UNIVERSITY OF LEEDS

Data Assimilation for Agent-Based Models

Mathematical and Computational Challenges

Jonathan Ward
j.a.ward@leeds.ac.uk

i-Sim 2017

Collaborators: Nick Malleson & Andy Evans (Leeds)

Goals

	Do DA	Don't do DA
Do ABMs		
Don't do ABMs		

Goals

	Do DA	Don’t do DA
Do ABMs	Let’s talk!	
Don’t do ABMs		

Goals

	Do DA	Don't do DA
Do ABMs		Intro to DA: It's useful!
Don't do ABMs		

Goals

	Do DA	Don’t do DA
Do ABMs		
Don’t do ABMs	ABM challenges. You could help!	

Goals

	Do DA	Don't do DA
Do ABMs		
Don't do ABMs		Enjoy!

Overview

Overview

- What are Agent-Based Models (ABMs)?

Overview

- What are Agent-Based Models (ABMs)?
- What is Data Assimilation (DA)?

Overview

- What are Agent-Based Models (ABMs)?
- What is Data Assimilation (DA)?
- DA is easy to implement…

Overview

- What are Agent-Based Models (ABMs)?
- What is Data Assimilation (DA)?
- DA is easy to implement…
- …but be careful

Overview

- What are Agent-Based Models (ABMs)?
- What is Data Assimilation (DA)?
- DA is easy to implement…
- …but be careful
- Some comments on DA for ABMs

Overview

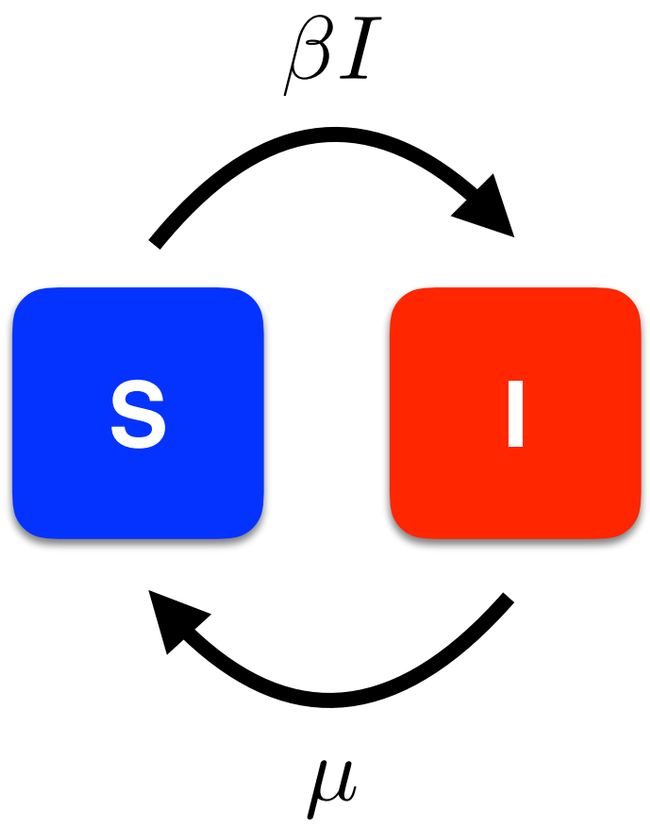
- What are Agent-Based Models (ABMs)?
- What is Data Assimilation (DA)?
- DA is easy to implement…
- …but be careful
- Some comments on DA for ABMs

ABMs

- Agents as objects, an environment
- Complexity, emergence
- ABMs to explain; multi-agent simulation to solve practical problems; individual Based Models: less/no interaction
- Informed by “data”
- Heterogeneity, interaction networks, learning,… (Kitchen sink approach)

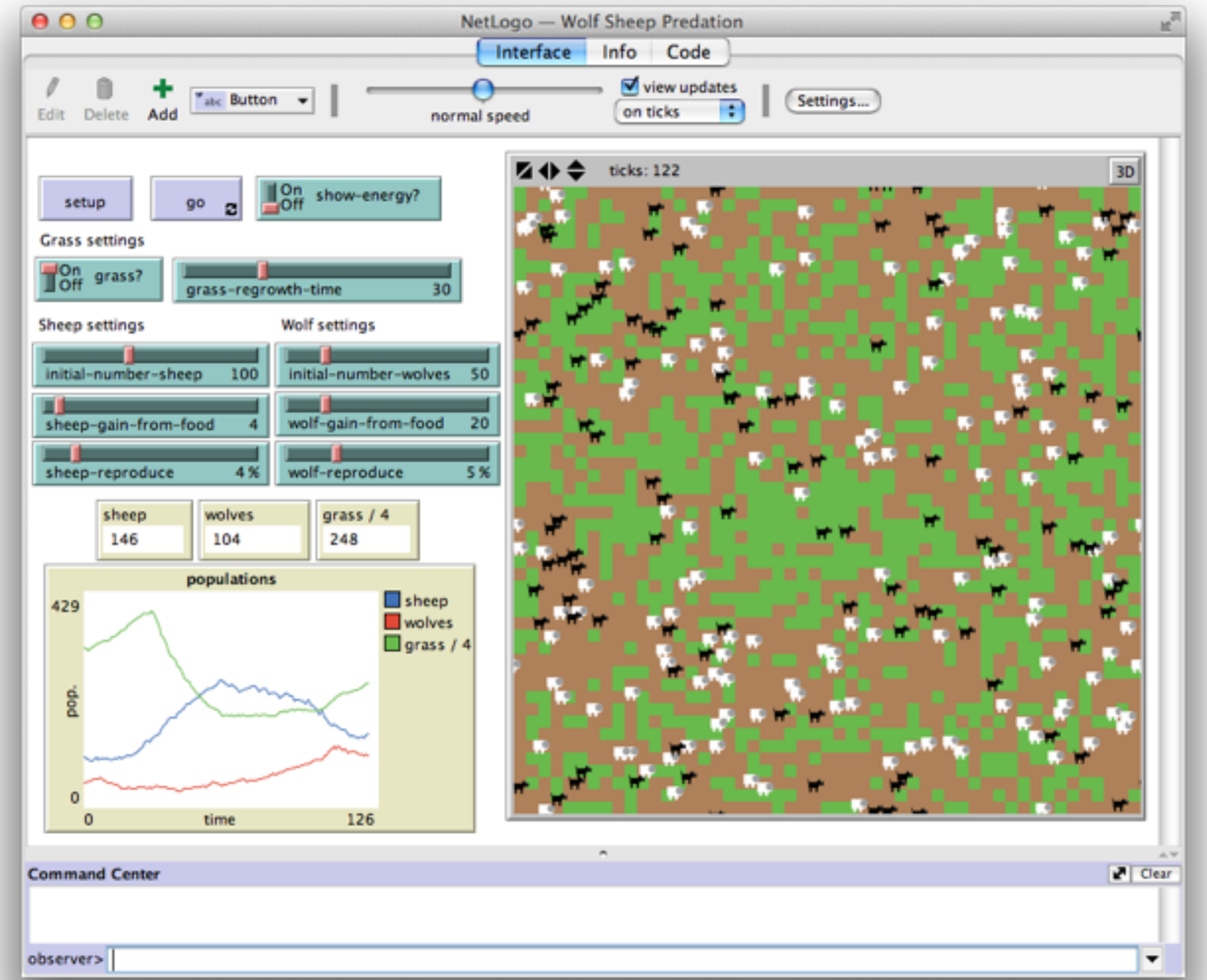
ABMs

- Epidemics
- Voting behaviour
- Innovation diffusion
- Statistical physics models
- Cellular Automata



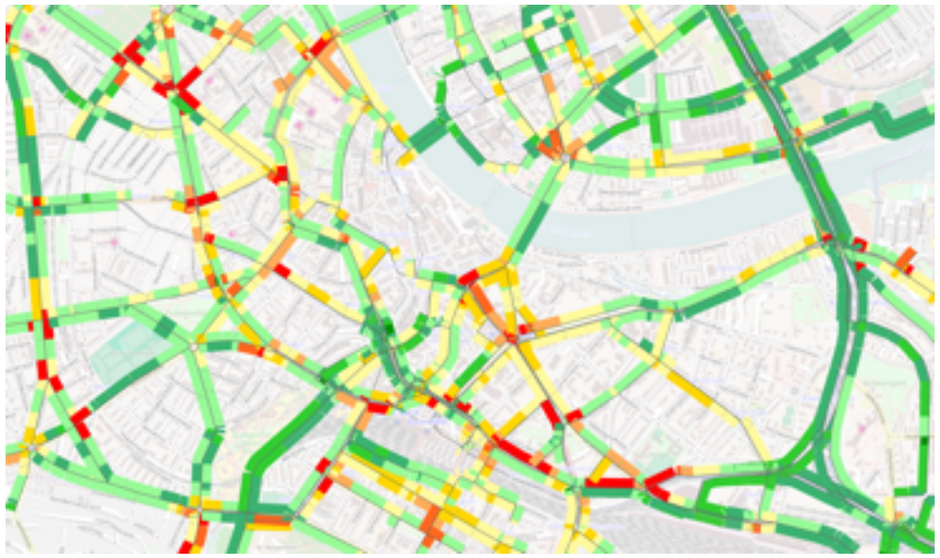
ABMs

- NetLogo: Ants, DaisyWorld, forest fires…
- Altruism
- Co-operation
- Social influence
- Segregation



ABMs

- Commercial planning models (multi-modal transport)



http://vision-traffic.ptvgroup.com

- Others? (Uber, Dunnhumby…)

Simple ABM Example

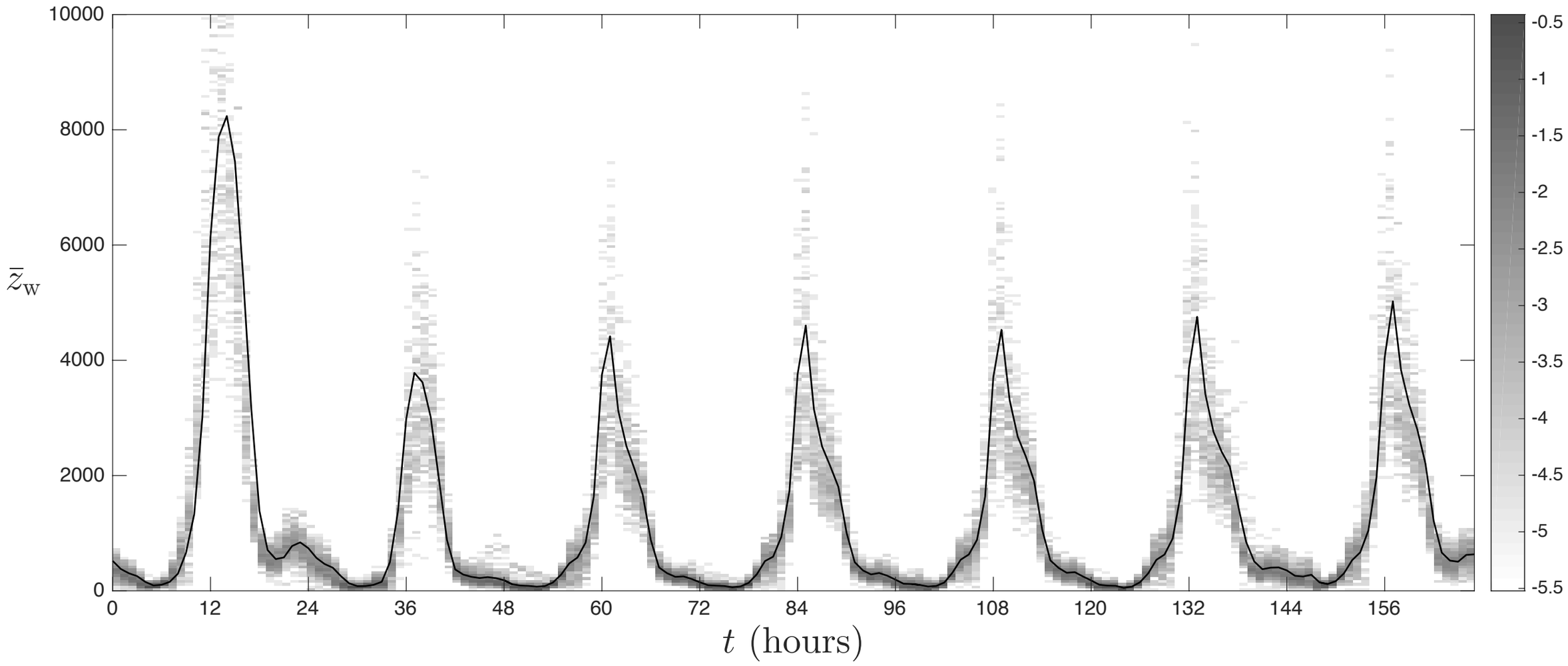
- Simple example to illustrate how the Ensemble Kalman Filter (EnKF) works

- Simple enough to know everything about the model dynamics and parameters

- Use the model to indicate issues that could arise

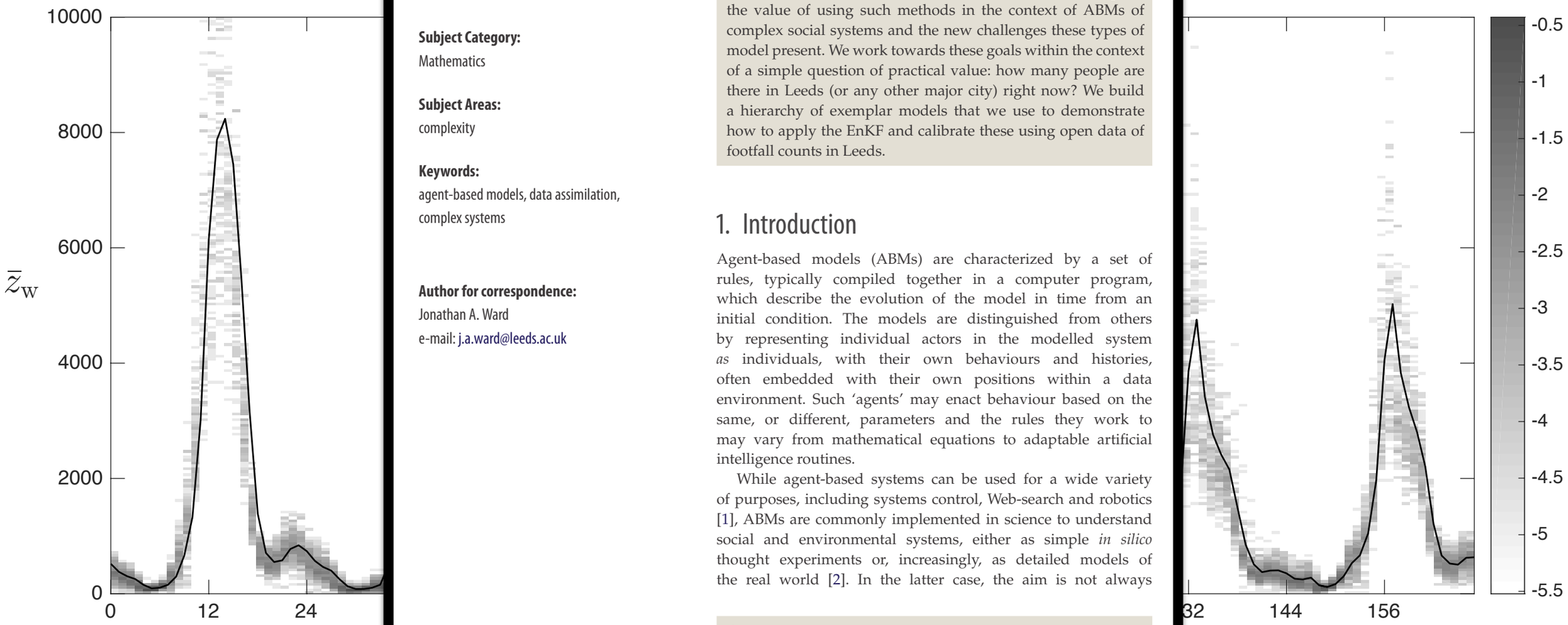
Model Motivation

- Footfall counts in Leeds UK city centre: https://datamillnorth.org



M on

- Footfall co



ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Ward JA, Evans AJ, Malleson NS. 2016 Dynamic calibration of agent-based models using data assimilation. *R. Soc. open sci.* **3**: 150703.
http://dx.doi.org/10.1098/rsos.150703

Received: 19 December 2015
Accepted: 11 March 2016

Subject Category:
Mathematics

Subject Areas:
complexity

Keywords:
agent-based models, data assimilation, complex systems

Author for correspondence:
Jonathan A. Ward
e-mail: j.a.ward@leeds.ac.uk

Dynamic calibration of agent-based models using data assimilation

Jonathan A. Ward[1], Andrew J. Evans [2] and Nicolas S. Malleson[2]

[1]School of Mathematics, and [2]School of Geography, University of Leeds, Leeds LS2 9JT, UK

A widespread approach to investigating the dynamical behaviour of complex social systems is via agent-based models (ABMs). In this paper, we describe how such models can be dynamically calibrated using the ensemble Kalman filter (EnKF), a standard method of data assimilation. Our goal is twofold. First, we want to present the EnKF in a simple setting for the benefit of ABM practitioners who are unfamiliar with it. Second, we want to illustrate to data assimilation experts the value of using such methods in the context of ABMs of complex social systems and the new challenges these types of model present. We work towards these goals within the context of a simple question of practical value: how many people are there in Leeds (or any other major city) right now? We build a hierarchy of exemplar models that we use to demonstrate how to apply the EnKF and calibrate these using open data of footfall counts in Leeds.

