Coupling Computer Models through Linking their Statistical Emulators

Ksenia Kyzyurova
Joint work with Jim Berger and Robert Wolpert.

Duke University, Department of Statistical Science

ksenia@stat.duke.edu

EPFL, Lausanne, Switzerland
April 7, 2016

Introduction

Shinmoedake, Japan, 2011.

Motivation - Volcano eruptions

Natural hazards concern.

Populated areas, air traffic routes.

Forecasting.

Little data from natural events.

Computer simulators.

Where does statistics come?

Computer models take long time: hours, days, etc.

Statistical ***emulators*** approximate computer simulators.

Combining computer models together is called ***coupling***.

Linking emulators.

Outline

Emulator

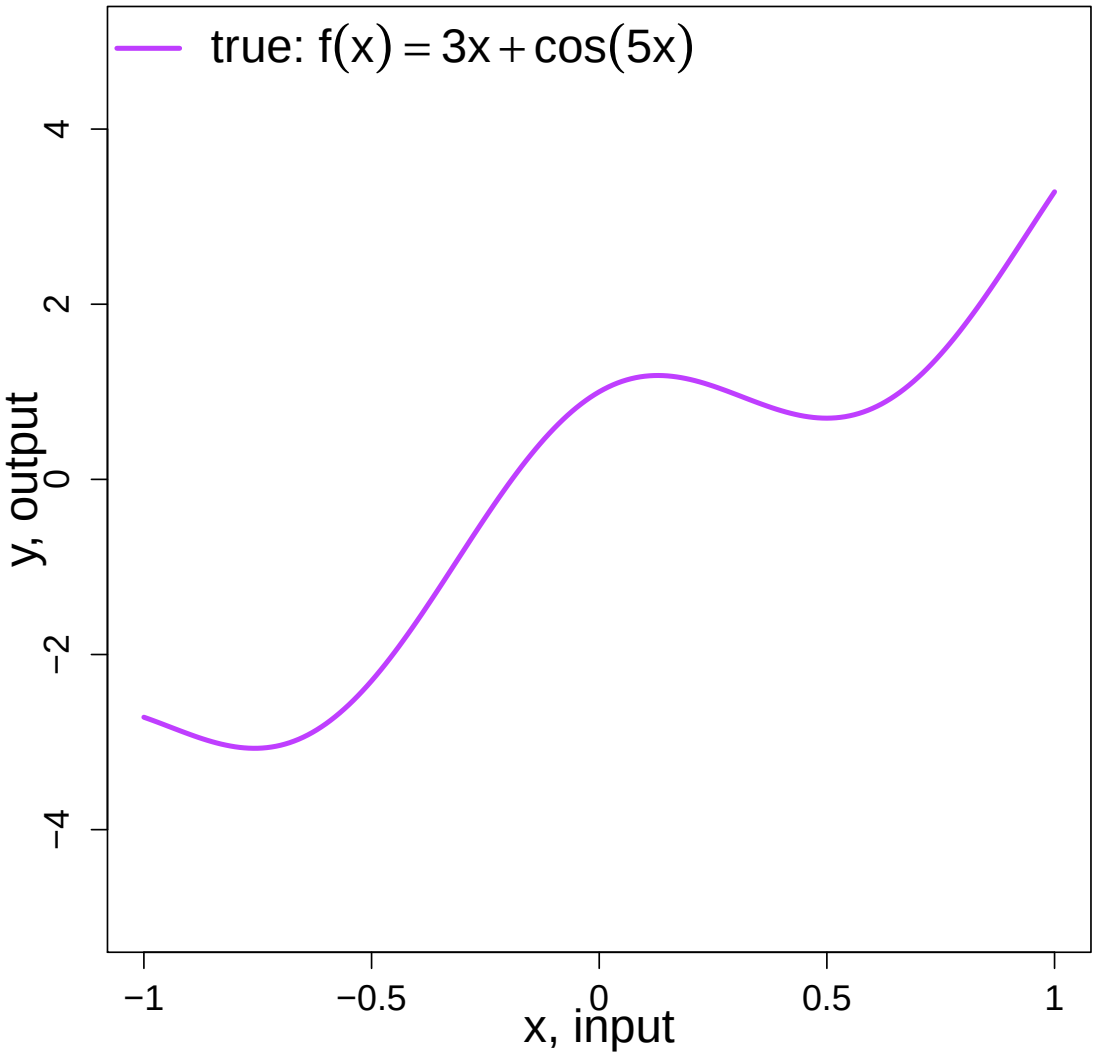
Linked emulator

Simulation

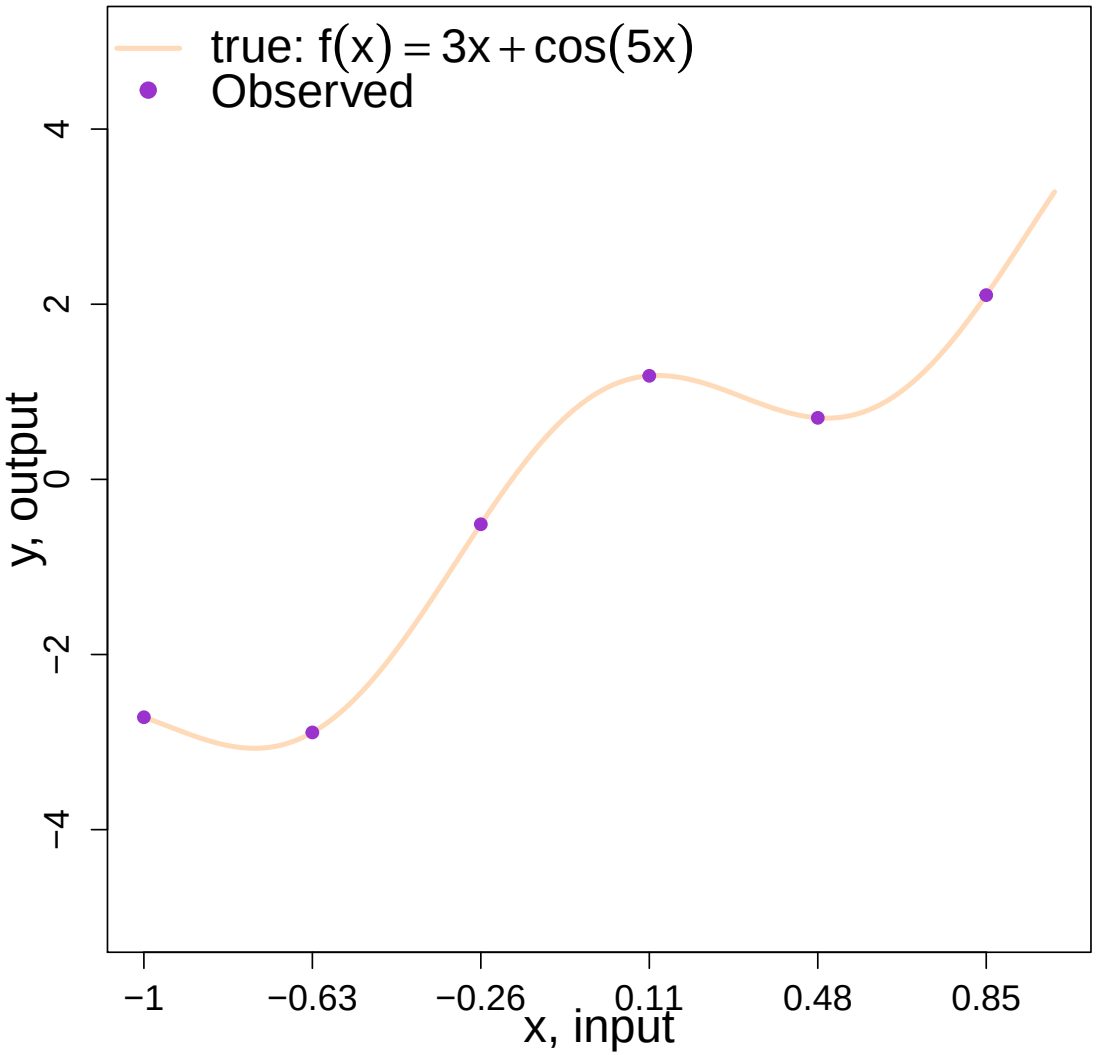
In practice

Conclusions

Example

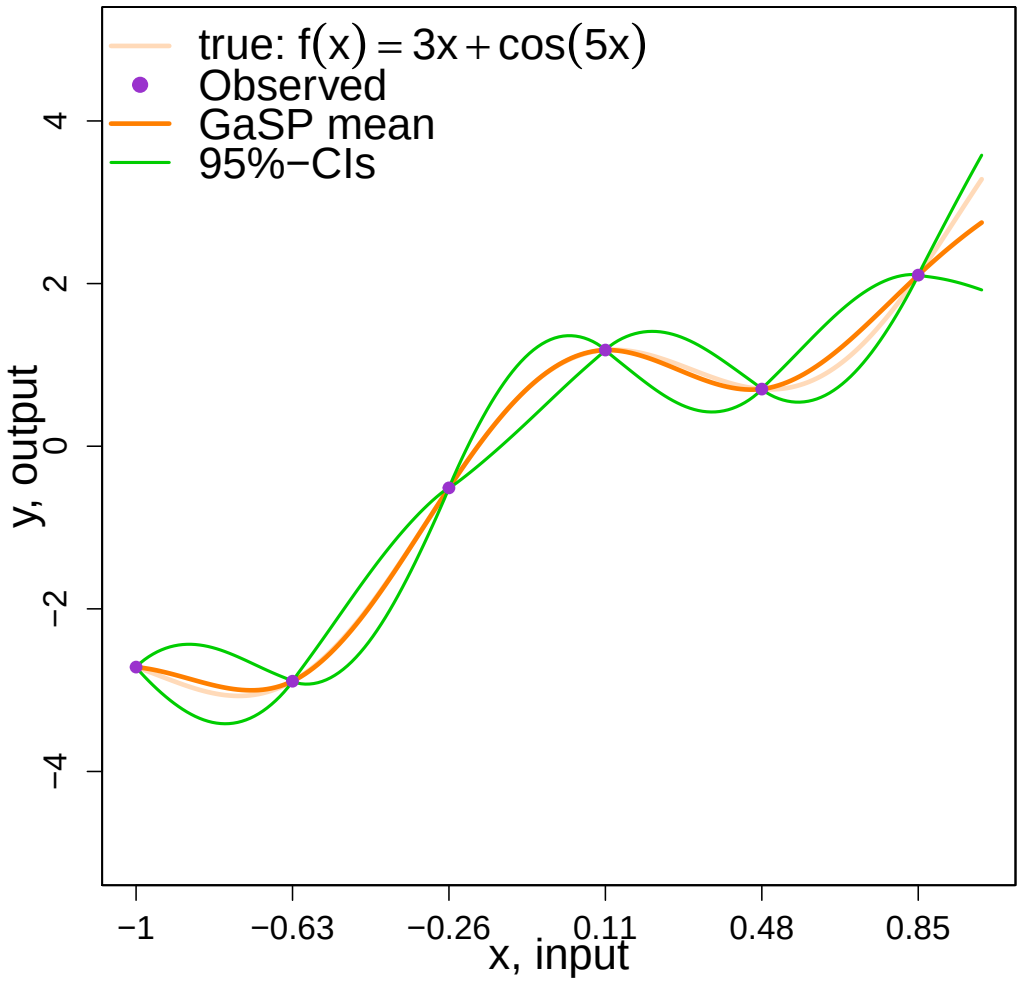


Example

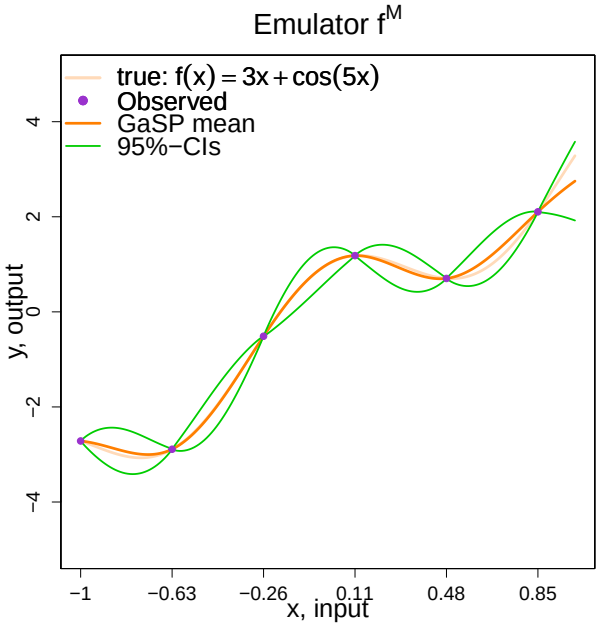

true: f(x) = 3x + cos(5x)
Observed
y, output
4
2
0
−2
−4
−1
−0.63
−0.26
0.11
0.48
0.85
x, input

