Gaussian Processes for calibration and exploration of complex computer models

C.Coleman-Smith
cec24@phy.duke.edu

March 2014

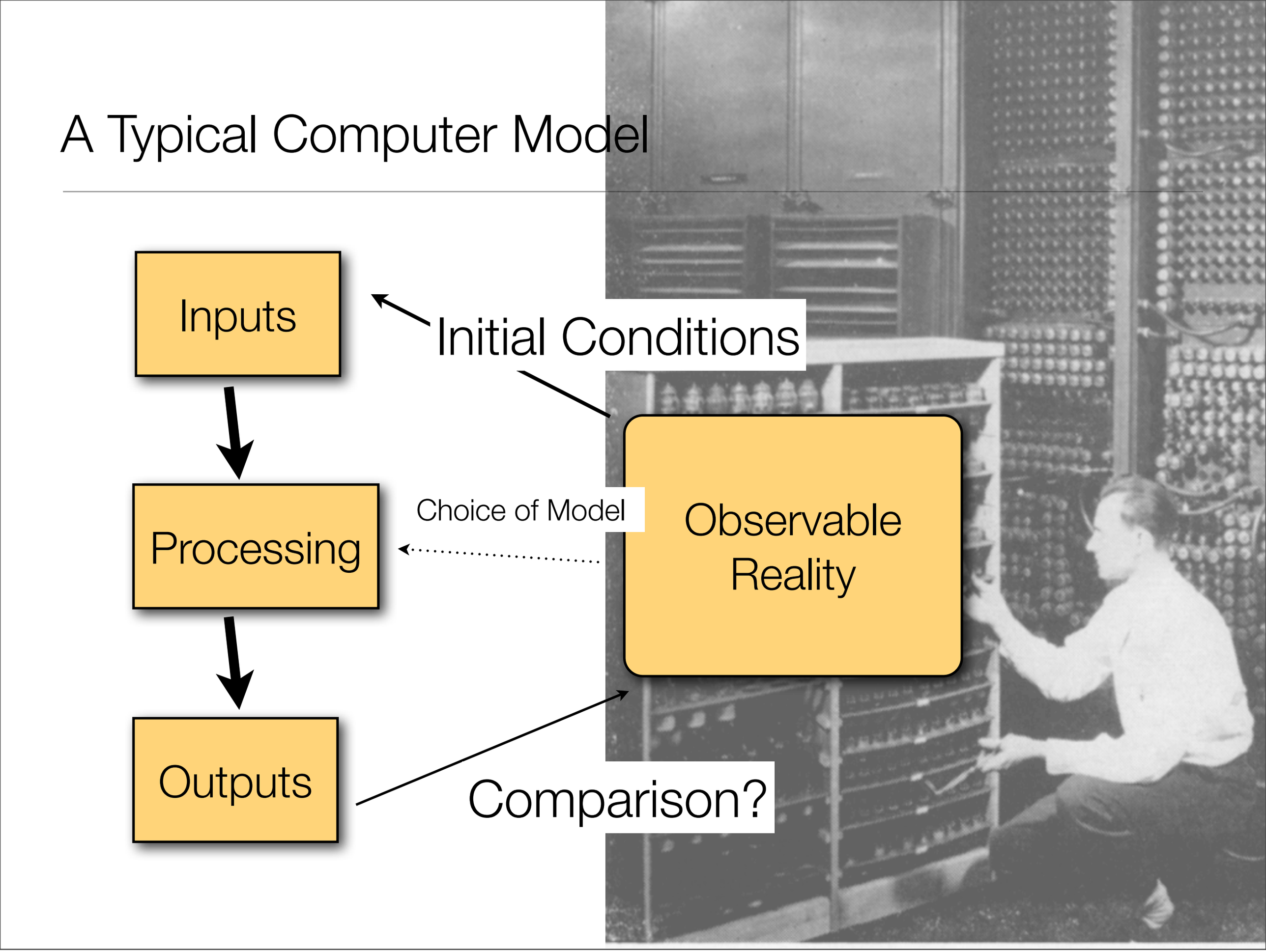
I stand between two worlds, am at home in neither, and in consequence have rather a hard time of it. You artists call me a commoner, and commoners feel tempted to arrest me ...

What Are We Trying to Do

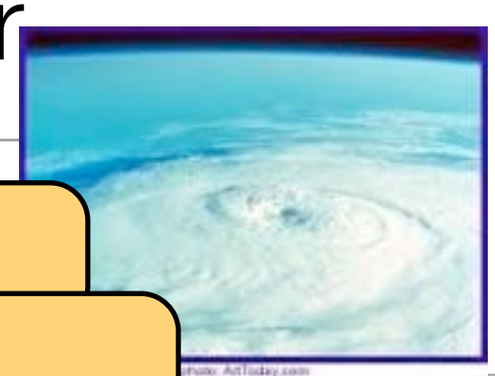
- Advanced science often requires advanced computer models.
- These are expensive, complicated with many inputs and output and may represent only our best guess at the underlying process
- To be useful we need to calibrate these models so that they can reproduce experimental data
- Understand uncertainties and errors in this process and build a systematic model-data comparison

- "How well does our model reproduce reality?"
- "What is the **true** value of a given parameter?"

A Typical Computer Model



A Typical Computer Model - The Weather



Inputs

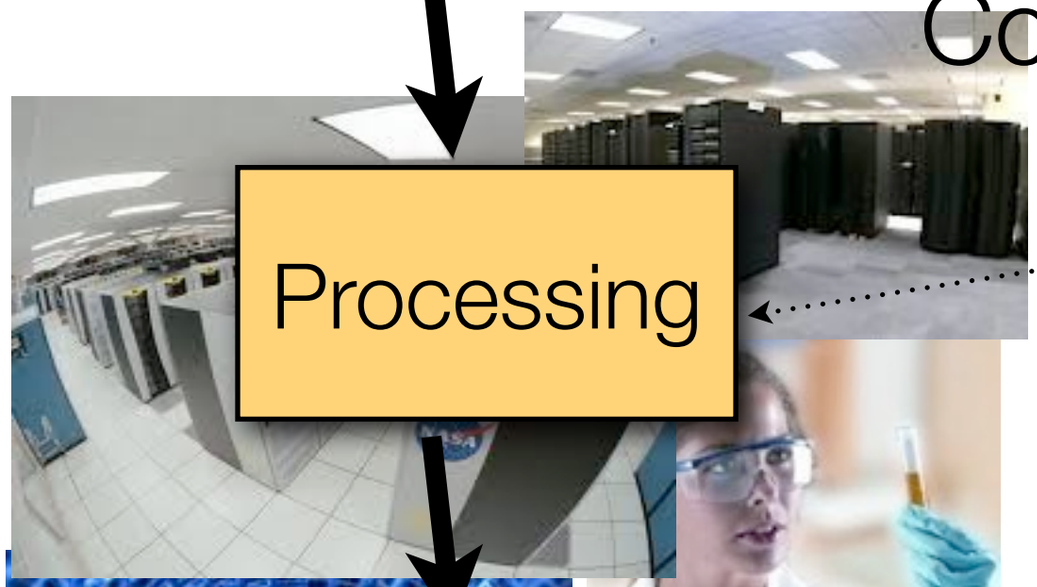
Temperatures

Pressures

Wind Speed

Geography

Initial Conditions



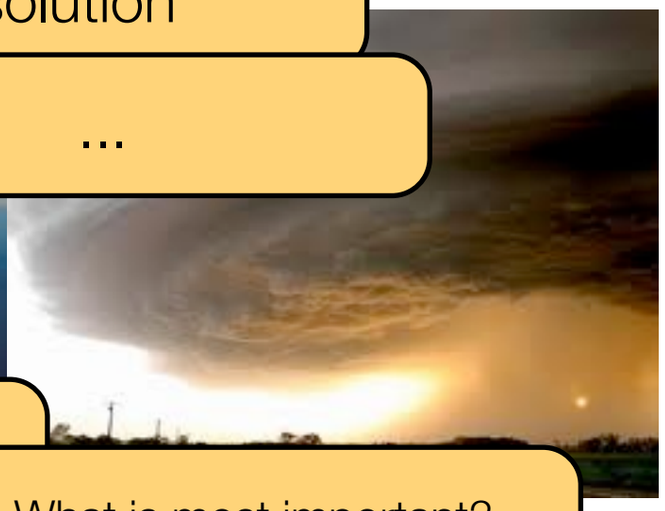
Processing

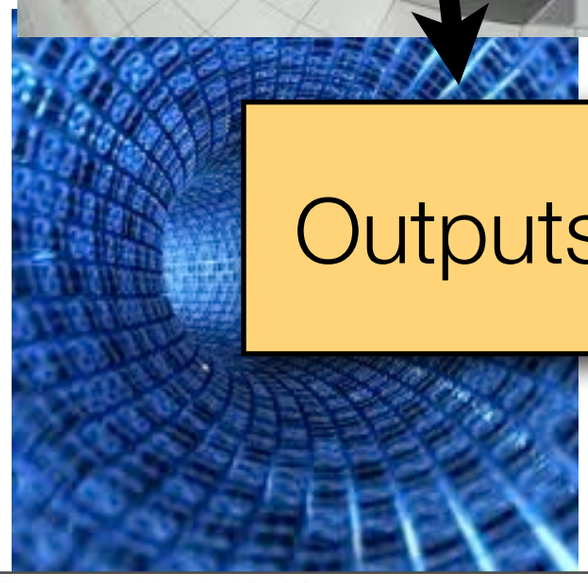
Turbulence?

Resolution

...

Choice of Model





Outputs

Did it work?

What is most important?

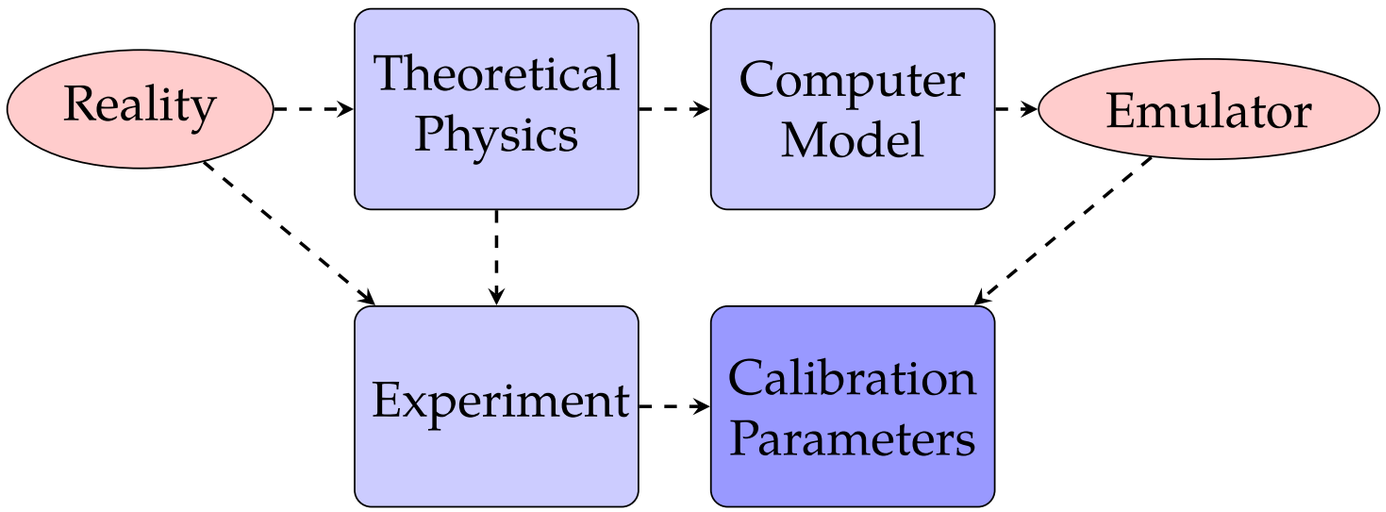
Where?

What didn't work?

Comparison?

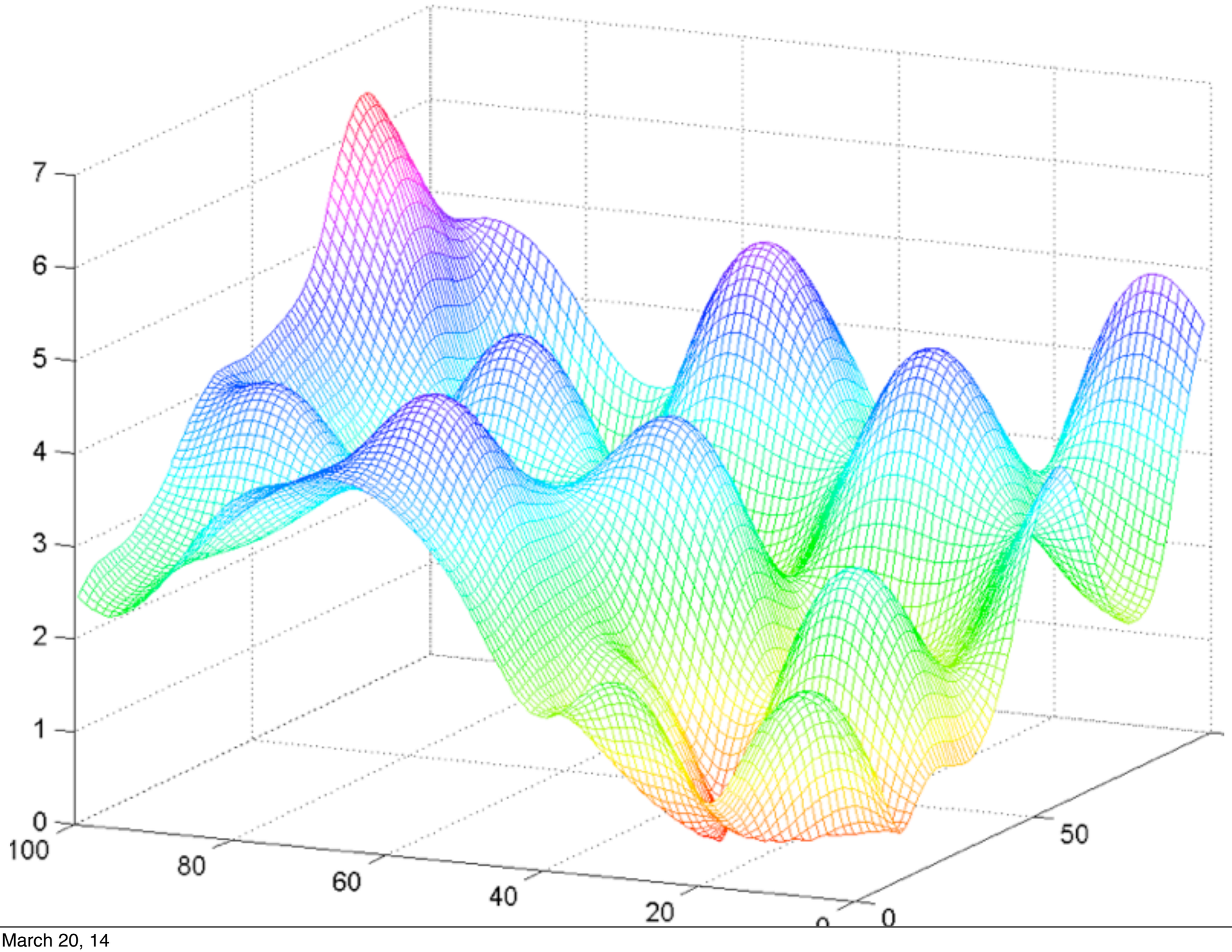
We are parameter space explorers!

- Validation and verification, calibration and sensitivity analysis all depend on knowing the model output throughout the parameter space.
- Fundamental tension: rigor <-> results
- Want to minimize the computational cost and maximizes the amount of useful information gained



- We can address the run-time issue by generating a statistical model of the code, an emulator

Exploration not Minimization



What do I mean by a computer model?

$$Y_f(x, u_\star) = Y_r(x, u_\star) + \epsilon_f(x)$$

$$Y_r(x, u_\star) = Y_m(x, u_\star) + b(x, u_\star),$$

What do I mean by a computer model?

Field Observations

Field Observations Errors

$$Y_f(x, u_\star) = Y_r(x, u_\star) + \epsilon_f(x)$$

$$Y_r(x, u_\star) = Y_m(x, u_\star) + b(x, u_\star),$$

What do I mean by a computer model?

Field Observations

'Reality'

Field Observations Errors

$$Y_f(x, u_\star) = Y_r(x, u_\star) + \epsilon_f(x)$$

$$Y_r(x, u_\star) = Y_m(x, u_\star) + b(x, u_\star),$$

What do I mean by a computer model?

Field Observations

'Reality'

Field Observations Errors

$$Y_f(x, u_\star) = Y_r(x, u_\star) + \epsilon_f(x)$$

$$Y_r(x, u_\star) = Y_m(x, u_\star) + b(x, u_\star),$$

Model Output

$Y_m(x, u) : \mathbb{R}^{p_x} \times \mathbb{R}^{p_u} \to \mathbb{R}^k$

'Bias'

scary

What do I mean by a computer model?

Field Observations

'Reality'

Field Observations Errors

$$Y_f(x, u_\star) = Y_r(x, u_\star) + \epsilon_f(x)$$

$$Y_r(x, u_\star) = Y_m(x, u_\star) + b(x, u_\star),$$

Observation Parameters

Experimentally Well Controlled

Calibration Parameters

Computationally Well Controlled

Model Output

$Y_m(x, u) : \mathbb{R}^{p_x} \times \mathbb{R}^{p_u} \to \mathbb{R}^k$

'Bias'

scary

What do I mean by a computer model?

Field Observations

'Reality'

Field Observations Errors

$$Y_f(x, u_\star) = Y_r(x, u_\star) + \epsilon_f(x)$$

$$Y_r(x, u_\star) = Y_m(x, u_\star) + b(x, u_\star),$$

Model Output

$Y_m(x, u) : \mathbb{R}^{p_x} \times \mathbb{R}^{p_u} \to \mathbb{R}^k$

'Bias'

scary

Observation Parameters

Experimentally Well Controlled

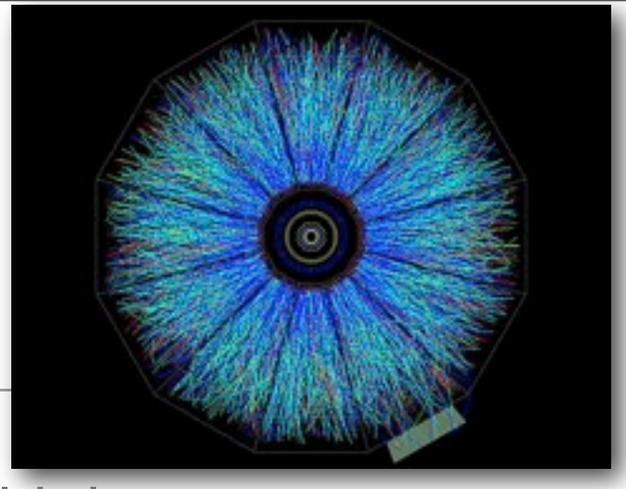
Calibration Parameters

Computationally Well Controlled

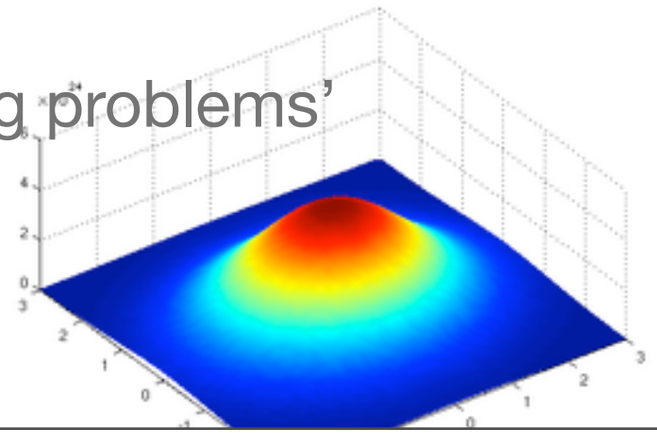
$u_\star$

True value, unknown, may **not** be best fitting value

Simple Vs Complex Codes



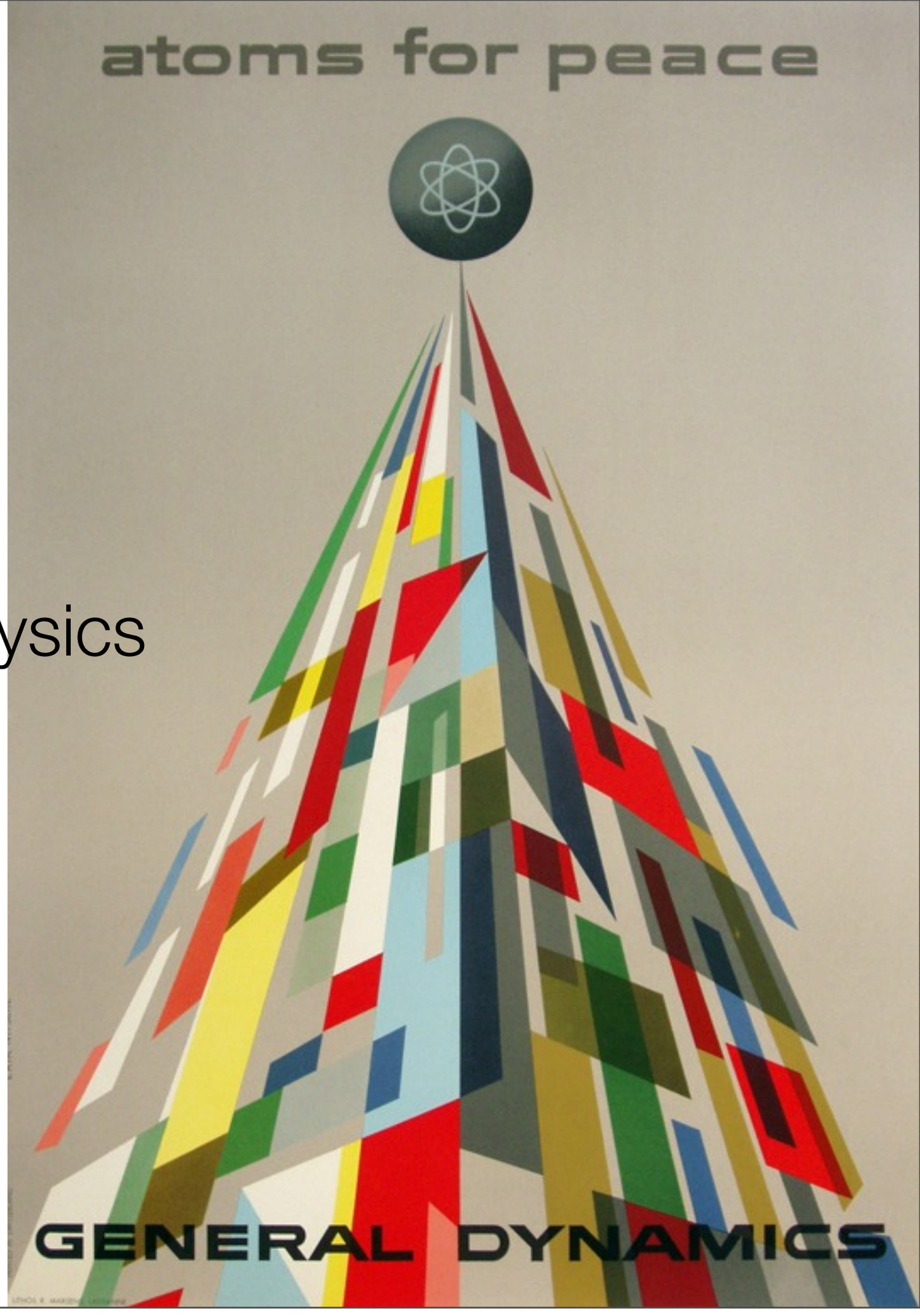
- Easy to make observations of the physical process
- Solving a set of simple equations, initial conditions well known
- Deterministic
- always **certain** we are solving the correct problem
- eg: heat equation, diffusion etc.
- ‘Engineering problems’



- Difficult to make multiple observations, expensive
- Complicated set of strongly interacting equations. Initial conditions may be uncertain.
- may be somewhat Stochastic
- may not be the correct formal description of the phenomenon!
- eg: galaxy formation, heavy ion collisions, climate
- ‘Research problems’

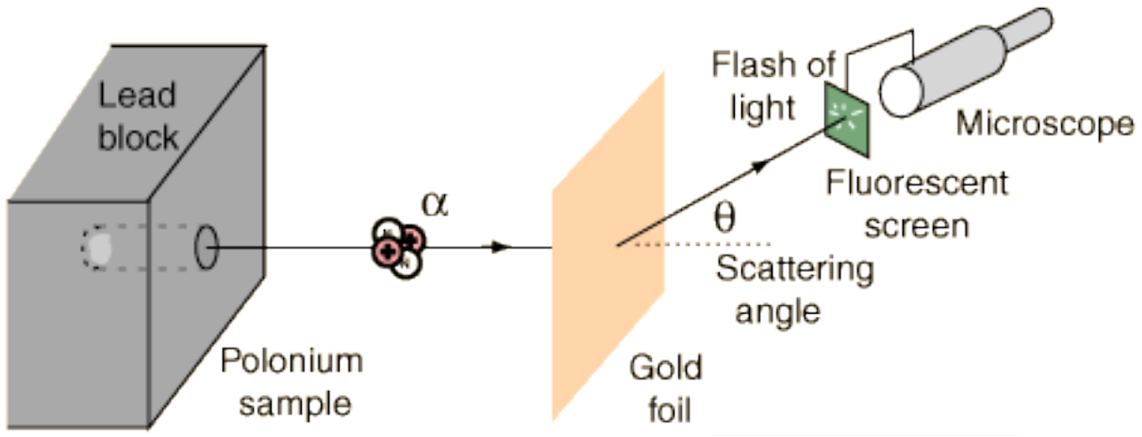
High Energy Nuclear Physics

With H.Petersen (FIAS), S.Pratt (MSU)