1. Introduction

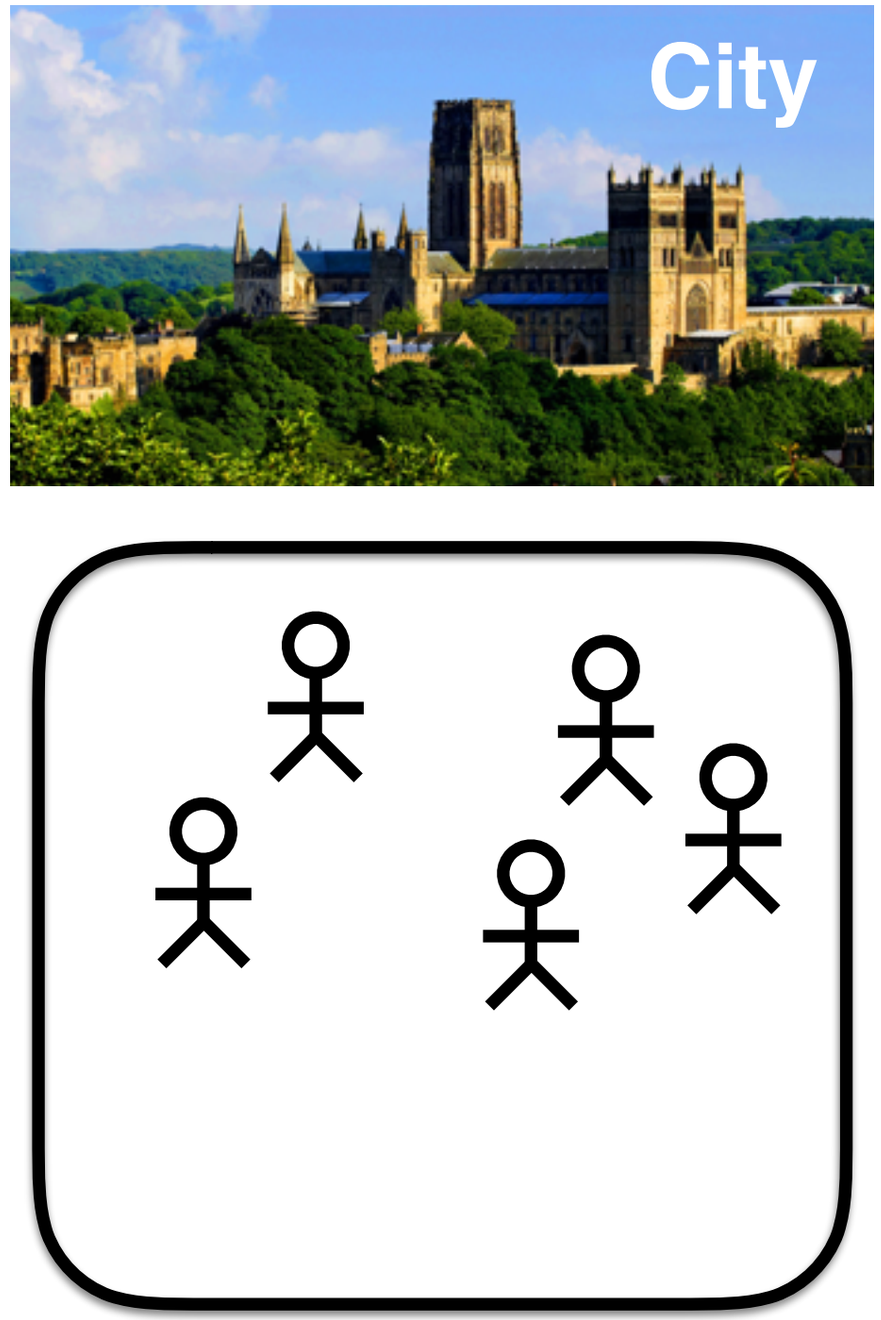
Agent-based models (ABMs) are characterized by a set of rules, typically compiled together in a computer program, which describe the evolution of the model in time from an initial condition. The models are distinguished from others by representing individual actors in the modelled system *as* individuals, with their own behaviours and histories, often embedded with their own positions within a data environment. Such 'agents' may enact behaviour based on the same, or different, parameters and the rules they work to may vary from mathematical equations to adaptable artificial intelligence routines.

While agent-based systems can be used for a wide variety of purposes, including systems control, Web-search and robotics [1], ABMs are commonly implemented in science to understand social and environmental systems, either as simple *in silico* thought experiments or, increasingly, as detailed models of the real world [2]. In the latter case, the aim is not always

THE ROYAL SOCIETY PUBLISHING

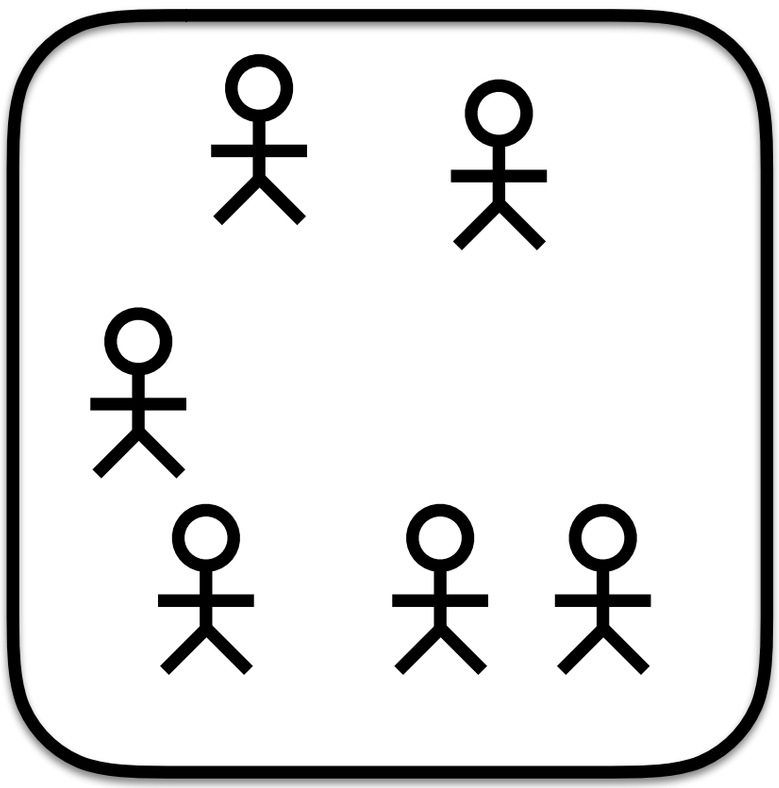
© 2016 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

