platt, 1964)
  - E.g. Discovery of DNA

Calibration/Parameter-fitting

- **Categorical calibration**: produce results within a category or range we deem acceptable
  - E.g., do we have collapses and regrowth in a fisheries model?
  - Is the period roughly that observed historically?
- **Best fit calibration:** Find the set of parameters that minimises an objective function
  - E.g., how closely can we fit the observed mean and standard deviation of the population of wild dogs?
  - Easier with deterministic models
- Increases in difficulty with the number of parameters

Calibration strategies

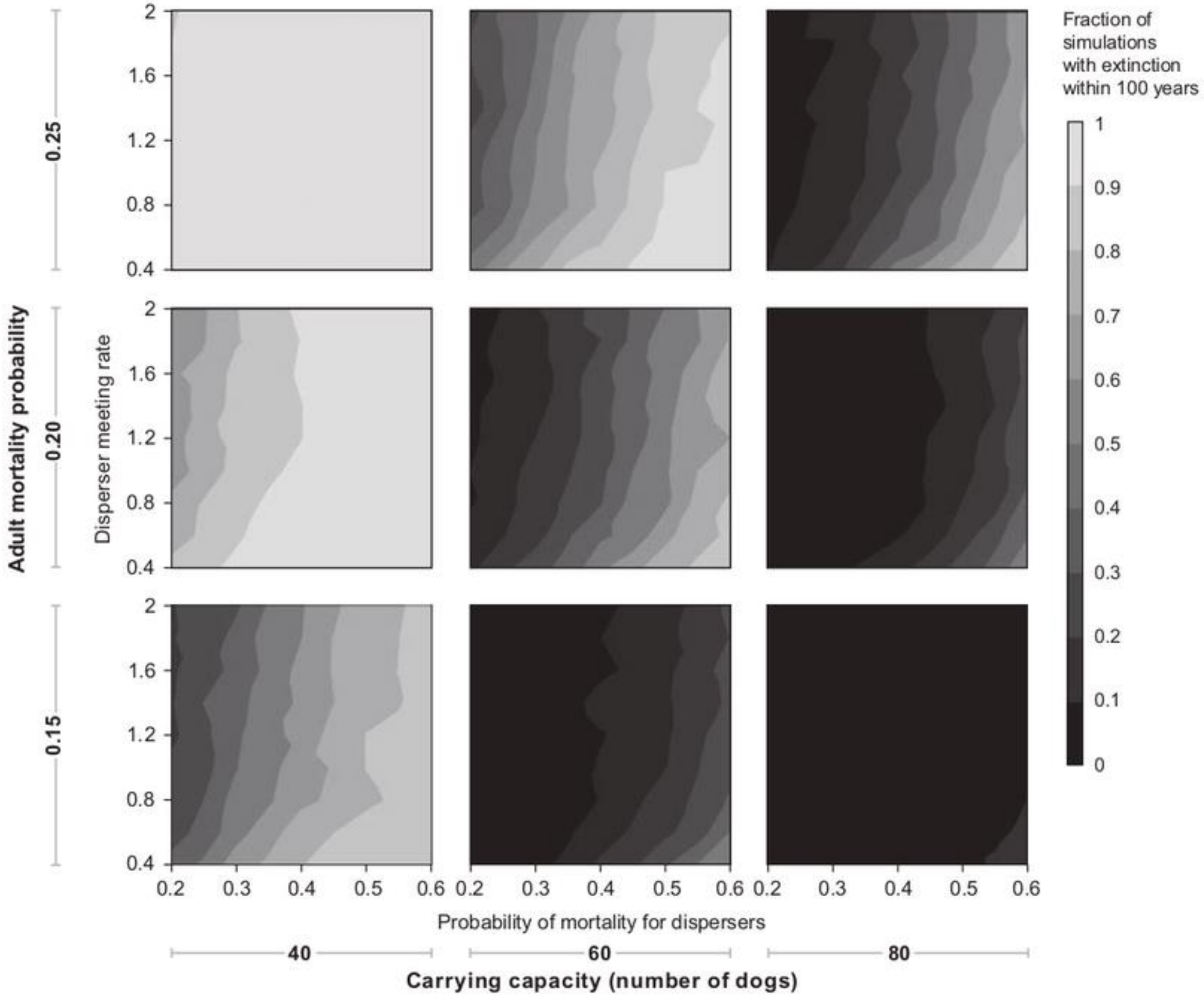
- Testing parameters takes time
- Focus on parameters that are **uncertain** and **important**
- **Sensitivity analysis** can help to determine which parameters are important
- Try to **make parameters identifiable**, i.e. parameters should have distinct effects
  - E.g., in wild dog model, suppose we had rate of dispersers meeting in the day-time and the rate of meeting at night-time
  - Essentially the same effect, but larger parameter space

Multiple parameters

- Suppose we want 400 replications for each of 10 values of one parameter
  - => 400*10 = 4000 runs needed
- Suppose we want 400 replications for all combinations of 10 values of two parameters
  - => 400*10*10 = 40000 runs needed
- Repeat for 3, 4 parameters...
- Exploring large parameter spaces is a challenge!

Strategies for large parameter spaces

- Avoid large parameter spaces if possible!
- 2 & 3 parameters: exhaustive search and visualisation
- Beyond 3 parameters:
  - Deterministic models: optimisation methods, e.g. gradient descent
  - Stochastic models: Monte-Carlo methods, e.g. Approximate Bayesian Computation



Railsback & Grimm (2019)

Summary

- Design concepts in the context of the African Wild dogs model
  - Collectives
  - Sensing
  - Stochasticity
  - Interaction
- Model analysis
  - Visualisation
  - Sensitivity analysis
- Parameterisation and callibration
  - Categorical
  - Best fit (parameter-fitting)