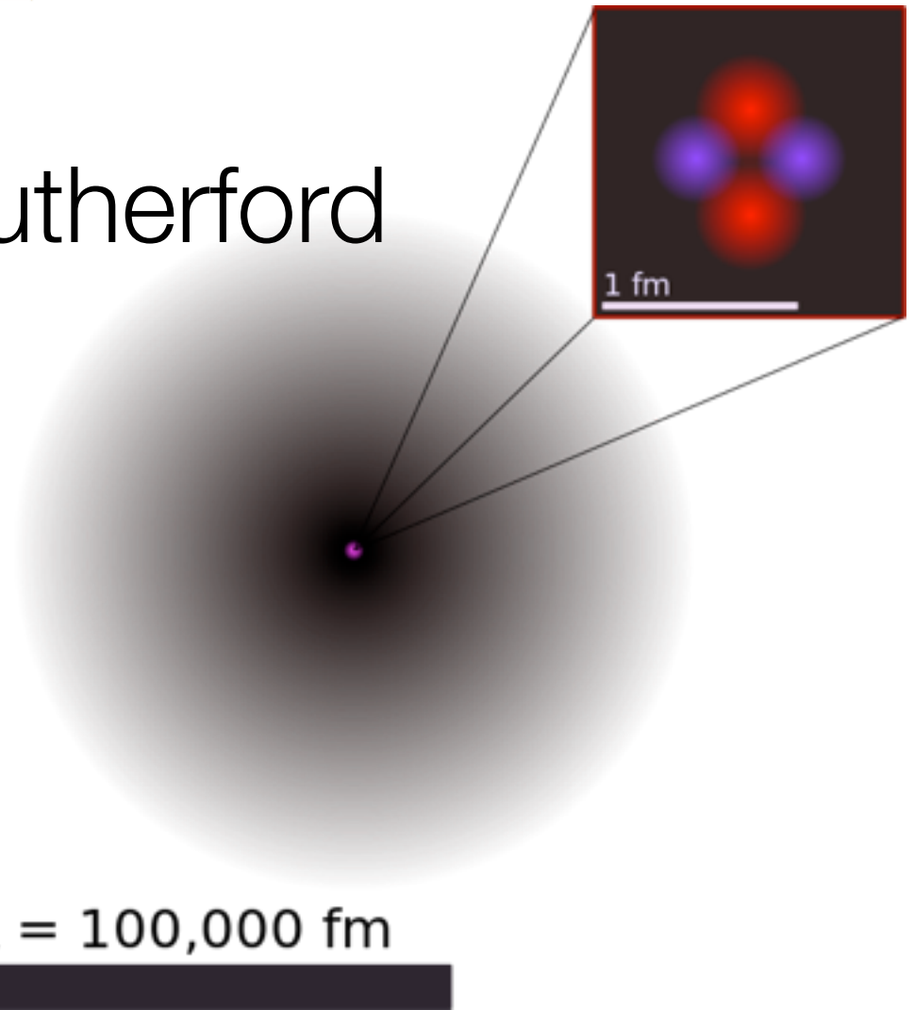


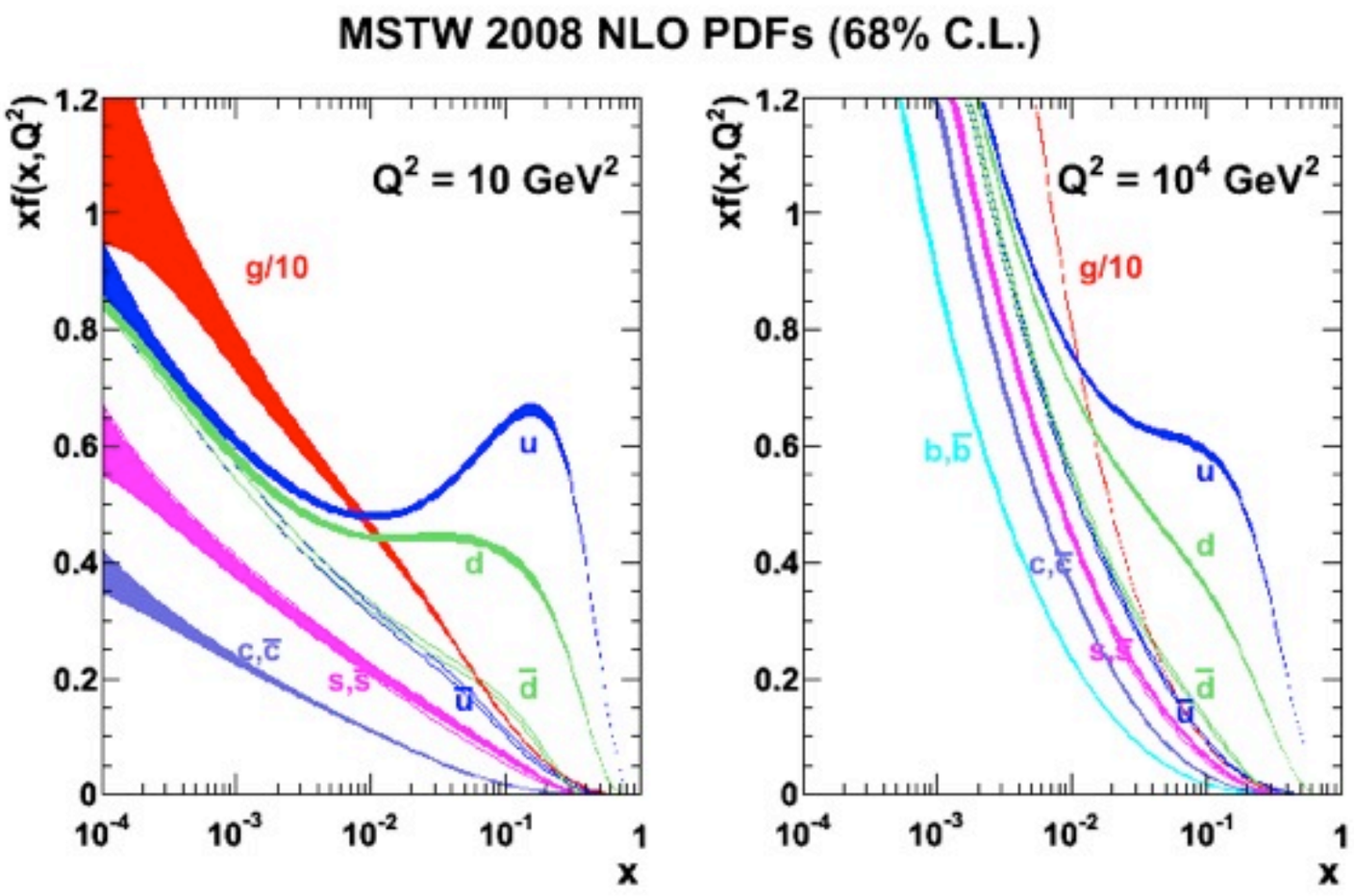
Nucleii Are Interesting

Nucleon structure is non trivial

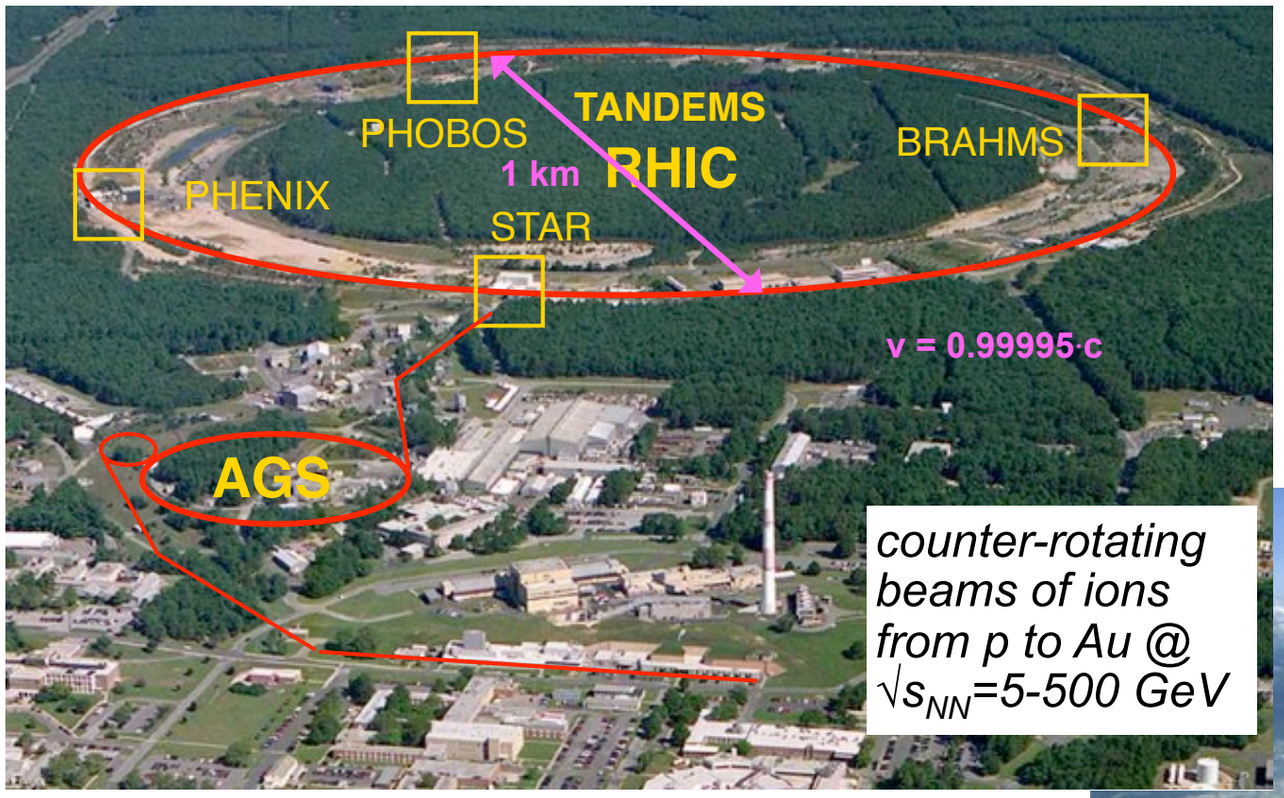


Rutherford



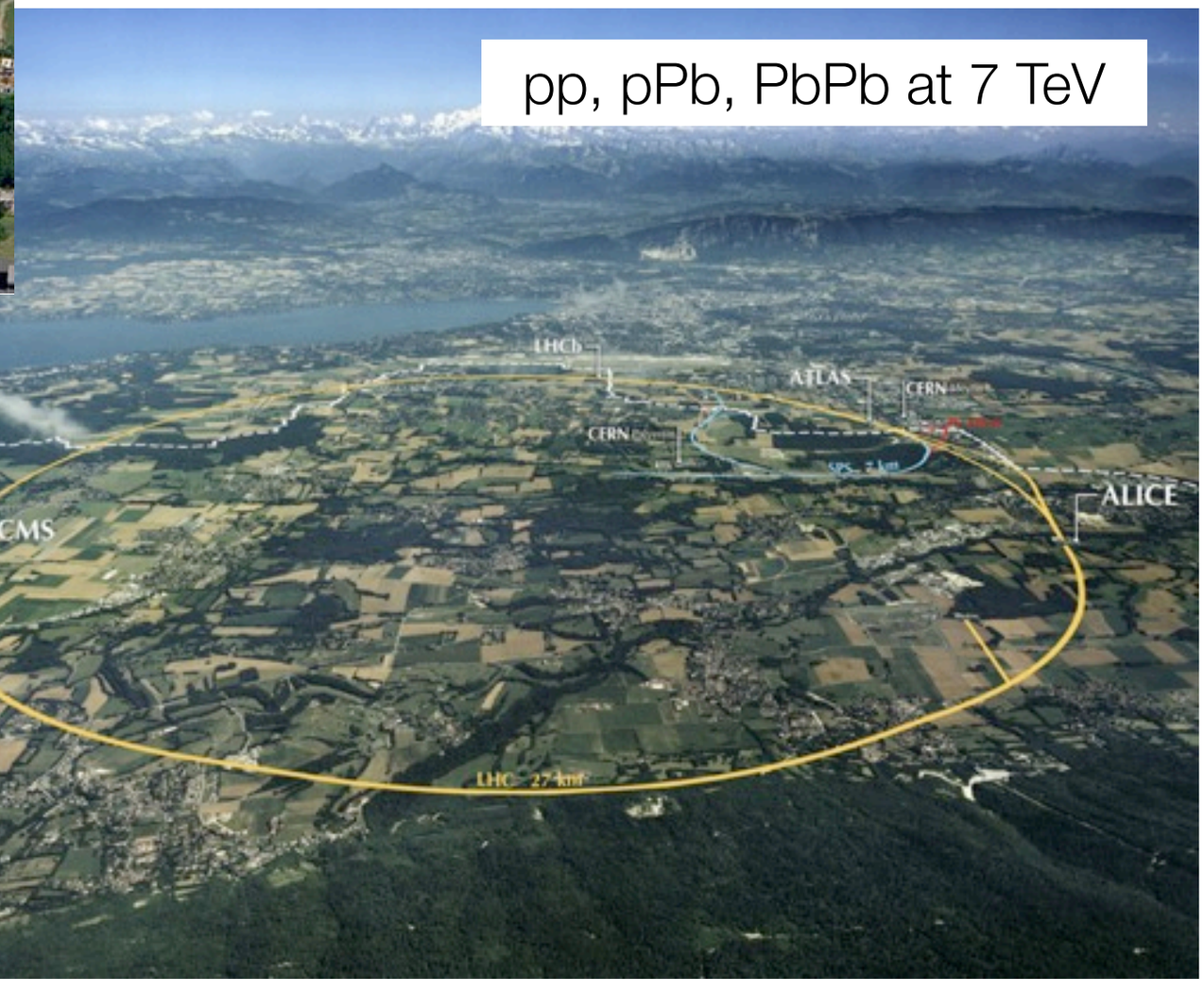


Let's Collide Nucleii At High Energies



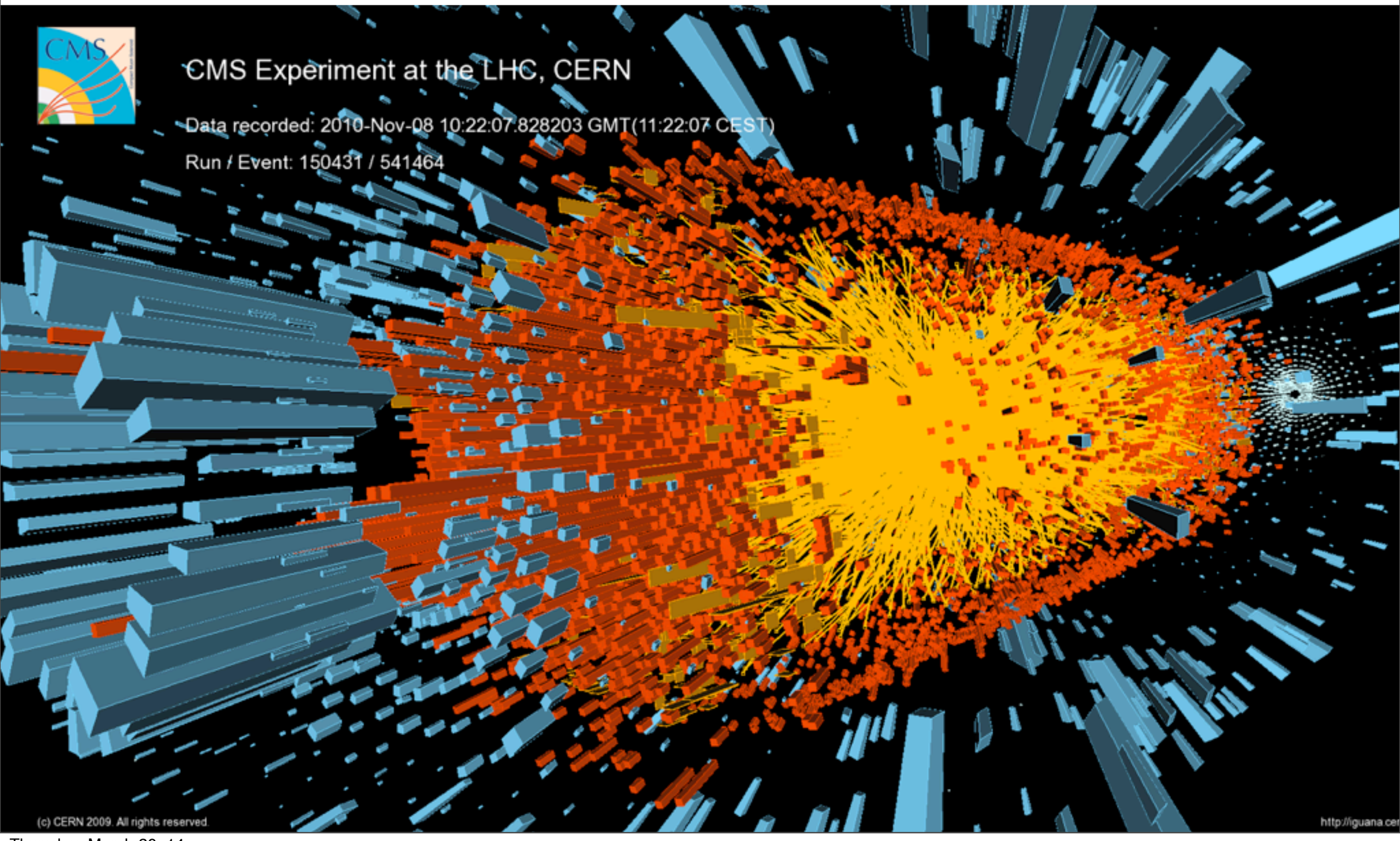
counter-rotating beams of ions from p to Au @ $\sqrt{s_{NN}}$=5-500 GeV

RHIC:
Brookhaven
National Lab NY



LHC: CERN

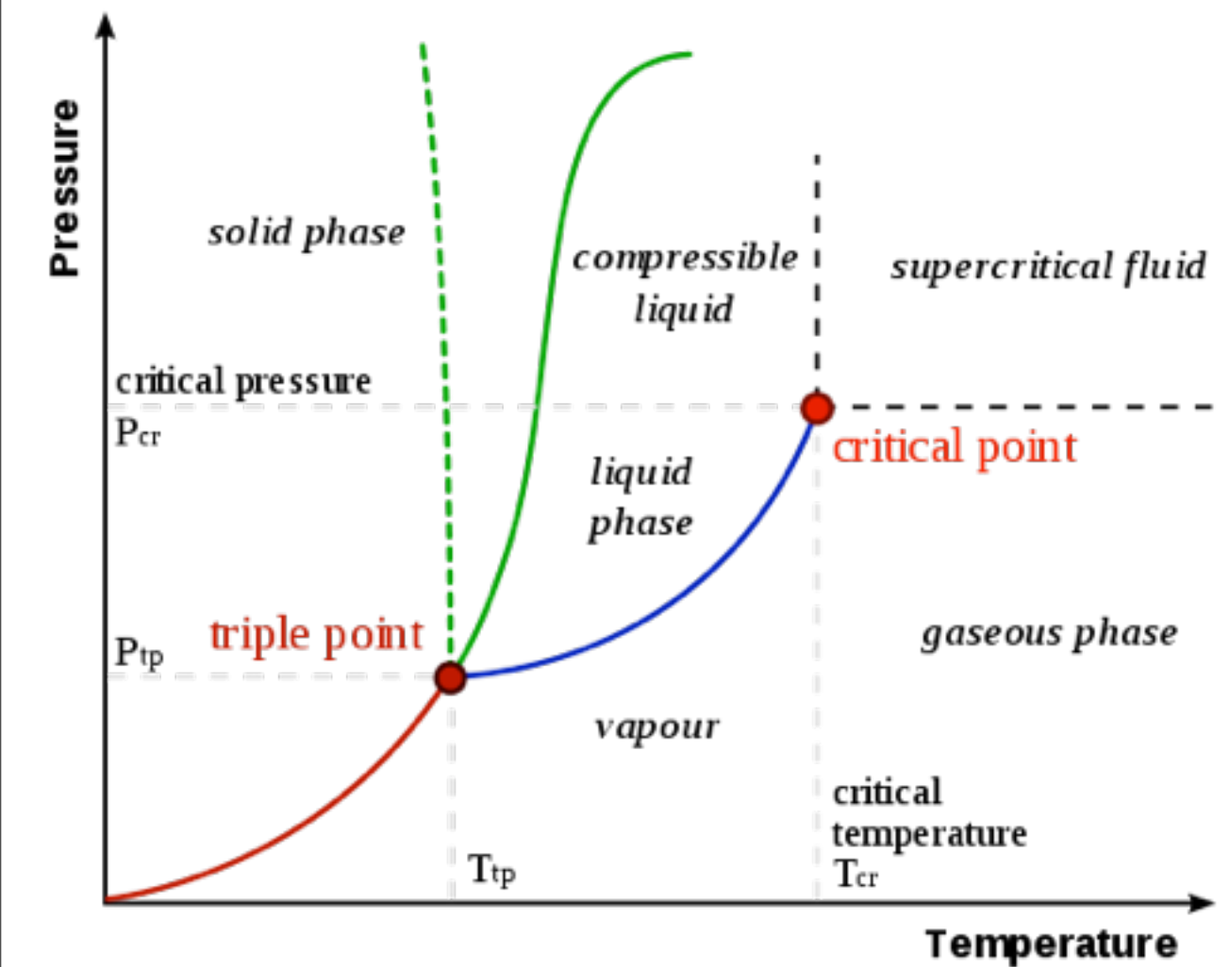
It goes bang


CMS
CMS Experiment at the LHC, CERN
Data recorded: 2010-Nov-08 10:22:07.828203 GMT(11:22:07 CEST)
Run / Event: 150431 / 541464
(c) CERN 2009. All rights reserved.

Matter Produced is a Quark Gluon Plasma

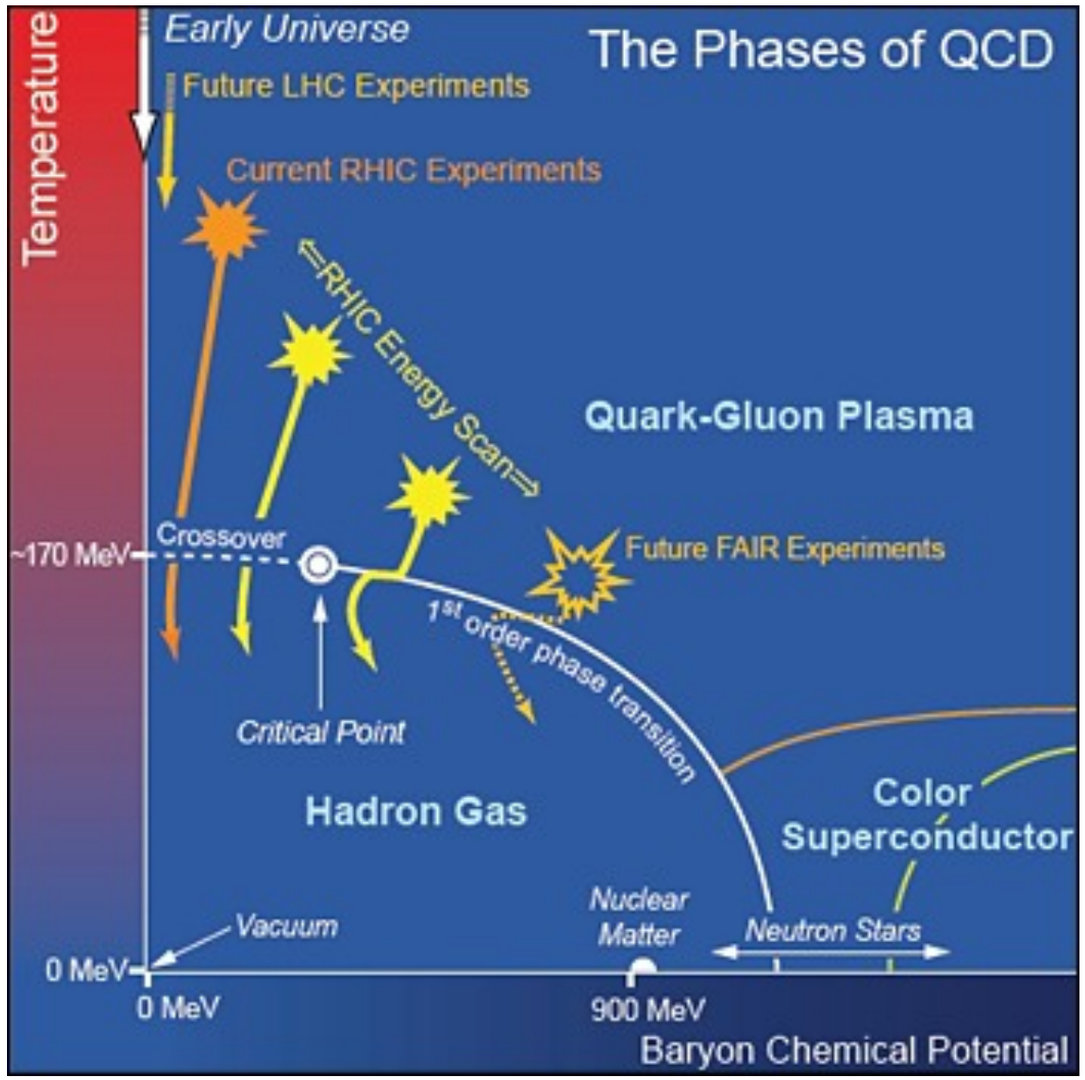
Ordinary Matter:

- phases determined by (electro-magnetic) interaction
- apply heat & pressure to study phase-diagram



Phases of QCD matter:

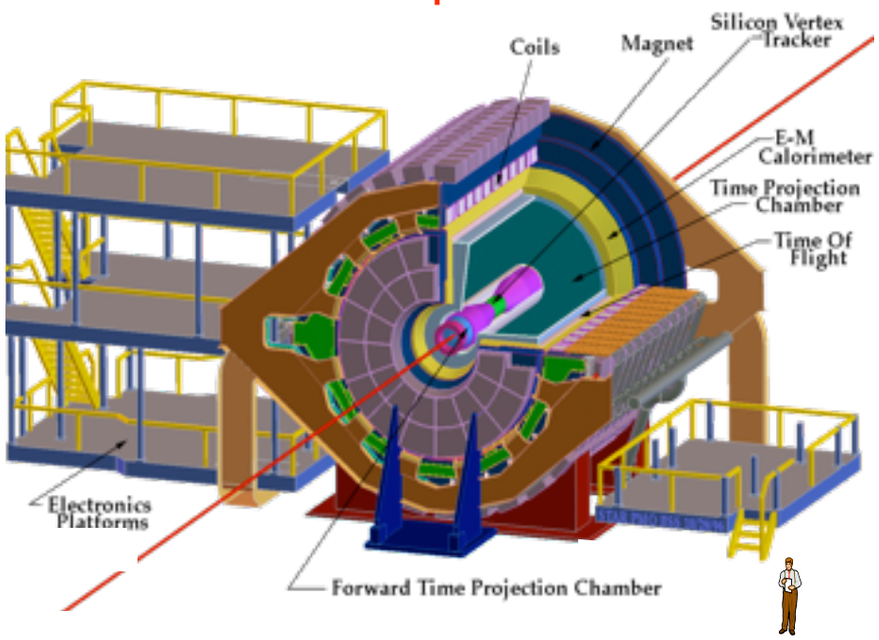
- heat & compress QCD matter:
  - collide heavy atomic nuclei
- numerical simulations:



Record This With Detectors

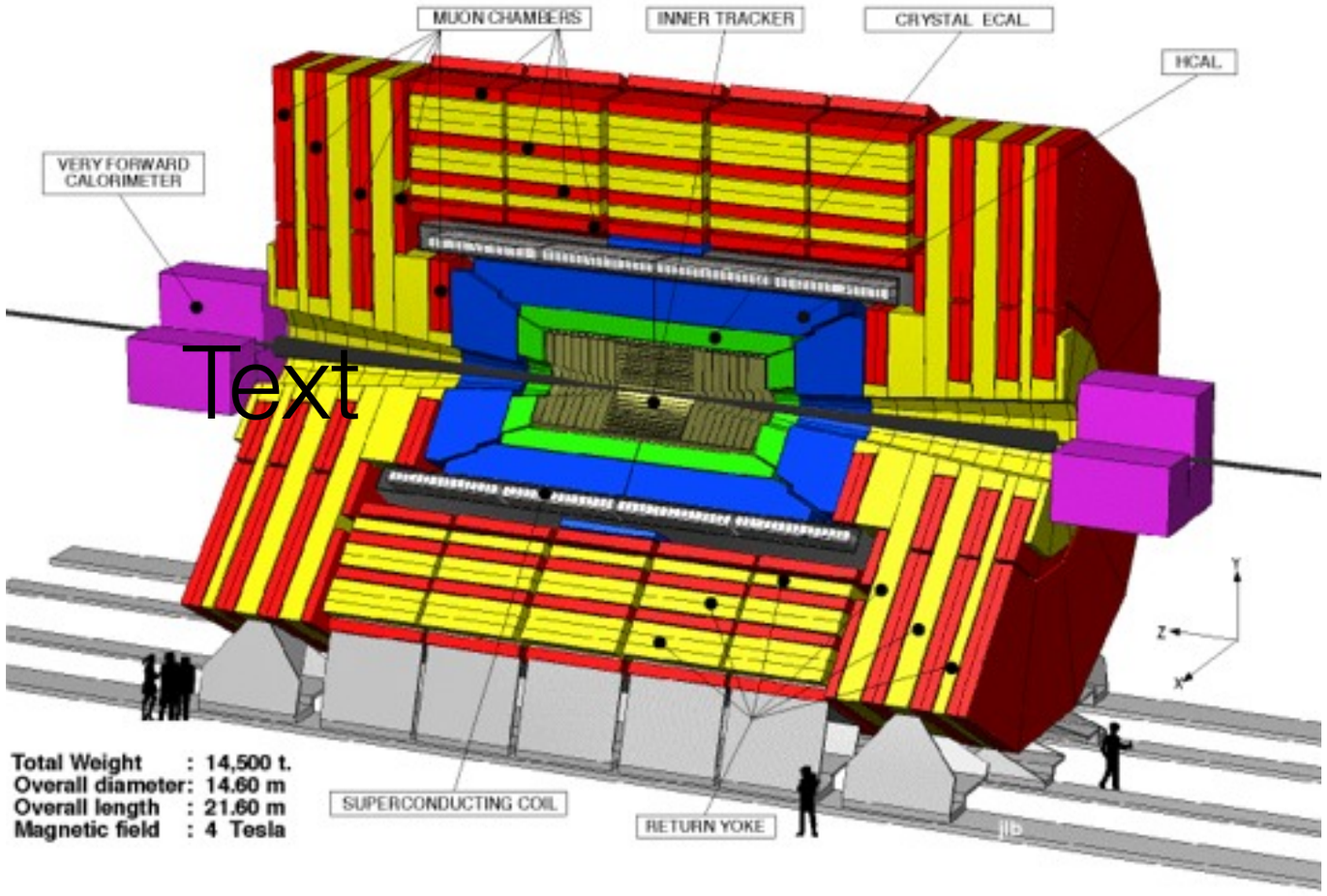
STAR

Solenoidal field
Large-Ω Tracking
TPC's, Si-Vertex Tracking
RICH, EM Cal, TOF
~420 Participants



- **Measurements of Hadronic Observables using a Large Acceptance**
- **Event-by-Event Analyses of Hadrons and Jets, Forward physics, Leptons, Photons**

CMS - LHC



some of the most complicated machines ever built

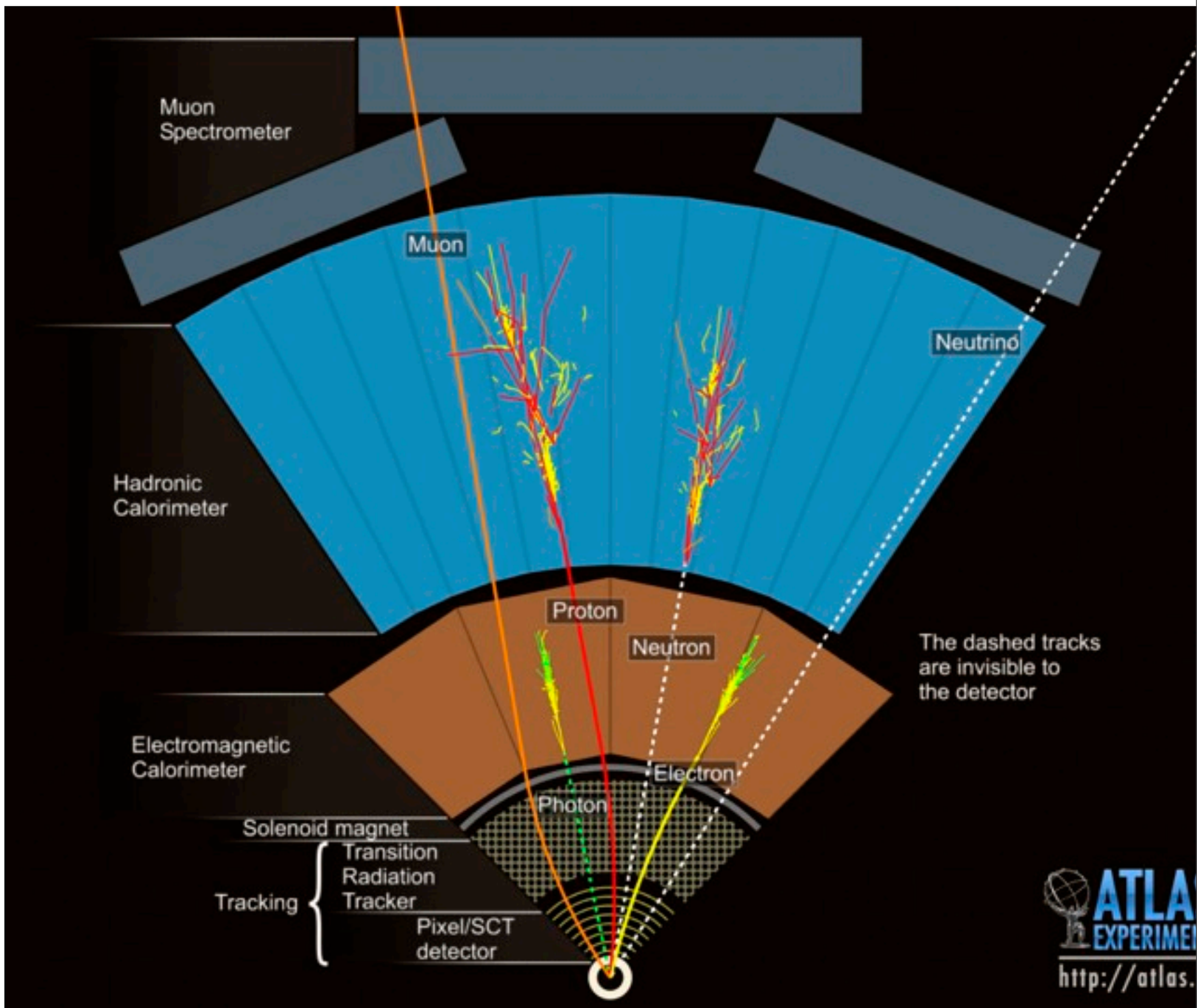
What do detectors detect? Particles

Raw Data Recorded

- Counts
- Energy measurements
- Hit positions & times
- Tracks

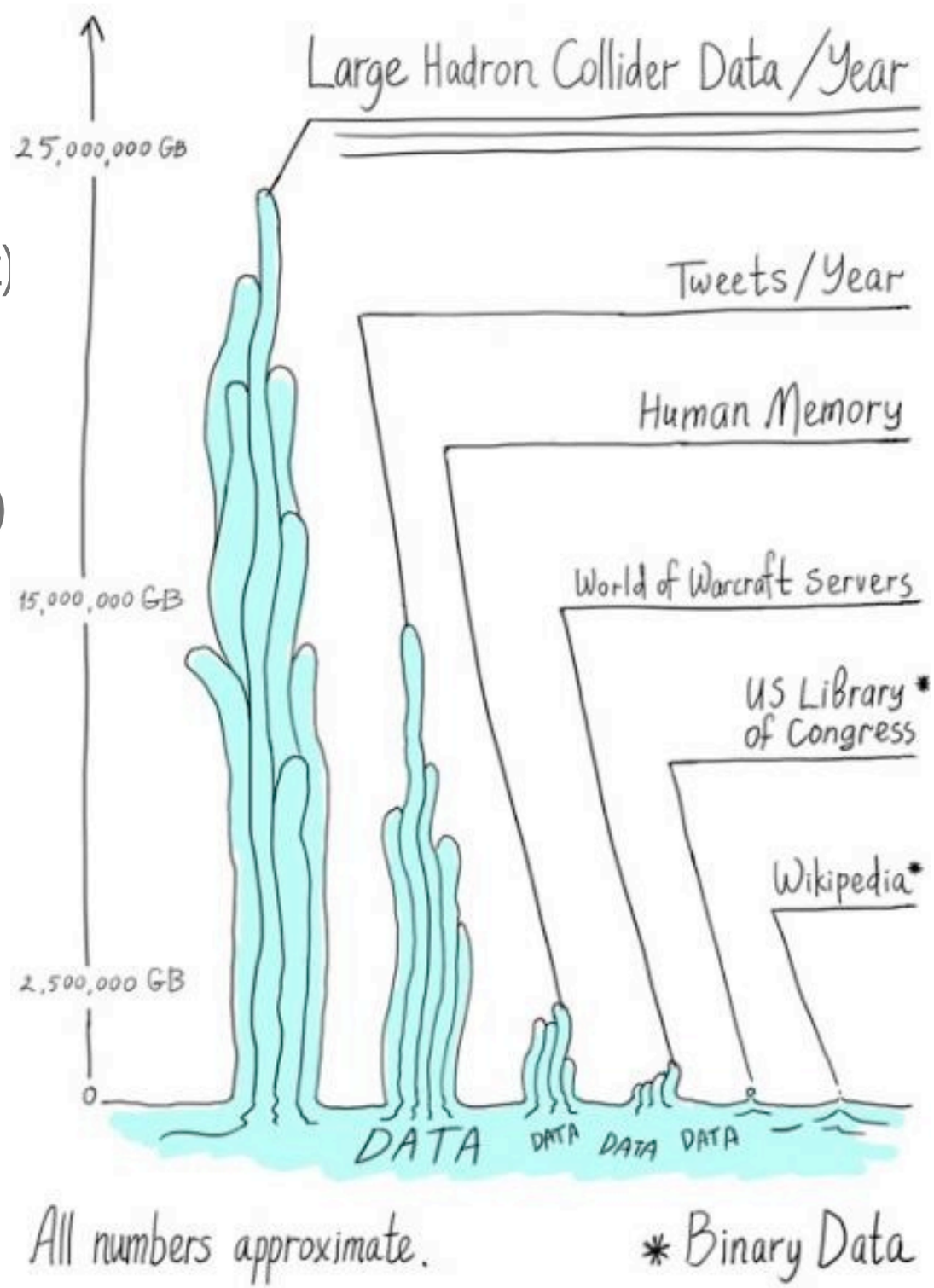
Reconstructed

- Particle Charges
- Particle Identities
- Particle Trajectories
- Particle Momenta

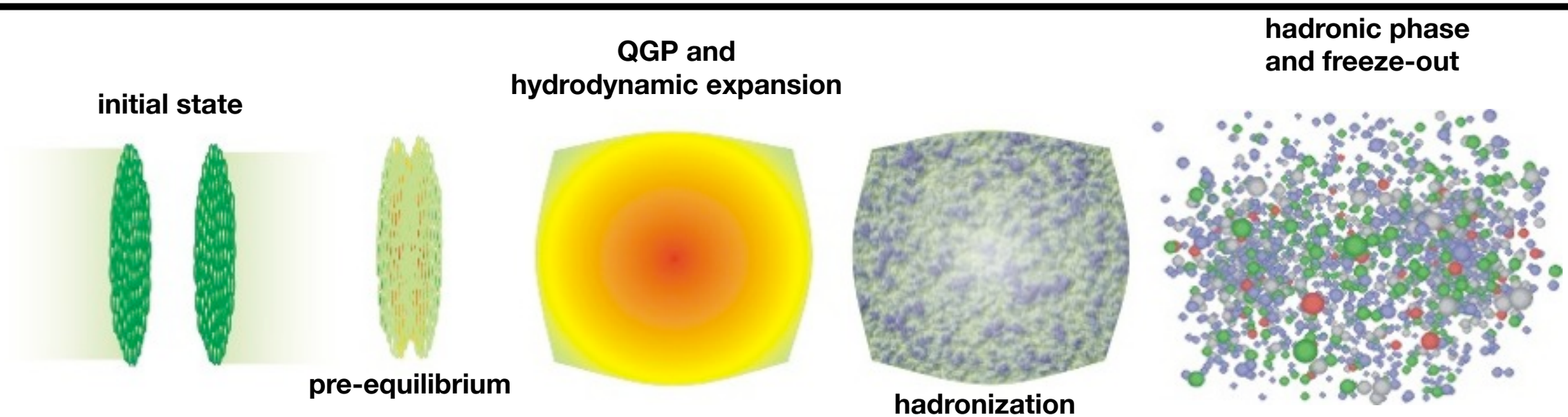


It's Big Data

- LHC **keeps** about 25 Pb of data a year
- 20 million PP collisions per second (40MHz)
- Use triggers (hardware/software) to cut down to keep an interesting subset (300Hz)
- Data rate to further processing ~ 500MB/s
- Tiered global distribution of the raw and progressively summarized data



nuclear collision process

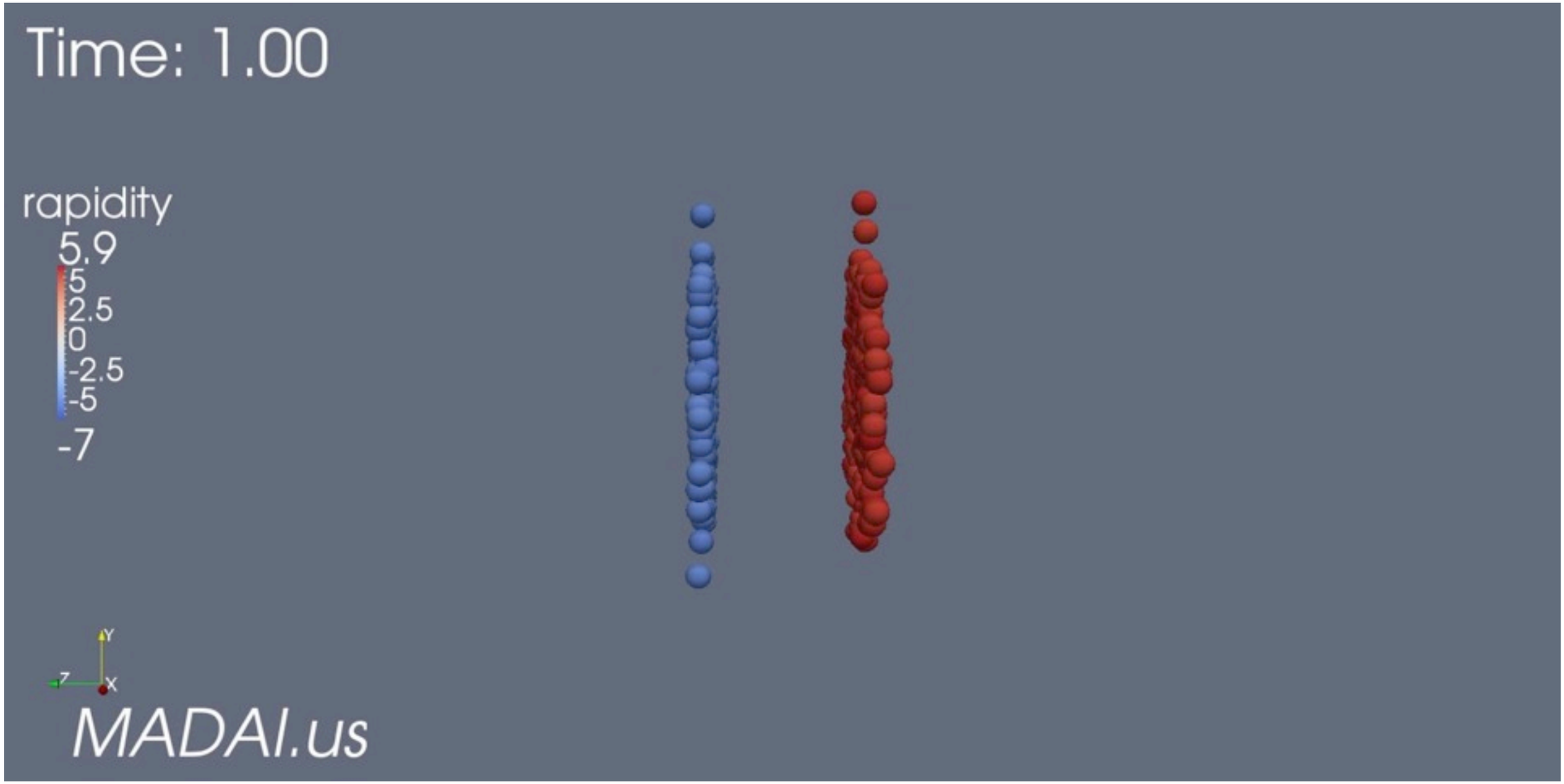


Challenges:

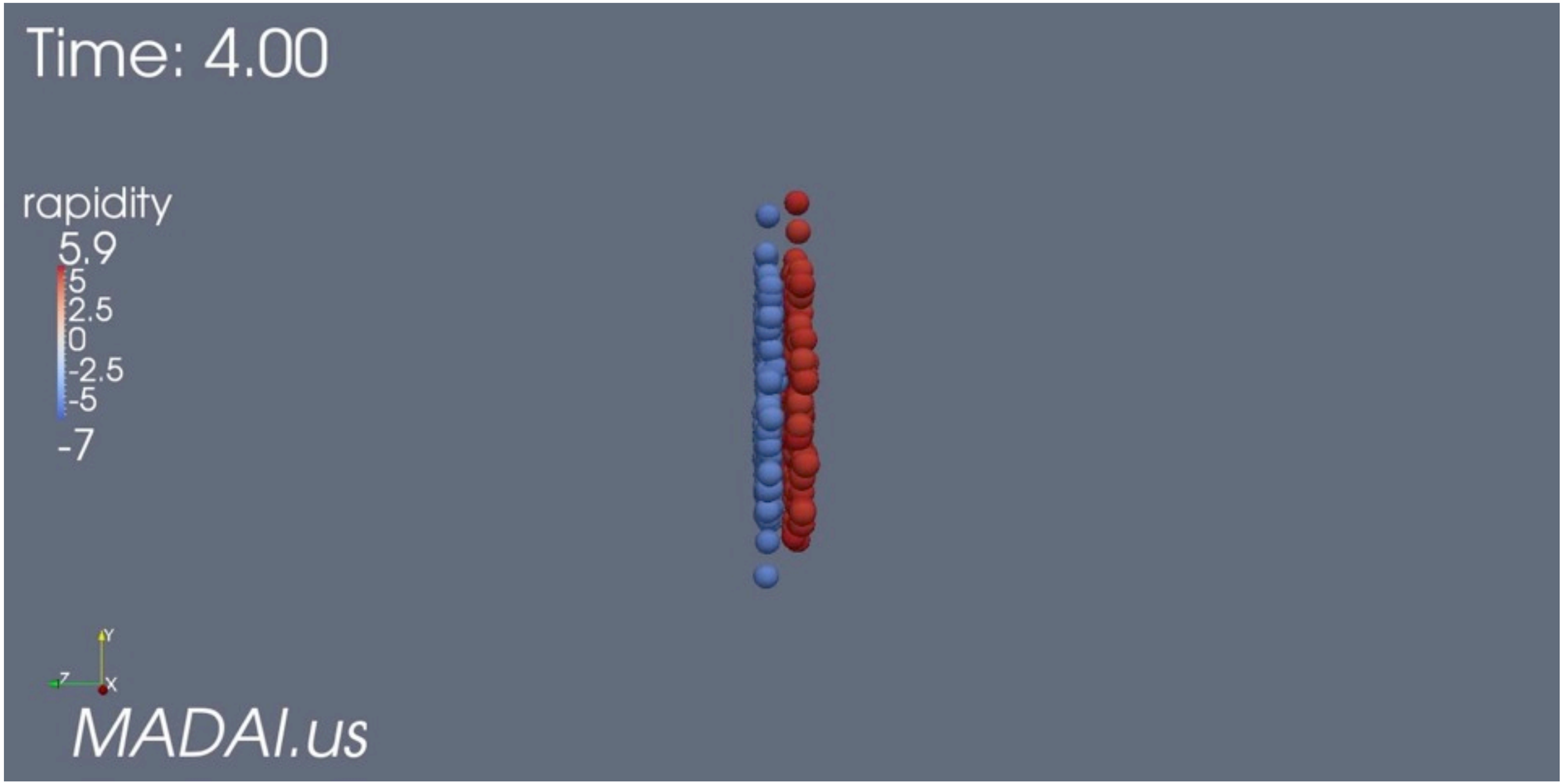
- time-scale of the collision process: 10^{-24} seconds! [too short to resolve]
- characteristic length scale: 10^{-15} meters! [too small to resolve]
- confinement: quarks & gluons form bound states @ hadronization, experiments don't observe them directly

- The matter produced is Very hot ~ 500,000 x Solar Core ~ e^27 K
- Energy contained in a RHIC collision ~ 30 TeV ~ 6μJ
- Energy density at RHIC is ~ 5GeV / fm^3, at this density the yearly US energy use would fit into a box 5μm on a side.