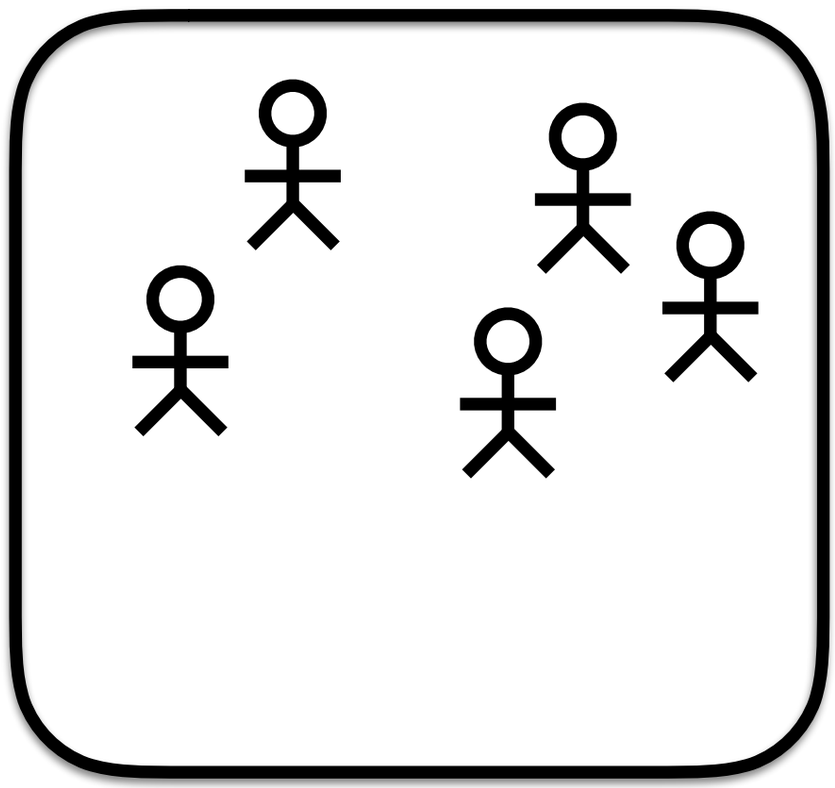
Simple ABM Example



Simple ABM Example

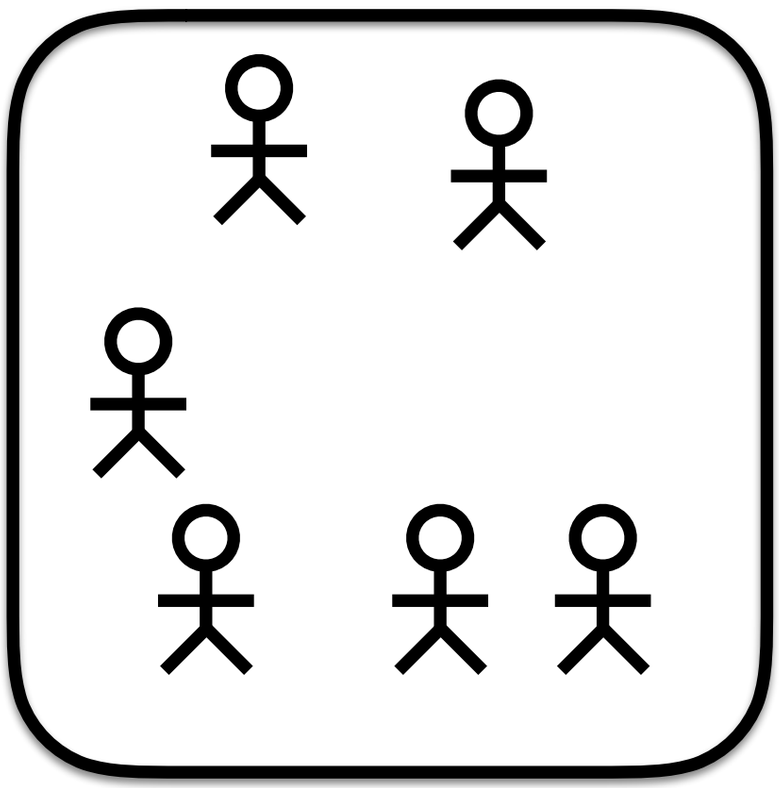
Agent pool

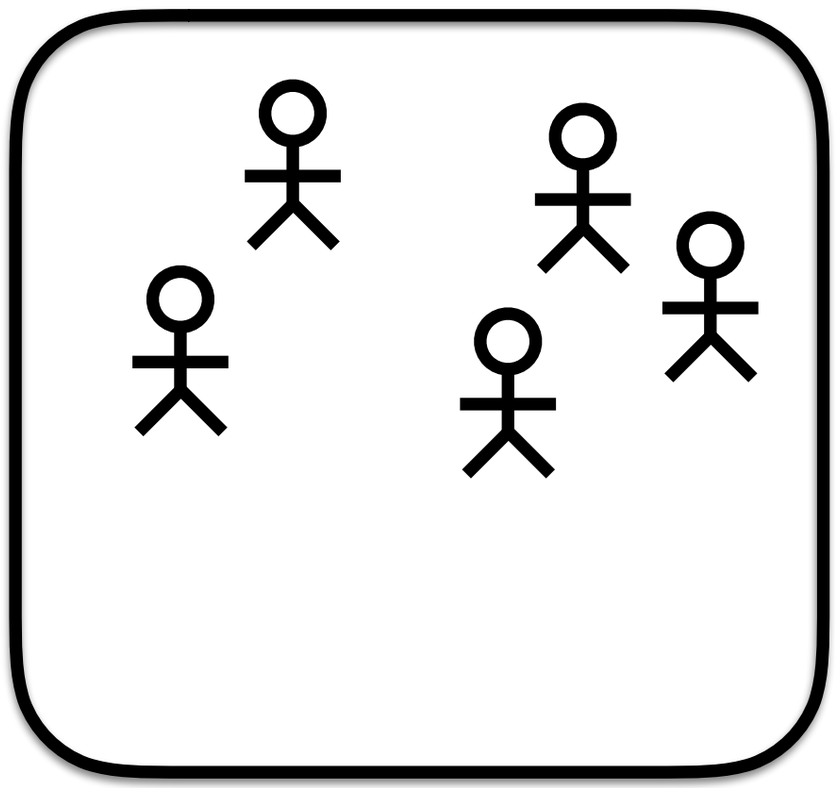




Simple ABM Example

Agent pool



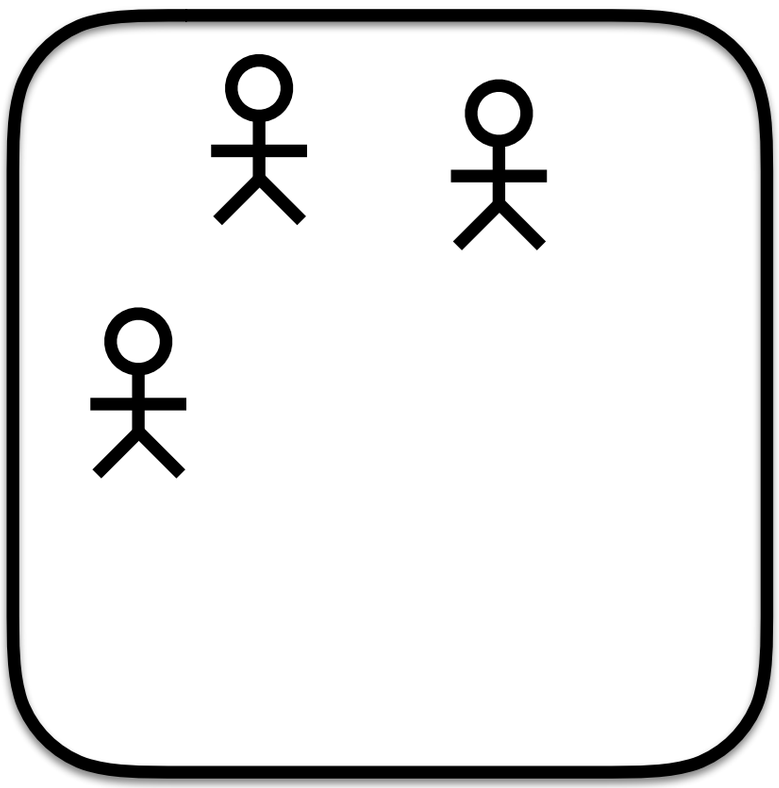


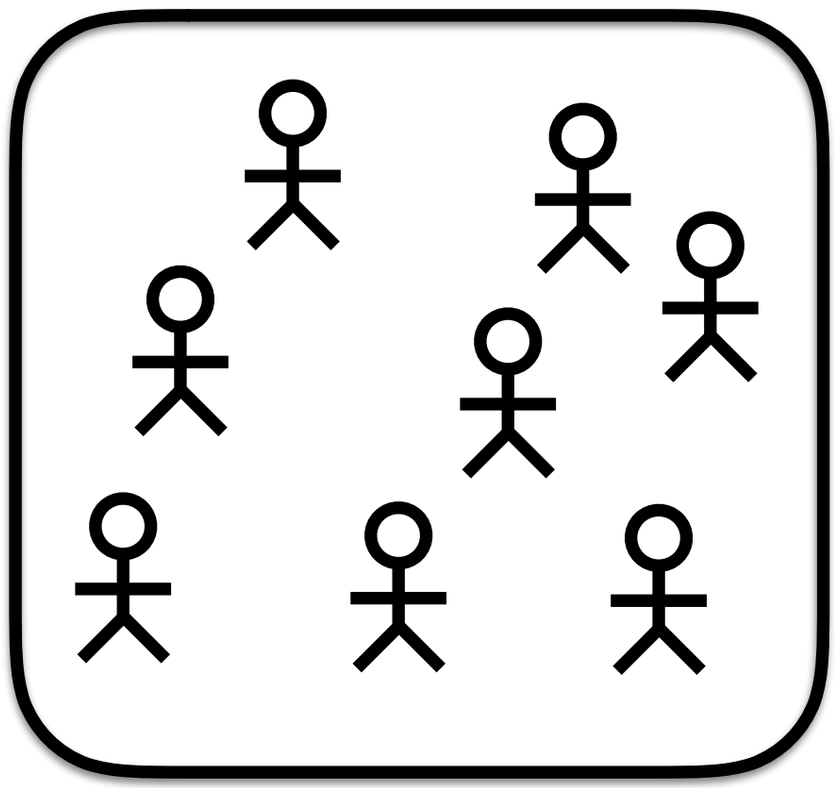
Simple ABM Example

Agent pool

Arrive

α



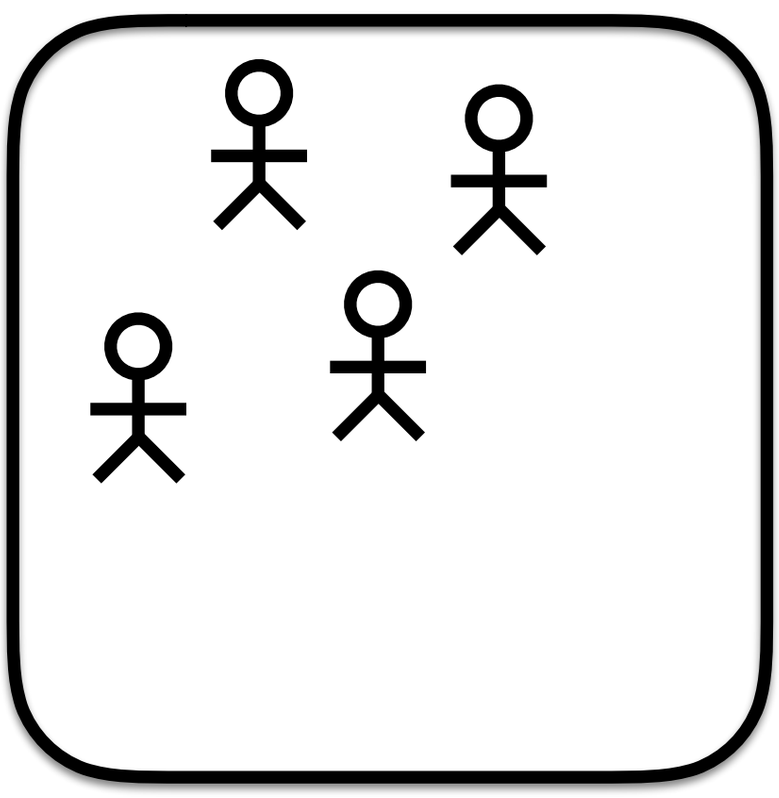


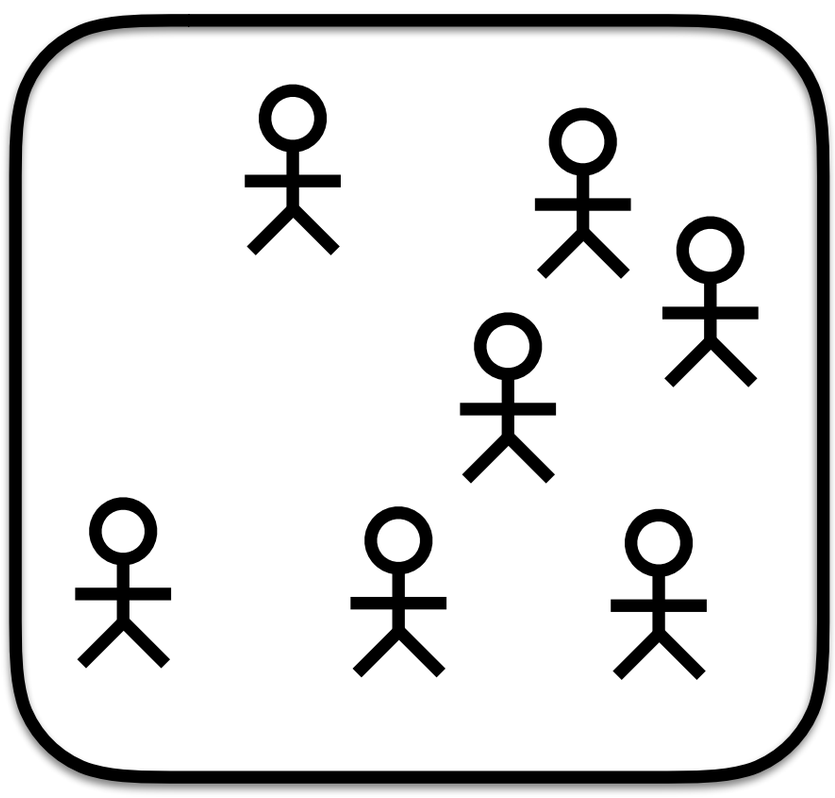
Simple ABM Example

Agent pool

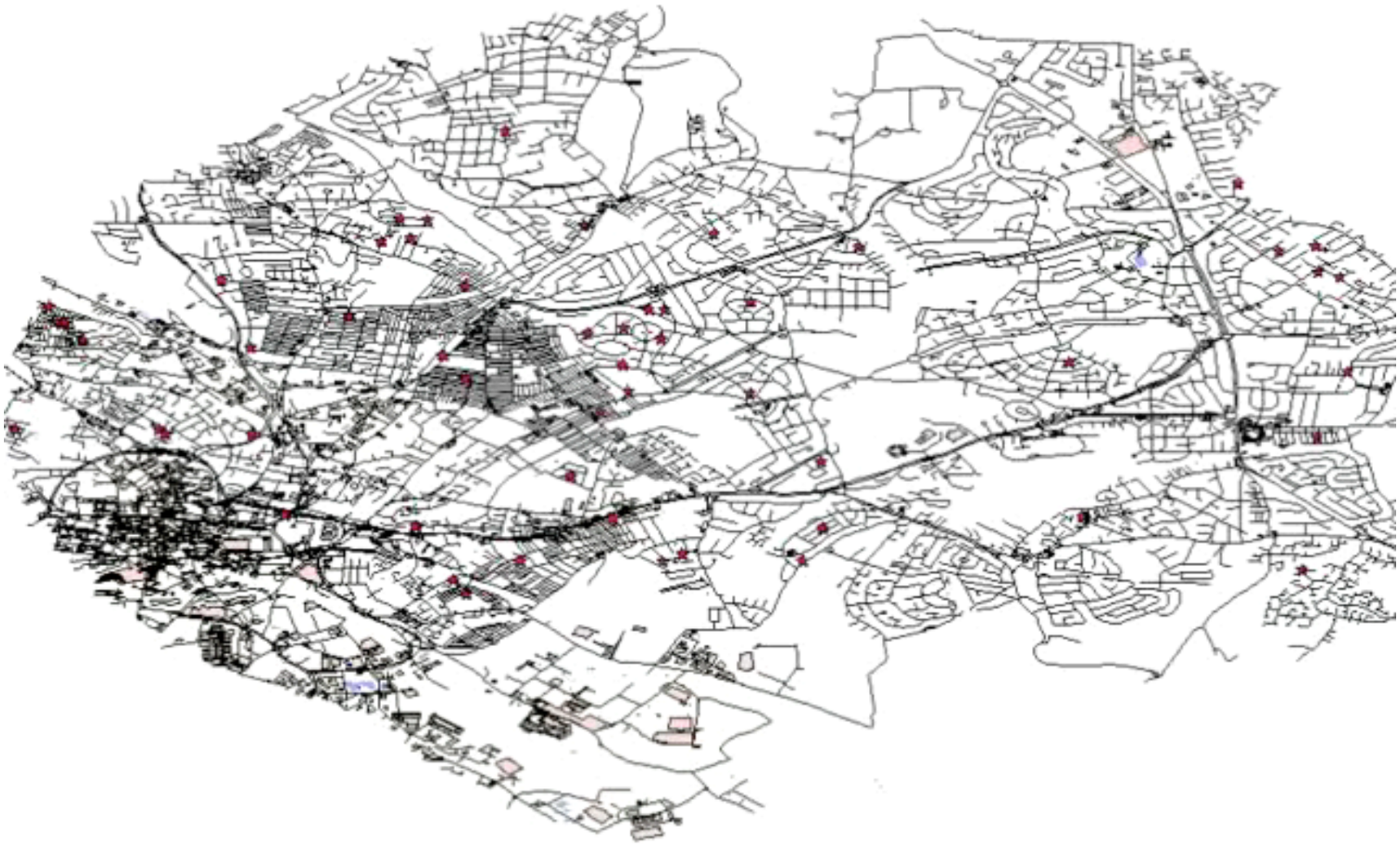
Depart

β



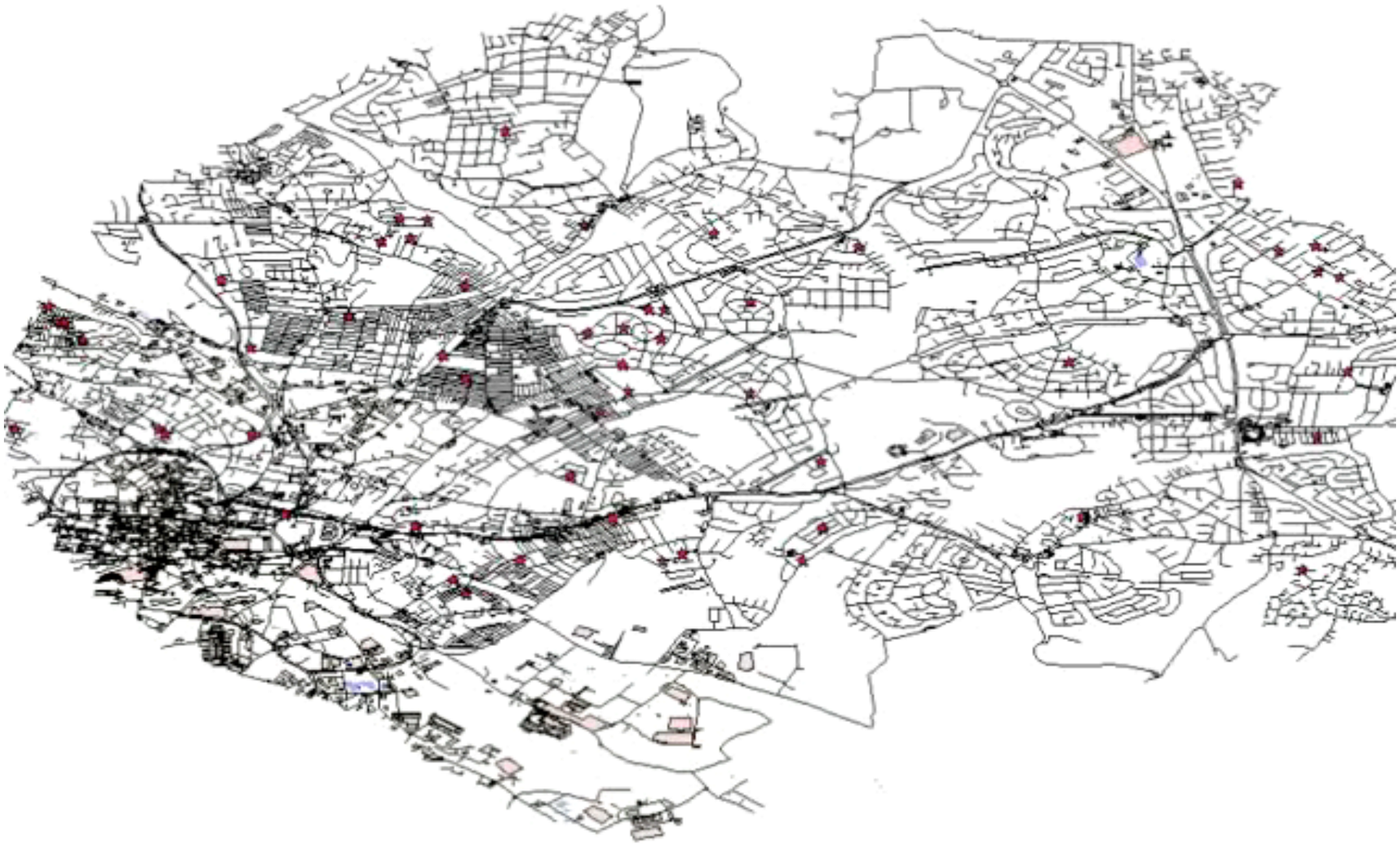


Adding Complexity



Malleson, Heppenstall, See, Evans; Environment and Planning B: Planning and Design (2013)

Adding Complexity



Malleson, Heppenstall, See, Evans; Environment and Planning B: Planning and Design (2013)

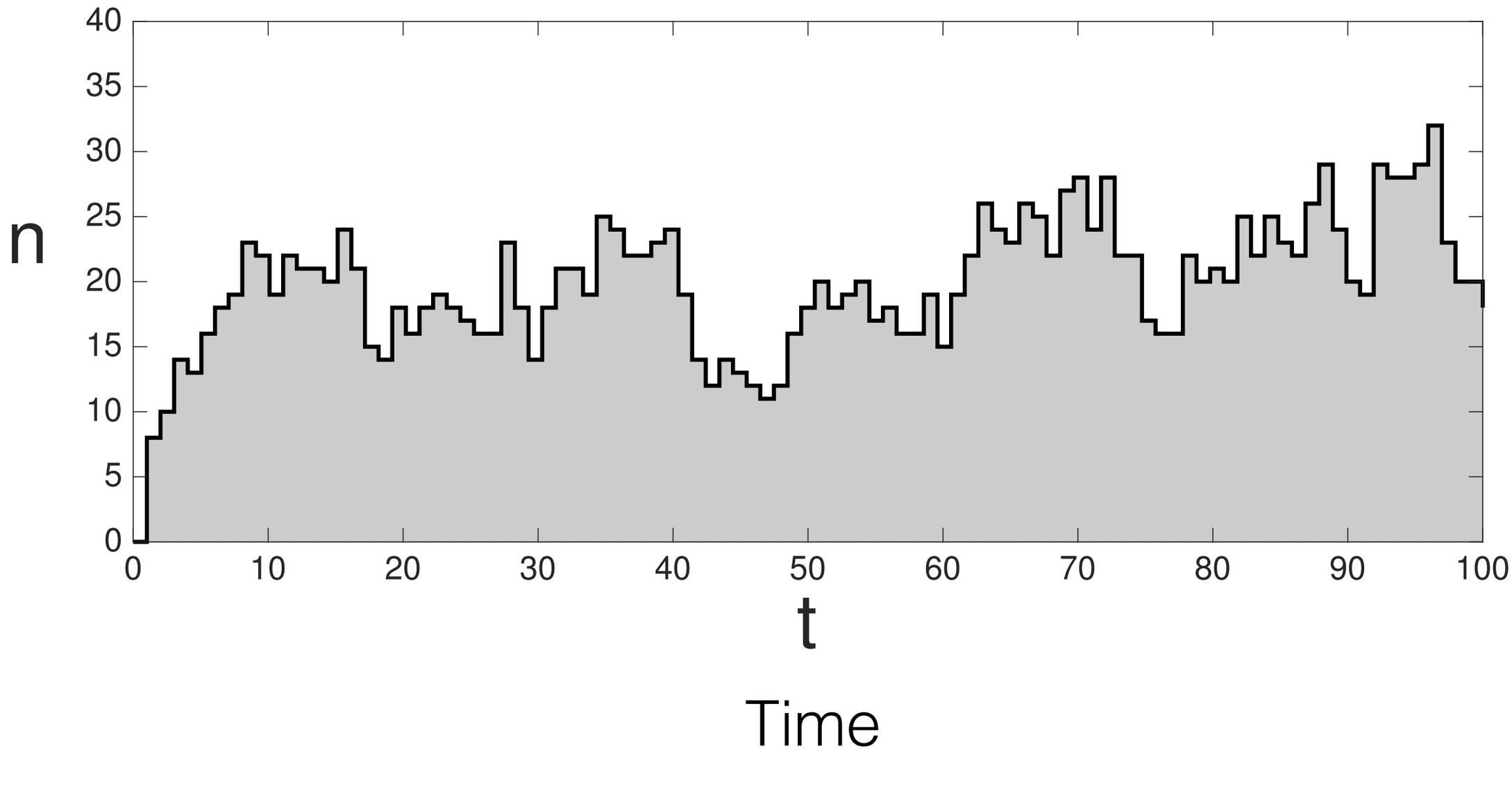
Adding Complexity

Malleson, Heppenstall, See, Evans; Environment and Planning B: Planning and Design (2013)

Simple ABM Example

- Agent states $x_i(t) \in \{0, 1\}$
- Population in city $n(t) = \sum_i x_i(t)$
- Simulate using Gillespie algorithm