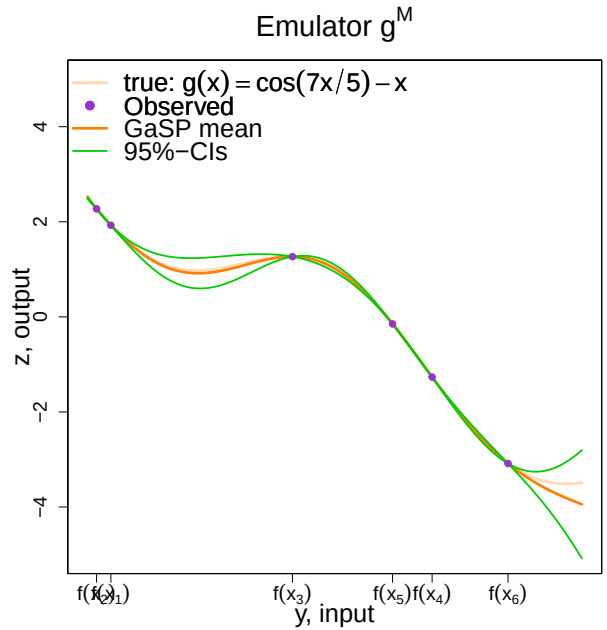
Emulator: GaSP

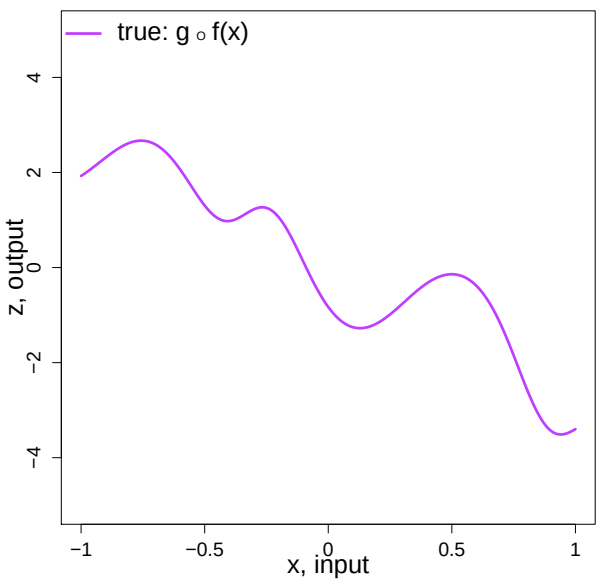
Gaussian Process prior. *Bayarri et al., 2007*.



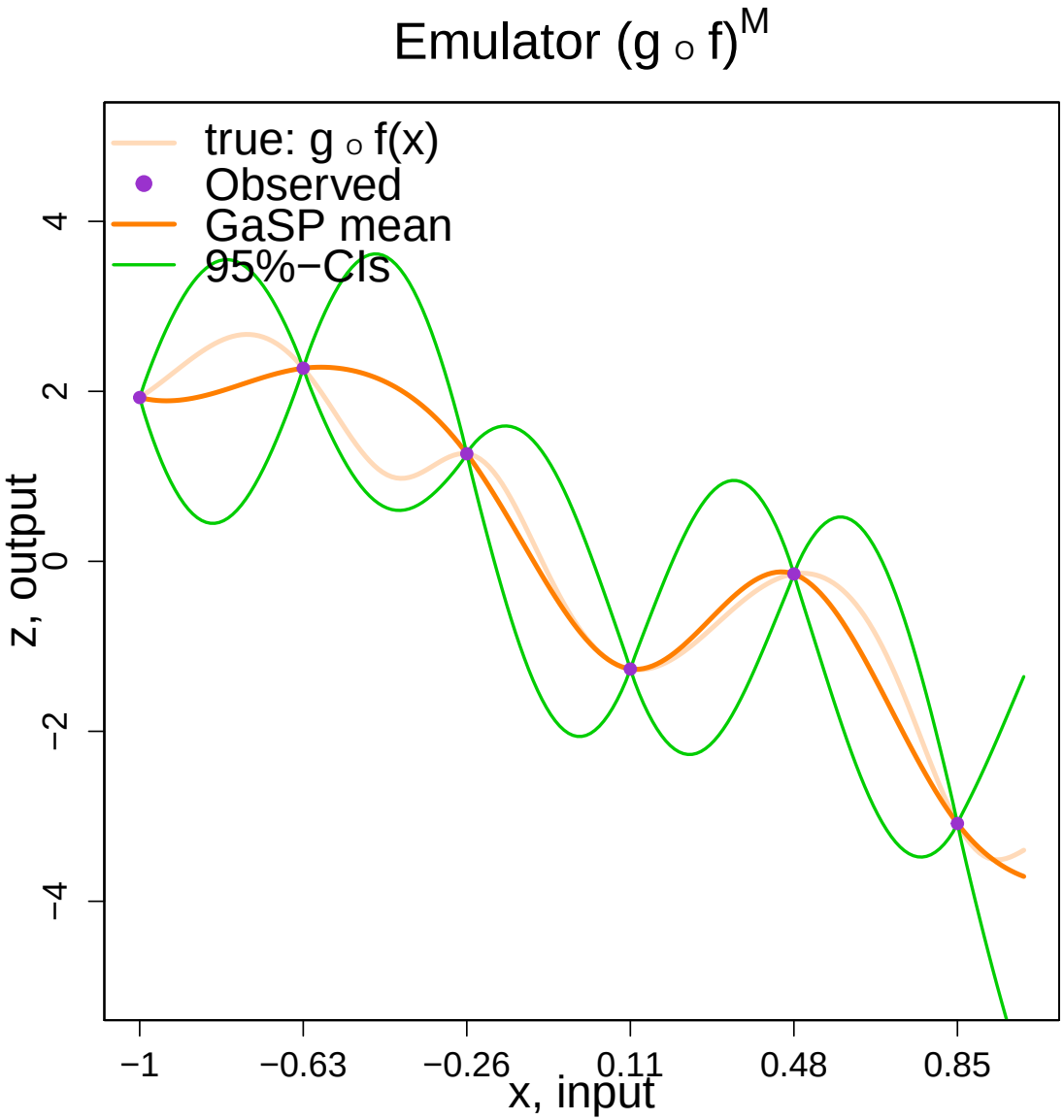
Example (cont.)







