MSc thesis Femke Keij
Supervisor:
George van Voorn

REPRODUCING REALITY

A Comparison of Methodologies and Machine Learning Algorithms for the Calibration of Agent-Based Models in Ecology

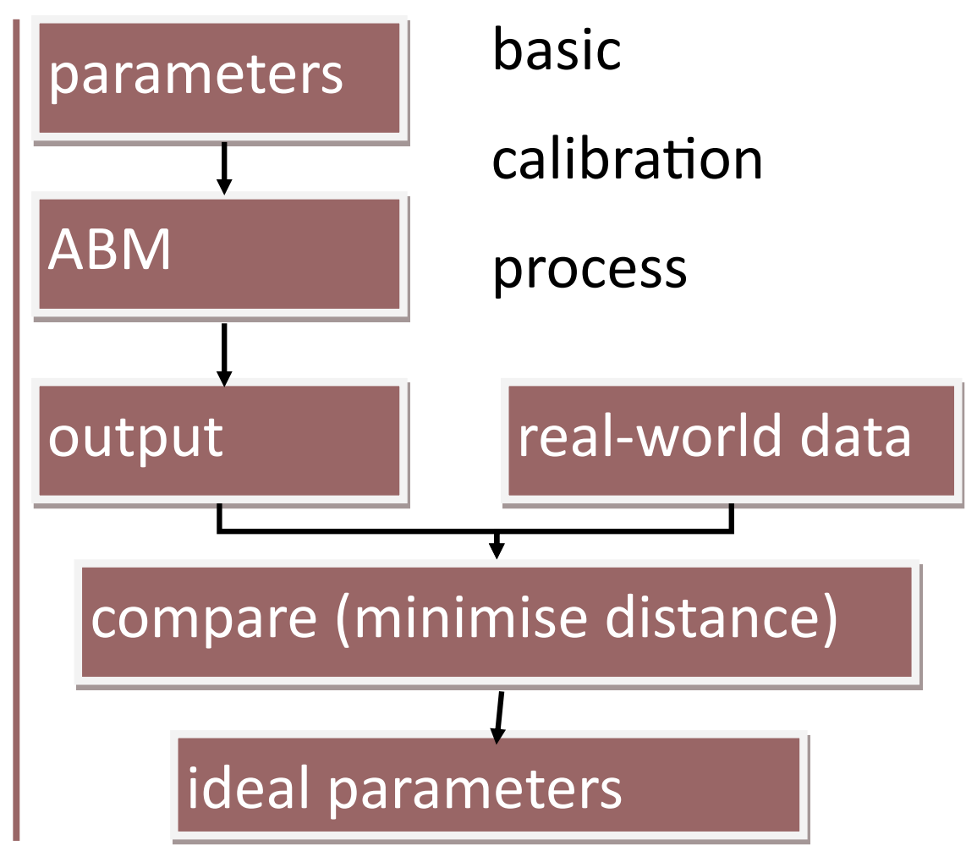
Universiteit Leiden
Wageningen University & Research

INTRODUCTION

Agent-Based Models (ABM) have recently gained popularity in the ecological modelling community. ABMs describe rules that govern the behaviour of individual organisms, as well as their interactions with their environment and other individuals, from which population dynamics emerge. ABMs more accurately represent the fact that real-world dynamics emerge from individual decision-making entities, which yields greater predictive power in novel conditions [1, 2, 3]. For ABMs to continue to advance, the associated methodologies need to be standardized [4, 5]. It is important that ABMs are capable of reproducing real-world systems dynamics. To this end, the model parameters are estimated in the calibration process. Due to the computationally expensive nature of ABMs, it is best to automate this process, but how exactly to do this is currently unknown.

Research questions:

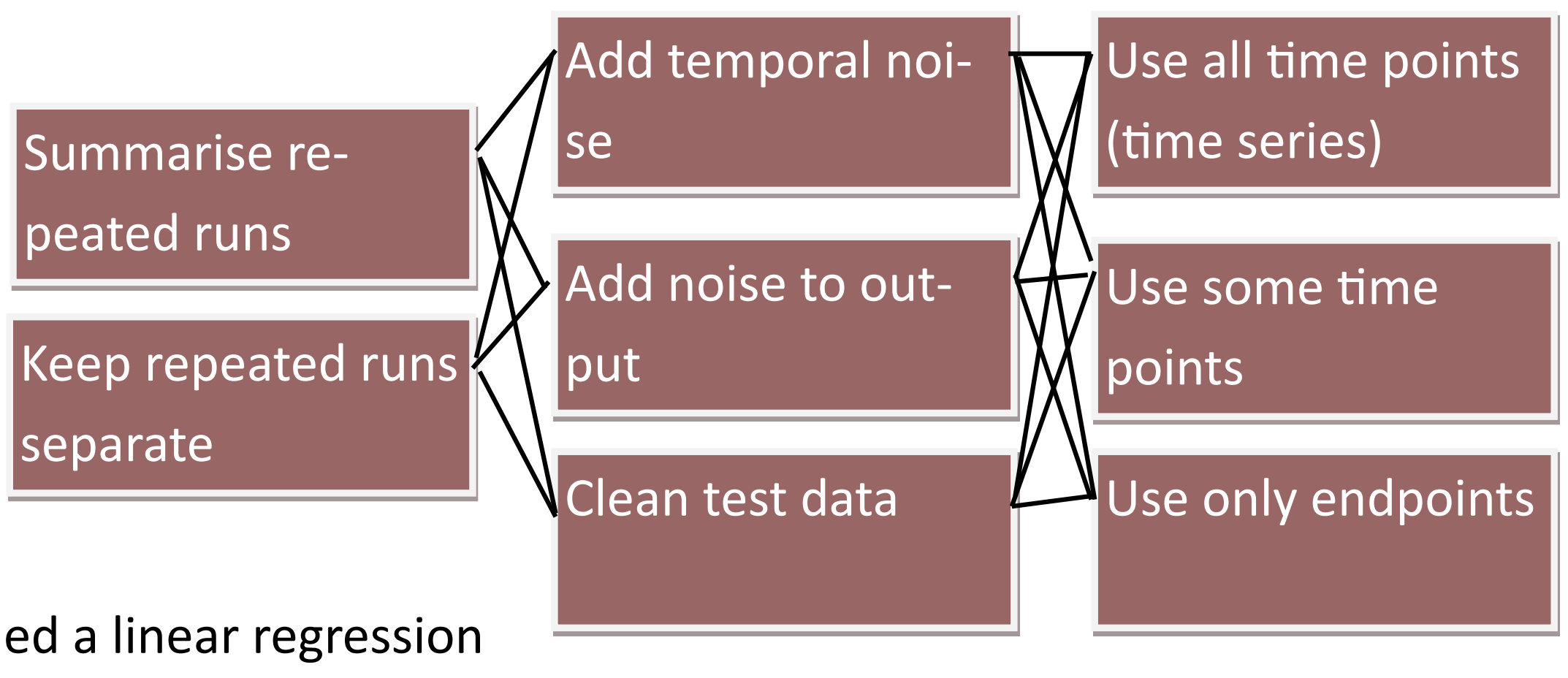
1. What is the influence of… on ABM calibration?
   - Using a full time series vs. using only time points
   - Summarising vs. not summarising repeated runs of the same parameter combination
   - Noise in the test data
2. Which algorithm most successfully calibrates ABMs?
   - A univariate or multivariate algorithm?
   - A linear or non-linear algorithm?


basic calibration process

parameters
ABM
output
real-world data
compare (minimise distance)
ideal parameters