Simple ABM Example



Mathematics

- Master equation

$$\dot{p}_n = \alpha p_{n-1} - (\alpha + \beta n) p_n + \beta (n+1) p_{n+1}$$

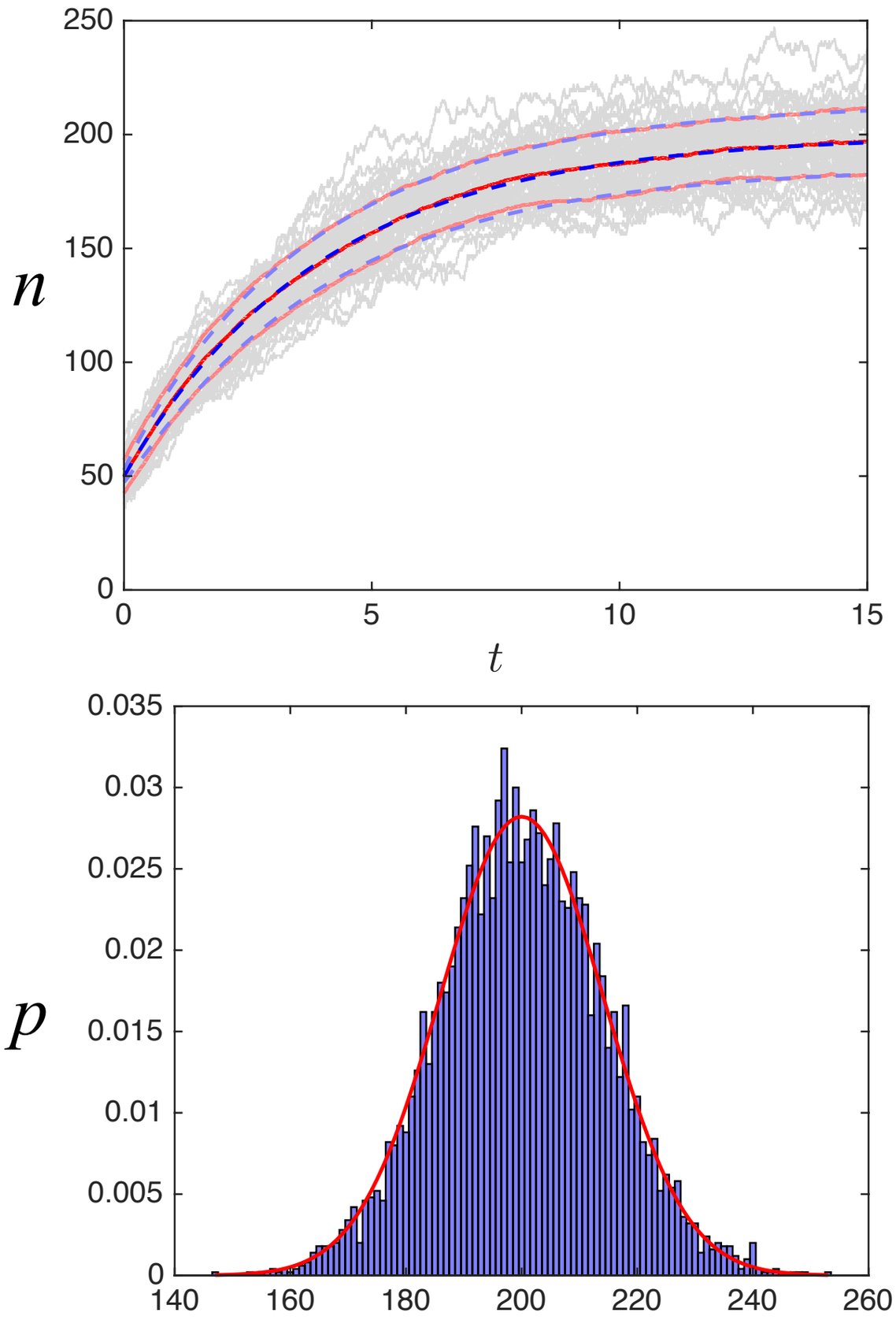
- Mean

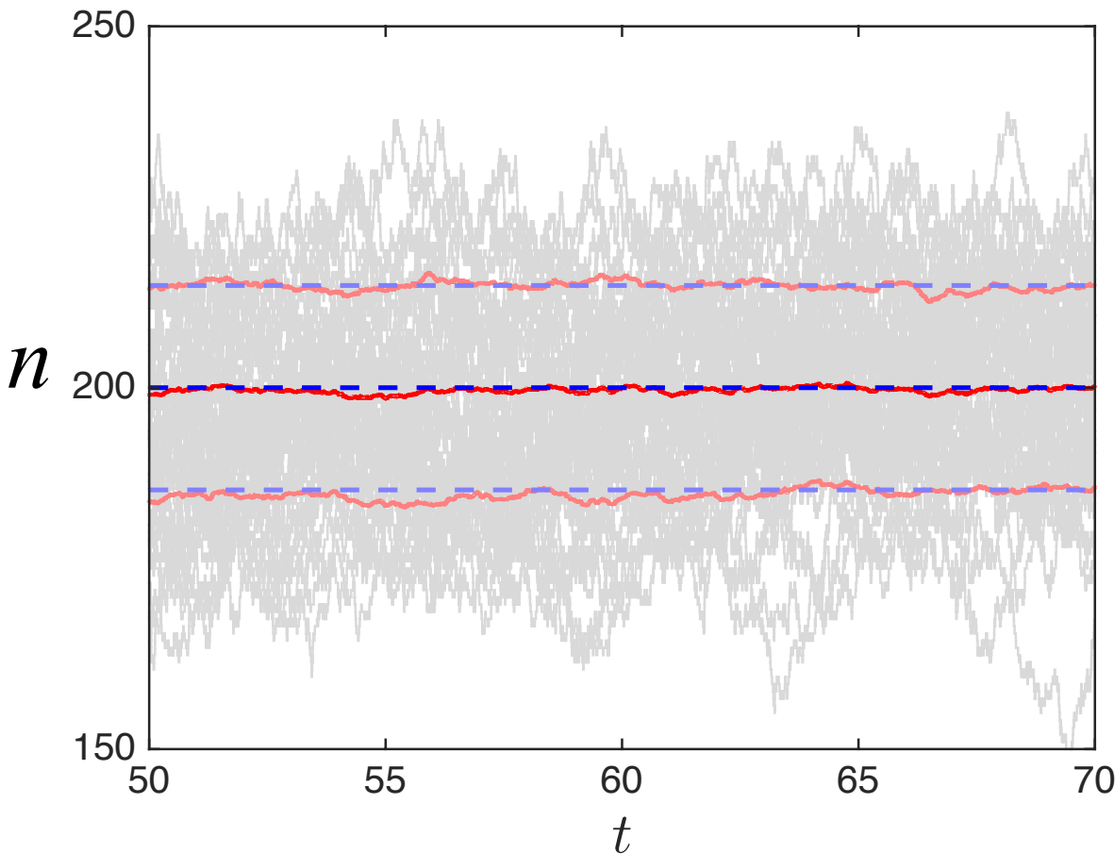
$$\lambda(t) = \frac{\alpha}{\beta} + \left(n_0 - \frac{\alpha}{\beta} \right) \mathrm{e}^{-\beta t}$$

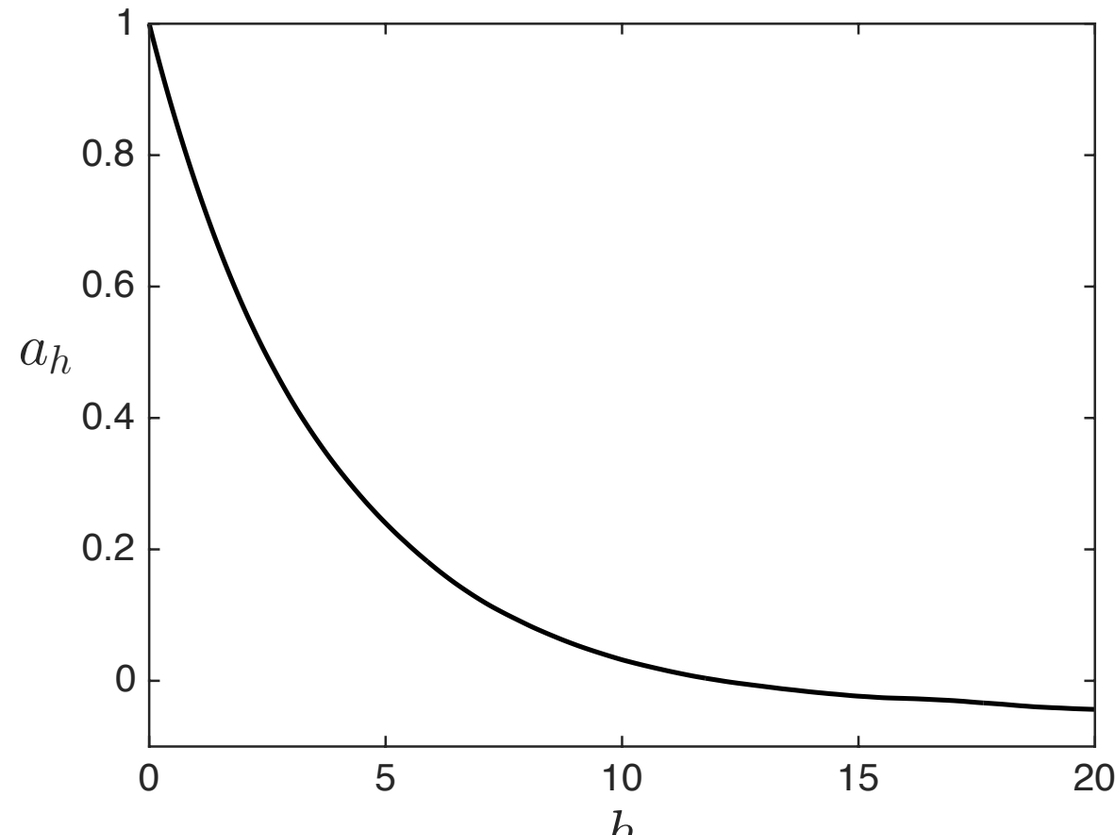
- Distribution

$$p_n(t) = \frac{\lambda(t)^n \mathrm{e}^{-\lambda(t)}}{n!}$$

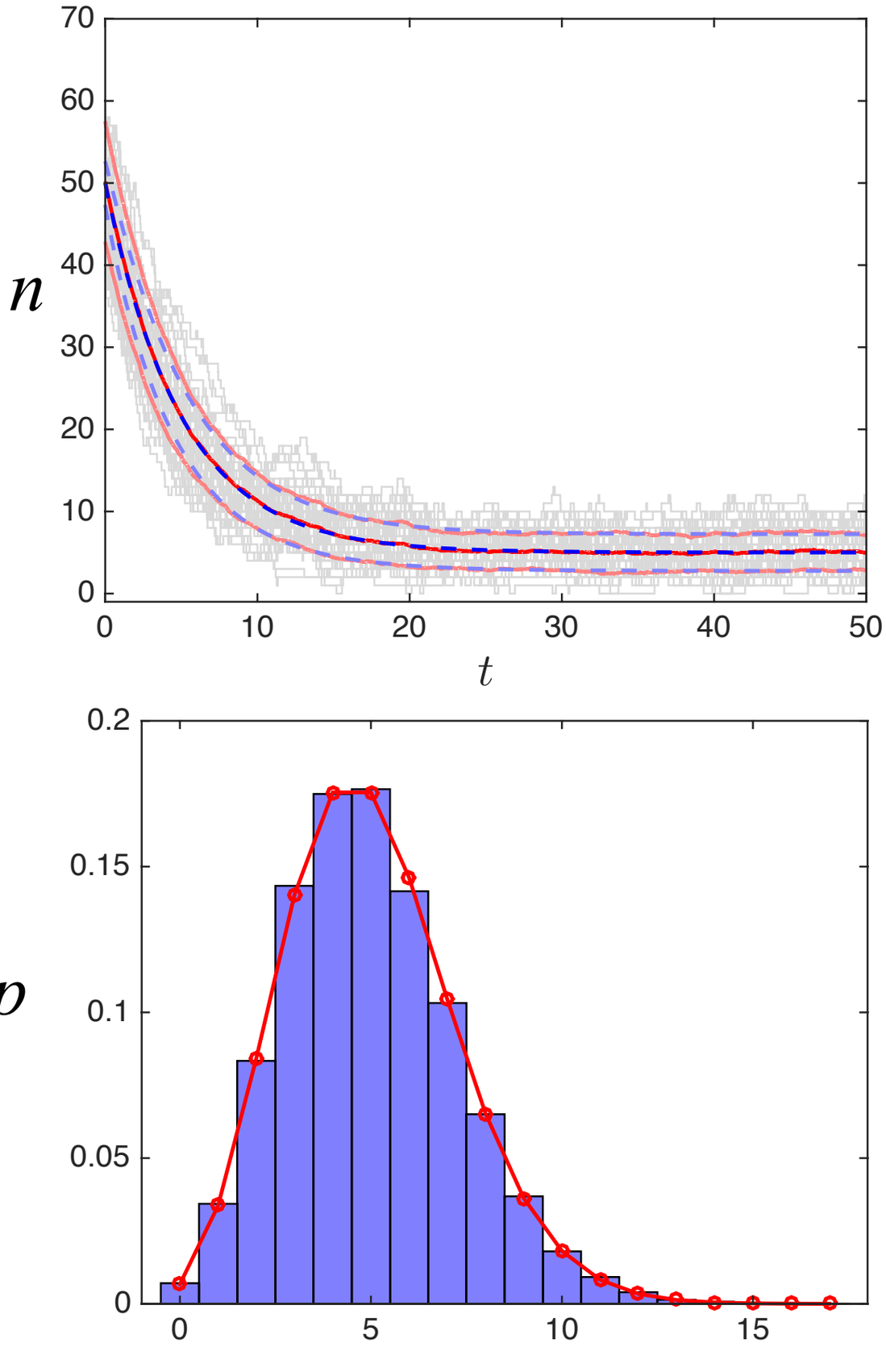
$$\alpha = 50, \quad \beta = 0.25$$

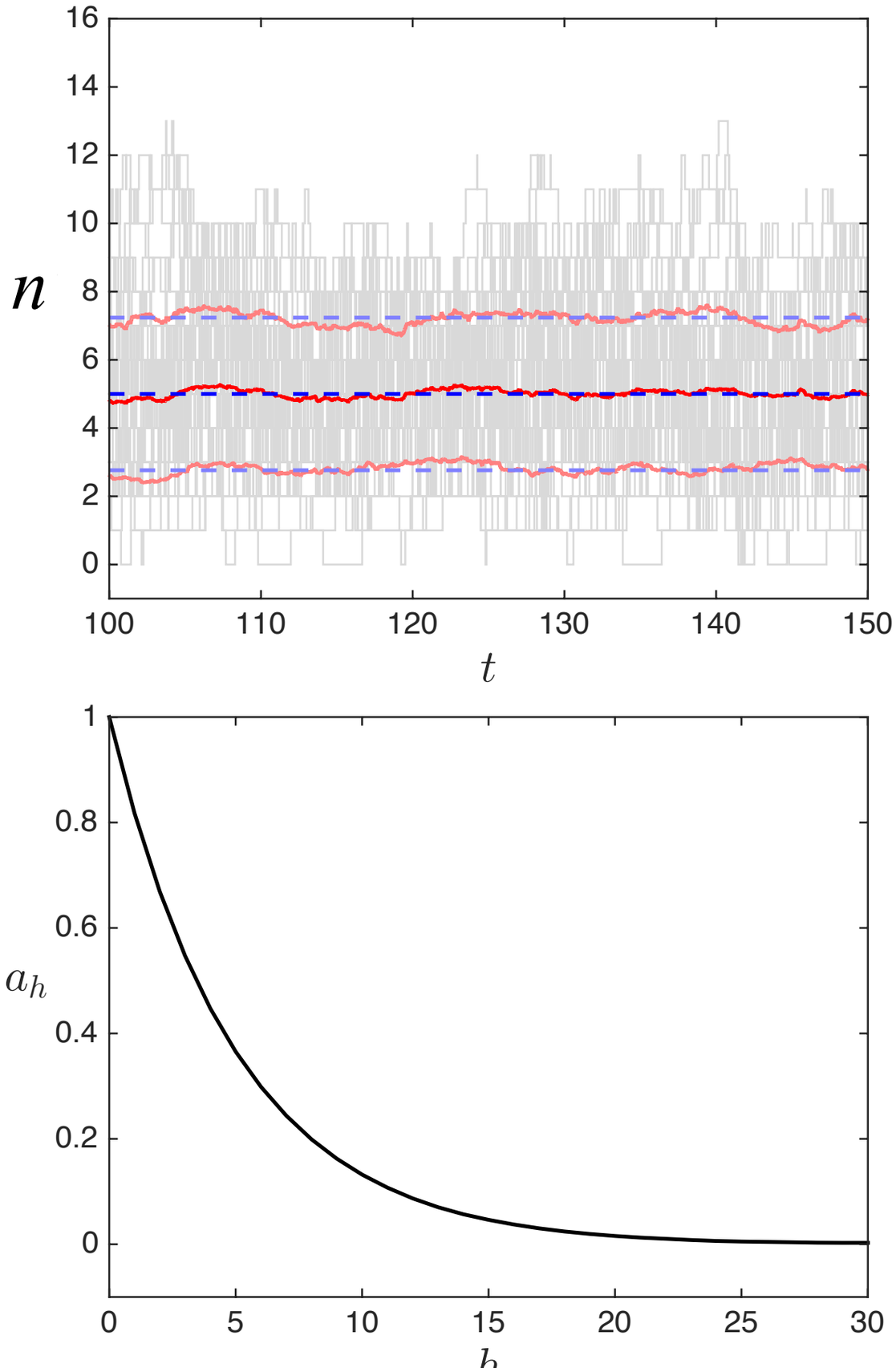






$$\alpha = 1, \quad \beta = 0.2$$





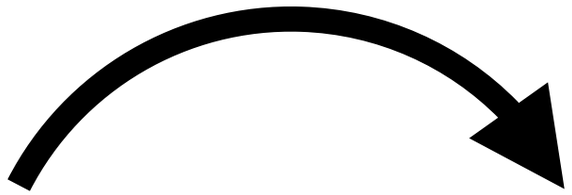
Overview

- What are Agent-Based Models (ABMs)?
- What is Data Assimilation (DA)?
- DA is easy to implement…
- …but be careful
- Some comments on DA for ABMs

Fitting to Data

- System state not observed directly
- System state typically much larger than observations
- Observations have error
- Goals:
  - Fit the model to observations
  - Use the model to make forecasts

Data Assimilation

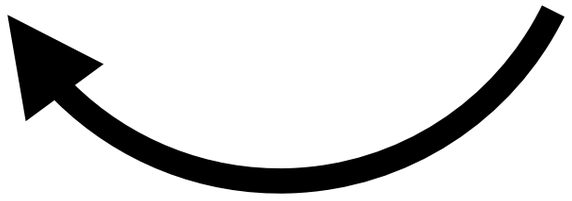


Forecast

- Run model(s) forward from estimated state(s)
- Compute forecast uncertainty

Analysis

- Use observations to adjust model states
- Compute Analysis uncertainty



Data Assimilation

- DA routinely used in weather forecasting
- Well established area with lots of methods:

  e.g. Kalman filters, variational methods, particle filters, hidden Markov models

- Typically used when forecasting

Overview

- What are Agent-Based Models (ABMs)?
- What is Data Assimilation (DA)?
- DA is easy to implement…
- …but be careful
- Some comments on DA for ABMs

Ensemble Kalman Filter

```
% Given: initial xa, x->y map H,
% obs times t_k, obs z, obs uncertainty R.

for t_k in {t_0,…,t_T}
   % Forecast
   for i in {1,…,ensemble_size}
     xf(i)=run model(xa(i),[t k, t k+1])

   xf_mean=mean(xf)% Best guess
   Pf=cov(xf)% Uncertainty

   % Analysis
   K=kalman_gain(Pf,H,R) % Solve: (H*Pf*H'+R)K=Pf*H'
   for i in {1,…,ensemble_size}
     xa(i)=xf(i)+K*(y(i)-H*xf(i))

   xa_mean=mean(xa) % Best guess
   Pa=cov(xa) % Uncertainty
```

Overview

- What are Agent-Based Models (ABMs)?
- What is Data Assimilation (DA)?
- DA is easy to implement…
- …but be careful
- Some comments on DA for ABMs

EnKF - Model

- Observation times $t_k \in \{t_0, t_1, \ldots, t_T\}$
- State vector $x_k \in \mathbb{R}^N$
- Consider models of the form

$$x_{k+1} = F(x_k) + \nu_k$$

Noise

EnKF - Model

- Observation times $t_k \in \{t_0, t_1, \ldots, t_T\}$
- State vector $x_k \in \mathbb{R}^N$
- Consider models of the form

$$x_{k+1} = M(x_k; \xi_k)$$

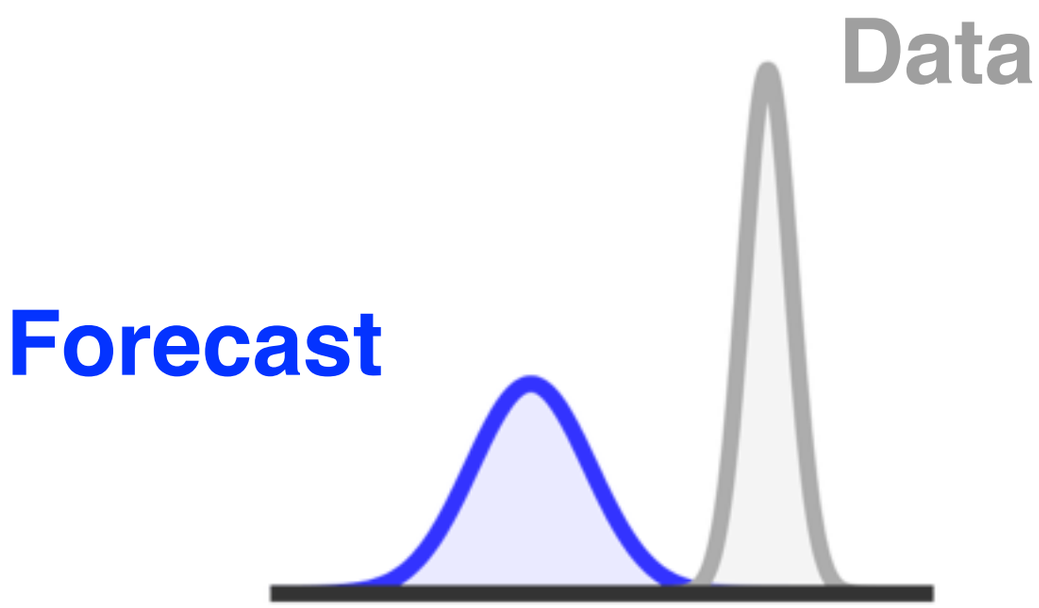
Random numbers

EnKF - Analysis

- Virtual observations:

$$z_k(i) = y_k + \nu_k(i)$$
$$\nu_k(i) \sim N(0, R)$$

- Analysis step:

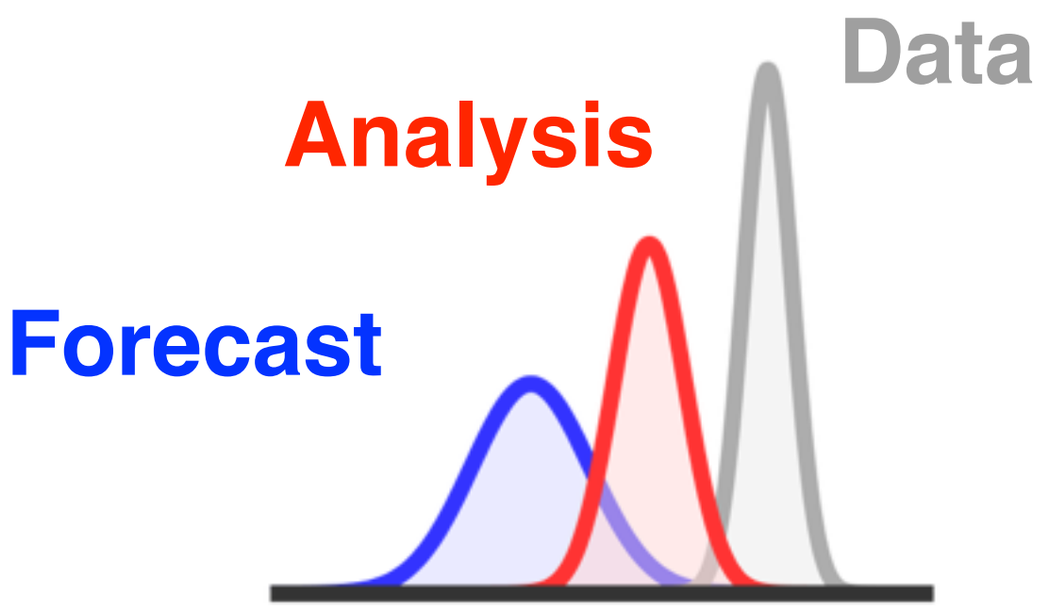


EnKF - Analysis

- Virtual observations:

$$z_k(i) = y_k + \nu_k(i)$$
$$\nu_k(i) \sim N(0, R)$$

- Analysis step:

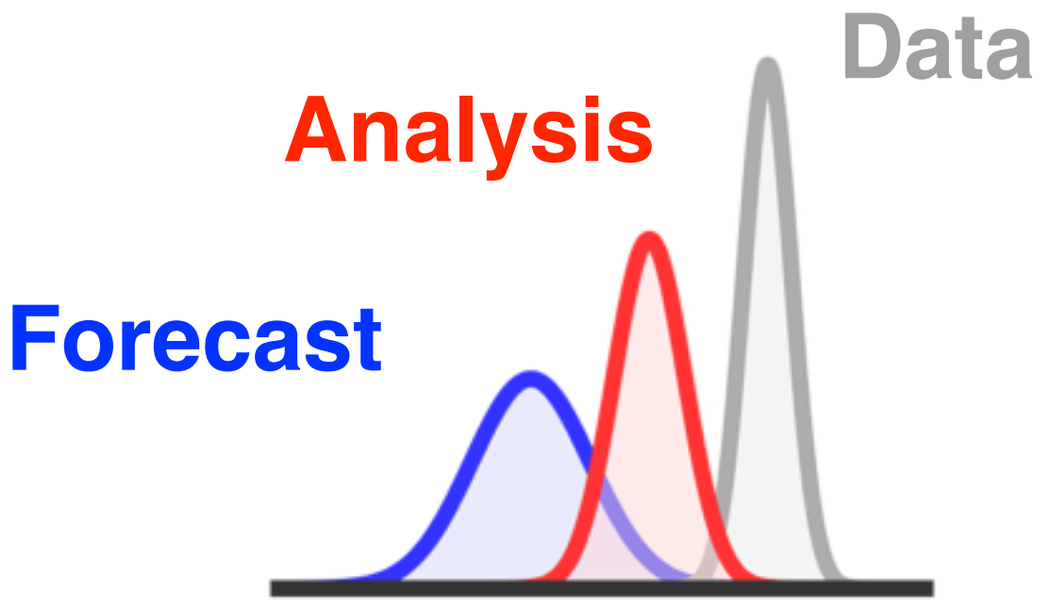


EnKF - Analysis

- Virtual observations:

$$z_k(i) = y_k + \nu_k(i)$$
$$\nu_k(i) \sim N(0, R)$$

- Analysis step:



Bayes

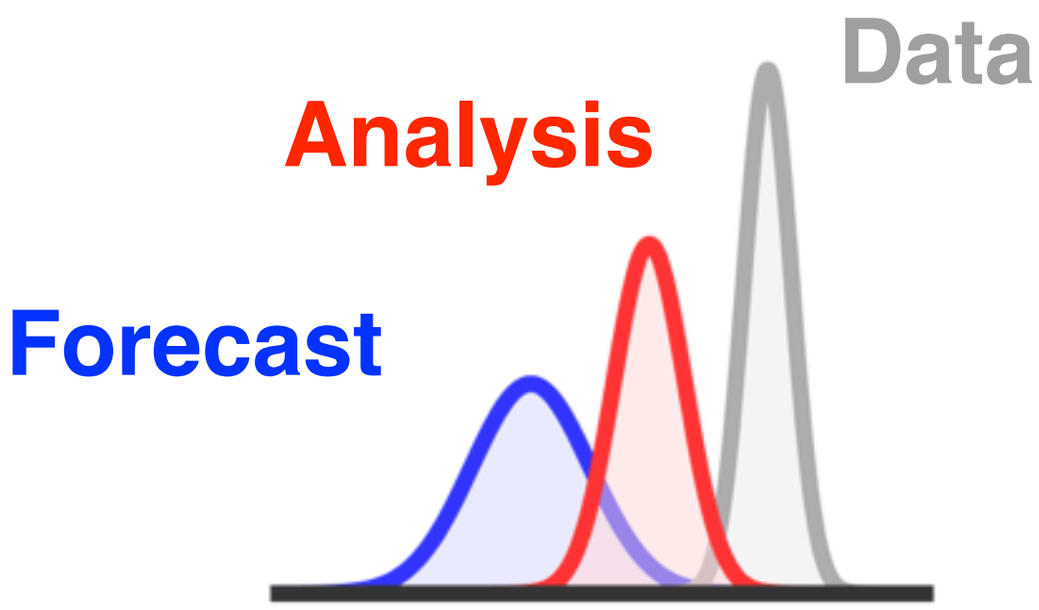
$$p(x|y) \propto p(y|x)p(x)$$

EnKF - Analysis

- Virtual observations:

$$z_k(i) = y_k + \nu_k(i)$$
$$\nu_k(i) \sim N(0, R)$$

- Analysis step:



Bayes

$$p(x|y) \propto p(y|x)p(x)$$

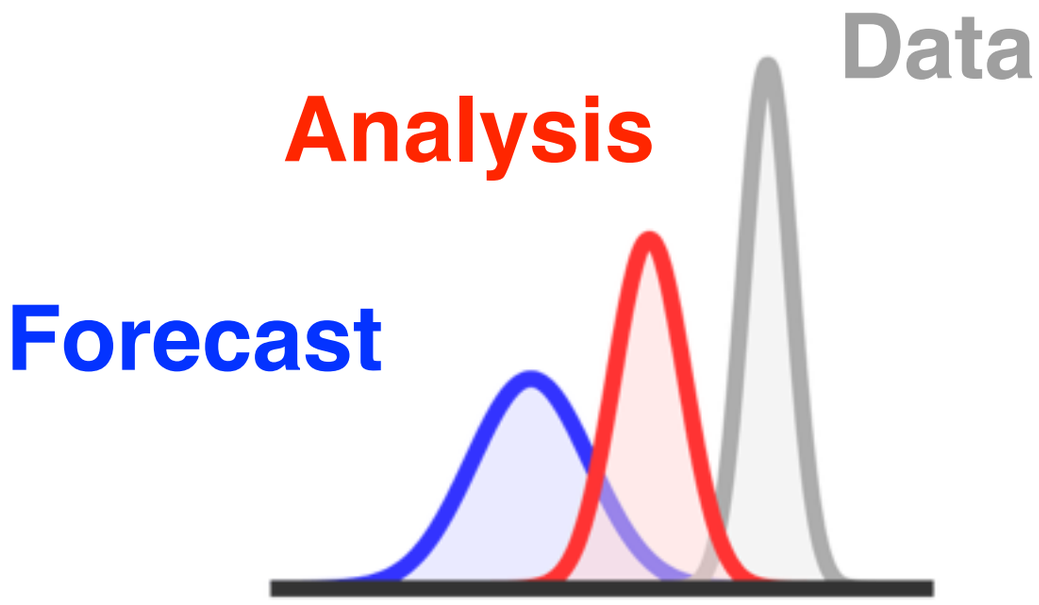
model stochasticity

EnKF - Analysis

- Virtual observations:

$$z_k(i) = y_k + \nu_k(i)$$
$$\nu_k(i) \sim N(0, R)$$

- Analysis step:



Bayes

$$p(x|y) \propto p(y|x)p(x)$$

observation error

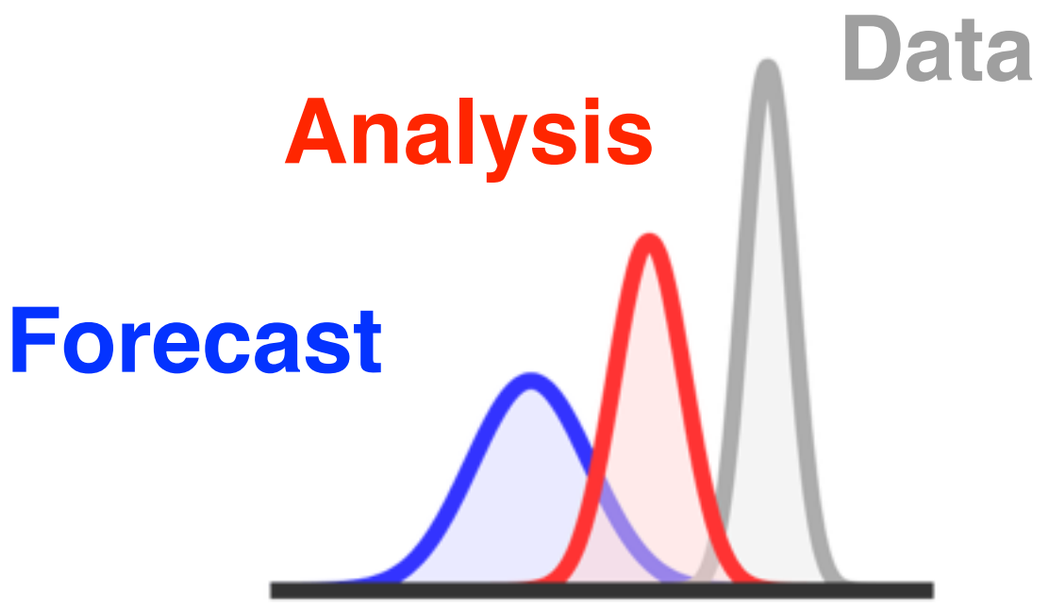
model stochasticity

EnKF - Analysis

- Virtual observations:

$$z_k(i) = y_k + \nu_k(i)$$
$$\nu_k(i) \sim N(0, R)$$

- Analysis step:



Bayes

$$p(x|y) \propto p(y|x)p(x)$$

observation error

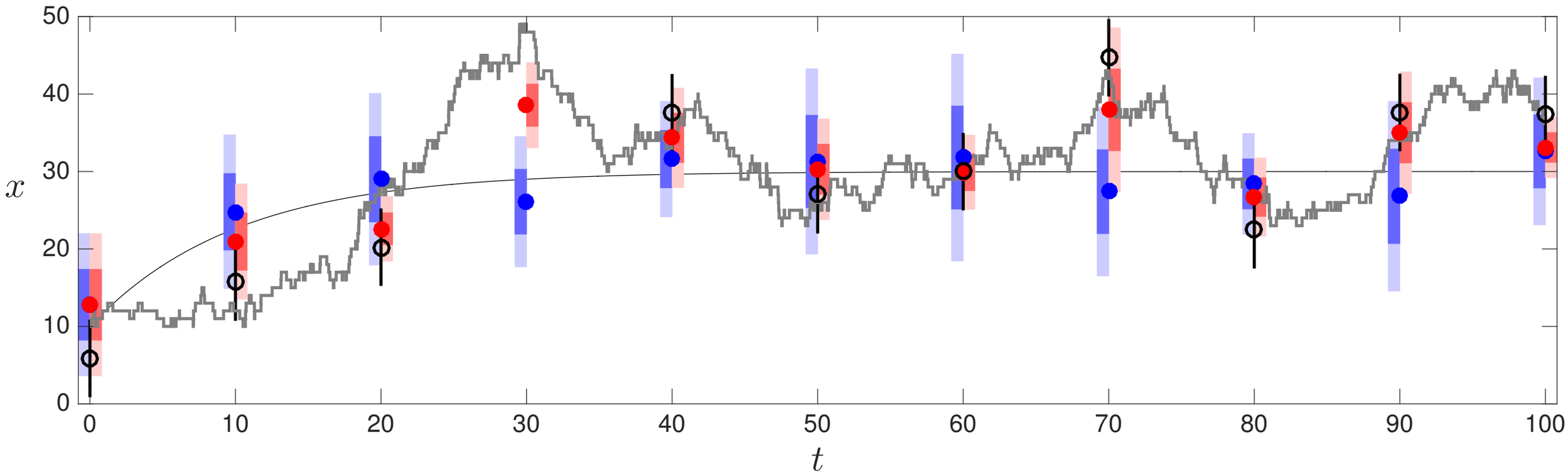
model stochasticity

Normal distributions

State Estimation

EnKF Example

$$N = 50, \quad n_0 = 10, \quad \alpha = 3, \quad \beta = 0.1, \quad R = 25, \quad N_{\text{real}} = 10$$

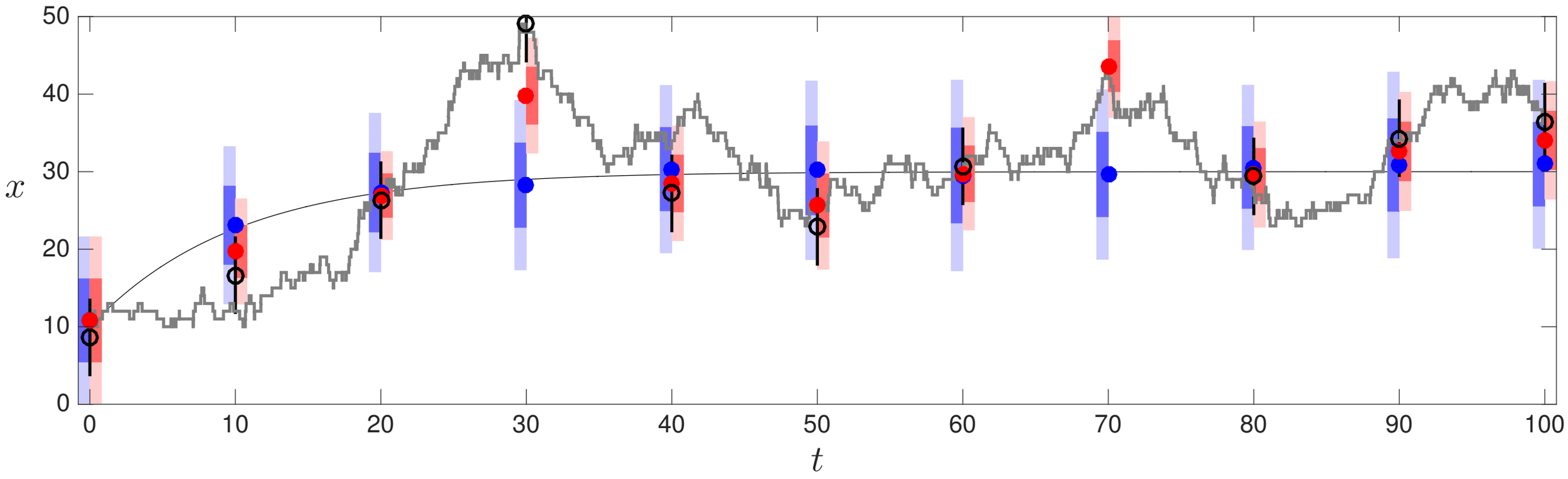


RMSEs

Steady state	Observations	Forecast	Analysis
8.20	4.32	9.55	4.87

EnKF Example

$$N = 50, \quad n_0 = 10, \quad \alpha = 3, \quad \beta = 0.1, \quad R = 25, \quad N_{\text{real}} = 100$$