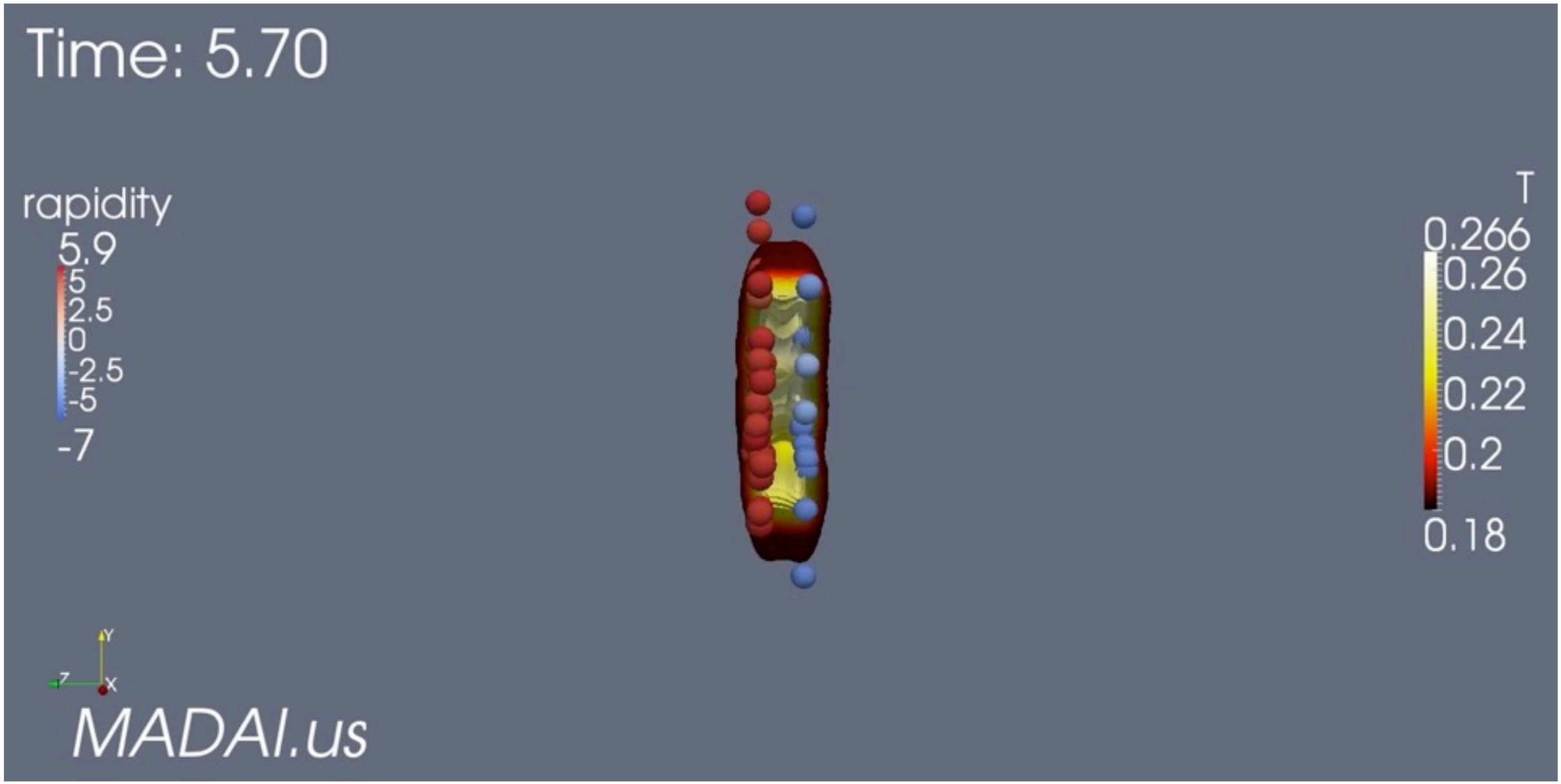
Simulation of a collision



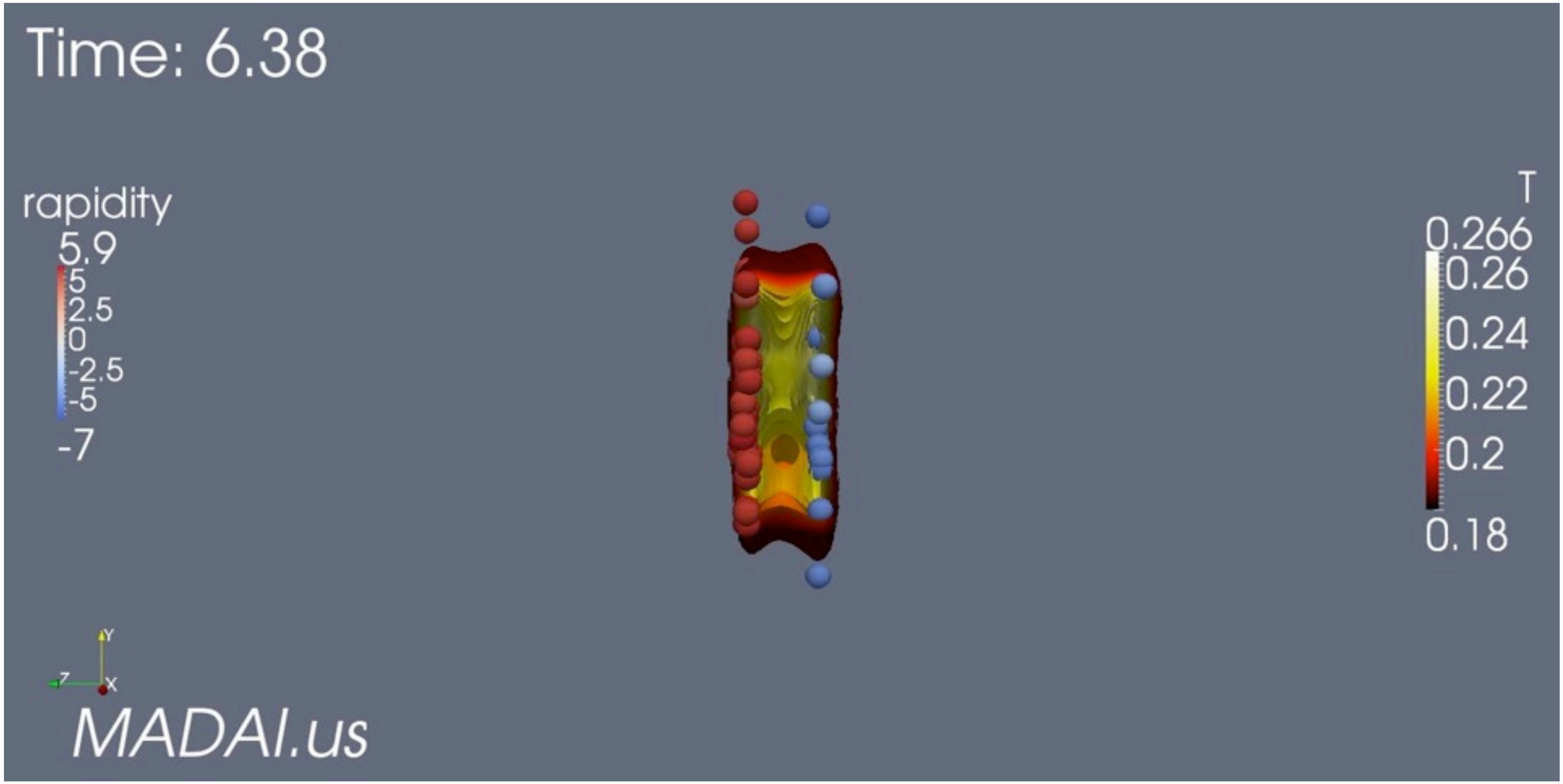
Simulation of a collision



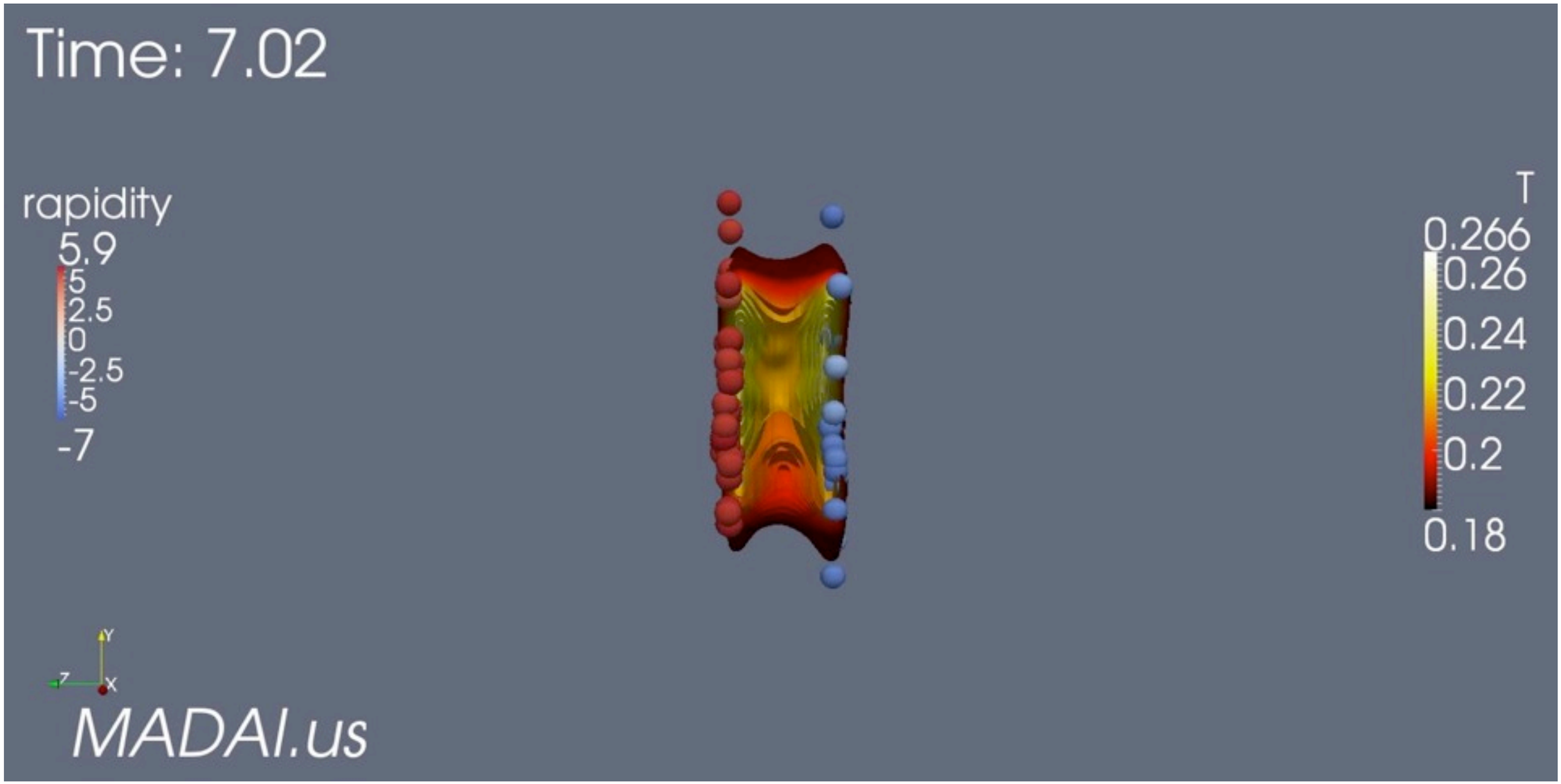
Simulation of a collision



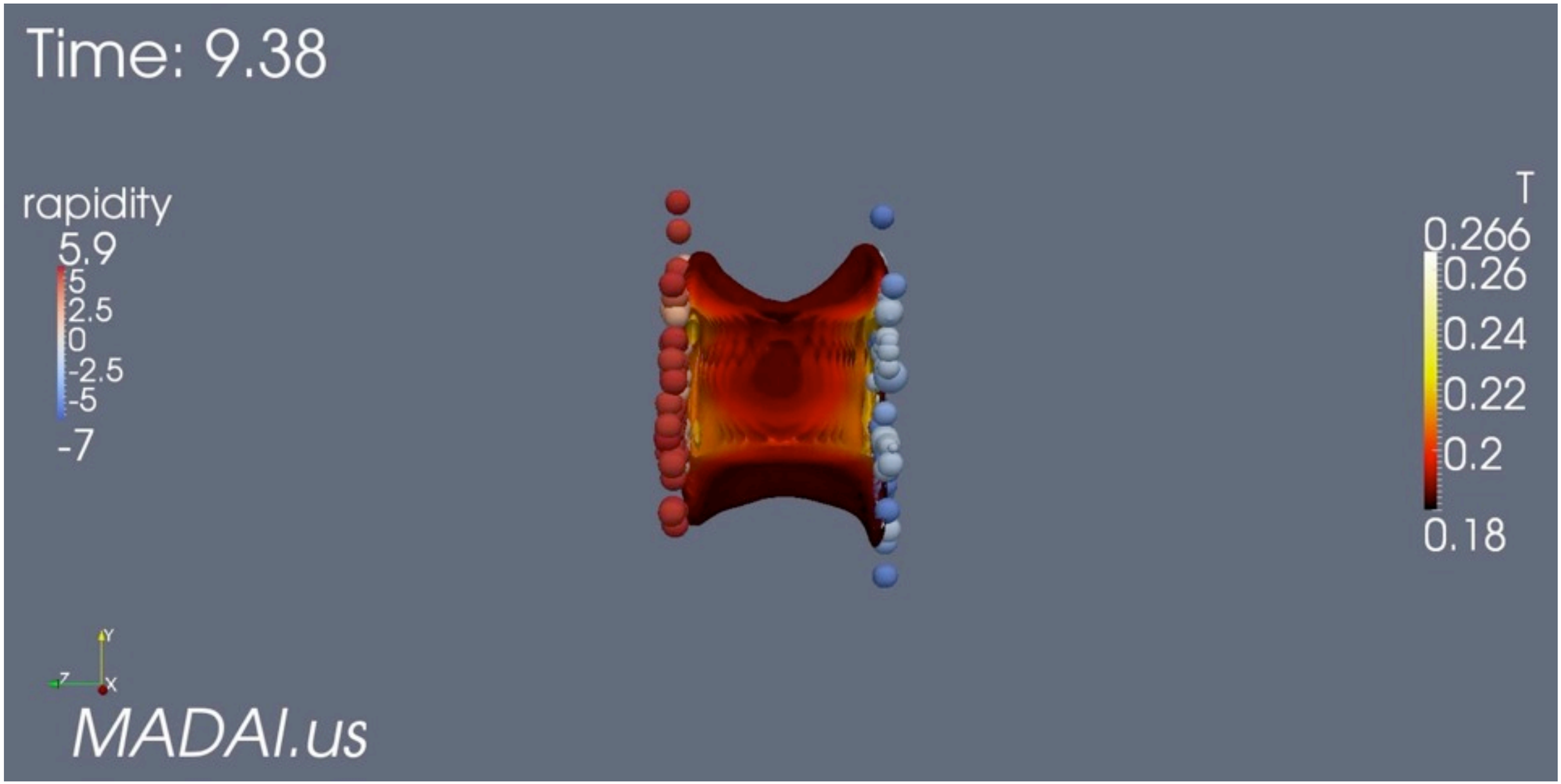
Simulation of a collision



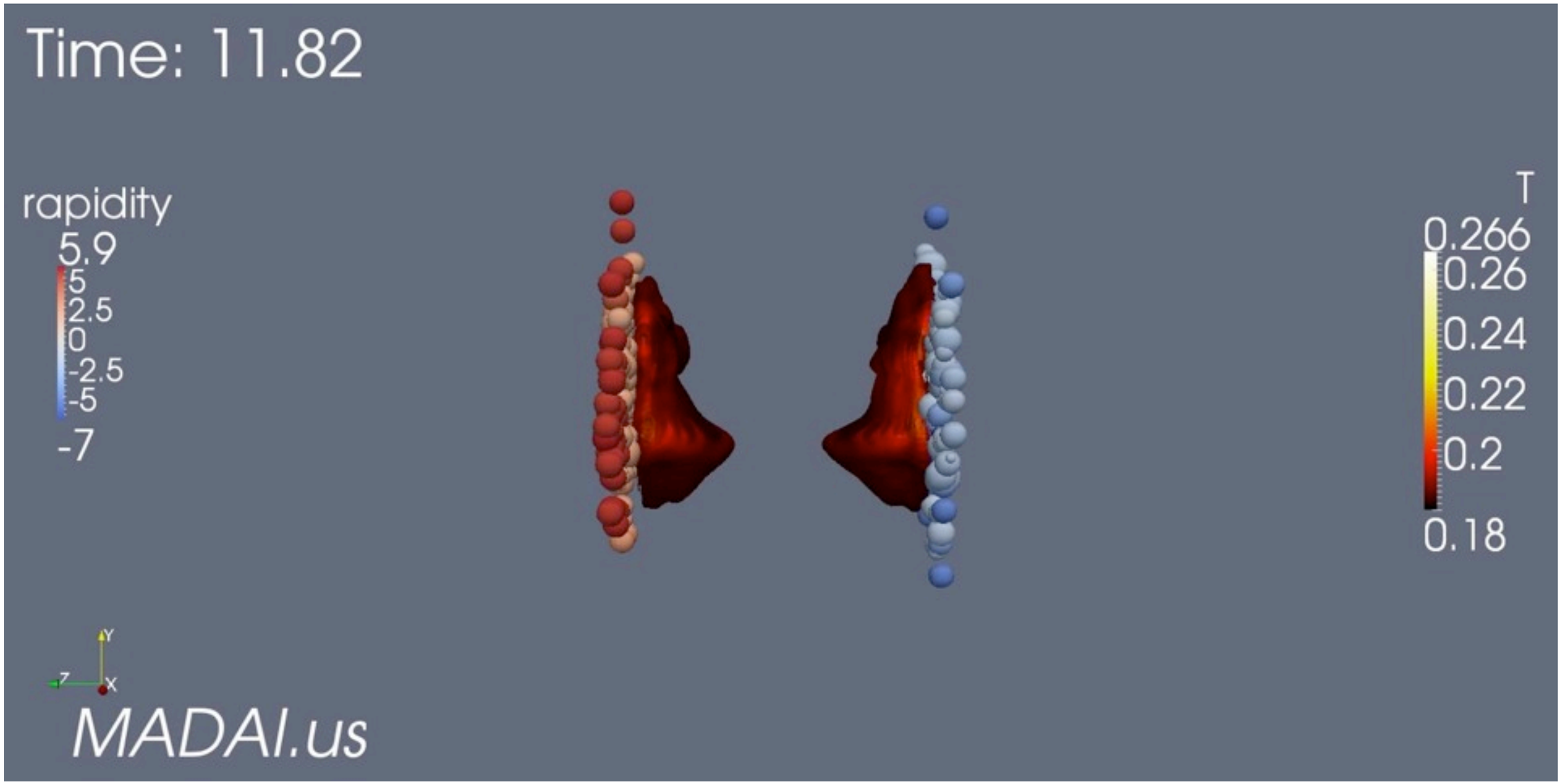
Simulation of a collision



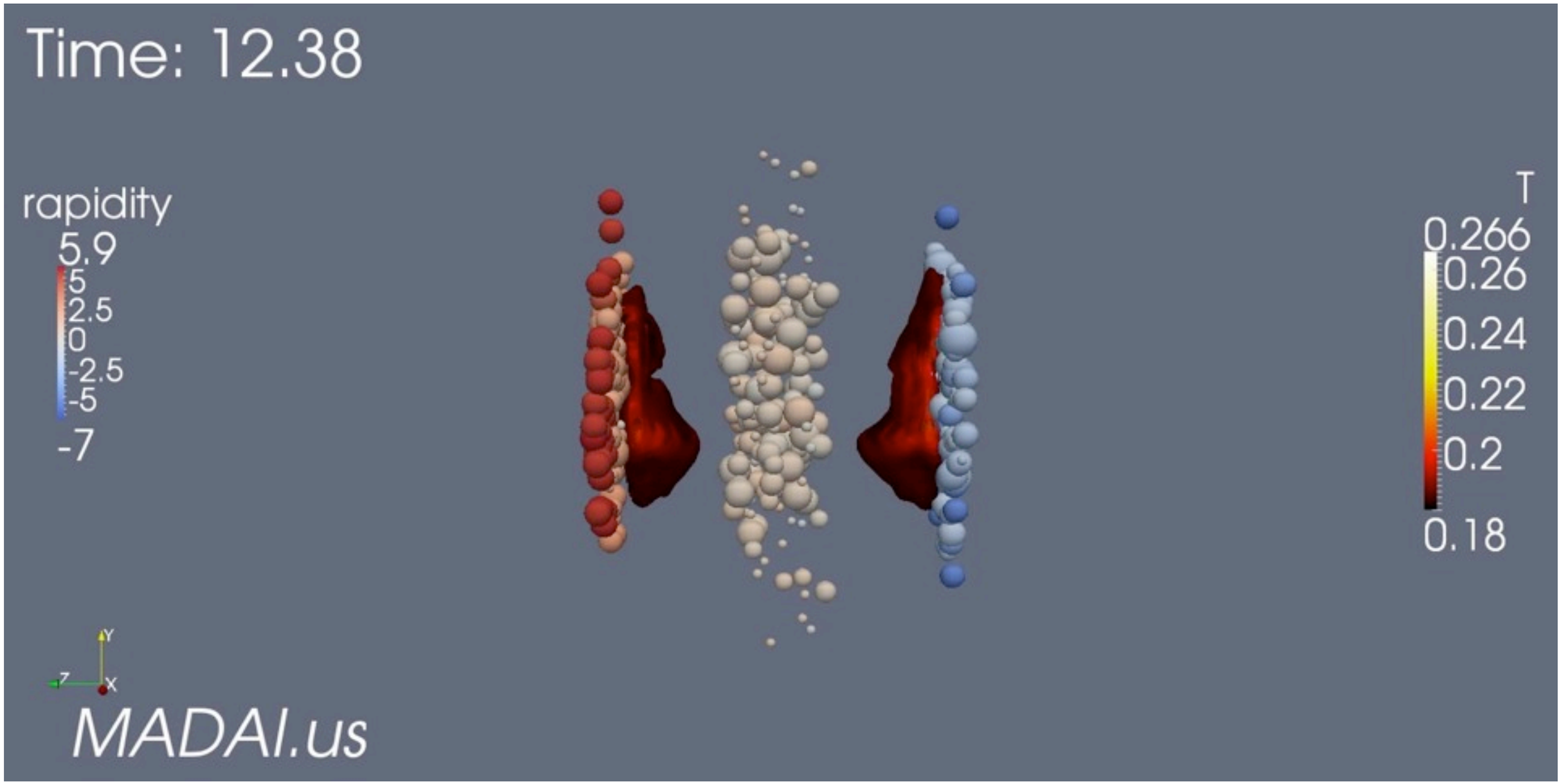
Simulation of a collision



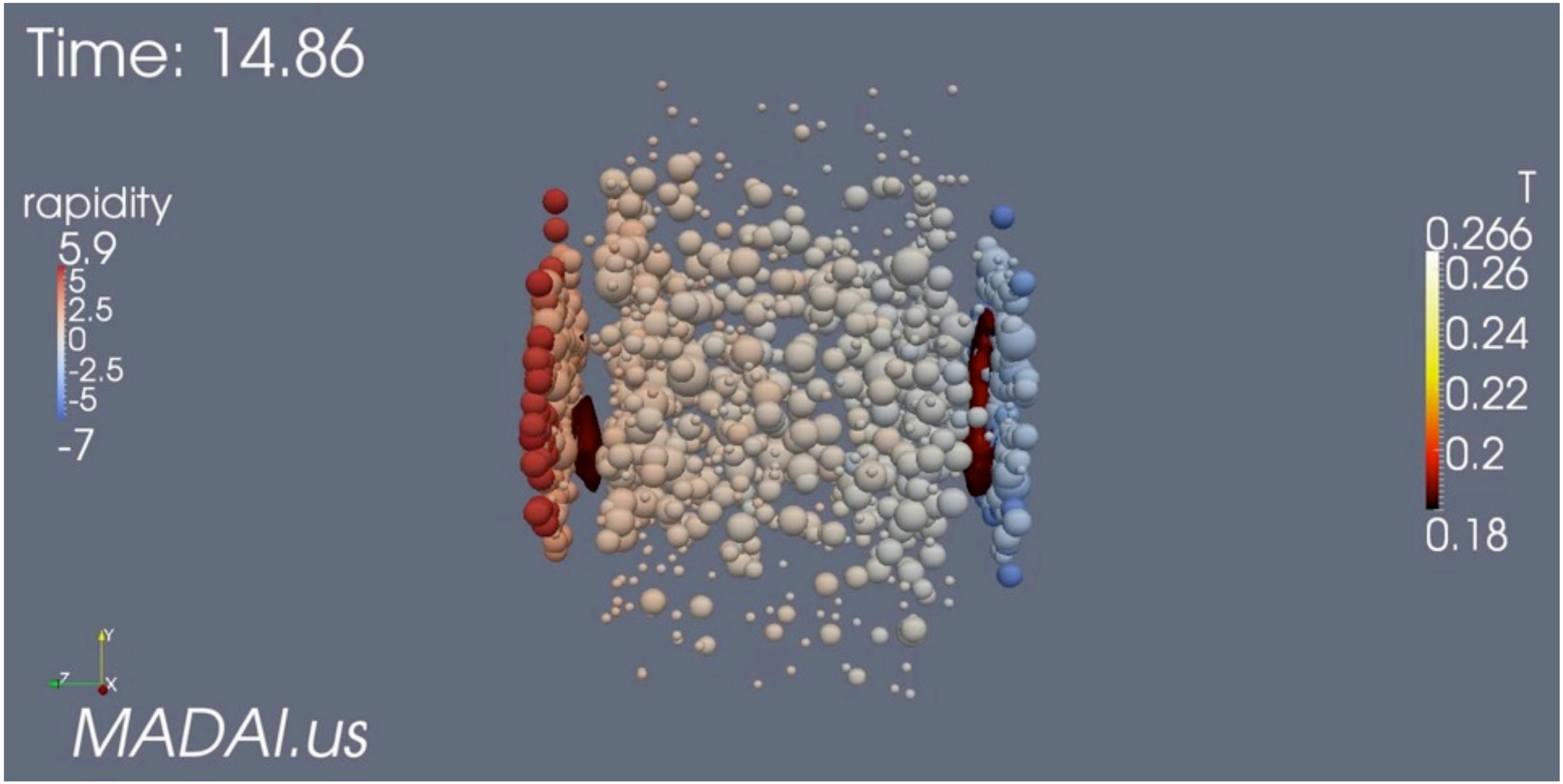
Simulation of a collision



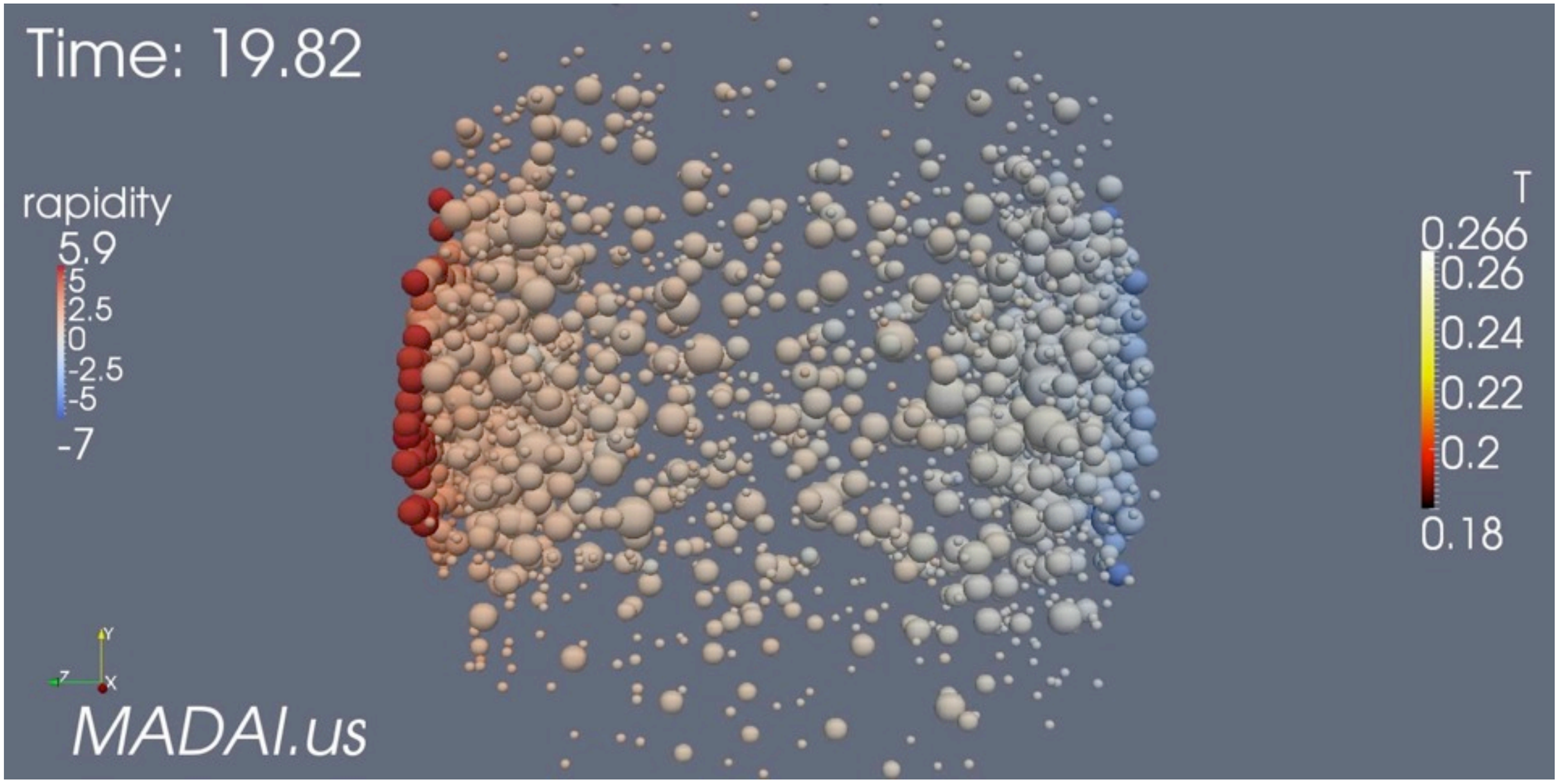
Simulation of a collision



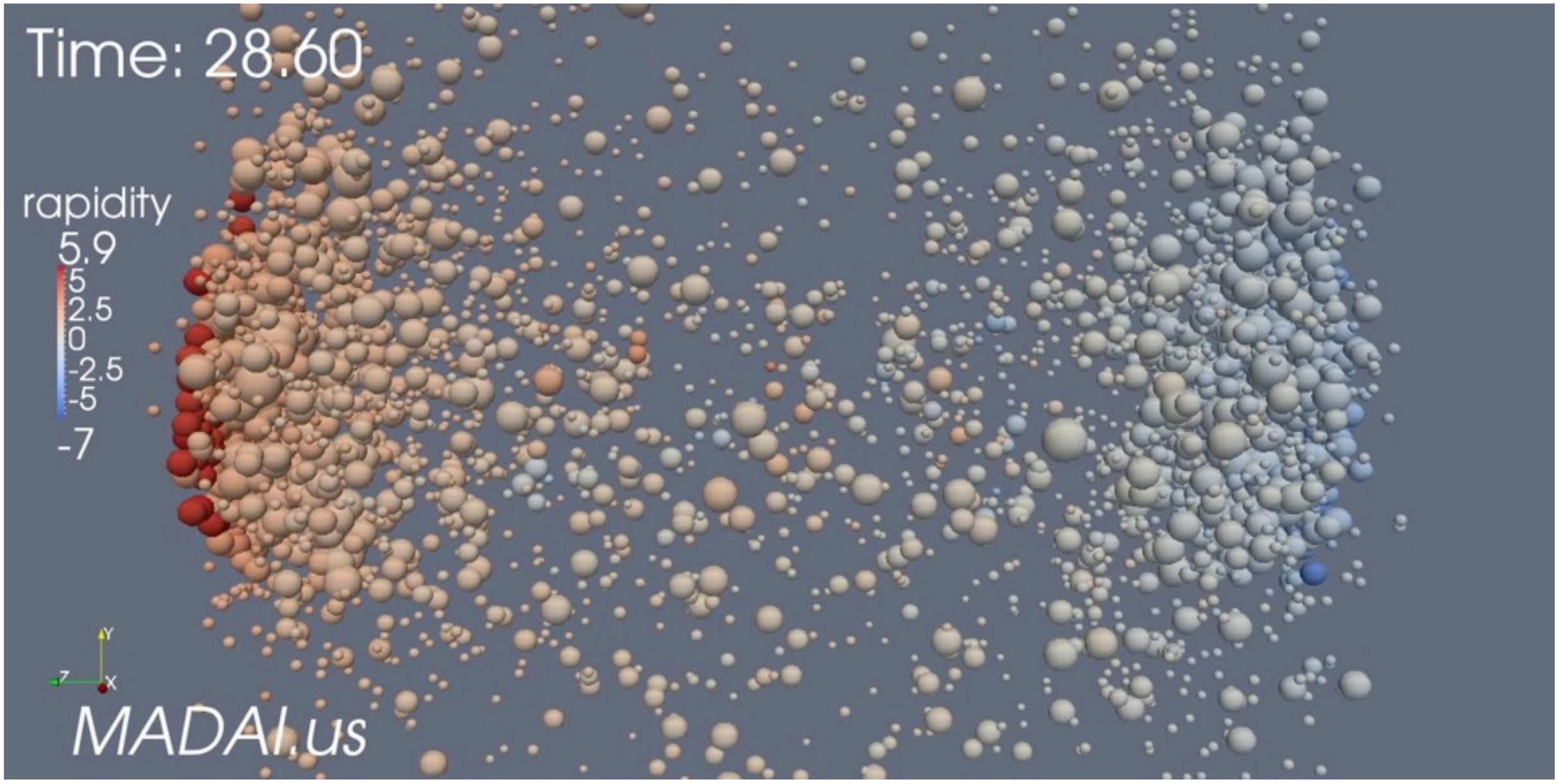
Simulation of a collision



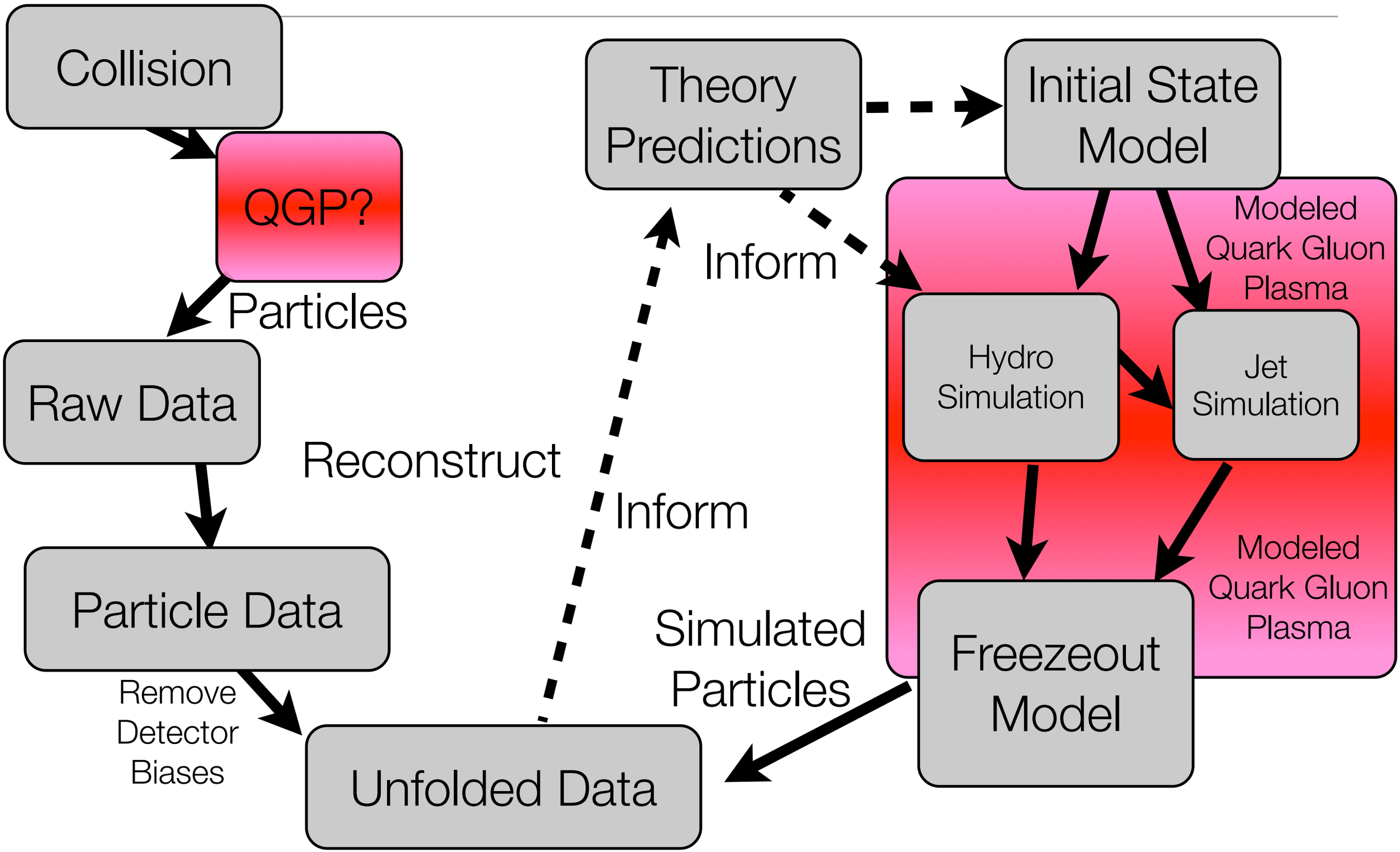
Simulation of a collision



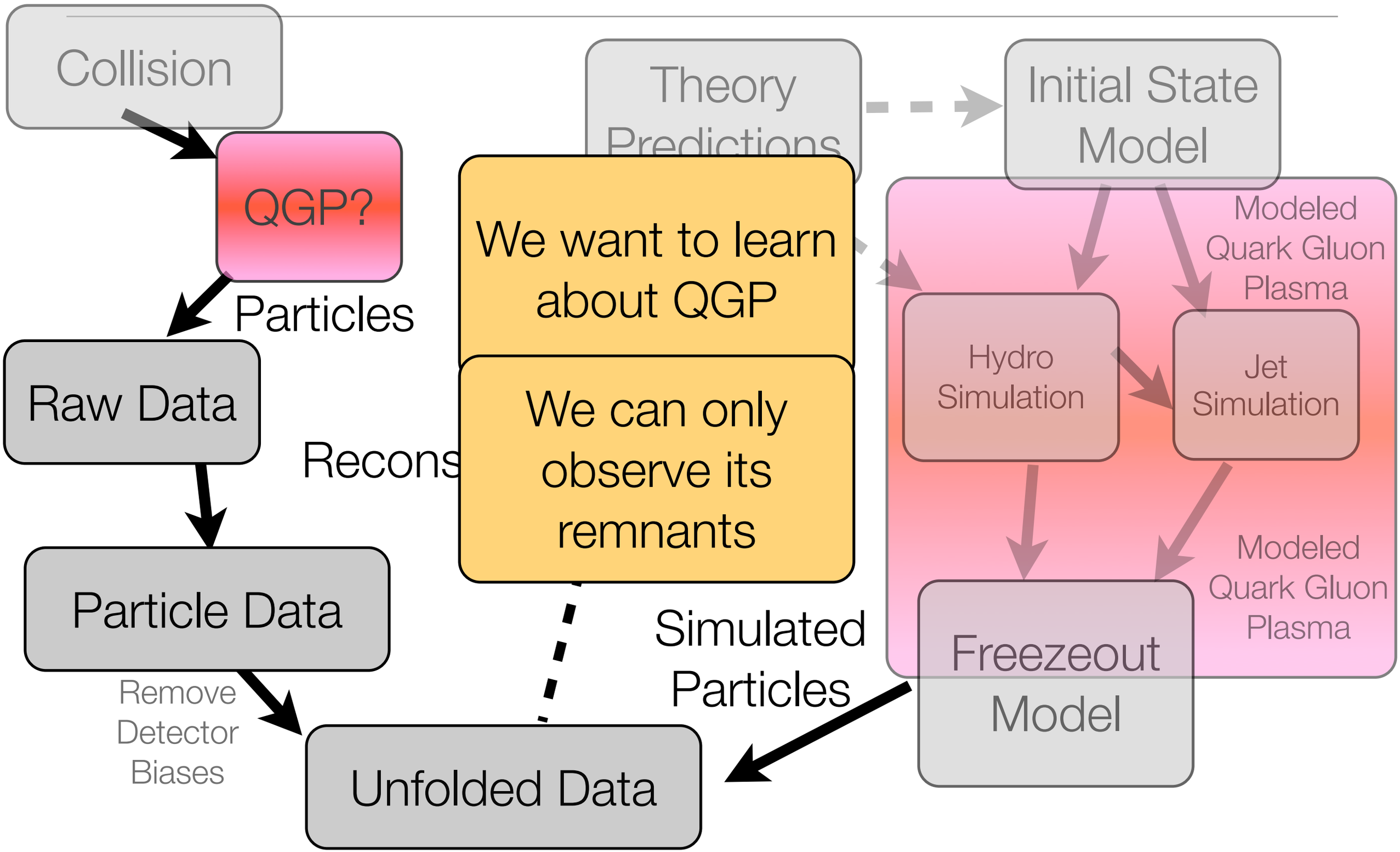