Direct emulator



Linked emulator

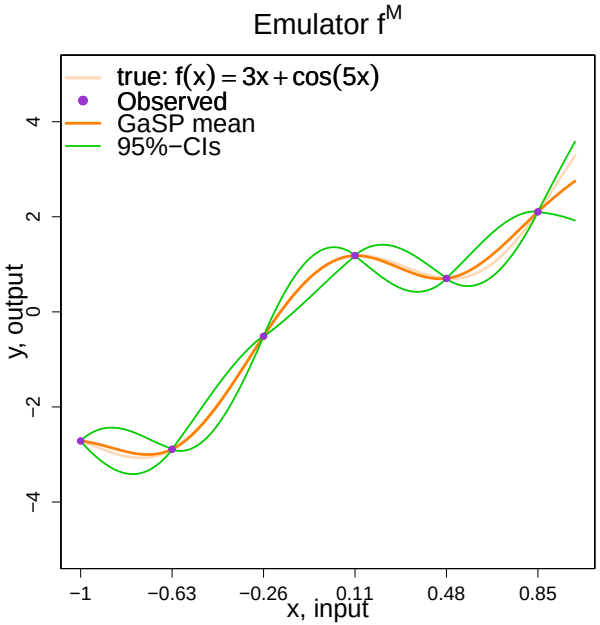
Computer models g and f with emulators g^M and f^M.

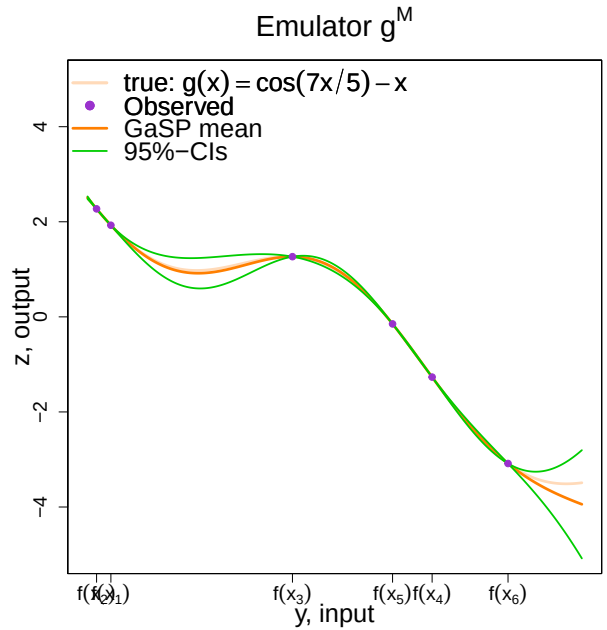
Direct emulator $p((g \circ f)^M(x_{unobs.})|(g \circ f)^M(x_{obs.}))$.

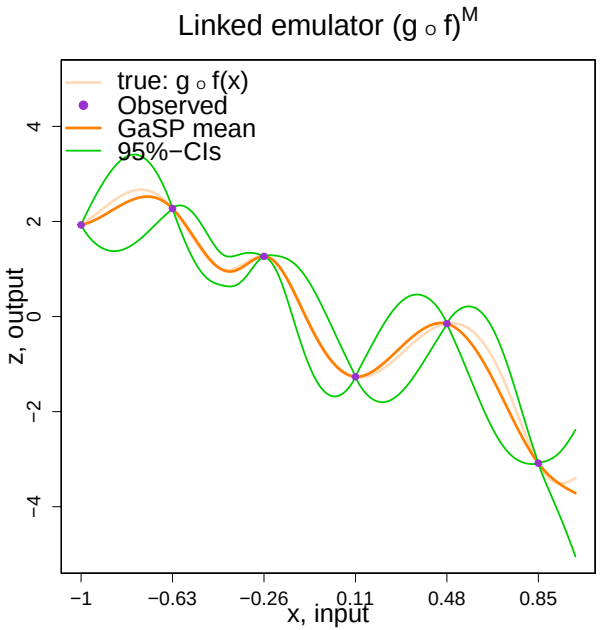
Linked emulator $p((g \circ f)^M(x_{unobs.})|g^M(y_{obs.}), f^M(x_{obs.}))$.

Not closed form.
The mean and variance are.
Normal approximation.