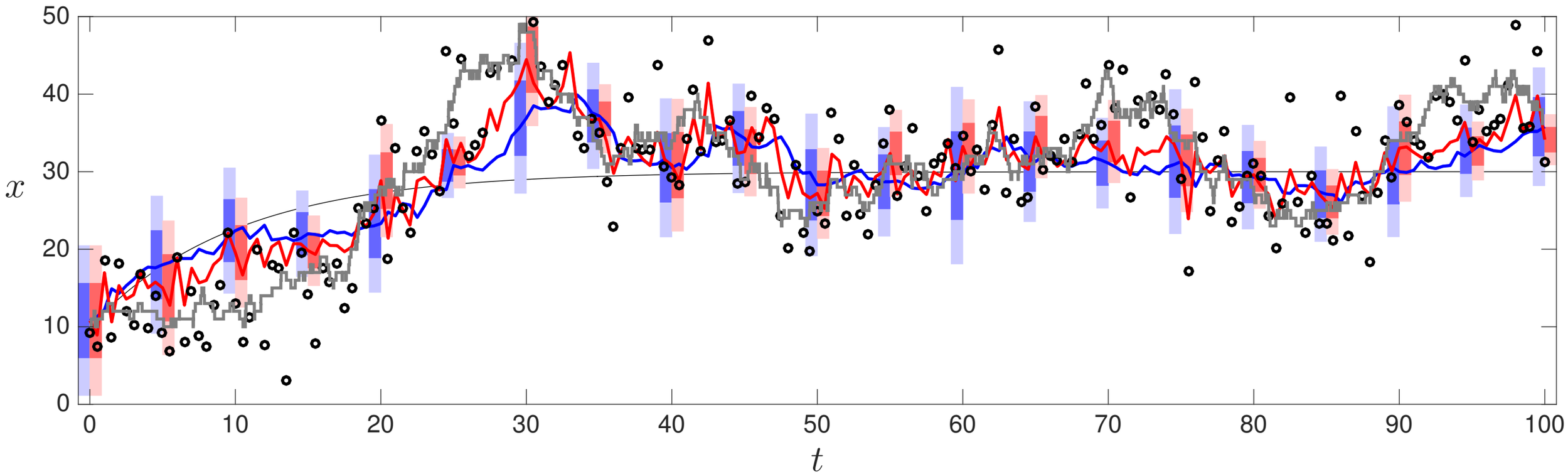


RMSEs

Steady state	Observations	Forecast	Analysis
8.20	5.01	8.13	3.39

EnKF Example

$$N = 50, \quad n_0 = 10, \quad \alpha = 3, \quad \beta = 0.1, \quad R = 25, \quad N_{\text{real}} = 10$$

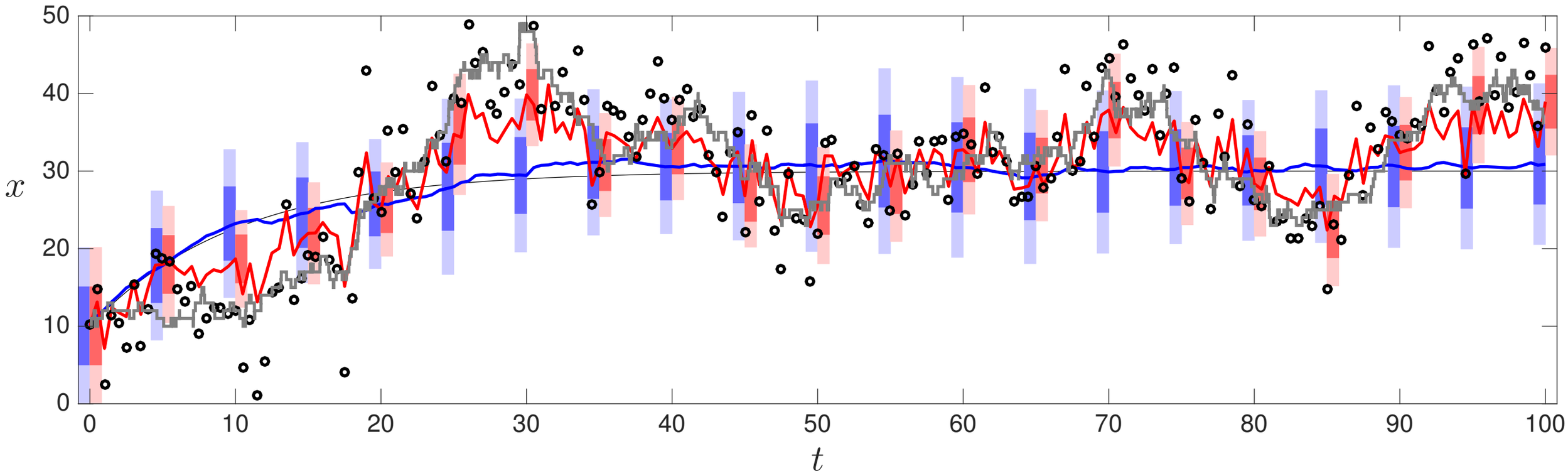


RMSEs

Steady state	Observations	Forecast	Analysis
7.17	5.10	5.62	4.29

EnKF Example

$$N = 50, \quad n_0 = 10, \quad \alpha = 3, \quad \beta = 0.1, \quad R = 25, \quad N_{\text{real}} = 100$$

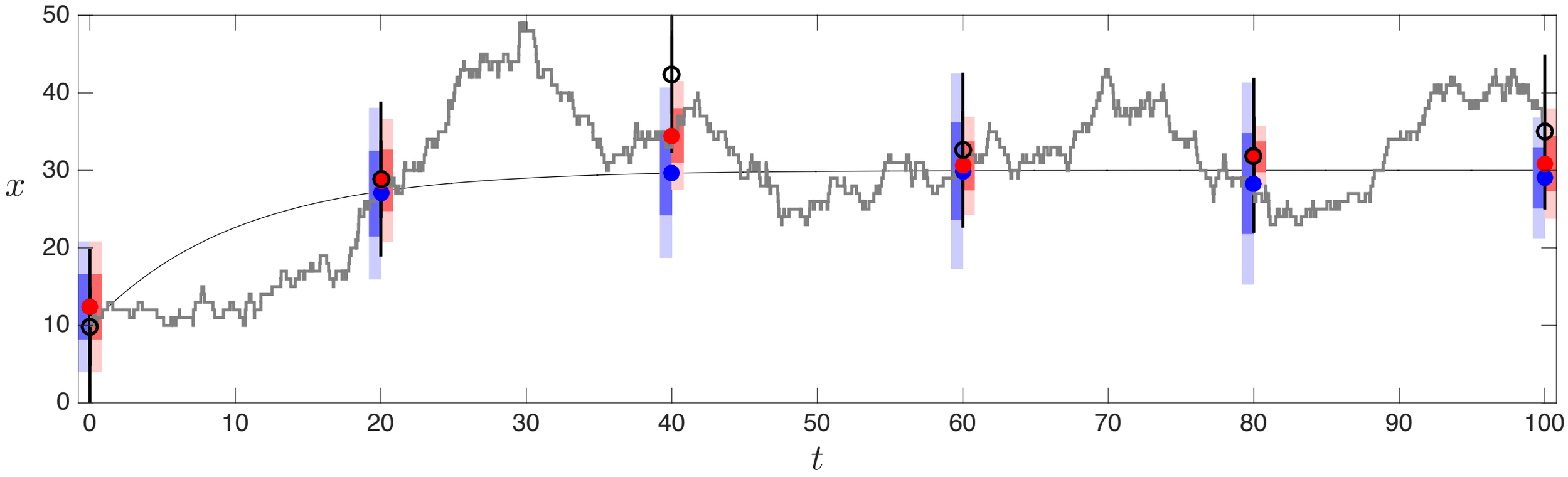


RMSEs

Steady state	Observations	Forecast	Analysis
7.17	5.11	6.90	4.19

EnKF Example

$$N = 50, \quad n_0 = 10, \quad \alpha = 3, \quad \beta = 0.1, \quad R = 25, \quad N_{\text{real}} = 10$$



RMSEs

Steady state	Observations	Forecast	Analysis
3.30	4.24	3.58	3.11

Comments

- If measurements accurate, analysis performance will be similar

- Forecasts will perform better if less noise

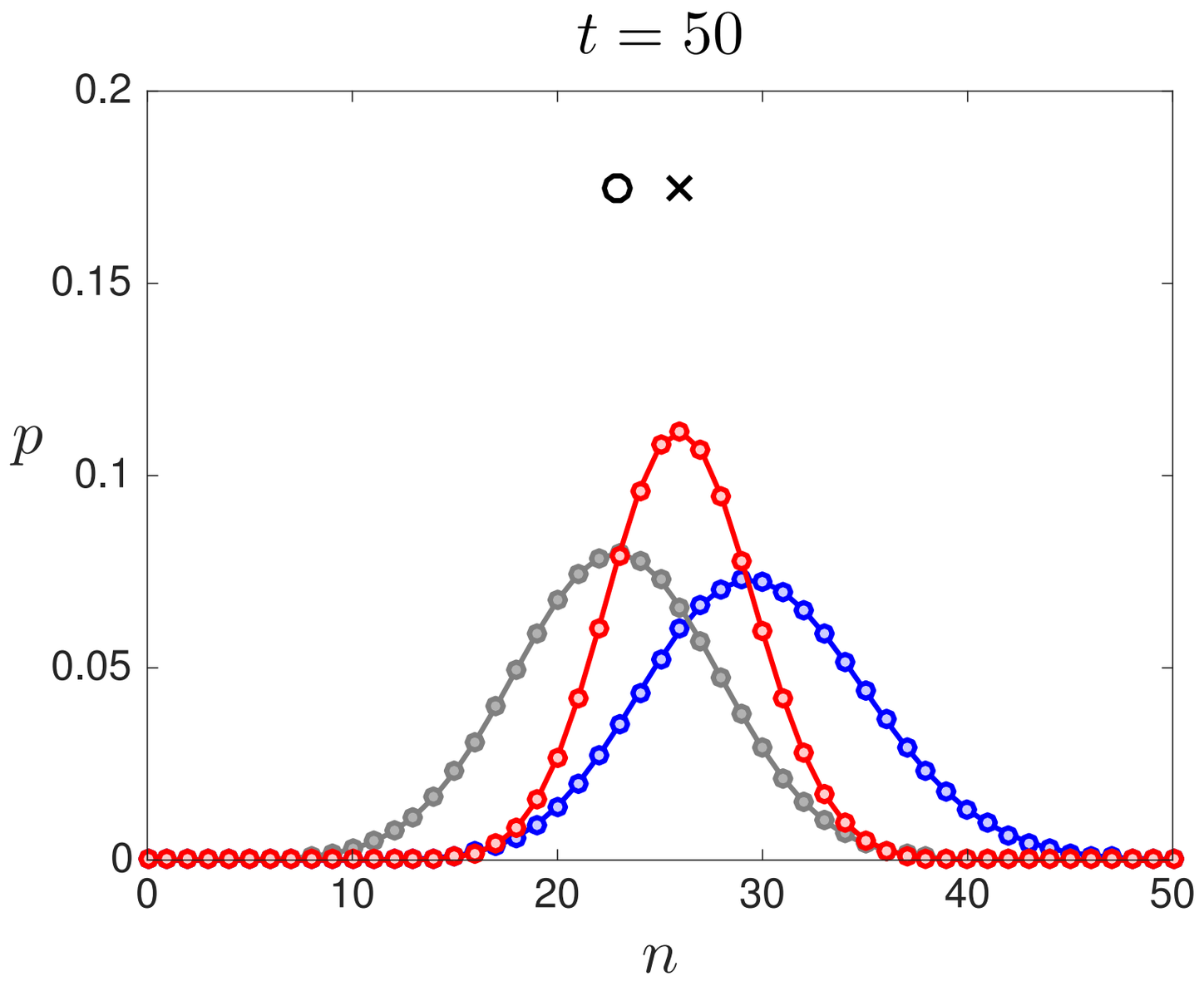
- Is the model a good forecast of the data?

- (Taking more history doesn't necessarily help since Markovian)

- Forecasting involves luck! (Hence errors important)