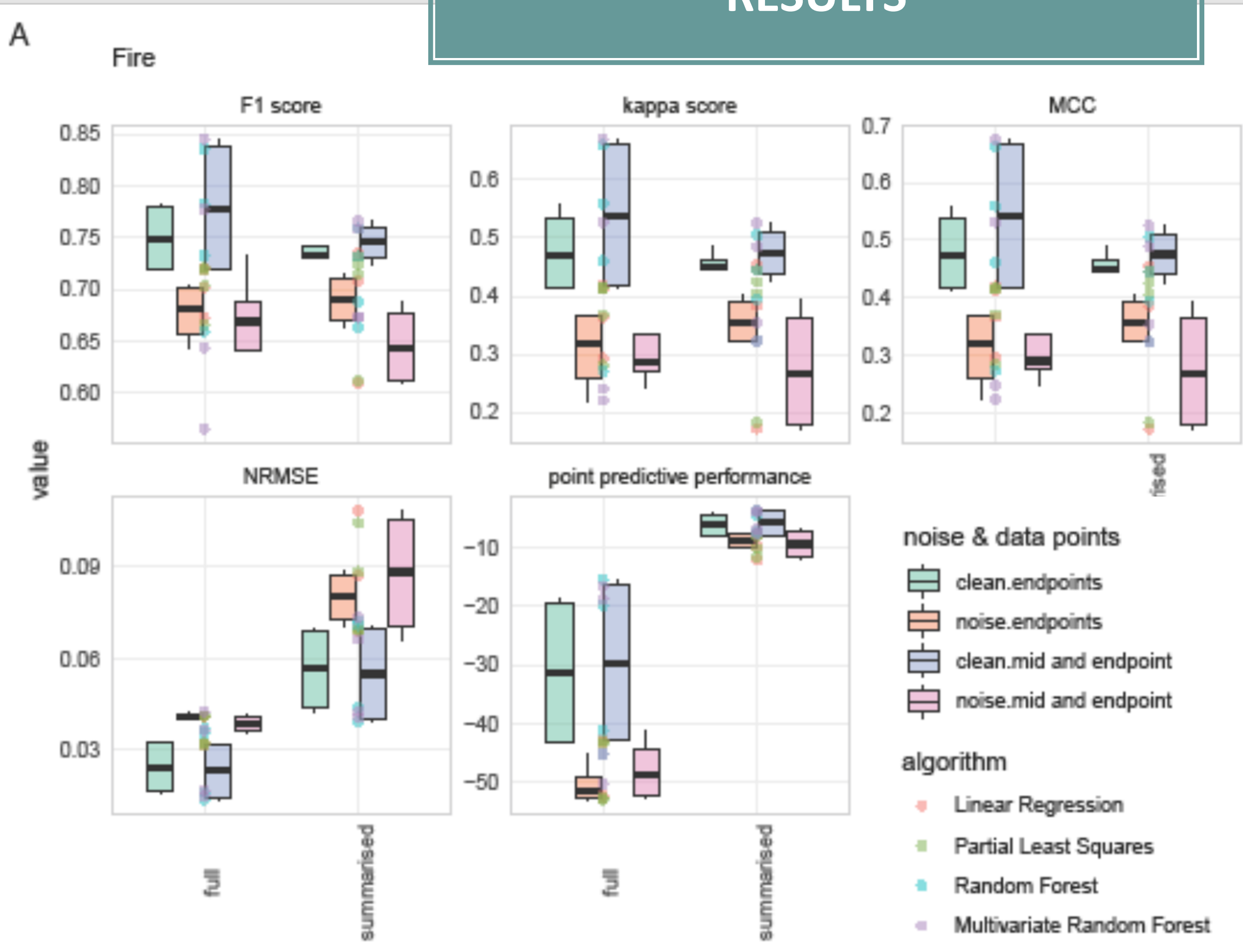
METHODS

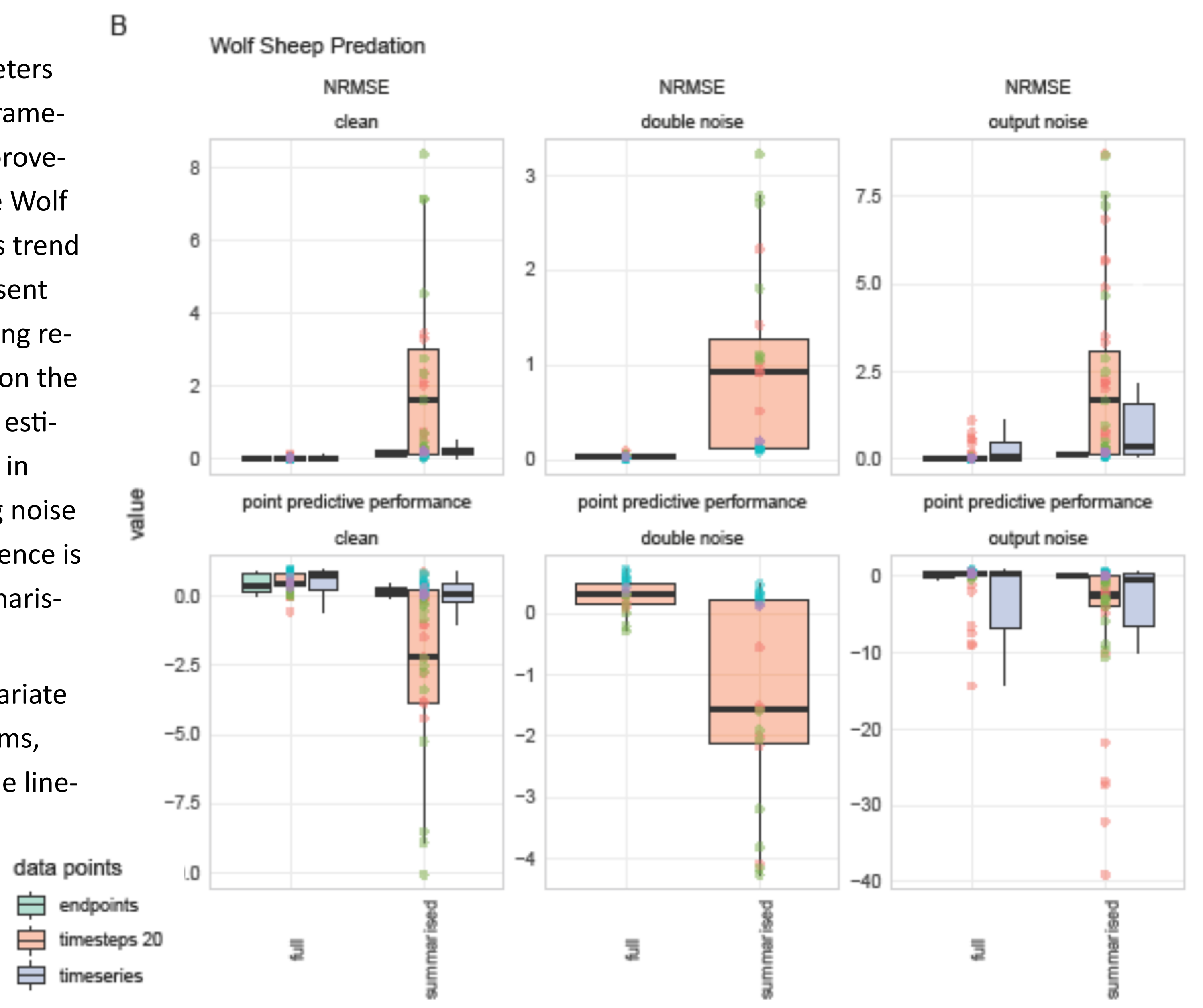
Fire (density & number of directions parameters) & Wolf Sheep Predation (initial number of sheep, initial number wolves, sheep gain from food, wolves gain from food, sheep reproductive probability, wolves reproductive probability, grass regrowth time parameters) ABMs were randomly sampled to obtain 200 parameter combinations. The training data were then optionally summarised, noise was optionally added, and the time series was optionally shortened to timepoints, creating the different data combinations listed below.


Summarise repeated runs
Keep repeated runs separate
Add temporal noise
Add noise to output
Clean test data
Use all time points (time series)
Use some time points
Use only endpoints

We trained a linear regression (LR), partial least squares (PLS), random forest (RF), and multivariate random forest (MRF). Performance was evaluated using F1 score, kappa, Matthew's Correlation Coefficient (for discrete parameter), NRMSE, and point prediction performance (for continuous parameter).

RESULTS





All scores are computed by using the parameters predicted by the algorithms and the true parameters given in de test data. There is some improvement when including more timesteps for the Wolf Sheep Predation ABM (see Figure B), but this trend is not consistent. The trend is completely absent from the Fire ABM (see Figure A). Summarising repeated runs does not have a large influence on the discrete parameter (A), but is detrimental to estimation of continuous parameters, especially in combination with LR and PLS (A & B). Adding noise to the test data is detrimental, but the difference is not as large as for summarising vs. not summarising the repeated runs (A & B).

There are no large differences between univariate (LR, RF) and multivariate (PLS, MRF) algorithms, though the latter are much slower to run. The linear LR and PLS are consistently outperformed by the non-linear RF and MRF. When estimating discrete parameters, the MRF scores best on all error metrics, although it also yields the worst score in one of the three cases. When estimating continuous parameters the RF scores best on the NRMSE measure, but not on the PPP. The LR / Logistic Regression approaches dominate the worst scores.