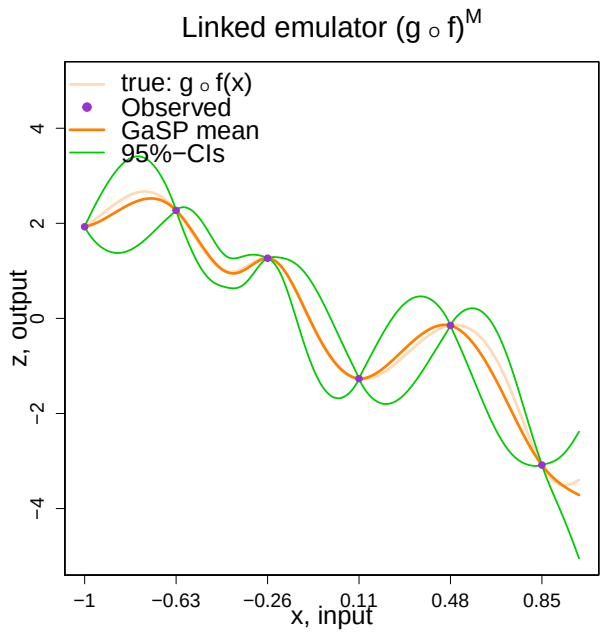
Simulation

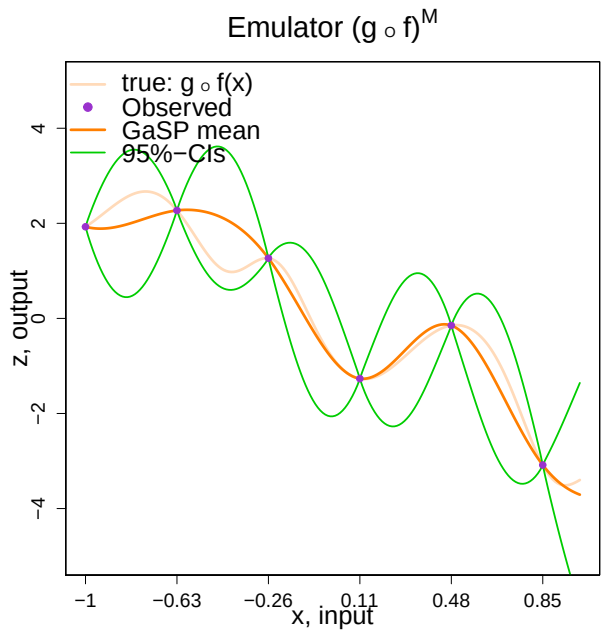

Emulator f^M
true: f(x) = 3x + cos(5x)
Observed
GaSP mean
95%−CIs
y, output
x, input
−1
−0.63
−0.26
0.11
0.48
0.85


Emulator g^M
true: g(x) = cos(7x/5) − x
Observed
GaSP mean
95%−CIs
z, output
y, input
f(x1)
f(x3)
f(x5)
f(x4)
f(x6)


Linked emulator (g ∘ f)^M
true: g ∘ f(x)
Observed
GaSP mean
95%−CIs
z, output
x, input
−1
−0.63
−0.26
0.11
0.48
0.85

Linked vs. direct composite emulators





In practice

Eyjafjallajökull eruption, Iceland, 2010.

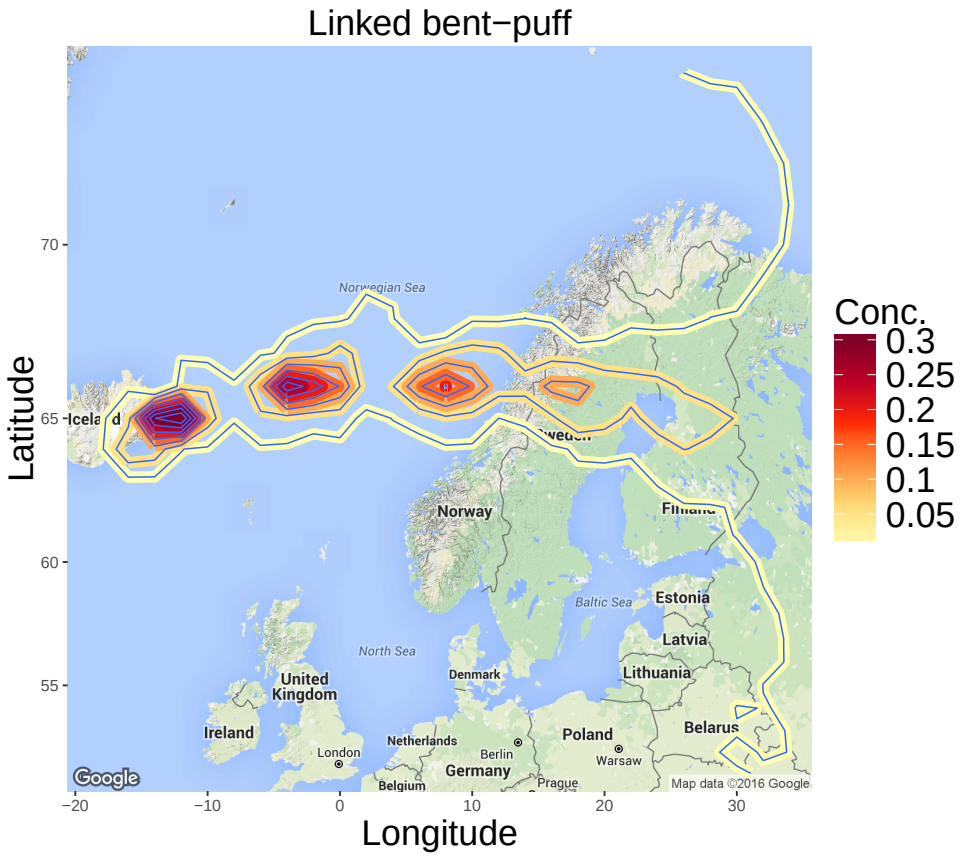
Model f, a **volcano eruption column**, Bent.