Simulation of a collision



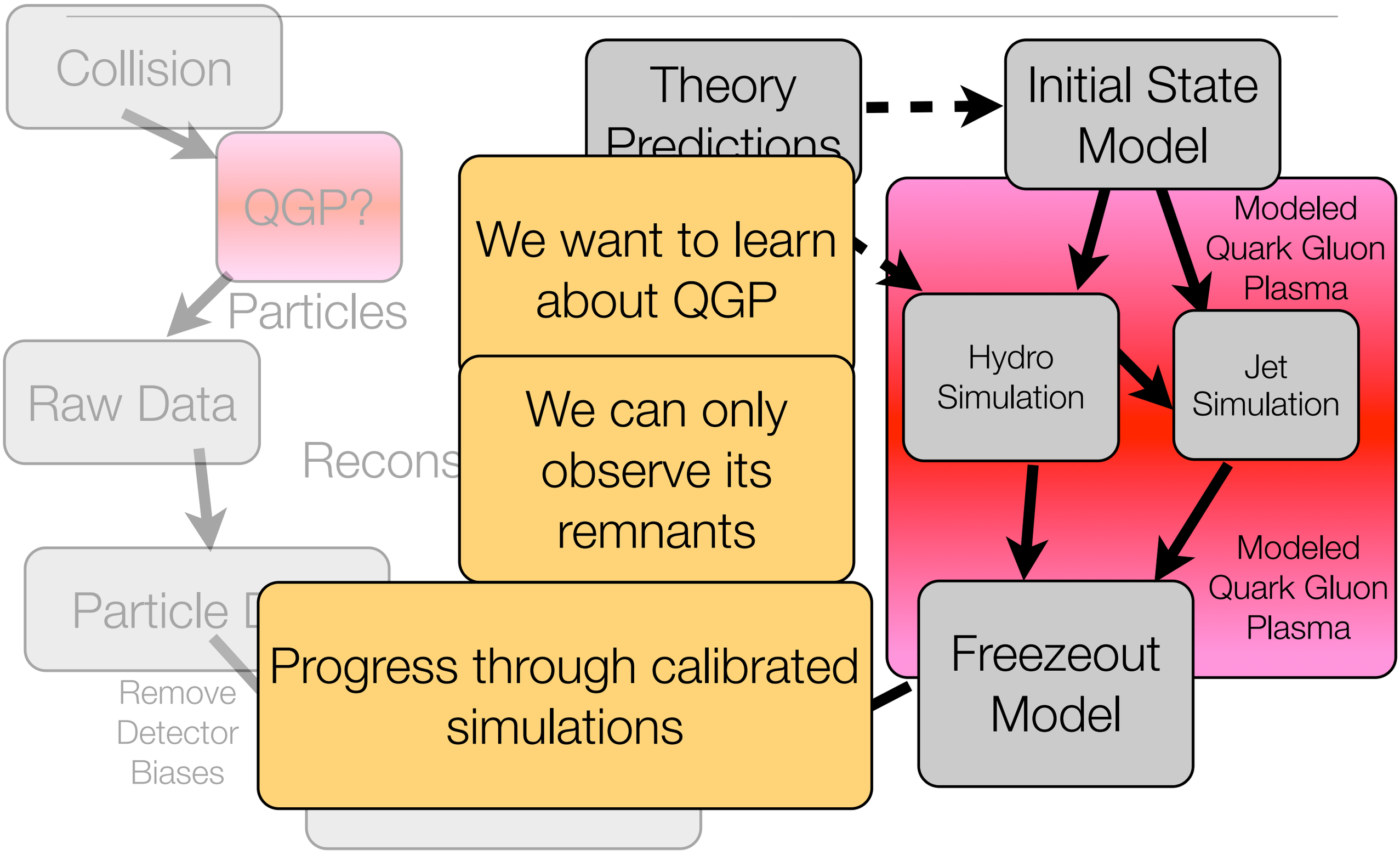
High Energy Nuclear Physics



High Energy Nuclear Physics



High Energy Nuclear Physics



Gaussian Process Emulators

Theory and Practice

Gaussian Processes

- A stochastic process

$$\mathrm{GP}\left(\mu(\cdot), \mathcal{C}(\cdot,\cdot)\right) : \mathbb{R}^n \to \mathbb{R},$$

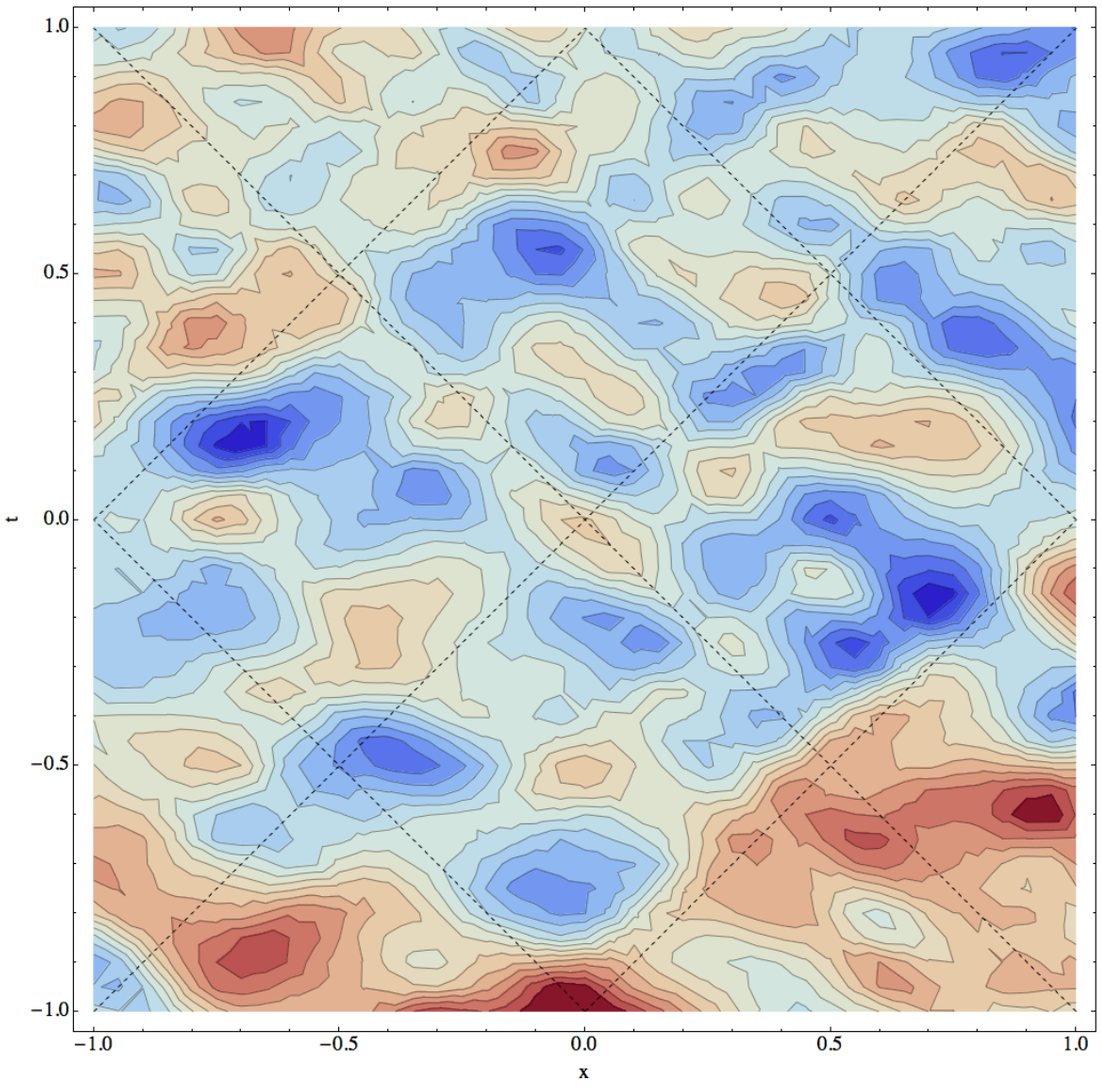
- Mean function: $\mu(\cdot)$

- Covariance function: $\mathcal{C}(\cdot,\cdot)$
  +ve def

- Any finite marginalization is MVN

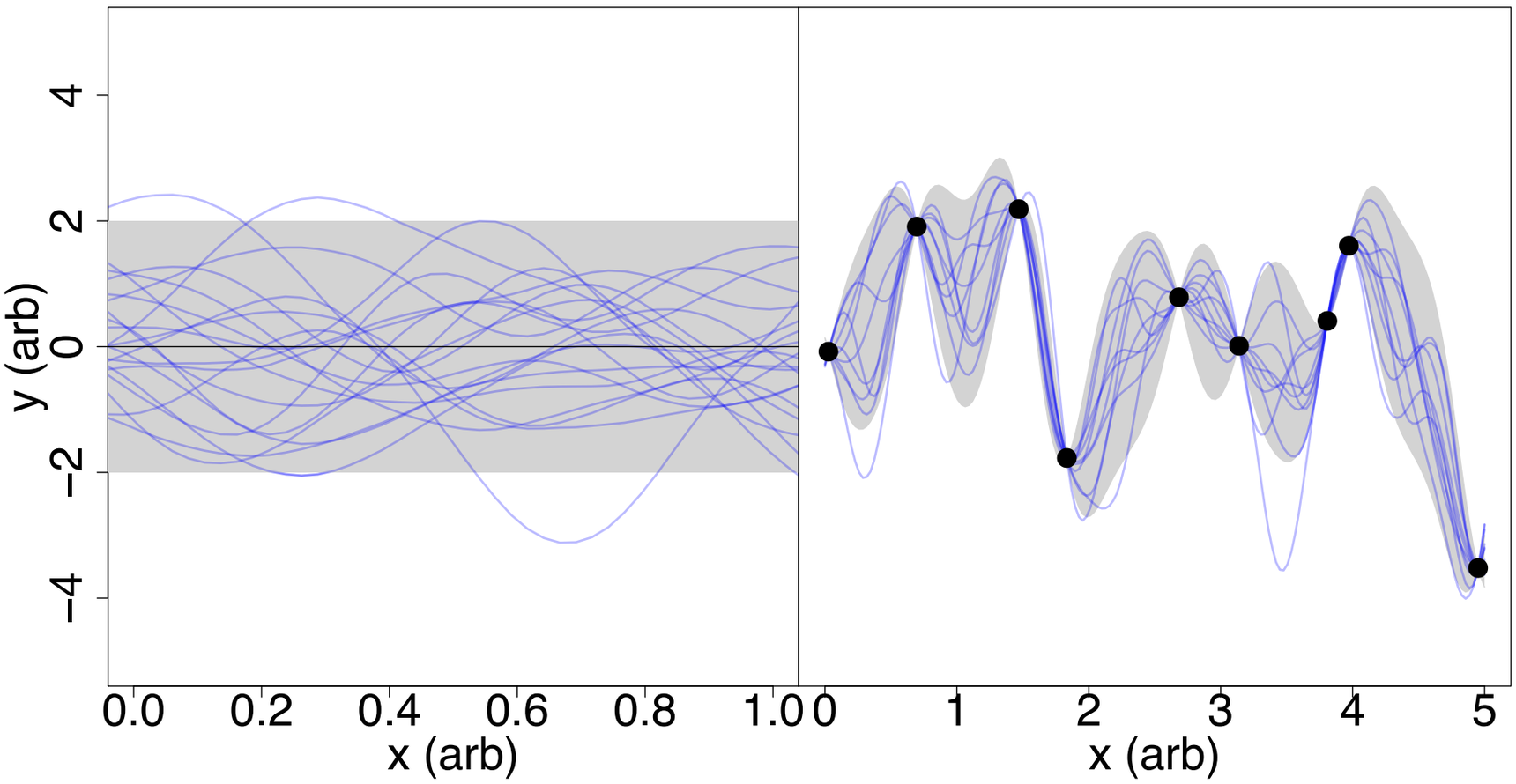
$$P(x_1, x_2, x_3) \sim \mathrm{MVN}(\vec{\mu}, K),$$
$$\vec{\mu} = (\mu(x_1), \mu(x_2), \mu(x_3))^\mathsf{T}$$
$$K_{i,j} = \mathcal{C}(x_i, x_j).$$