An overview of calibration issues and the calibration steps at which they occur:

Step	Issues	Solutions
General	Steps are combined and not studied separately; Cannot use likelihood	Study the effect of choices at each step separately
Sampling	Cannot run full parameter space; Many sampling strategies; Unclear effects of sample size	Use HPC or parallel computing, sample parameter space; Study effect of different sampling strategies
Summarising ABM output	ABMs are stochastic; ABMs produce a lot of data; Implicit spatial summarising Explicit temporal summarising; Summarising stochastic runs	Run each parameter combination multiple times; Summarise output data; Study effect of summarising output data
Obtaining real-world data	Data is unavailable; Uncertainty in data; Mismatch in scale	Study effect of noise on calibration
Choosing a distance function	Optimising multiple outputs simultaneously	Use multivariate methods or weighted summaries
Minimising distance function	Local minima & overfitting; Interdependent parameters; Non-linear relationship between parameters and output; Parameter identifiability; Equifinality; Large samples, long run times; Algorithms not accessible to non-statisticians	Use stochastic / multiple start algorithms; Study effect of noise on calibration; Study calibration performance of linear vs. non-linear algorithms; Acknowledge and take into account identifiability and equifinality before calibration; Use efficient algorithms; Make a clear overview of available algorithms and how to use them

DISCUSSION & CONCLUSION

1. What is the influence of…
   - Using a full time series vs. using only time points? *Depending on the algorithm, there may be improvement in calibration (RF, MRF) or not (LR, PLS).*
   - Summarising vs. not summarising repeated runs of the same parameter combination? *Although mean summarising is very common, this study shows that it can be detrimental to ABM calibration.*
   - Noise in the test data? *Noise in the real-world data (here proxied by noise in the test data) is detrimental to ABM calibration, although that is currently not taken into account.*
2. Which algorithm most successfully calibrates ABMs?
   - A univariate or multivariate algorithm? *There is no large difference between univariate (LR, RF) and multivariate (PLS, MRF) algorithms, but multivariate algorithms are much slower.*
   - A linear or non-linear algorithm? *Non-linear algorithms (RF, MRF) performed better than linear algorithms (LR, PLS) in most cases.*
   - *Although differences are small, MRF performs best and LR performs worst.*

Overall calibration performance was very poor. The main issue encountered was runtime limitations, these severely limited the expanse of the study. The methodology of this study can be re-used when further studying methodological options and different algorithms for ABM calibration. Ideas for future research:

- Using Genetic Algorithms or Convolutional Neural Networks.
- Using meta-models to improve sampling efficiency.
- Studying the effects of sample size, repeated runs for the same parameter combination, summarising runs by using a different summary measure than the mean, using a sliding window to obtain time point data.
- Adding more ABMs for a more complete comparison.
- Utilizing the temporal and autocorrelation in the ABMs instead of treating it as a nuisance factor and summarising it out.

Conclusion: ABM calibration needs to be standardized, many choices regarding methodology and optimisation algorithms are understudied. This study showed that the common practice of summarising repeated runs and summarising time series are potentially problematic. Since overall calibration performance was poor, further investigating other optimisation algorithms may be worthwhile. Further study of ABM calibration is recommended.

GitHub repository: https://github.com/FemkeKeij/ABM_Calibration_Thesis.

References: 1) Elske van der Vaart, Alice S.A. Johnston, and Richard M. Sibly. Predicting how many animals will be where: How to build, calibrate and evaluate individual-based models. Ecological Modelling, 326:113–123, 4 2016. 2) Rapha¨el Duboz, David Versmisse, Morgane Travers, Eric Ramat, and Yunne Jai Shin. Application of an evolutionary algorithm to the inverse parameter estimation of an individualbased model. Ecological Modelling, 221:840–849, 3 2010. 3) C. C.M. Chen, C. C. Drovandi, J. M. Keith, K. Anthony, M. J. Caley, and K. L. Mengersen. Bayesian semi-individual based model with approximate bayesian computation for parameters calibration: Modelling crown-of-thorns populations on the great barrier reef. Ecological Modelling, 364:113–123, 11 2017. 4) J. Gareth Polhill, Jiaqi Ge, Matthew P. Hare, Keith B. Matthews, Alessandro Gimona, Douglas Salt, and Jagadeesh Yeluripati. Crossing the chasm: a 'tube-map' for agent-based social simulation of policy scenarios in spatially-distributed systems. GeoInformatica, 23:169–199, 4 2019. 5) Steven Manson, Li An, Keith C. Clarke, Alison Heppenstall, Jennifer Koch, Brittany Krzyzanowski, Fraser Morgan, David O'sullivan, Bryan C. Runck, Eric Shook, and Leigh Tesfatsion. Methodological issues of spatial agent-based models. JASSS, 23, 1 2020.