- Inputs: vent radius, axial velocity, particle grain size (mean and standard deviation).
- Outputs: characteristics of a volcano eruption column.

Model g, a **volcano ash cloud concentration in time**, Puff.

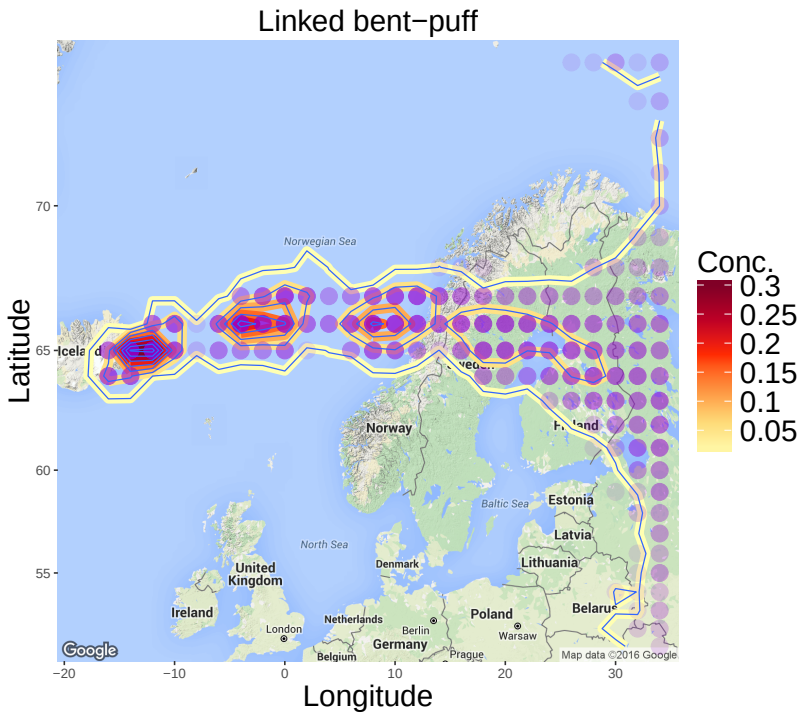
Apply the methodology to emulate $g \circ f(x)$.

