



Why computational modeling and simulation? Do's and Don'ts.

Ivo F. Sbalzarini



I. Introduction

The promise

Why Computers ?

Amount of data /
Reproducibility



Manual evaluation too slow
or too unreliable

Complexity



System behavior not
apparent from description

Time/length scales



Too big/small, slow/fast for
experimental measurement

Ethics



No living beings involved

Controllability



Variables controllable

Observability



Variables measurable

Computation in Biology

Computer science will be
for 2020 biology what
mathematics is for today's
physics.



2. Methods

How is it usually done?

Definitions

A **system** is a potential source of data

A system has:

- a boundary (we can define what is inside and what is outside the system)
- inputs (the environment influencing the system)
- outputs (the system influencing the environment)

Do not confuse in/output with in/outflows!

Definitions

An **experiment** is the process of extracting data from a system

By:

- observing the system inputs (observation) or exciting/ changing the system inputs (perturbation experiment)
- observing the trajectories of the outputs

Usually, not all inputs are controllable / not all outputs observable!

Definitions

A **model** for a system, and a specific set of experiments, is anything to which an experiment can be applied in order to answer questions about the system.

- each model is itself a system (hierarchy)
- models need not be mathematical
- no model is valid for all experiments (and all models are valid for the null experiment)

All models have a finite experimental frame!

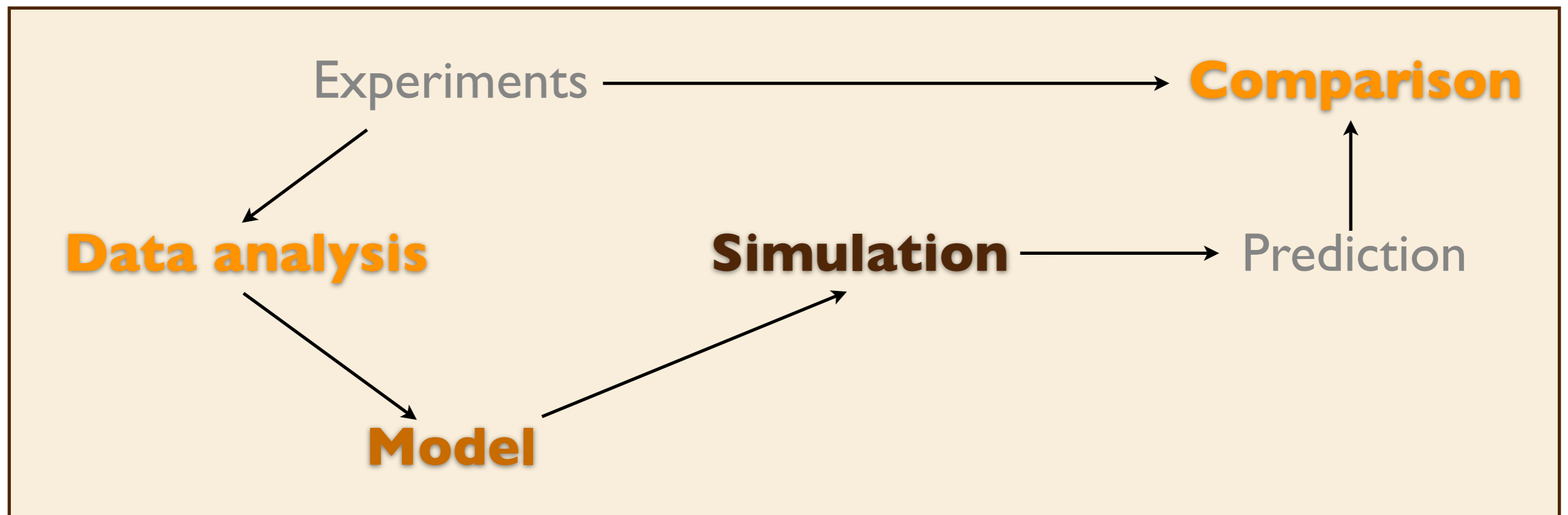
Definitions

A **simulation** is an experiment performed on a model.

- simulations need not be computational
- careful that the simulation does not violate the experimental frame of the model!
- model definition and simulation specs should be modularly separated

Typical workflow

- Understanding of function with minimal assumptions
- Data analysis and simulation using computers



What can be shown?

- **Simulation:** is the model sufficient to explain the data?
(sufficiency in both ingredients and their interaction mechanisms)
- **Perturbation Experiment:** is a predicted mechanism necessary?
(knock-out experiment)

Note: reconstitution experiments show sufficiency of ingredients, but not of interaction mechanisms.

Ockham's razor

The **simplest** model that is sufficient to explain the data is the one that should be used (i.e., from which we can learn the most).

G.E.P. Box:

“All models are wrong, but some are useful”



Computational modeling

1. Mathematical description of the system/experiment
2. Discretization (numerics)
3. Implementation in a programming language
4. Simulate
5. Verify and validate
6. Analyze the results

Verification / Validation

- **Verification:** am I simulating the model correctly?
(algorithm and code correctness / convergence plots / comparison to simplified models with analytical solutions)
- **Validation:** am I simulating the correct model
(comparison to data / cross-validation)



Questions to answer

- What is the proper level of detail for the model?
(QM, MM, Focker-Planck, Boltzmann, continuum?)
- Which physics/constraints are known/valid?
- How can the model be verified and validated?
(models that cannot be validated are useless)
- What is the experimental frame of the model?
- What hypothesis/mechanism shall the models show sufficiency of?
- Could the same output be obtained more directly?

Frequent problems

- Governing equations unknown
- Multiple scales present at once
(Is there a separation of scales, so multi-scale methods could apply?)
- Insufficient prior knowledge about the system
(Dimensional analysis could still help...)
- Simulations of the resulting models are computationally infeasible
- The model has a lot of free parameters (that need to be identified **ROBUSTLY**)

The crux of validation

- Whenever a model is needed to get access to non-observable variables, direct validation is impossible.
- Validation thus is mostly indirect, via the observable subset of outputs.
- If the observable outputs are sensitive to variations in the non-observable ones, validation is stronger
- Usually, infinitely many hidden dynamics can produce the same observables. Choose the simplest model that reproduces the data (Ockham's razor)!

Model classes

	continuous	discrete
deterministic	PDEs	interacting particles
	diffusion	molecular dynamics
stochastic	SDEs	random events
	reaction-diffusion with low molecule numbers	population dynamics

Phenomenological
Physical



3. Discussion

Do's and Don'ts

Do's

- **Specify the desired accuracy**
(As with experiments, you need a hypothesis)
- **Make sure the model adds value to your project**
- **Start with the simplest model possible and add more as needed**
- **Make sure the model can be validated and verified**
- **Think of a model as a hypothesis you design!**
- **Choose the most efficient simulation algorithm and the best implementation**
(You are going to run the simulation more often than you think)

Don'ts

- Aim for “complete” or “correct” models. Accept that models are (deliberately) wrong
- Start modeling if you don't have to (and know why!)
- Validate on data used to build the model
- Consider modeling as something you can quickly do on the side
- Give up if the model does not confirm your expectations
- Build a model without having a hypothesis