Exact Analysis

Analysis

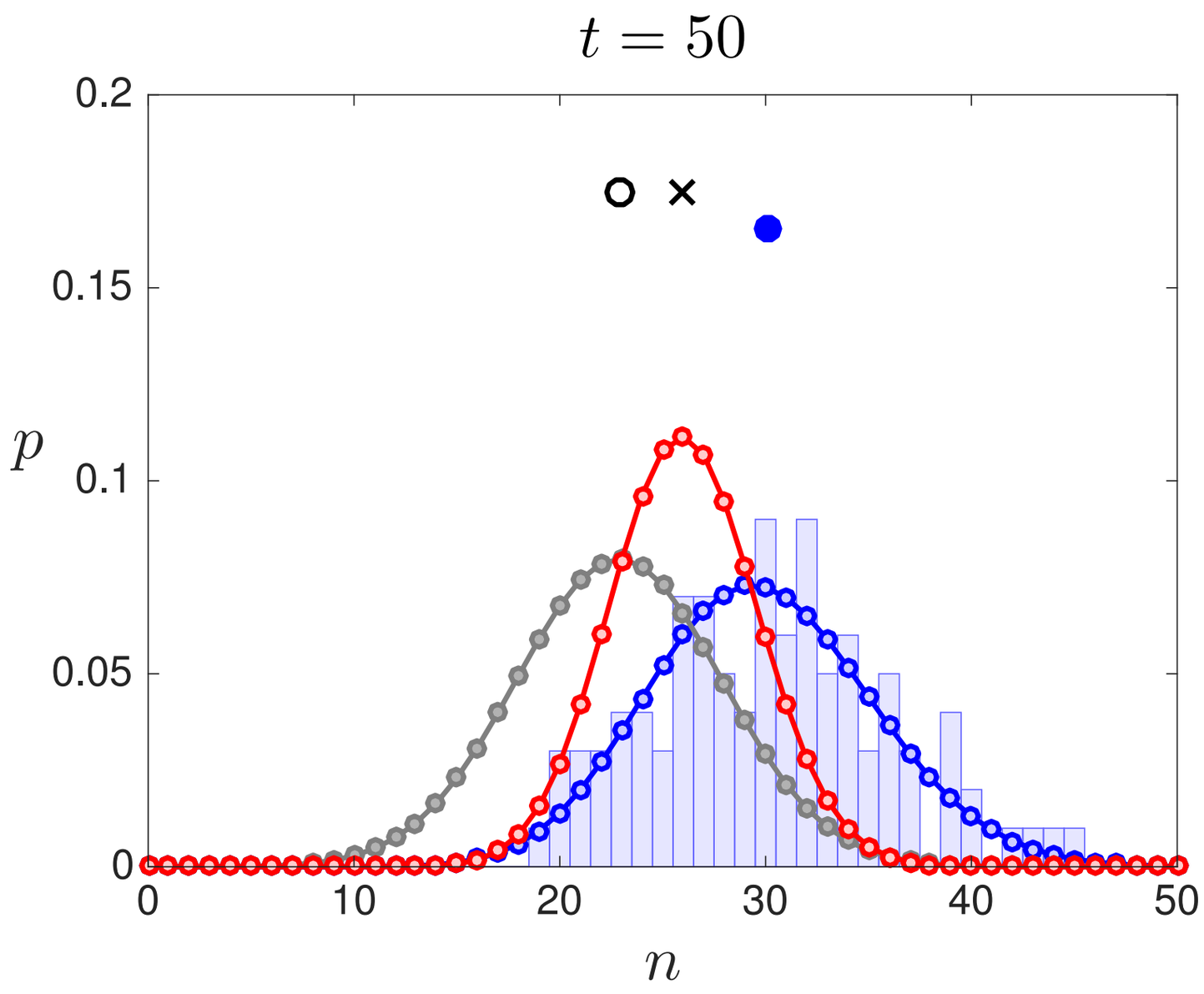


Data

Forecast

Exact Analysis

Analysis

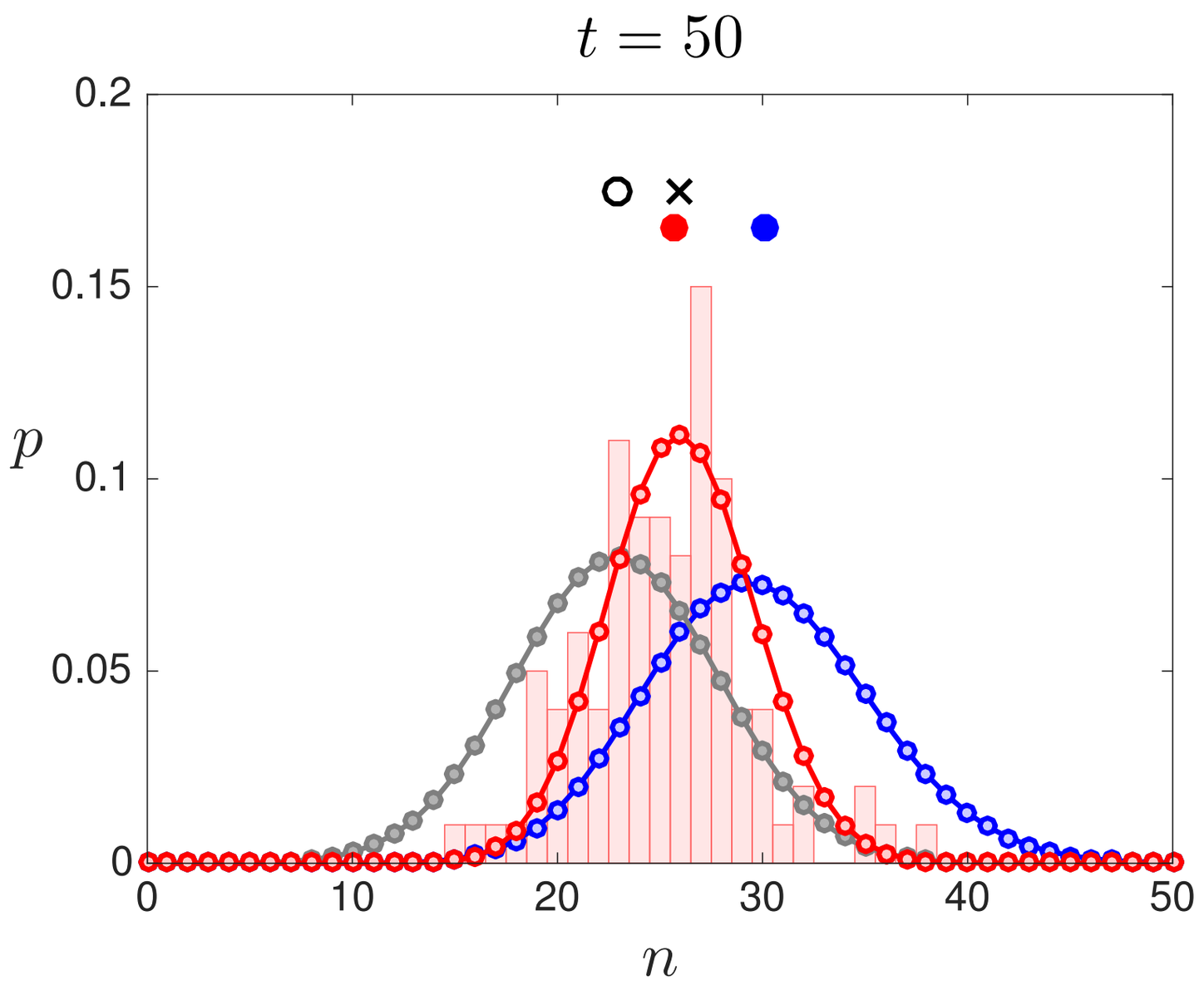


Data

Forecast

Exact Analysis

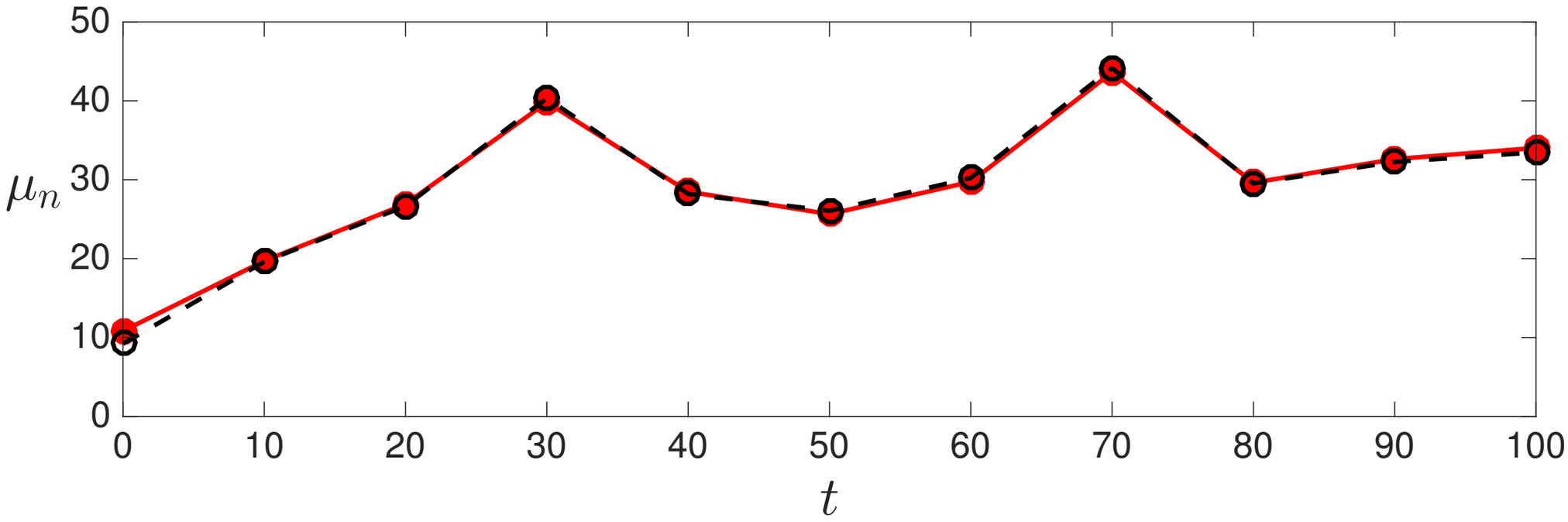
Analysis

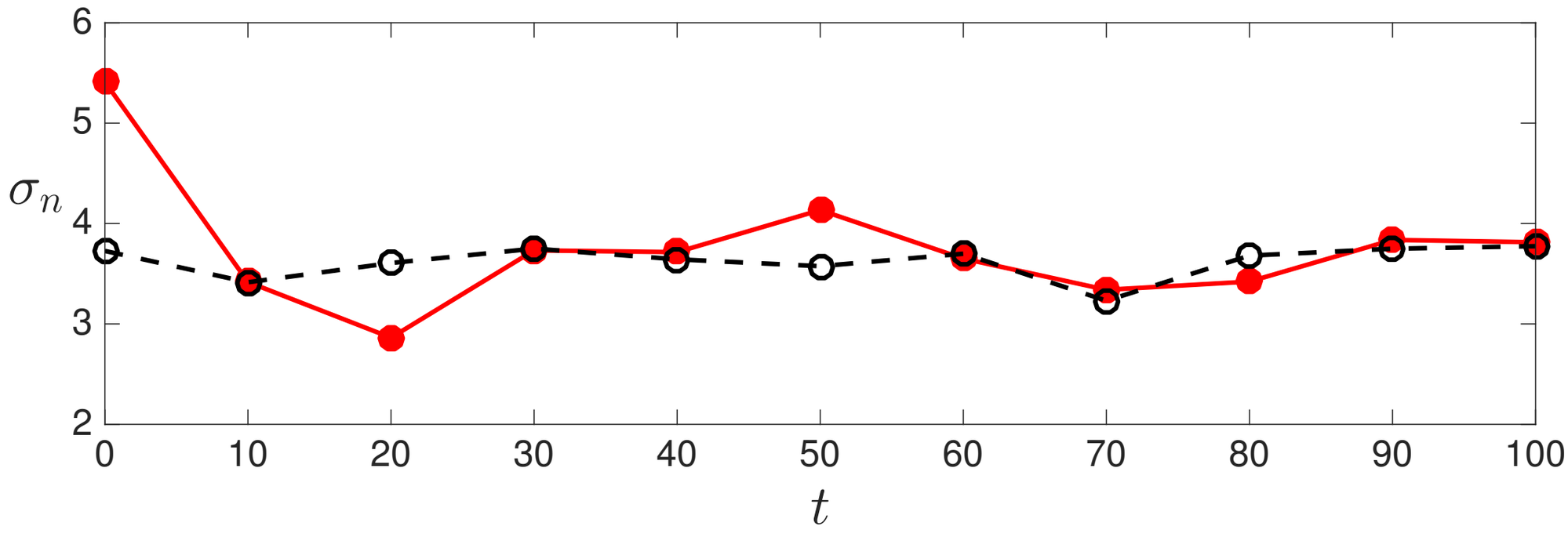


Data

Forecast

Exact Analysis

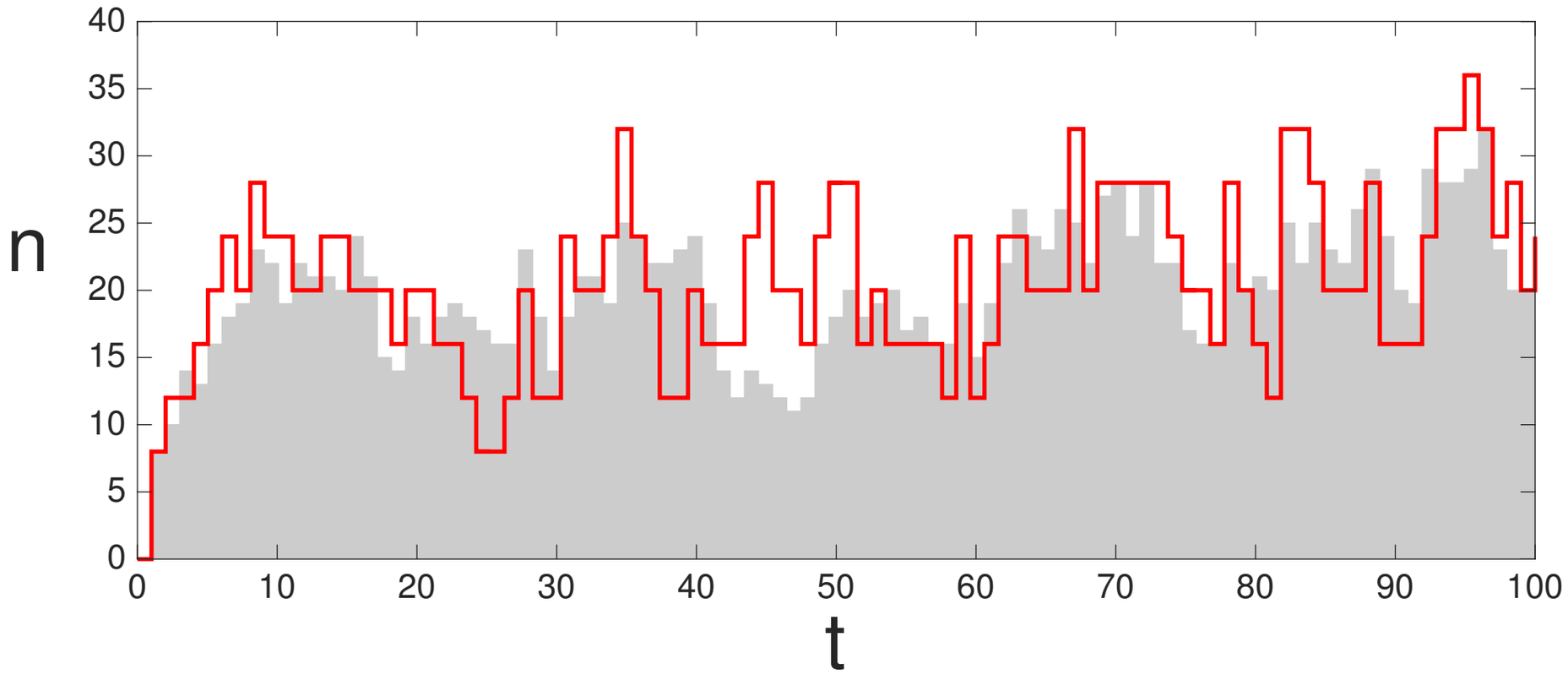




Observations

- GPS trackers on a sample of agents

$$y(t) = \frac{N}{|S|} \sum_{i \in S} x_i(t)$$



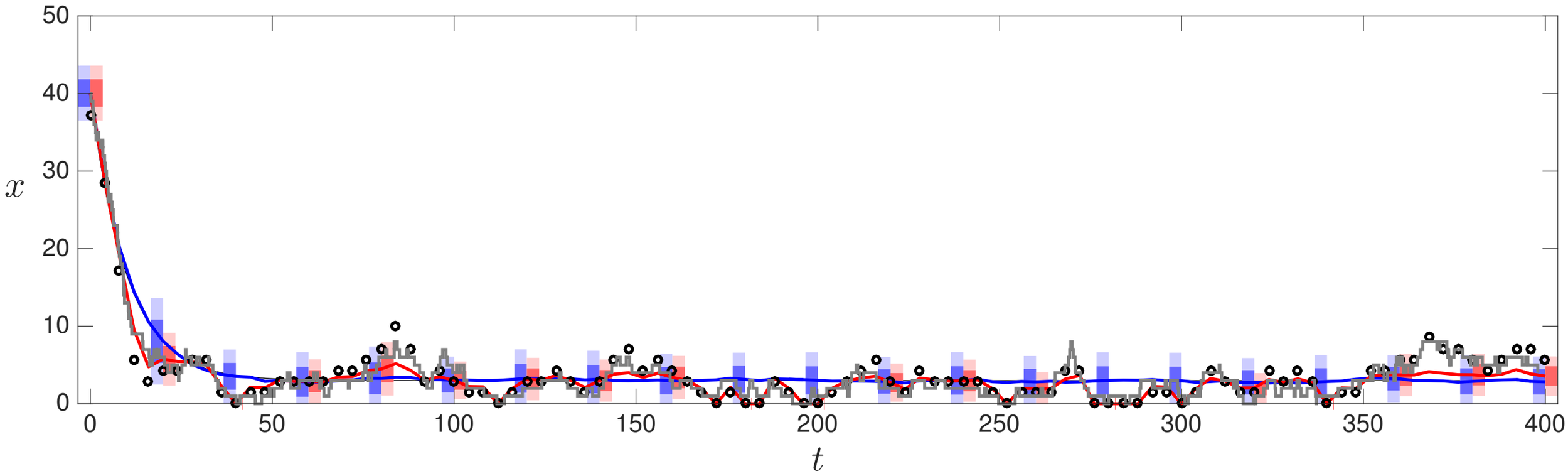
Error Statistics

- Variance quadratic in n, but n unknown

- Resample approximation at each time step

Non-Normality

$$N = 50, \quad n_0 = 40, \quad \alpha = 0.3, \quad \beta = 0.1, \quad M = 35, \quad N_{\text{real}} = 100$$

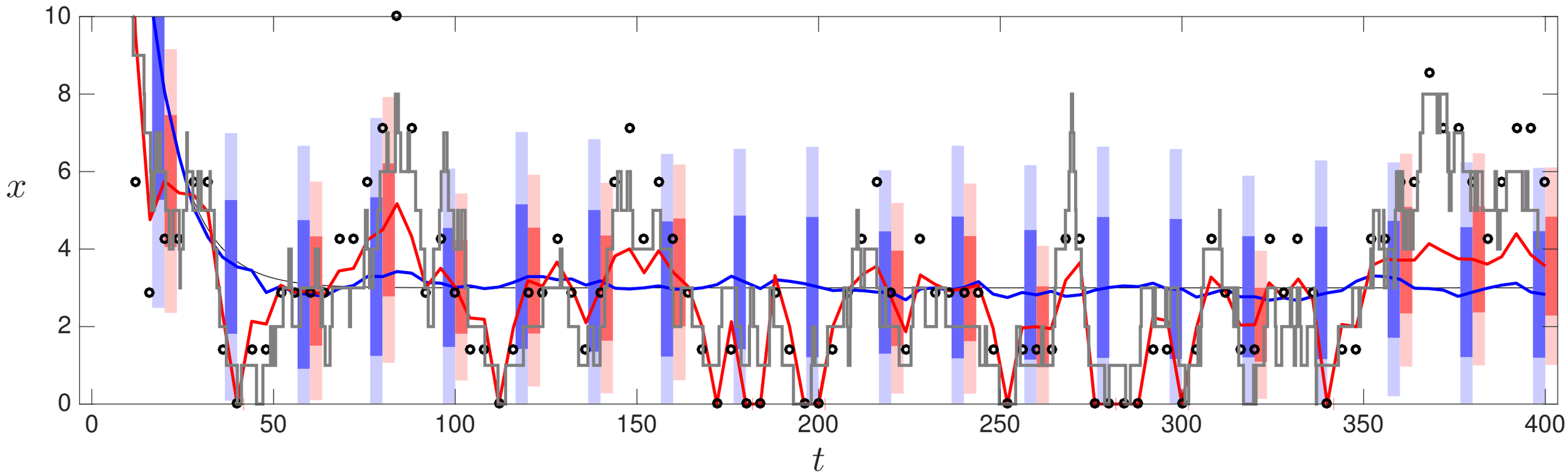


RMSEs

Steady state	Observations	Forecast	Analysis	Posterior
1.97	1.13	1.94	1.18	0.83

Non-Normality

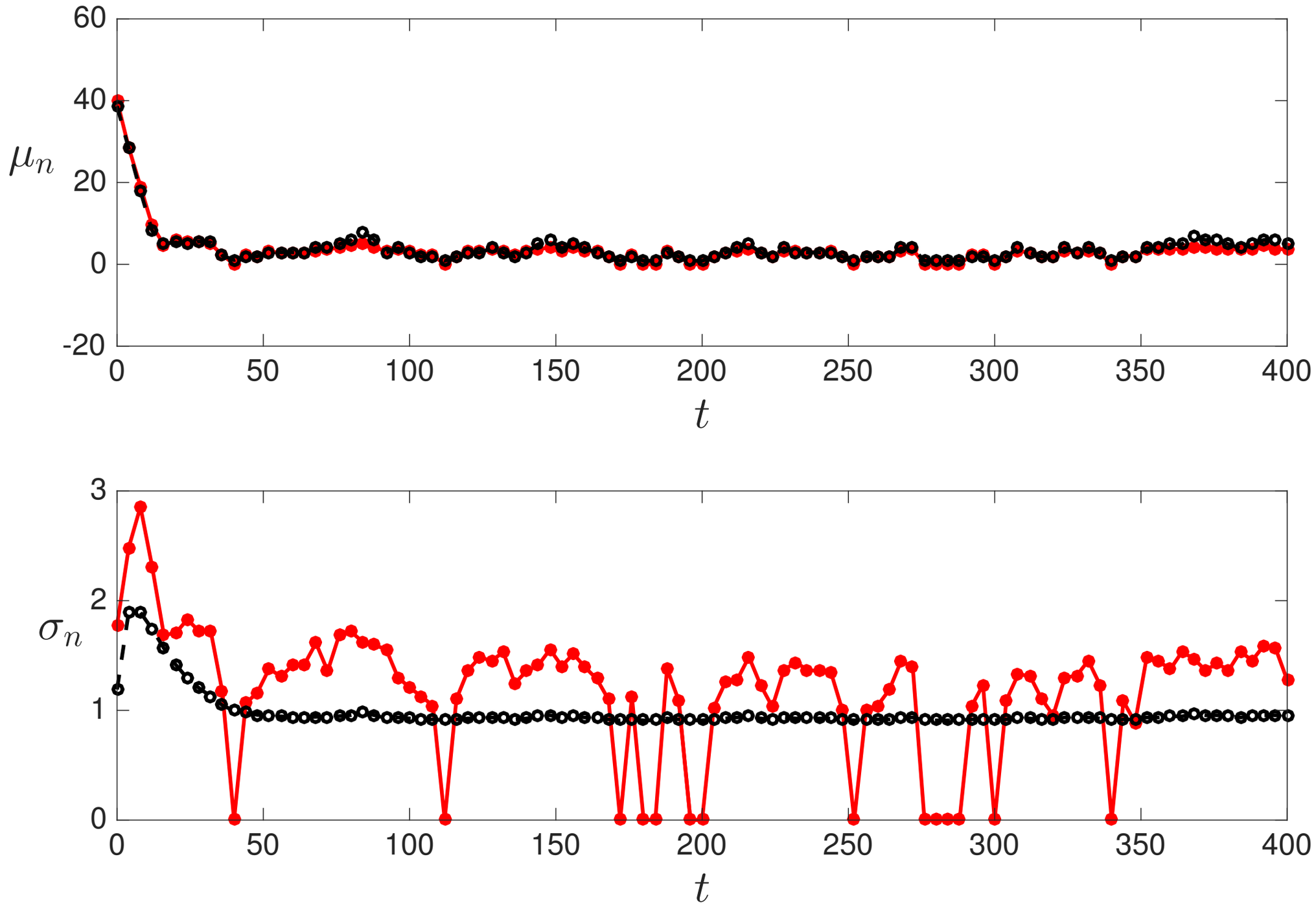
$$N = 50, \quad n_0 = 40, \quad \alpha = 0.3, \quad \beta = 0.1, \quad M = 35, \quad N_{\text{real}} = 100$$