- very flexible



$$\mathcal{C}(x,y) = \frac{1}{2}\left(iG^+(x,y) + iG^-(x,y)\right),$$
$$G^\pm(x,x') = \frac{-i}{4\pi^2}\frac{1}{\Delta x_\pm^2}, \quad \Delta x_\pm^2 = (t - t' \pm i\epsilon)^2 - \|x - x'\|^2$$

GP Emulators



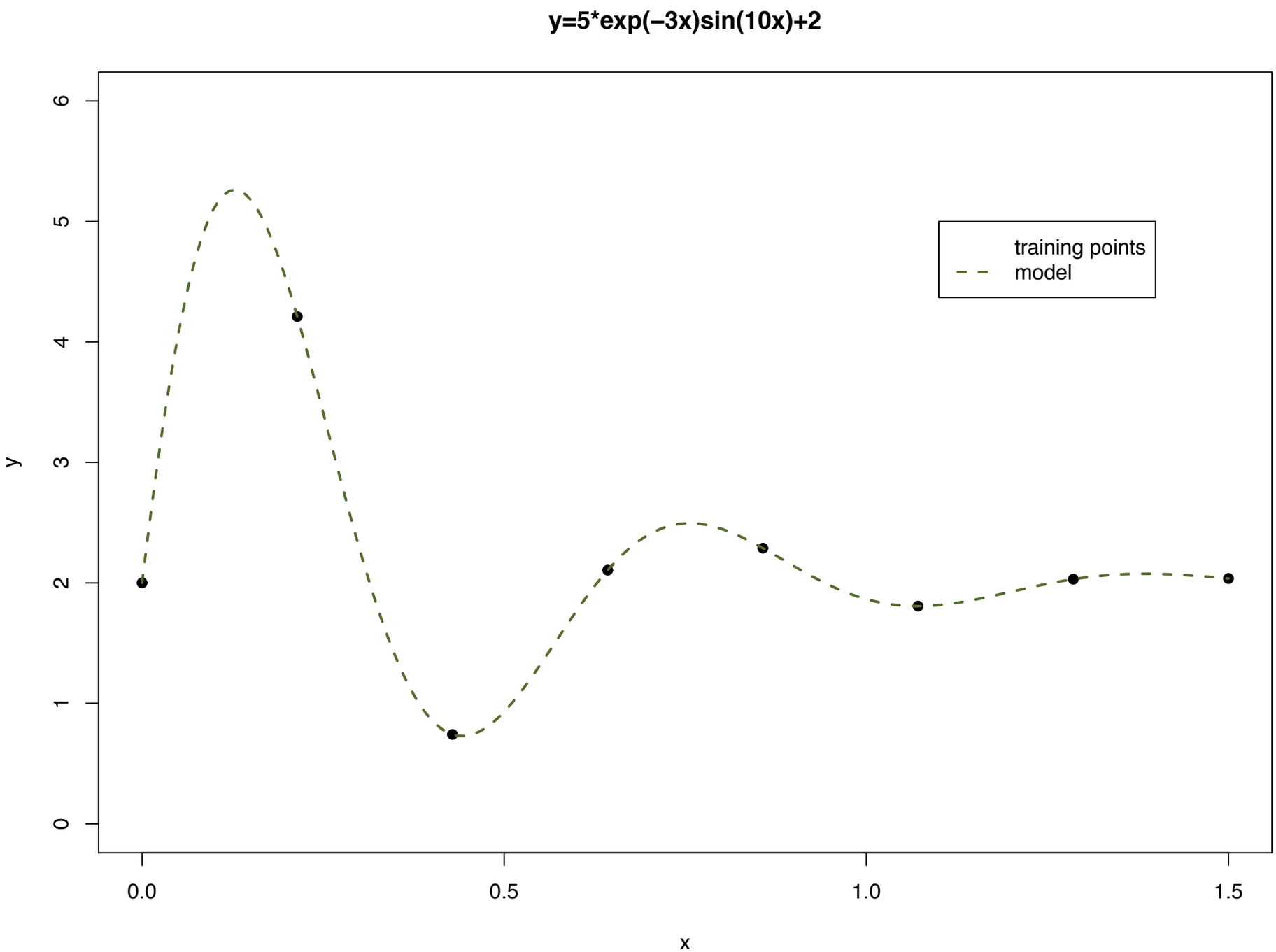
Unconditioned Draws

Conditioned on training data

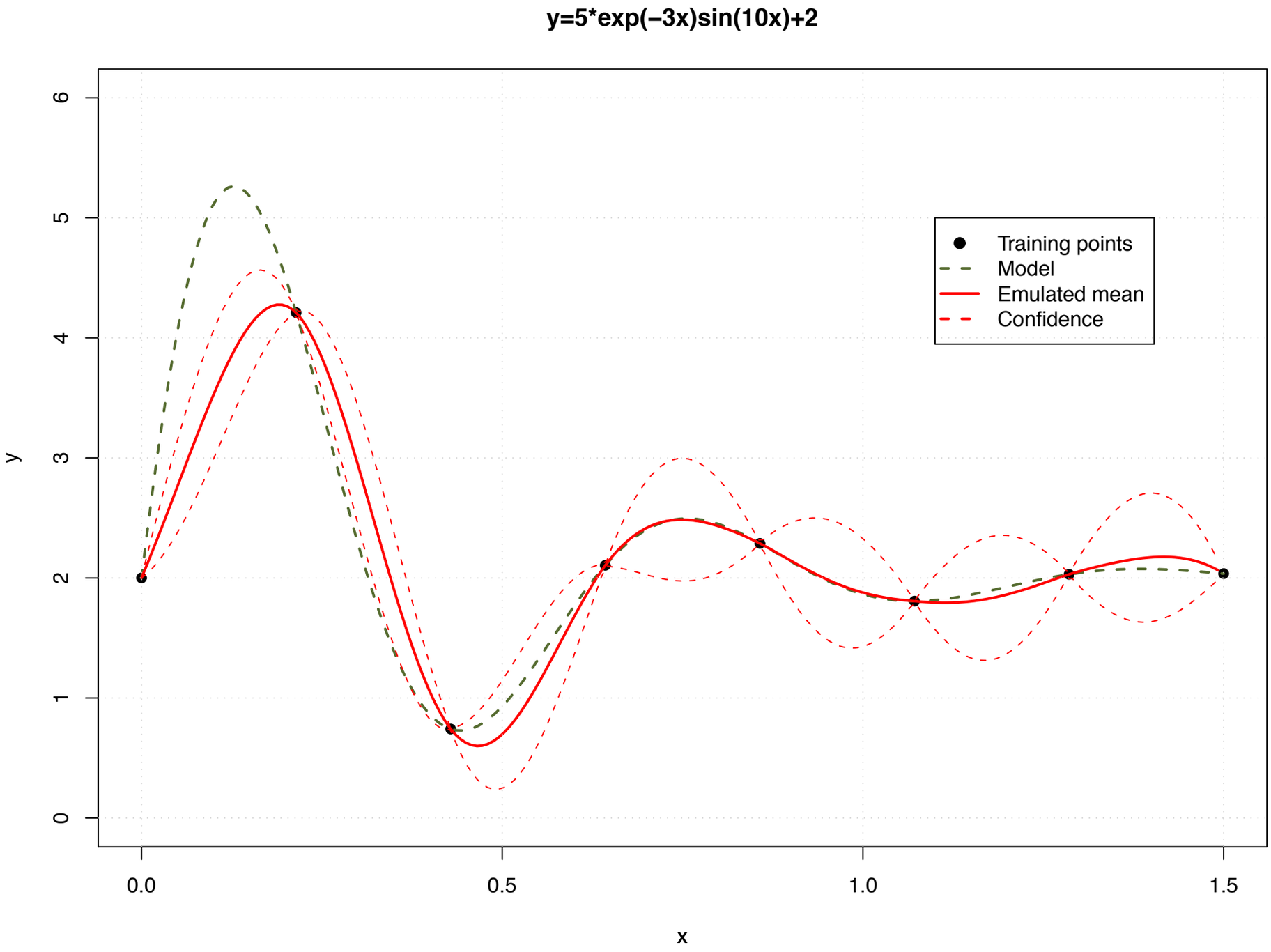
- A Gaussian Process emulator is a statistical approximation to a function.
- Condition a probability distribution for a family of functions (GP) to produce samples which pass through a set of training points.
- draws are then a statistical approximation to the simulator

Rasmussen & Williams, “Gaussian Processes for Machine Learning” MIT Press

Simple Example



Simple Example

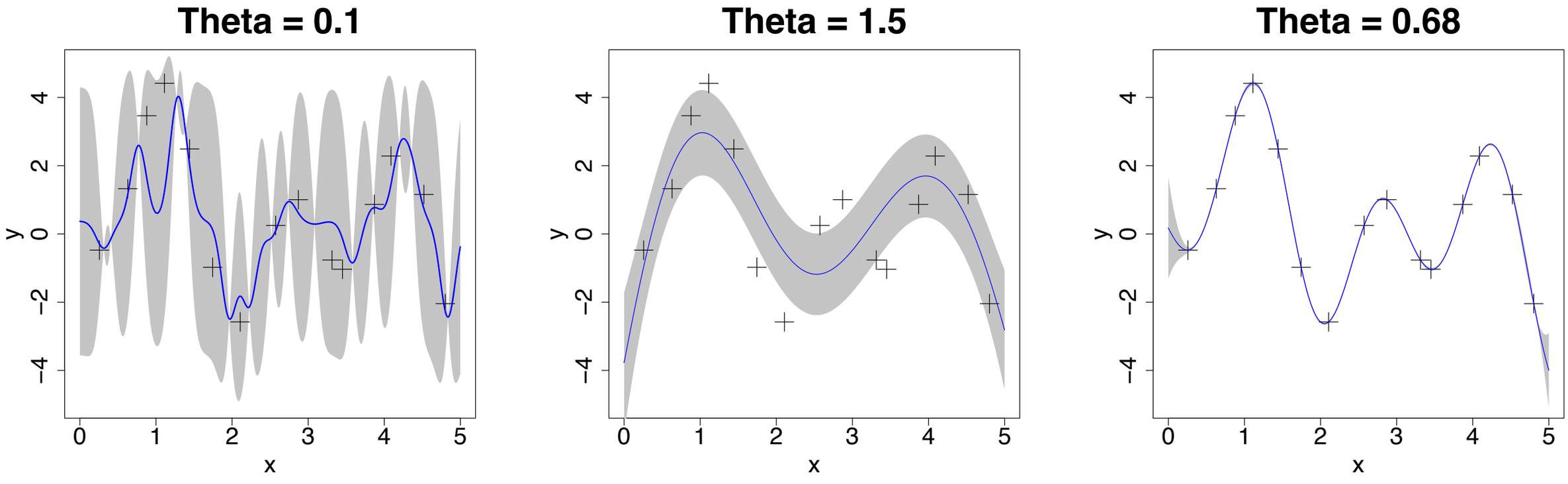


Essential Details

- Condition a Gaussian Process on observations of model
- Prior Mean from linear regression model, Prior Covariance from observations
- Covariance function sets “weights” of nearby training points.
  power exponential:

$$C_f(\mathbf{x}, \mathbf{x}', \Theta) = \theta_1 \exp\left(-\sum_i \frac{(x_i - x_i')^\alpha}{\theta_i^\alpha}\right)$$

- Length scale θ_i sets sensitivity to fluctuations in model output

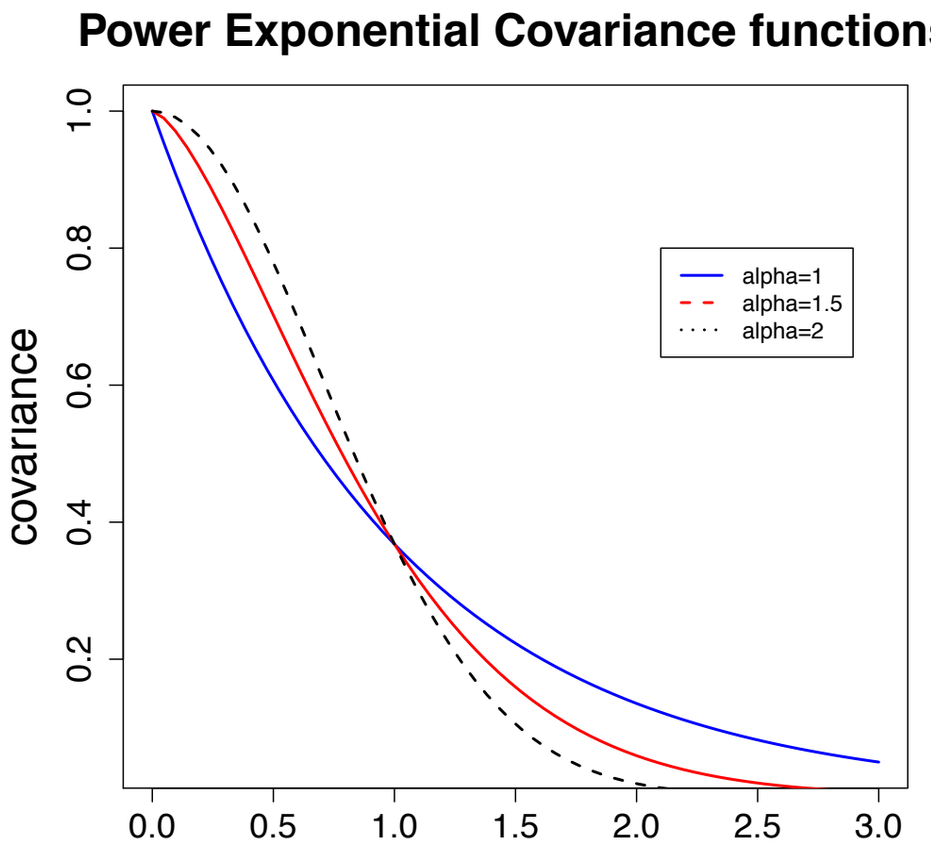


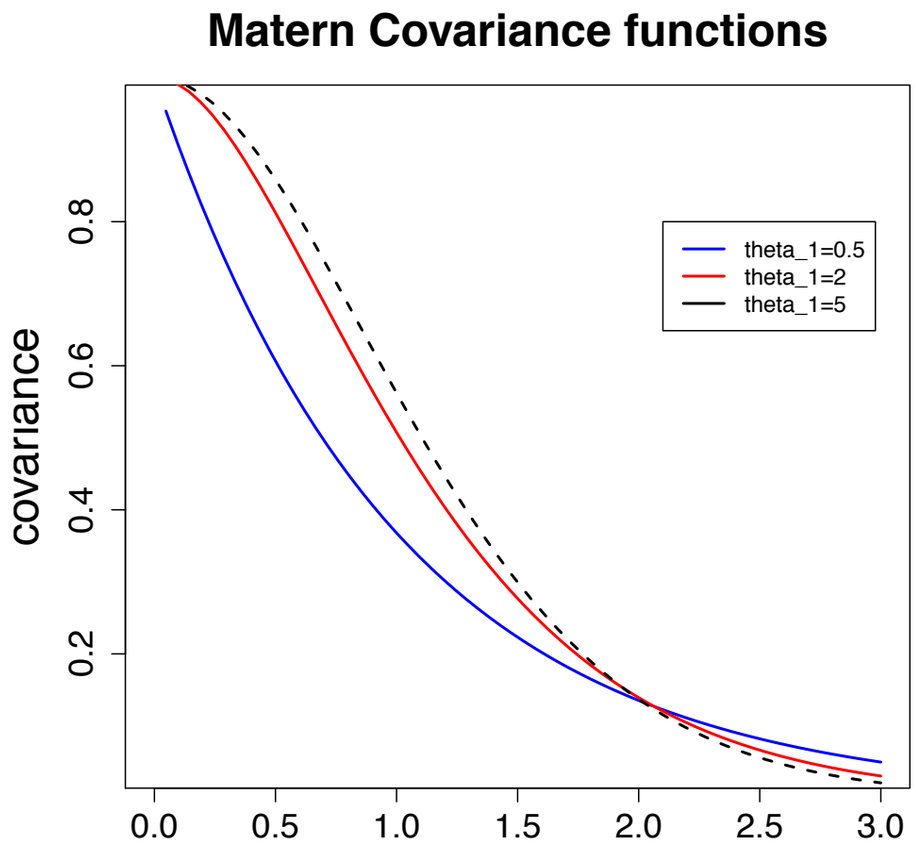
Covariance Function

- Choice of C has influence on the shape of the predictive distribution.

- Power Exponential: $$C_f(\mathbf{x}, \mathbf{x}', \Theta) = \theta_1 \exp\left(-\sum_i \frac{(x_i - x_i')^\alpha}{\theta_i^\alpha}\right)$$

- Matern Class: $$C_f(r, \Theta) = \frac{2^{1-\theta_1}}{\Gamma(\theta_1)}\left(\frac{\sqrt{2\theta_1}r}{\theta_2}\right)^{\theta_1} K_{\theta_1}\left(\frac{\sqrt{2\theta_1}r}{\theta_2}\right)$$





The posterior distribution (how it actually works)

- GP Prior: mean from linear regression, covariance defined by choice of covariance fn and design.
- After conditioning on training data -> MVN posterior distribution.

$$\hat{m}(\mathbf{x}) = \mathbf{h}(\mathbf{x})^T \hat{\beta} + \mathbf{c}(\mathbf{x}) C^{-1} \left(\mathbf{y} - H\hat{\beta} \right) \qquad \mathcal{P} \sim \mathrm{MVN}\left(\hat{m}(\mathbf{x}), \hat{\Sigma} \right)$$

- Mean mixes broad trends (linear regression) and local fluctuations around these (correlation term).

$$\hat{\Sigma}(\mathbf{x_i}, \mathbf{x_j}) = c(\mathbf{x_i}, \mathbf{x_j}) - \mathbf{c}(\mathbf{x_i})^T C^{-1} \mathbf{c}(\mathbf{x_j}) + \Gamma(\mathbf{x_i}, \mathbf{x_j})$$

- Prior variance is restricted by a correlation term, gamma arises from linear regression model.

$$C_{ij} = c(\mathbf{x}_i, \mathbf{x}_j)$$

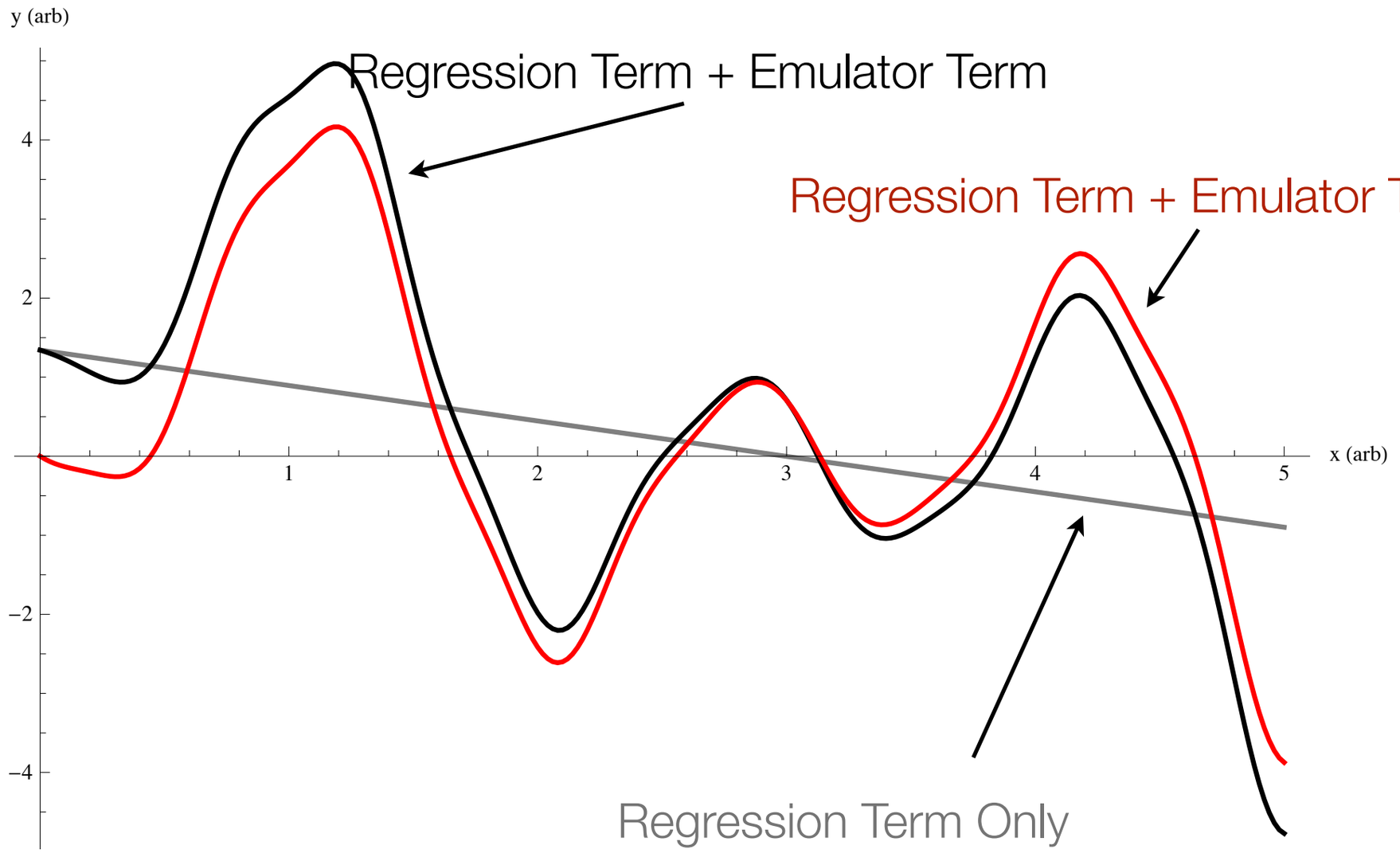
$$\mathbf{c}(\mathbf{x})^T = (c(\mathbf{x}_1, \mathbf{x}), \ldots, c(\mathbf{x}_n, \mathbf{x}))$$

"Bayesian Calibration of computer models" Kennedy & O'Hagan. J.R.Statist.Soc. B 63, 2001

The posterior distribution

- Mean mixes broad trends (linear regression) and local fluctuations around these (correlation term).