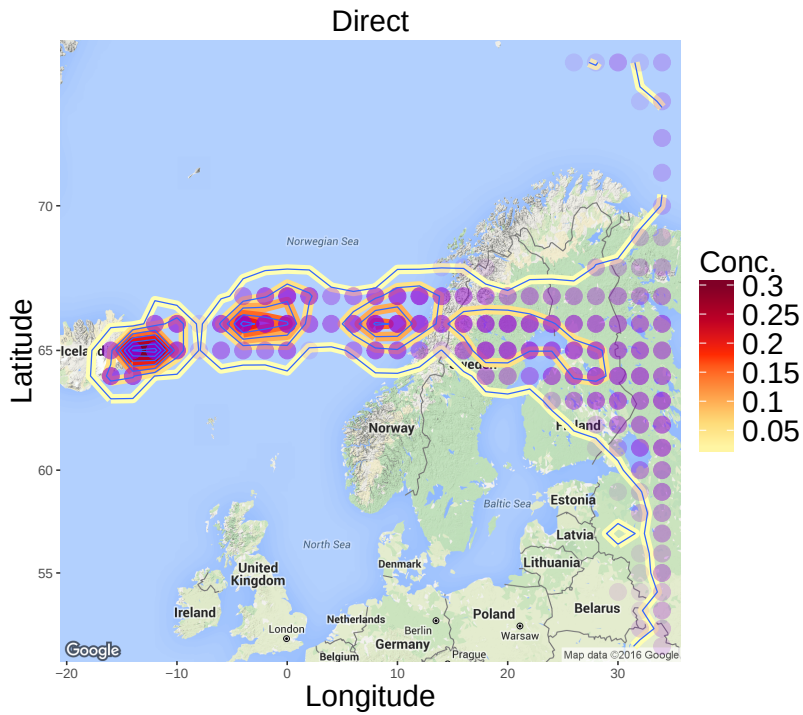
Linked emulator performance



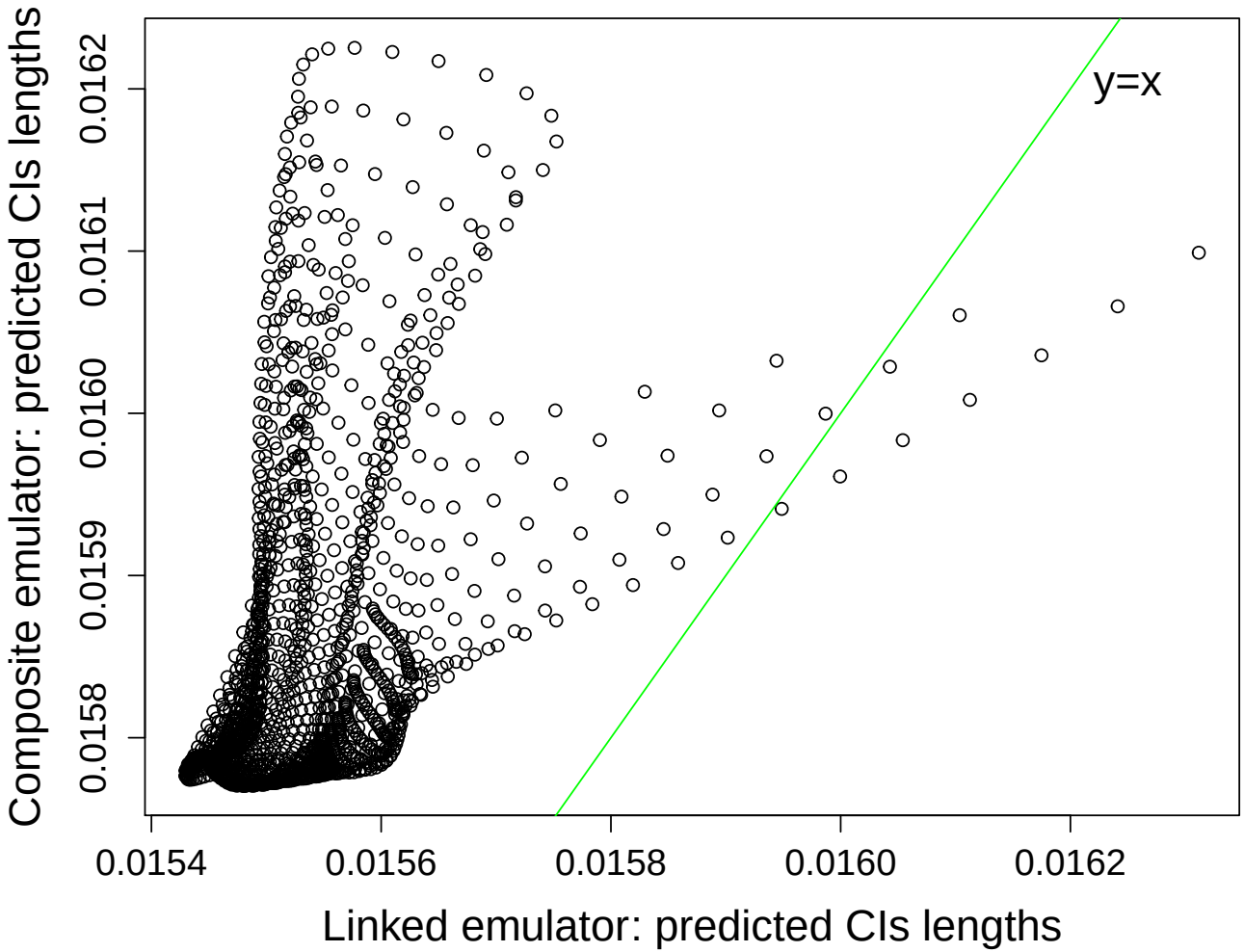
Play movies.

Accuracy: out of sample validation





Epistemic uncertainty



Conclusions

Linking statistical emulators is an alternative to direct coupling of computer models.

Linking allows usage of separately developed emulators of the submodels.

The linked emulator results in a smaller epistemic uncertainty than a direct emulator of the coupled computer model would have (if such a model were available).

The method provides a framework for developing an emulator of a system of emulators of multiple computer models.

Future work. Collaborations

Scalability of the approach.

Data from other applications.

ksenia@stat.duke.edu

Future work. Collaborations

Scalability of the approach.

Data from other applications.

ksenia@stat.duke.edu

Thank you!

Questions.