RMSEs

Steady state	Observations	Forecast	Analysis	Posterior
1.97	1.13	1.94	1.18	0.83

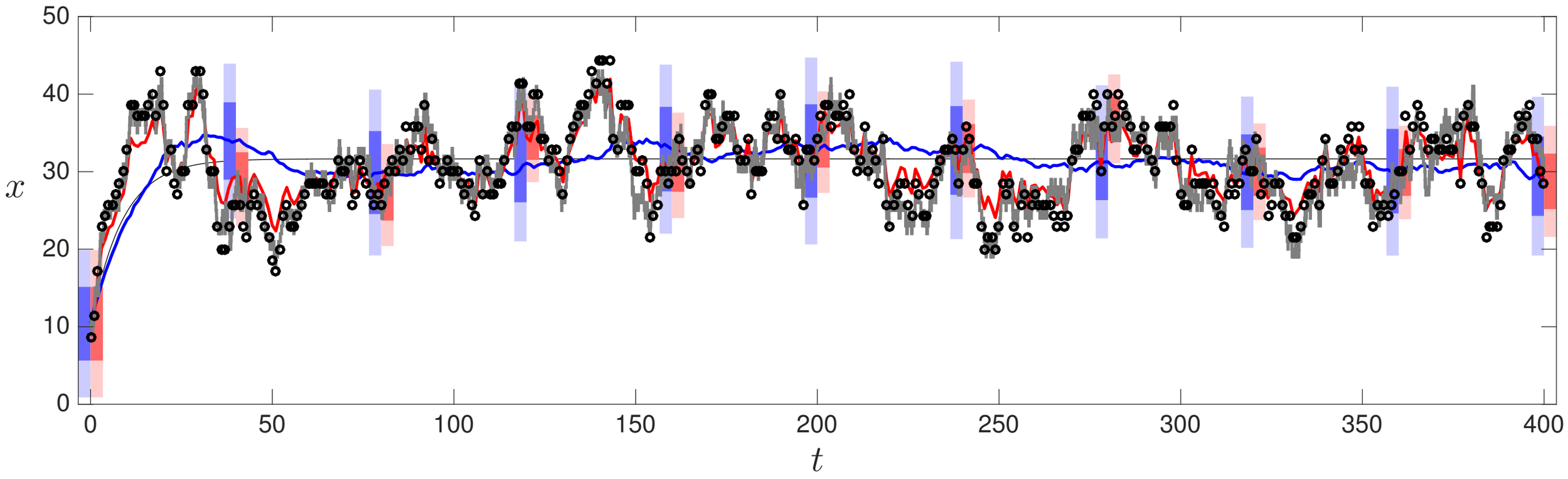
Non-Normality



Parameter Estimation

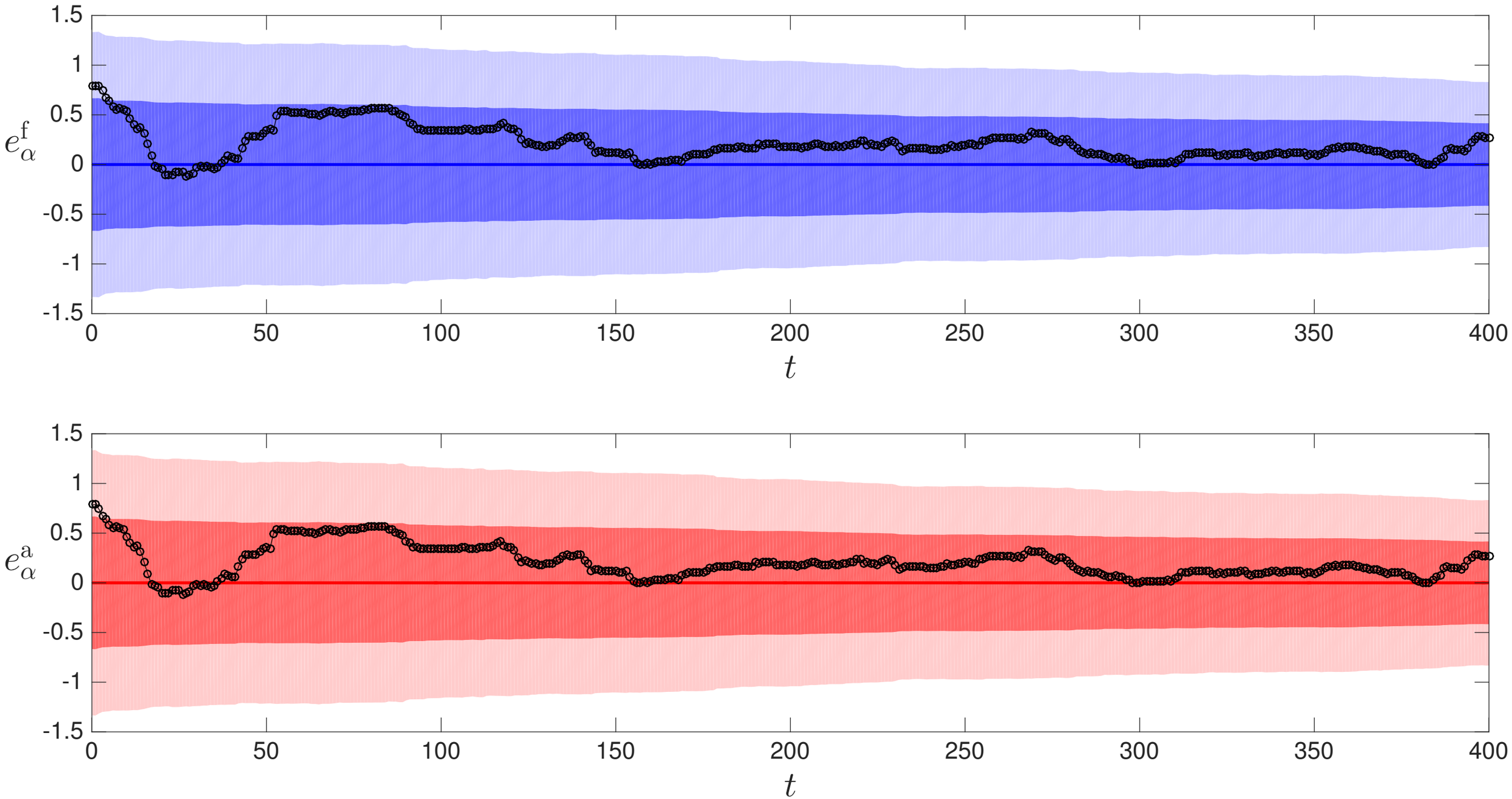
Parameter Estimation

$$N = 50, \quad n_0 = 10, \quad \alpha = 3.65, \quad \beta = 0.12, \quad M = 35, \quad N_{\text{real}} = 100$$

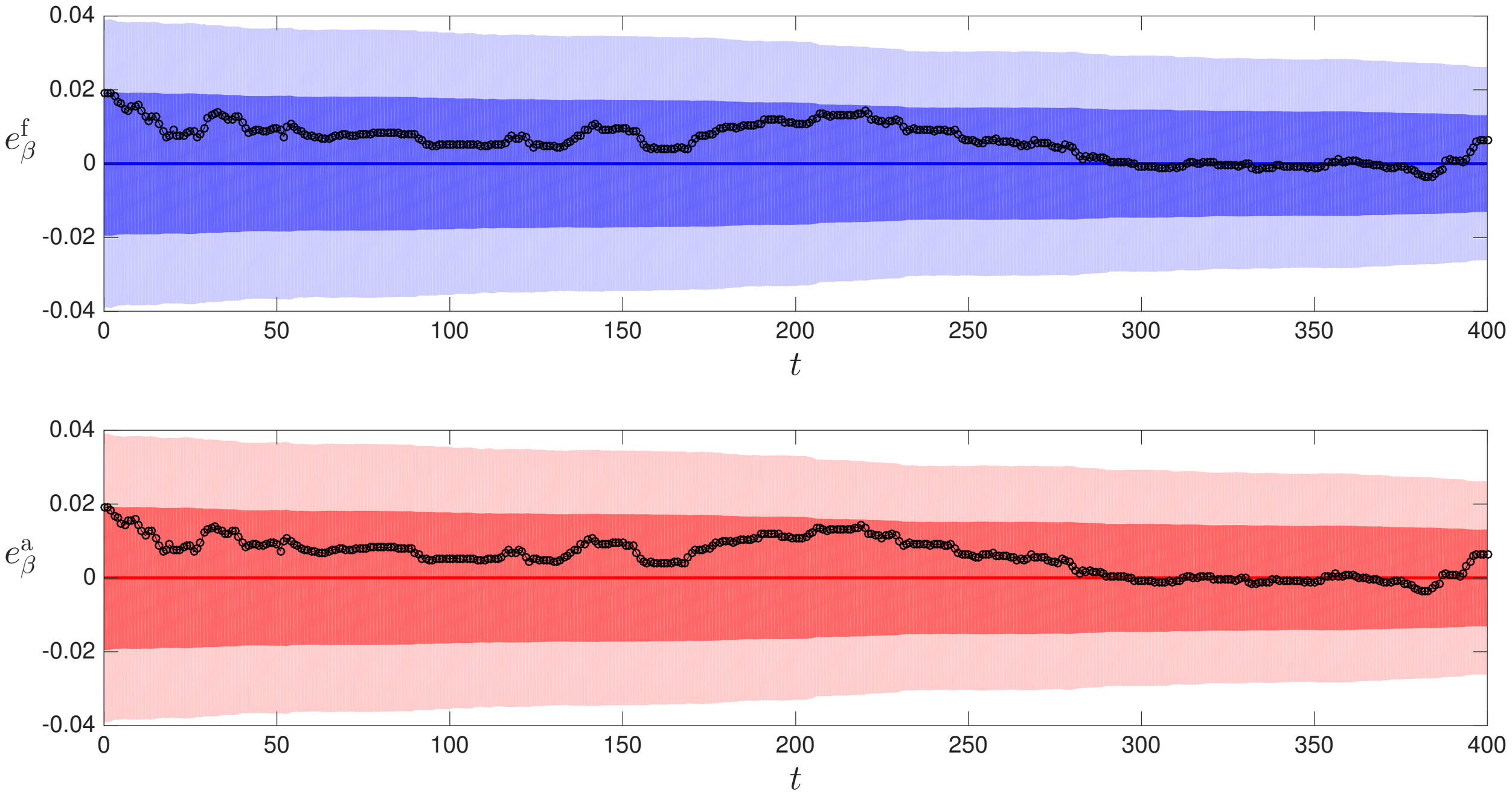


Steady state	Observations	Forecast	Analysis
5.11	1.84	5.26	2.26

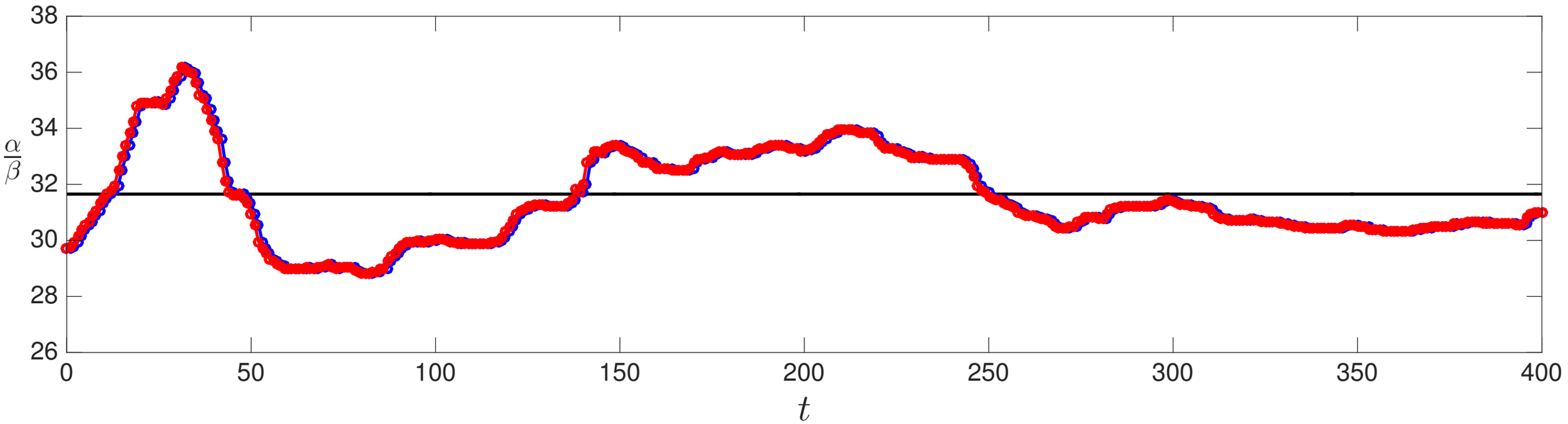
Parameter Estimation



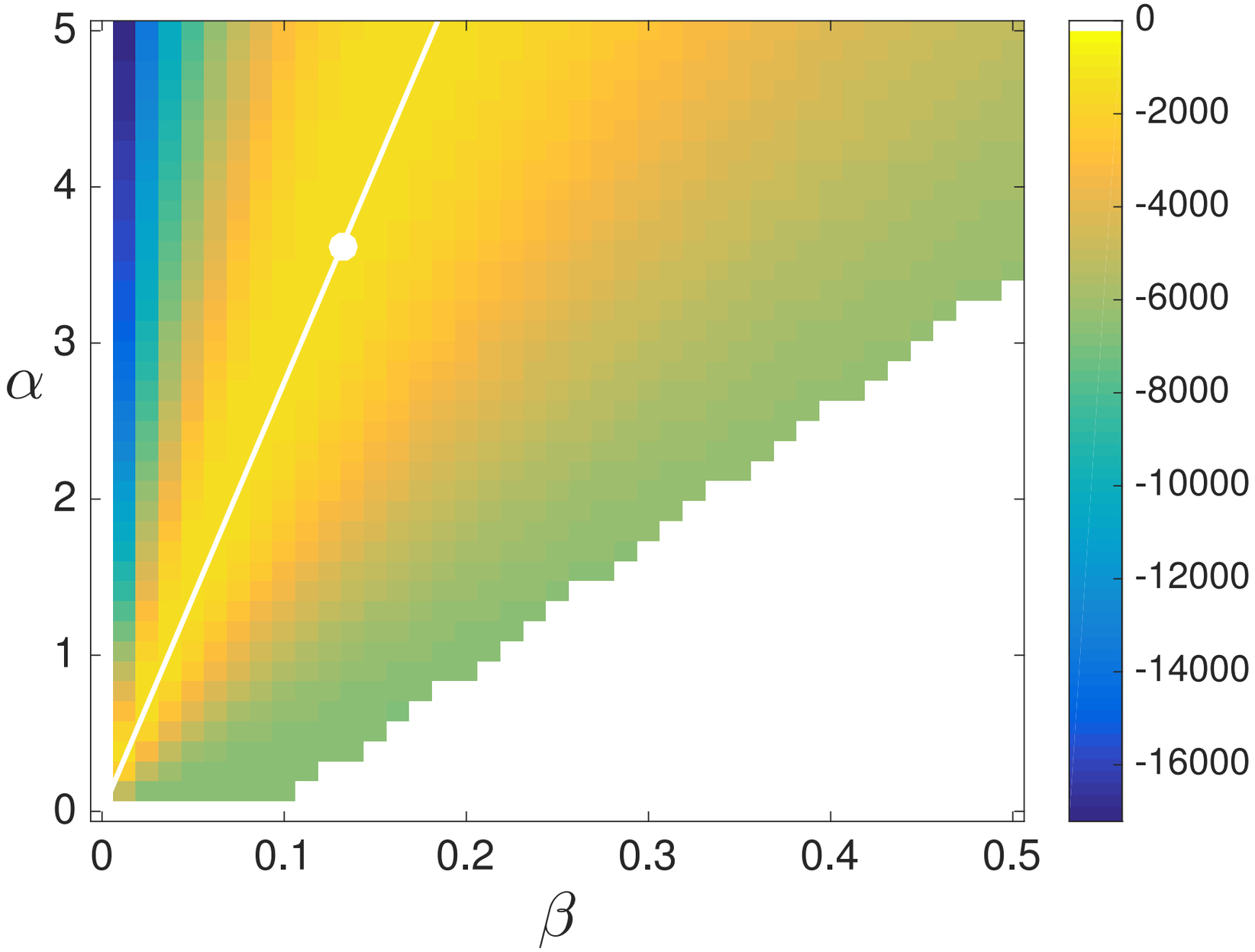
Parameter Estimation



Parameter Estimation



Parameter Estimation



Other Issues

- Agent heterogeneity
- Exogenous events & dynamic adaptation
- Single shot calibration vs dynamic DA
- Fitting the wrong model
- Non-Markovian models
- (Variable rescaling, rounding errors, practical issues)

Overview

- What are Agent-Based Models (ABMs)?
- What is Data Assimilation (DA)?
- DA is easy to implement…
- …but be careful
- Some comments on DA for ABMs

Calibration

Model validation
Data assimilation

Data Analysis

Statistical properties
Benchmark forecasts

Model Mechanisms

Avoid `black swans' via
multiple mechanisms
How do you combine them?

Model hierarchy

Analysis of simple models to
inform higher complexity models
Dimension reduction

Mapping Parameter Space

Identify dynamical regimes
via equation free methods
Parameter sensitivity

Conclusions

- Understanding model dynamics is important
- ABMs to forecast?!
- Gaussian mixtures, mixed discrete/continuous?
- Identifying when it’s not working
- Translation of expertise from developed disciplines
- Can ABMs compare with machine learning?

Thanks!