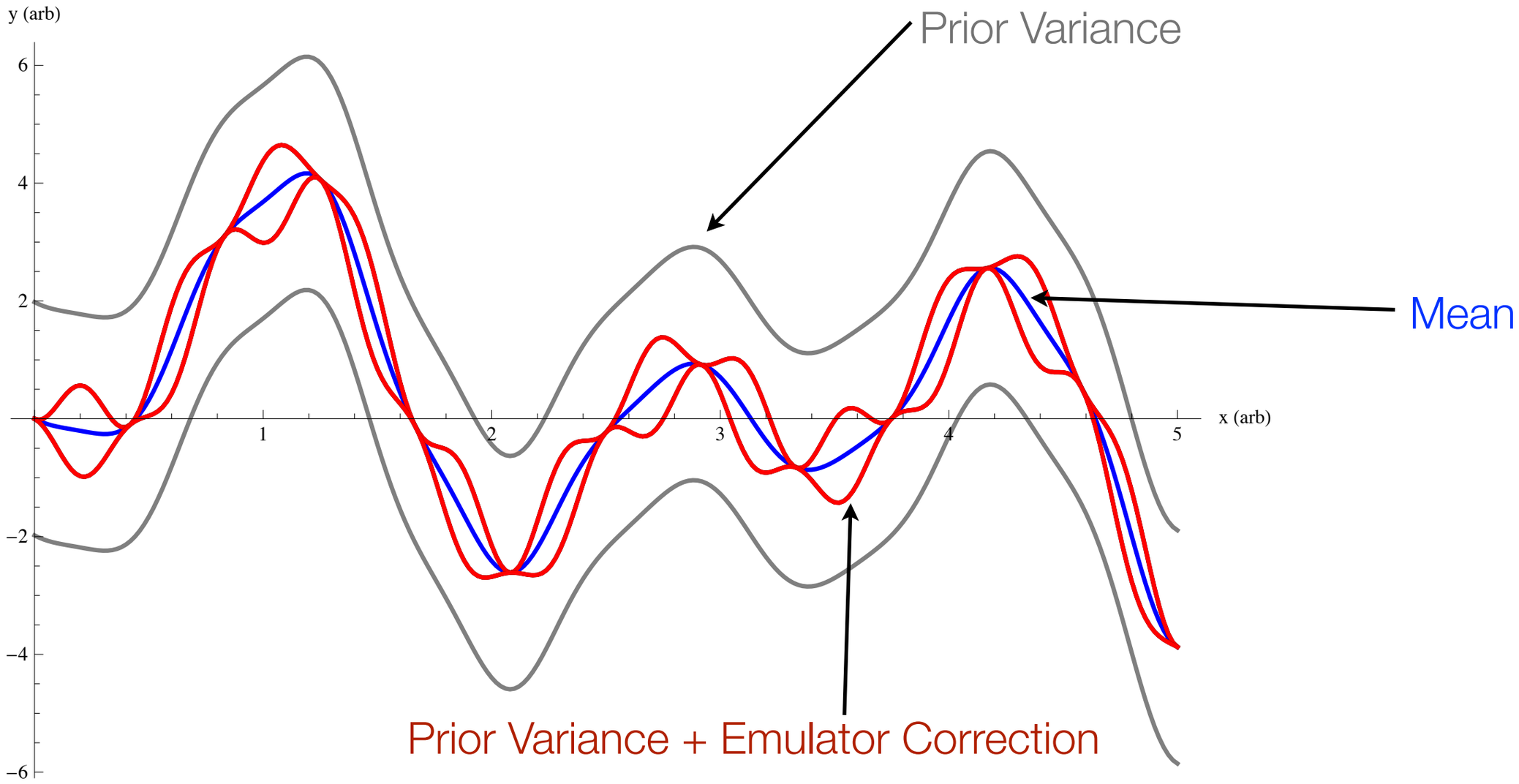
$$\hat{m}(\mathbf{x}) = \mathbf{h}(\mathbf{x})^T \hat{\beta} + \mathbf{c}(\mathbf{x}) C^{-1} \left(\mathbf{y} - H\hat{\beta} \right)$$



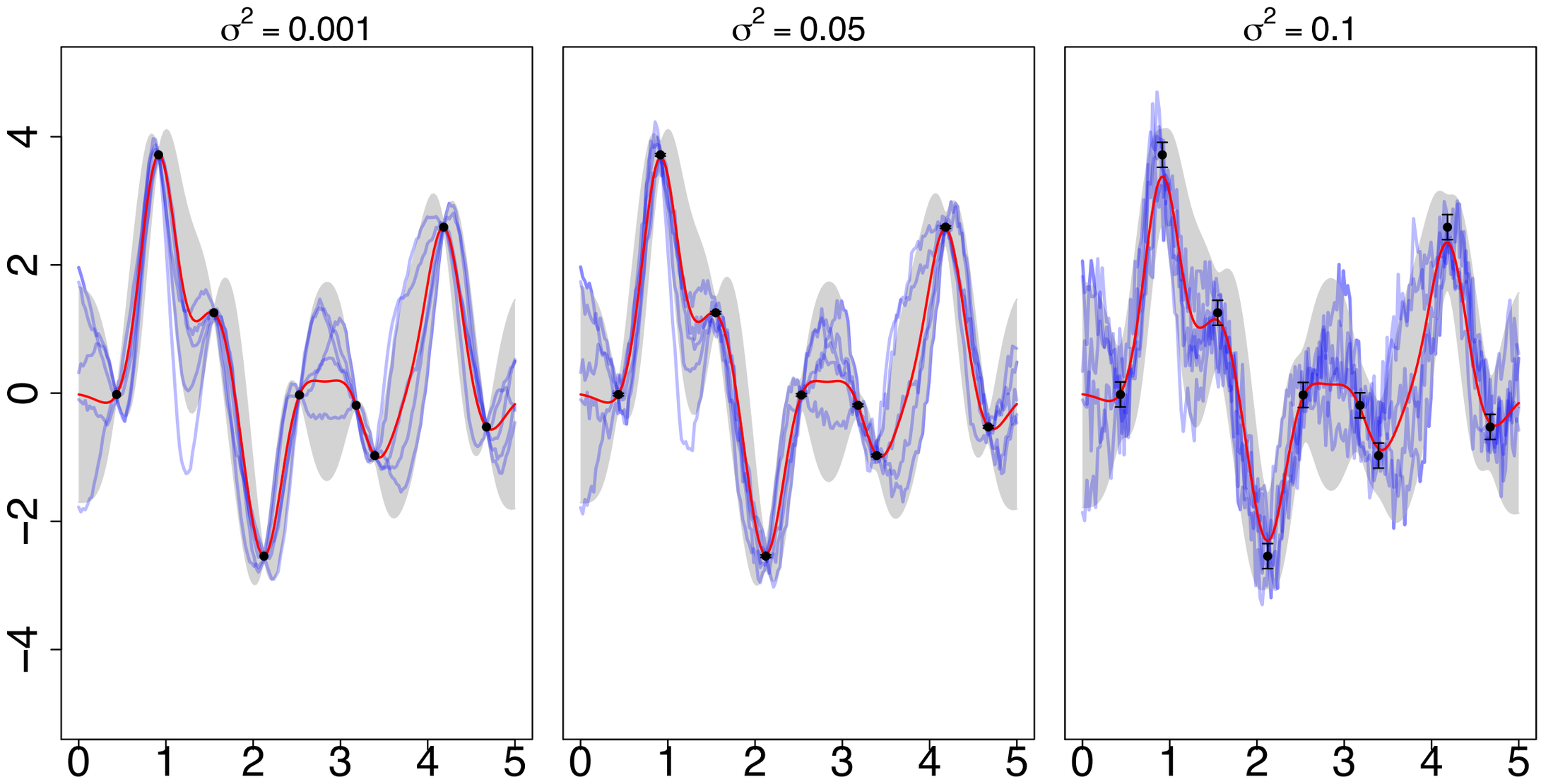
The posterior distribution

$$\hat{\Sigma}(\mathbf{x}_i, \mathbf{x}_j) = c(\mathbf{x}_i, \mathbf{x}_j) - \mathbf{c}(\mathbf{x_i})^T C^{-1} \mathbf{c}(\mathbf{x_j}) + \Gamma(\mathbf{x}_i, \mathbf{x}_j)$$

- Prior variance is restricted by a correlation term, gamma arises from linear regression model.



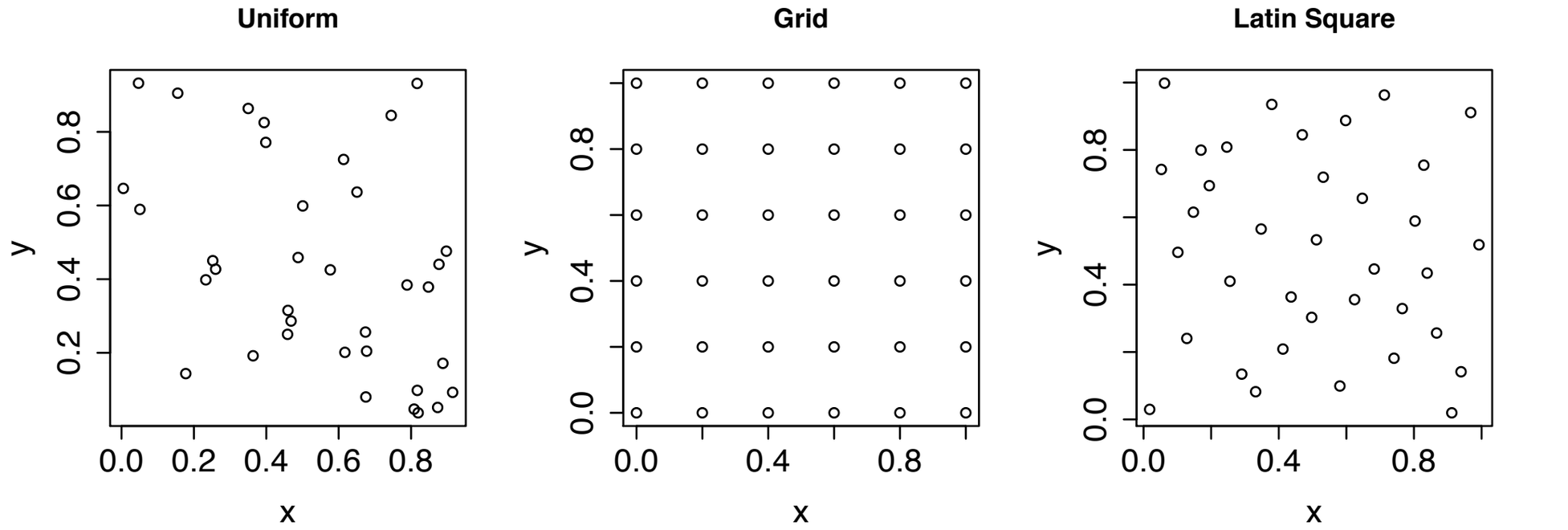
Influence of observation-noise / nugget



$$Y_m(x) = f(x) + z, \quad \mathbb{R}^n \to \mathbb{R}, \quad z \sim \mathrm{N}(0, \sigma^2).$$

$$f(x) = \sin(x) + 2\sin(2x) - 2\sin(4x)$$

Picking a design



- Should be space filling and efficient (i.e not a grid) Details are not terribly important. Latin square based designs are typical

M. D. McKay, R. J. Beckman, and W. J. Conover, Technometrics **21**, pp. 239 (1979).

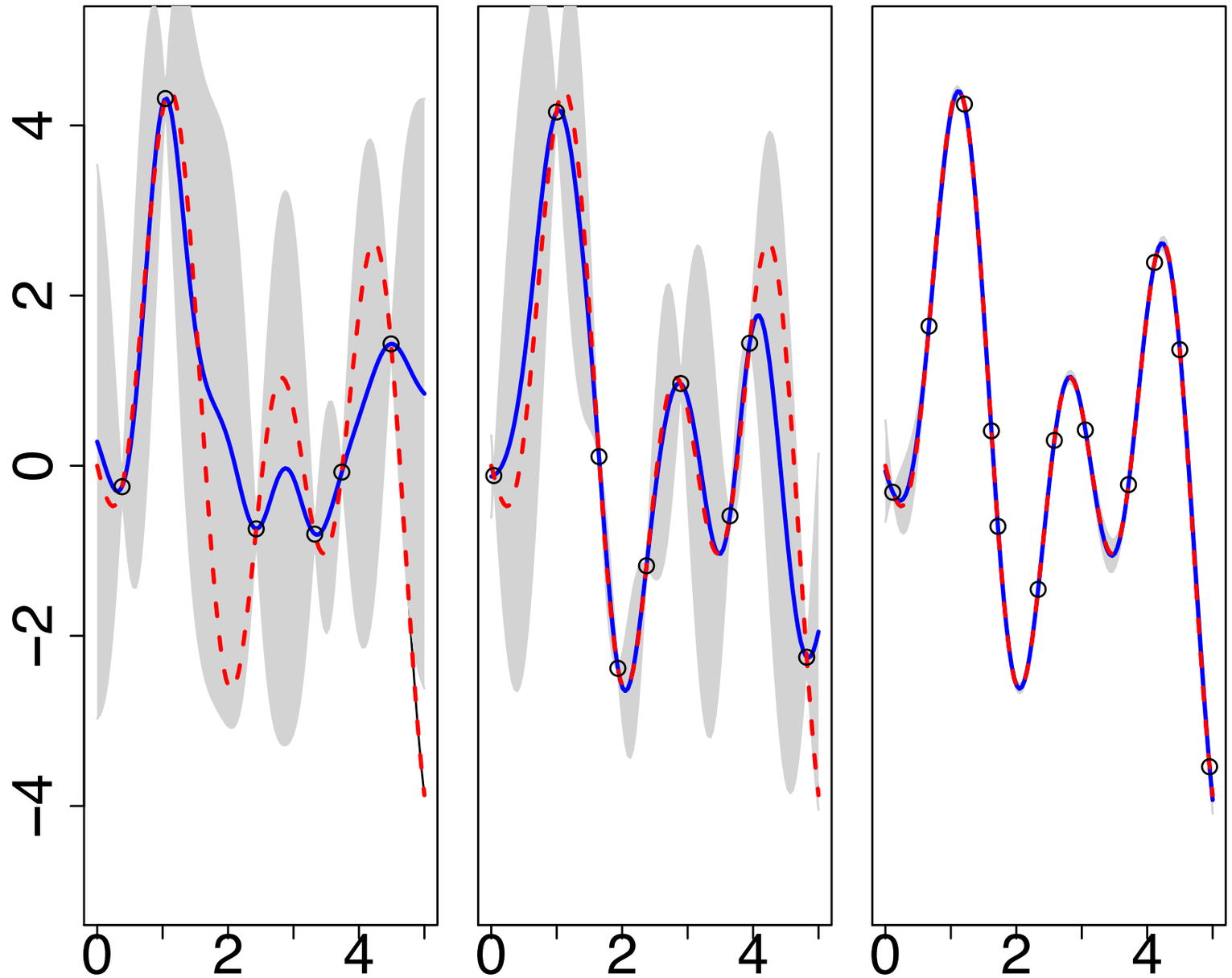
$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \\ 3 & 1 & 4 & 2 \\ 4 & 3 & 2 & 1 \end{pmatrix}$$

- Rough rule of thumb is at least 10 points per dimension.

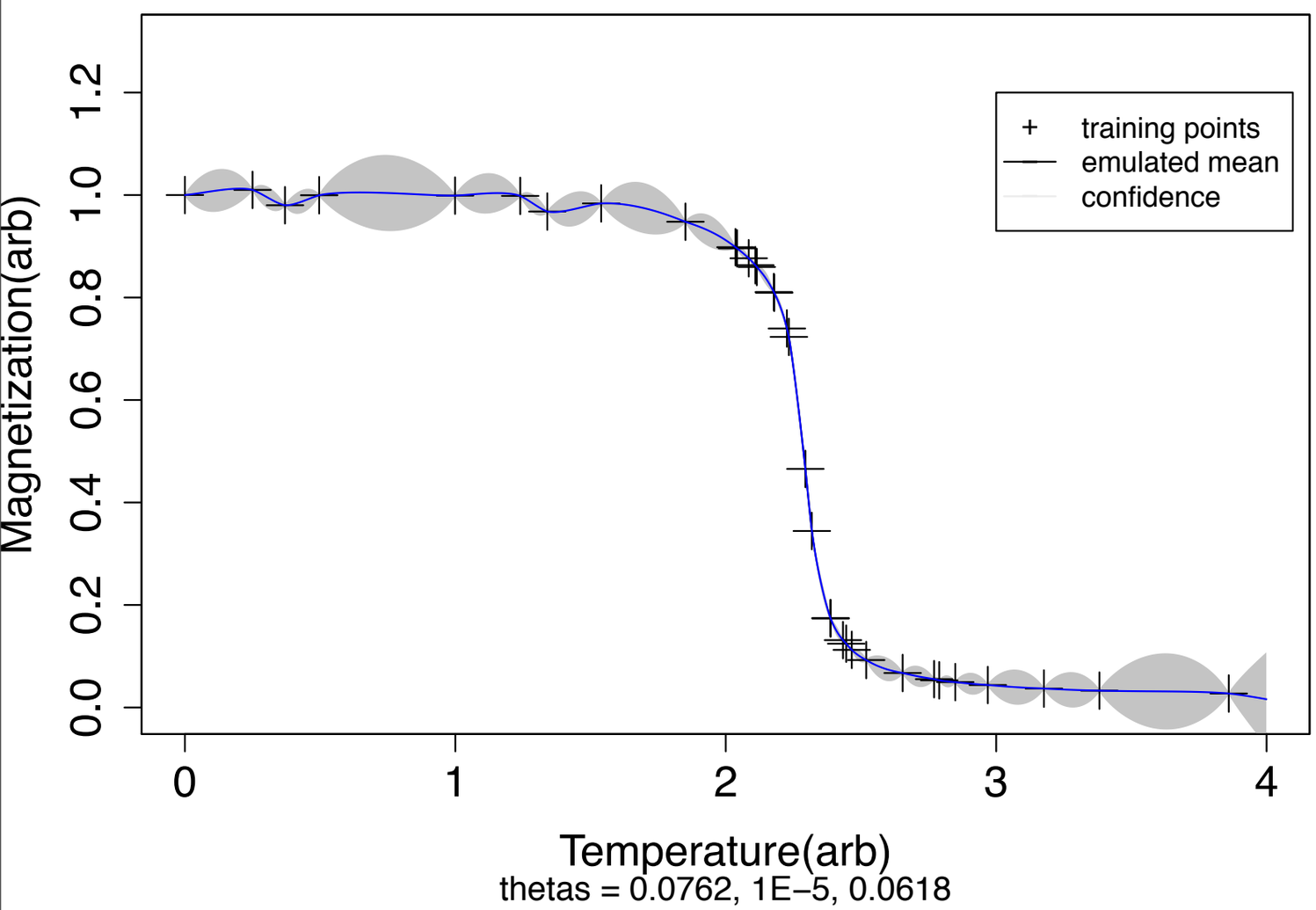
J. L. Loeppky, J. Sacks, and W. J. Welch, Technometrics **51**, 366 (2009).

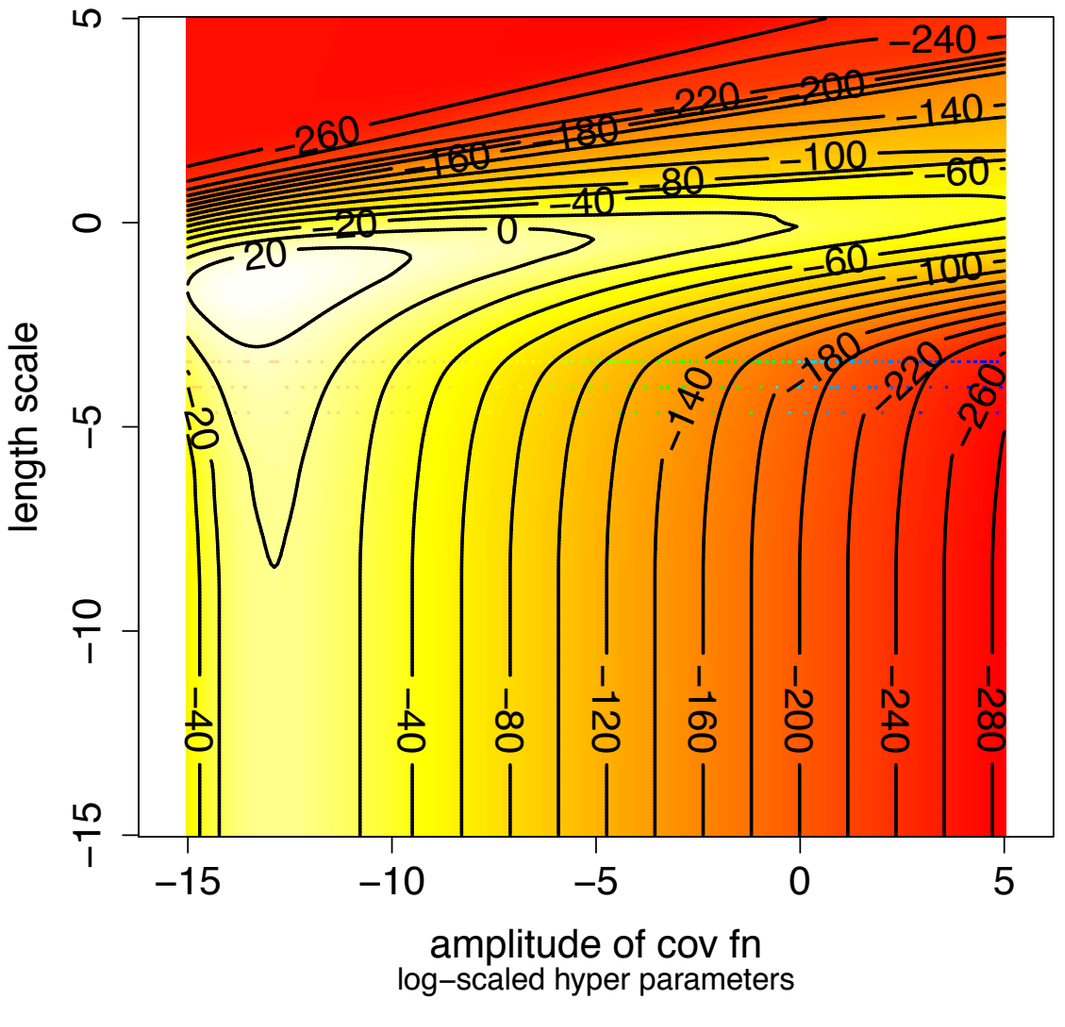
T. J. Santner, B. J. Williams, and W. Notz, *The Design and Analysis of Computer Experiments* (Springer Verlag, New York, NY, 2003).

Varying the number of design points



Toy Example - The Ising Model





- Models the interaction of spins on a lattice.
- Simple computational system, exhibits a phase transition.

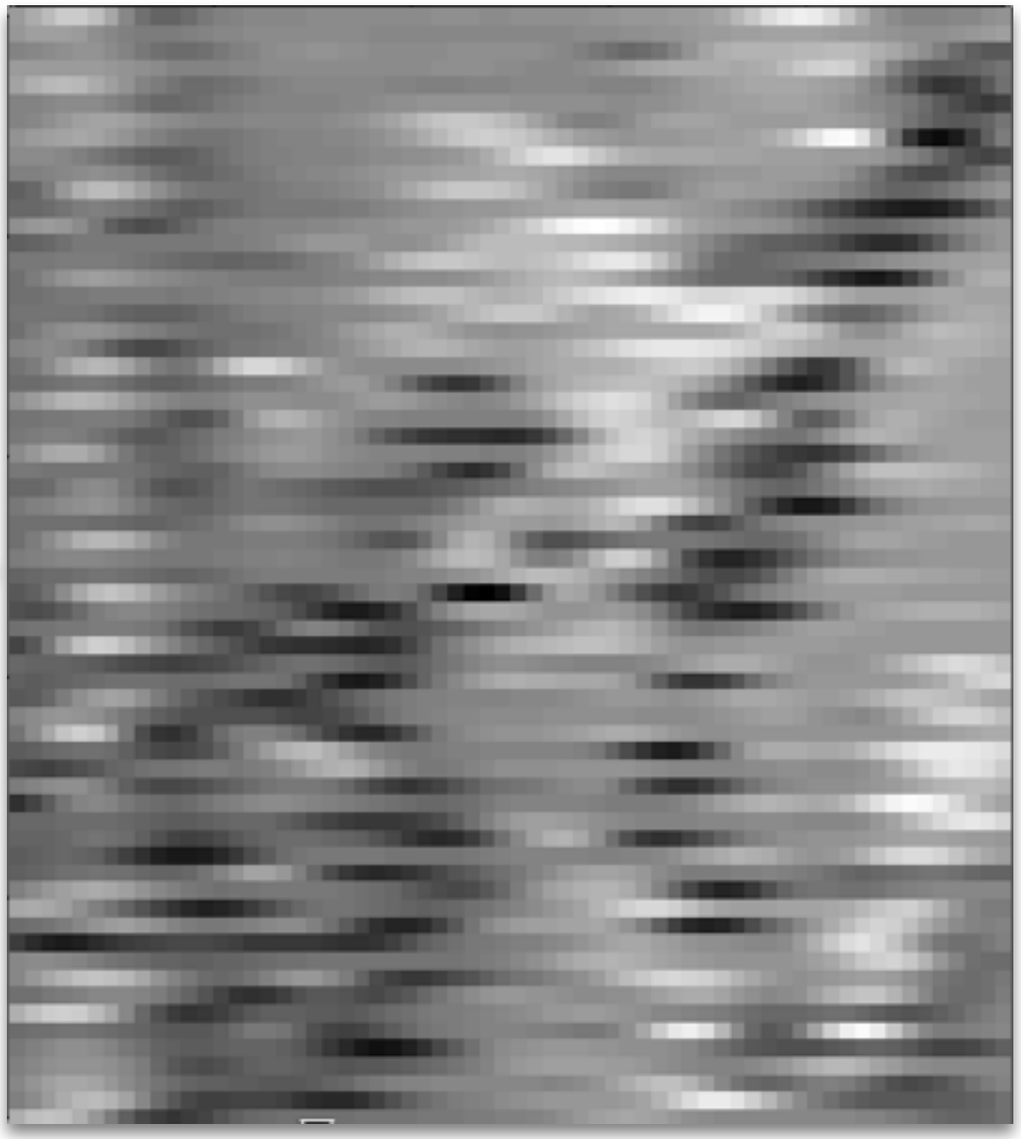
$$H = \sum_{i \neq j} J_{ij} S_i S_j$$

Exploring the limits

- How complicated a scalar function can we emulate? An image
- How many training points can we deal with?
- Limited by numeric linear algebra ~ O(N^3)

512x512 8 bit grayscale image “lenna”

1 hour



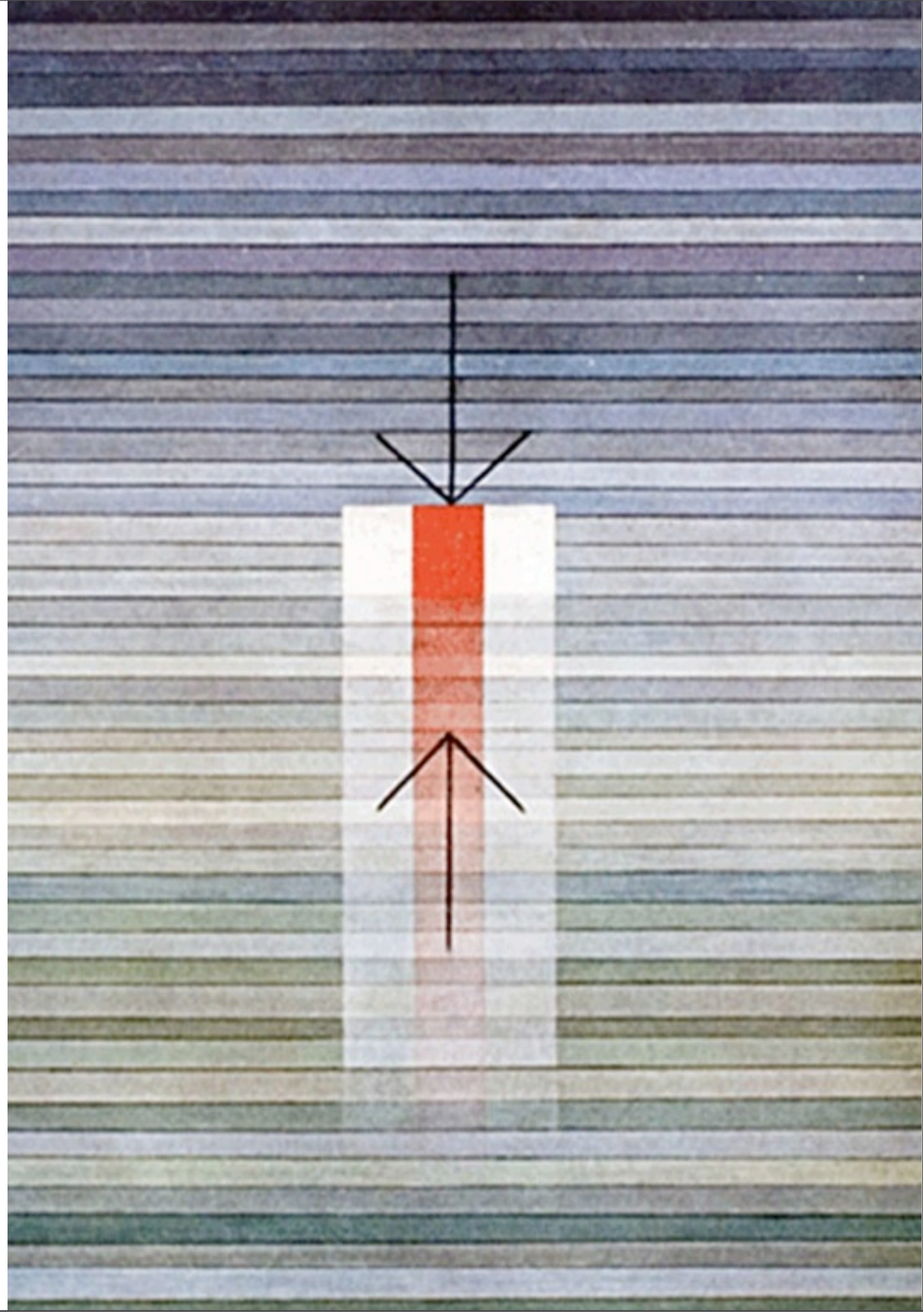
Emulator with 512 training points

GP Resources

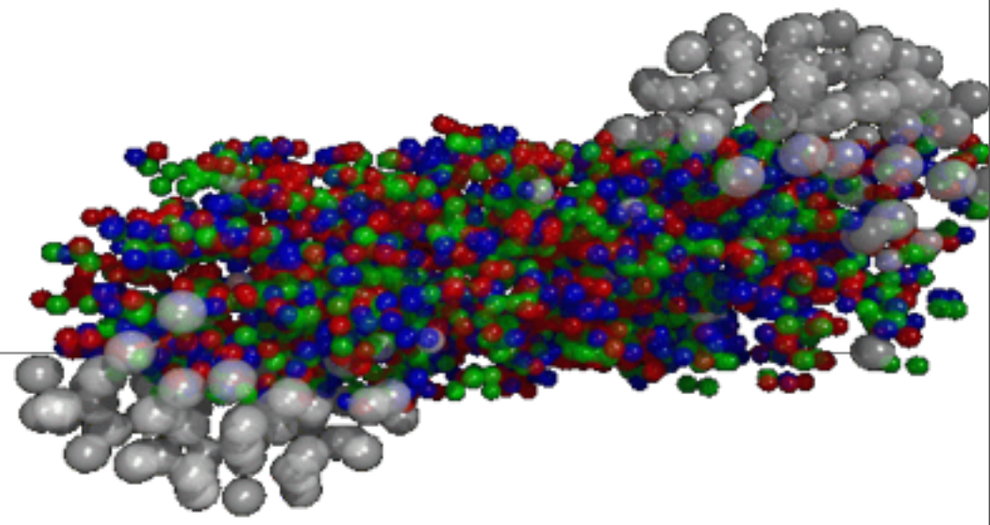
- Many R Packages on CRAN:
  - MLEGP - fairly fast very easy to use, well documented, includes sensitivity analysis
  - gausspr - GP regression from the R core team
  - tgp - Tree-GP's an approach which allows several spatial scales, think multi-scale-grids in FEM
  - plgp, gpfit, gptk, &c &c ...
  - SAVE: full package for Bayesian emulation, calibration and validation of computer models

- The book: "Gaussian Processes for Machine Learning" by C.Rasmussen, fulltext available free online.
  - GPML code for GP regression and classification in matlab (from the book)

A Simple Example: UrQMD



UrQMD - A real analysis



- Hybrid simulation of heavy ion collisions. Microscopic transport coupled to hydrodynamical model of QGP.
- Switch from hadronic transport model to hydro code at some time: T_{start}
- Particle distributions are smoothed into inputs for hydro, Gaussian smoothing width: σ
- How does code output depend on these choices?
- Typical runs ~ 3 hours for 1 event, need 100 events per design point.
- Elicited appropriate ranges, sampled a latin-hypercube design of 30 points.
  - 3 days runtime on OSG ~ 9-10 kHrs of computation

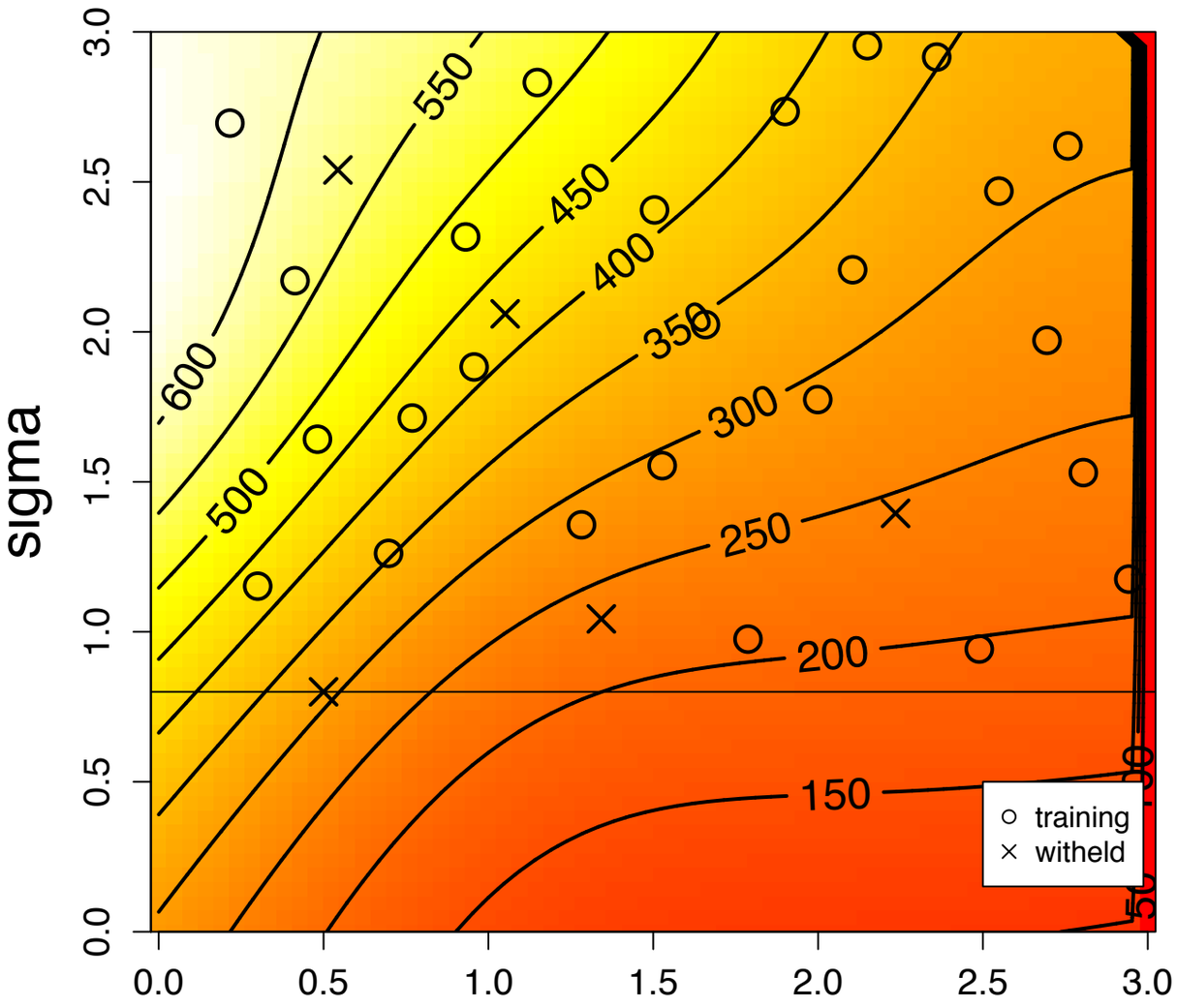
http://www.urqmd.org http://www.opensciencegrid.org Petersen.H et al. ***Phys Rev.C*** 78 (2008)

UrQMD

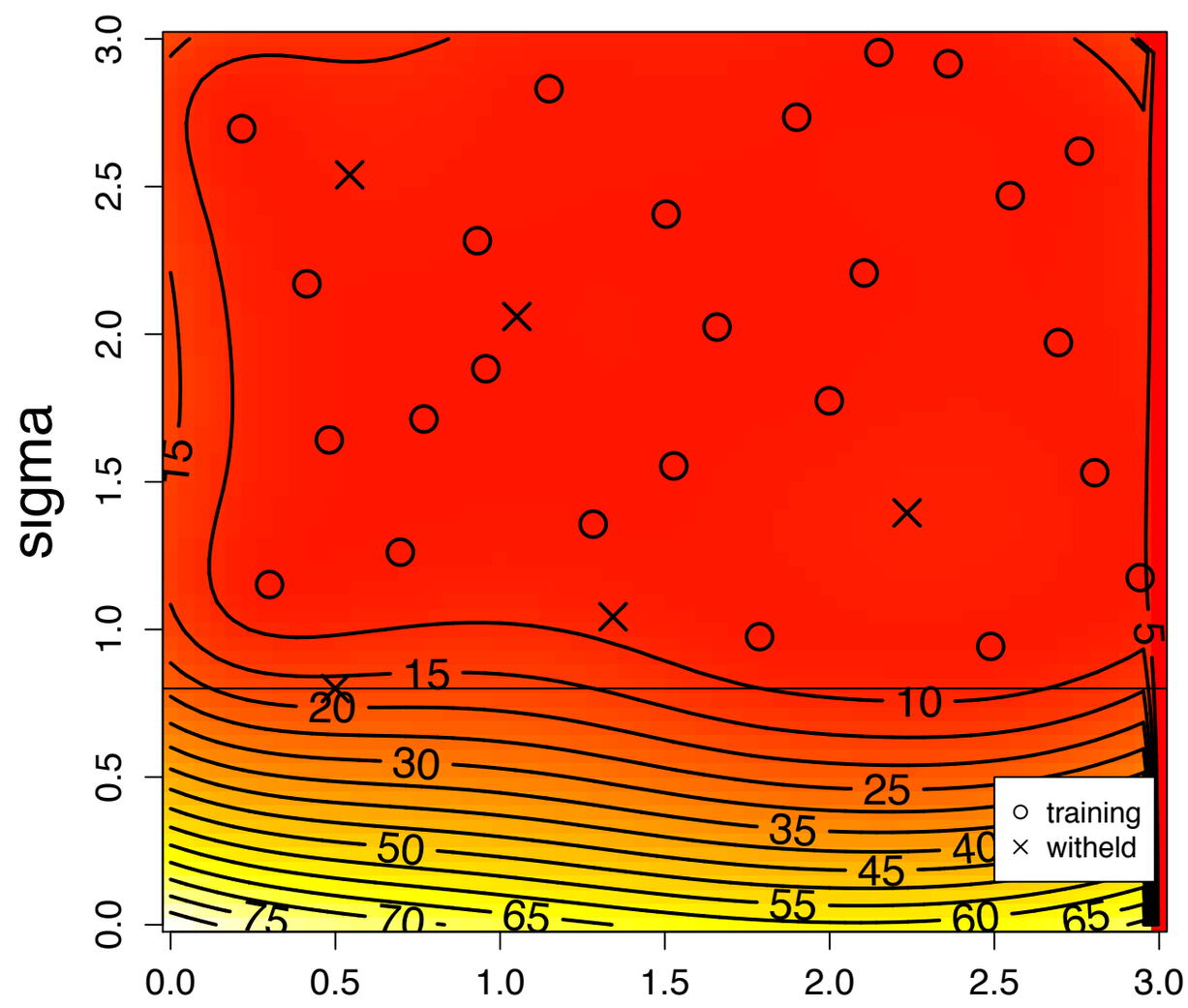
Constraining the initial state granularity with bulk observables in Au+Au collisions at $\sqrt{s_{\rm NN}} = 200$ GeV

Hannah Petersen[1], Christopher Coleman-Smith[1,2], Steffen A. Bass[1] and Robert Wolpert[2]

Emulated n(pi+) at mid rapidity



Sqrt(var)



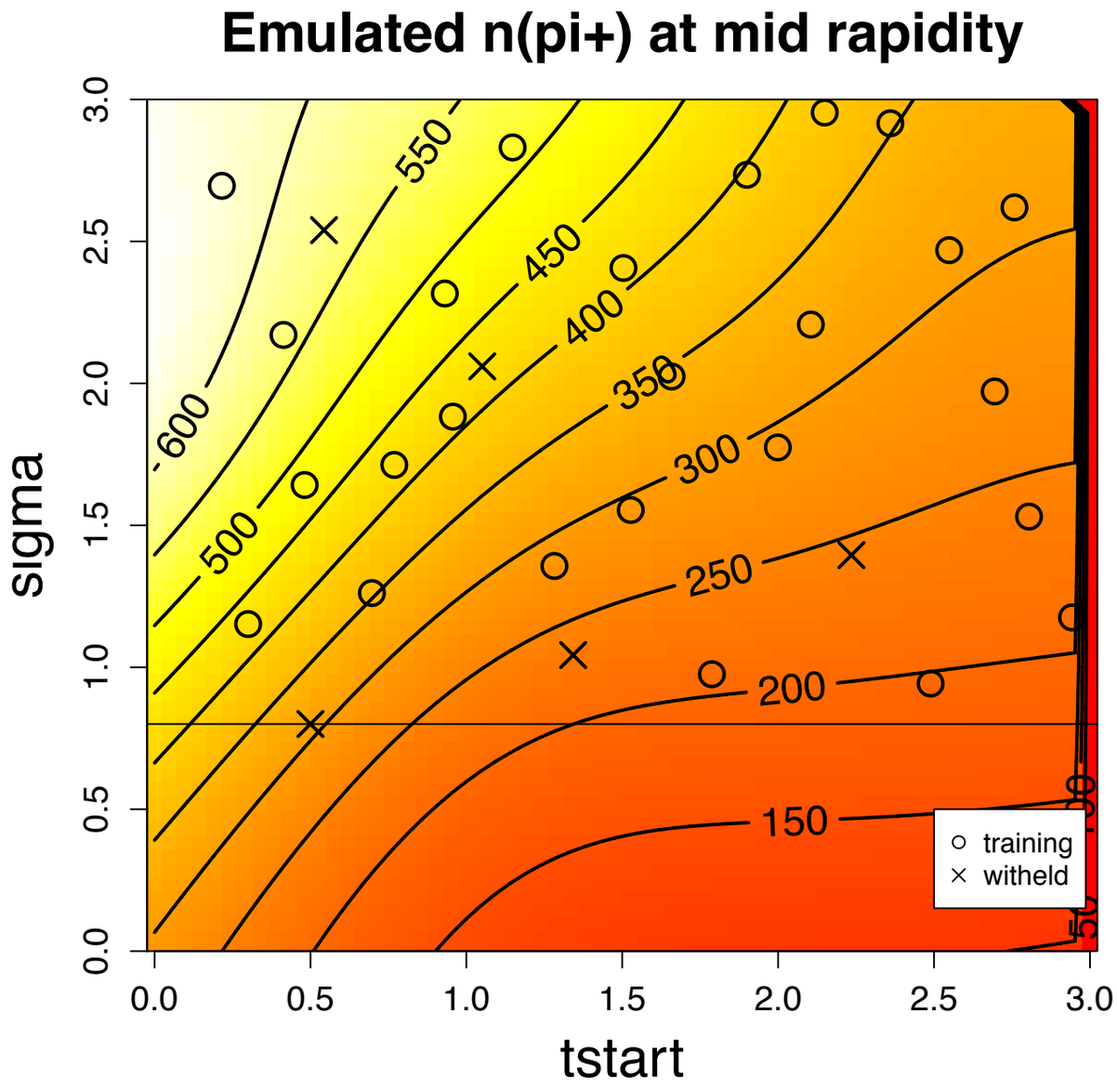
2d parameter space, scalar output.

Expect n(pi+) ~ 300 there is a large band satisfying this.

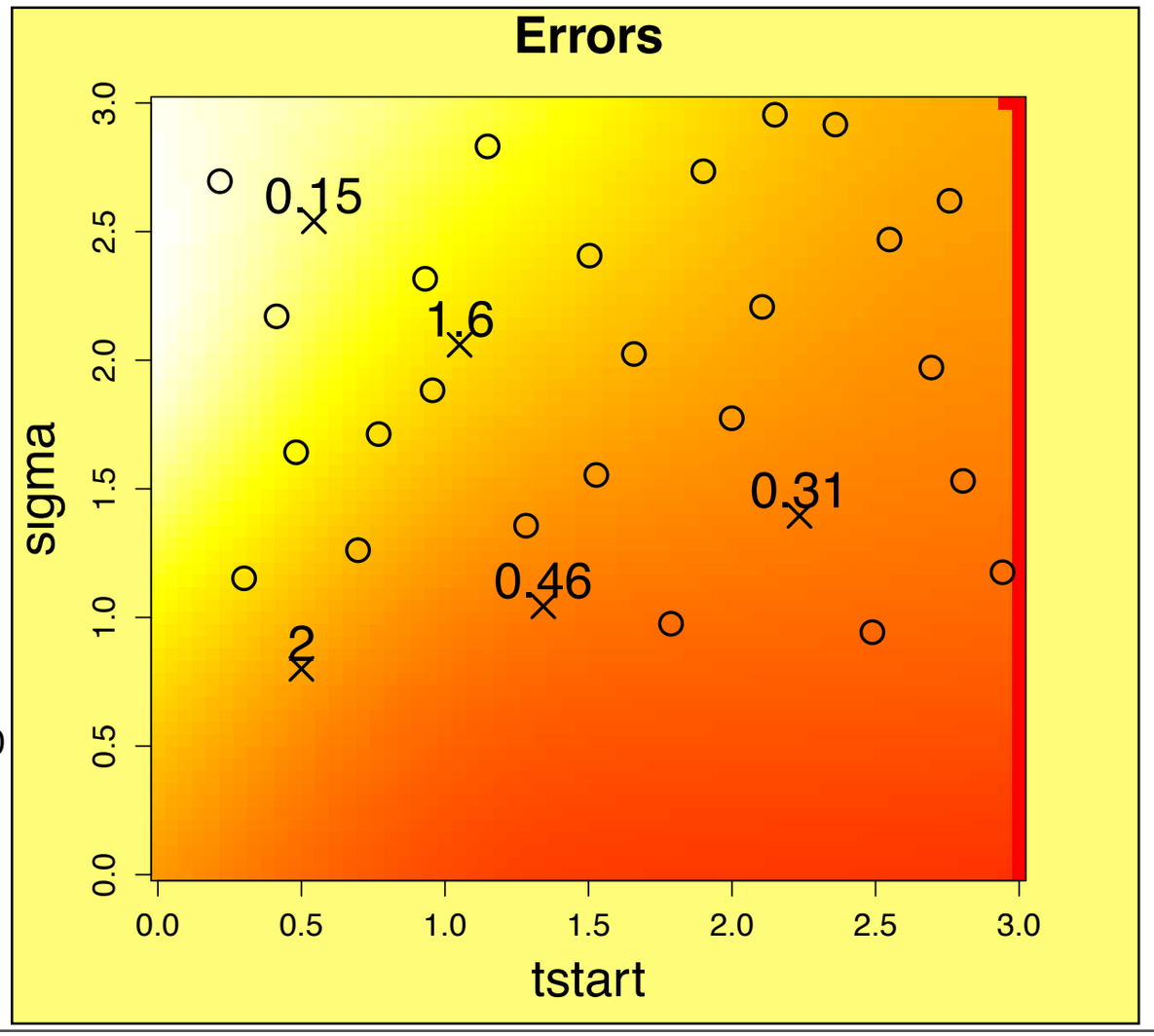
Published: Petersen et al, J.Phys.G38: 045102, (2011)

UrQMD - Validation

- Repeat analysis, withhold 5 of the 30 training points.

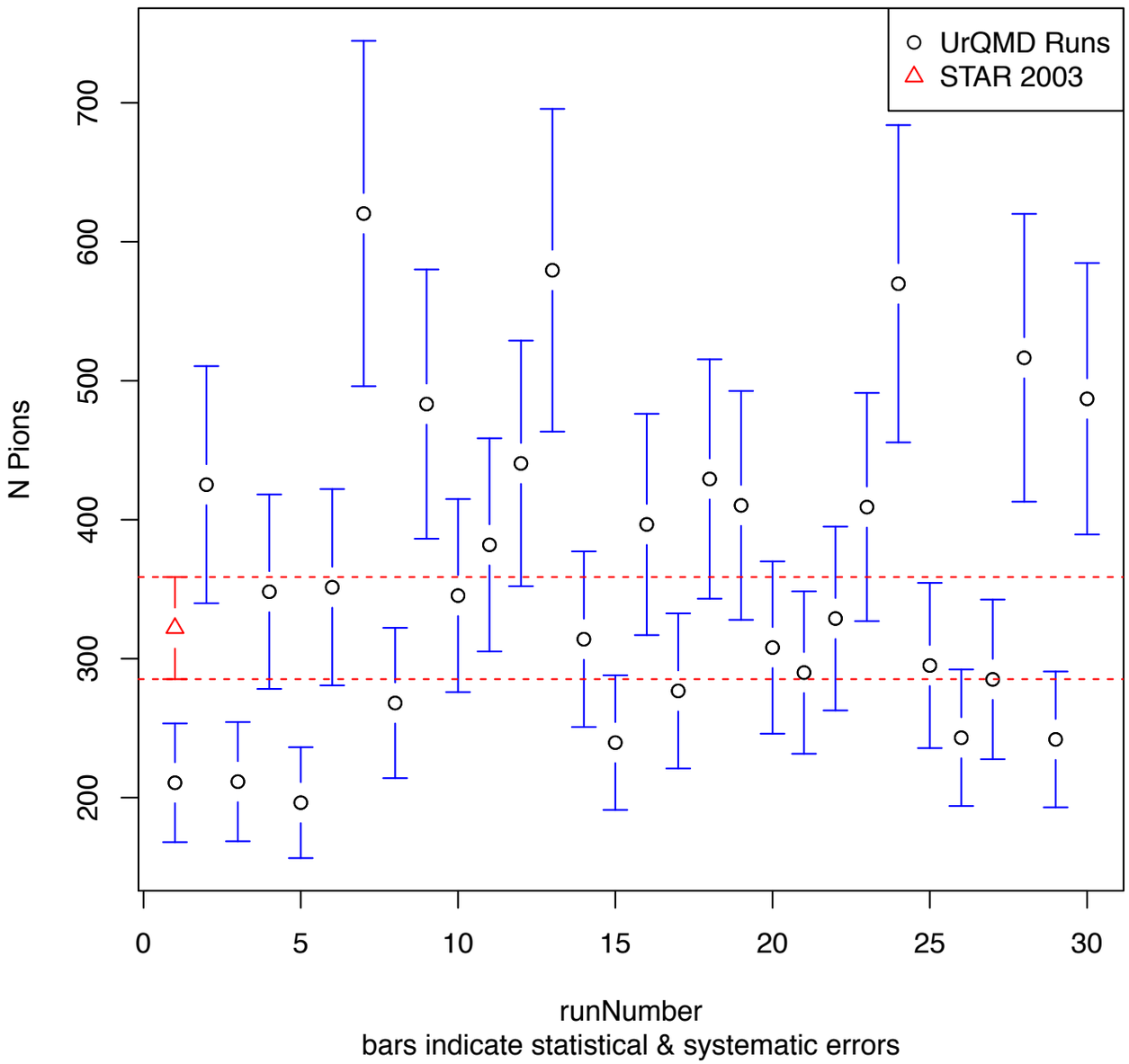


$$Err(y_i) = \frac{y_i - E(\eta(x_i)|y_i)}{\sqrt{V(\eta(x_i)|y_i)}}$$

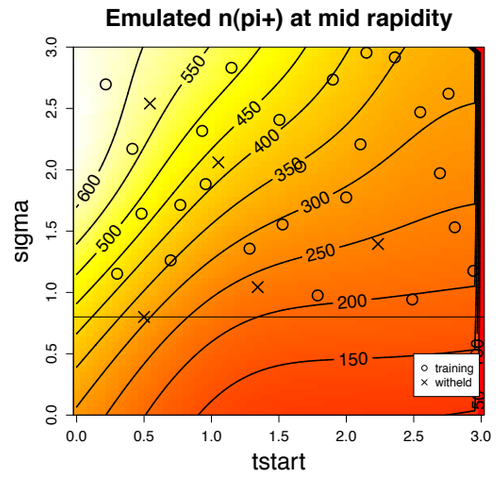


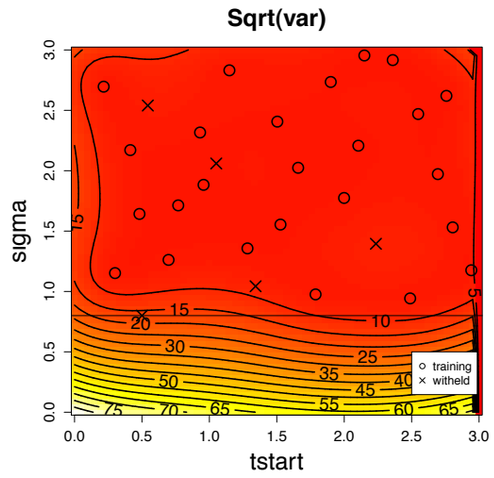
Expect deviations to be Student-*T* distributed

UrQMD Comparison with Experimental Data



Adams et al (STAR 2003), nucl-ex:0310004v1



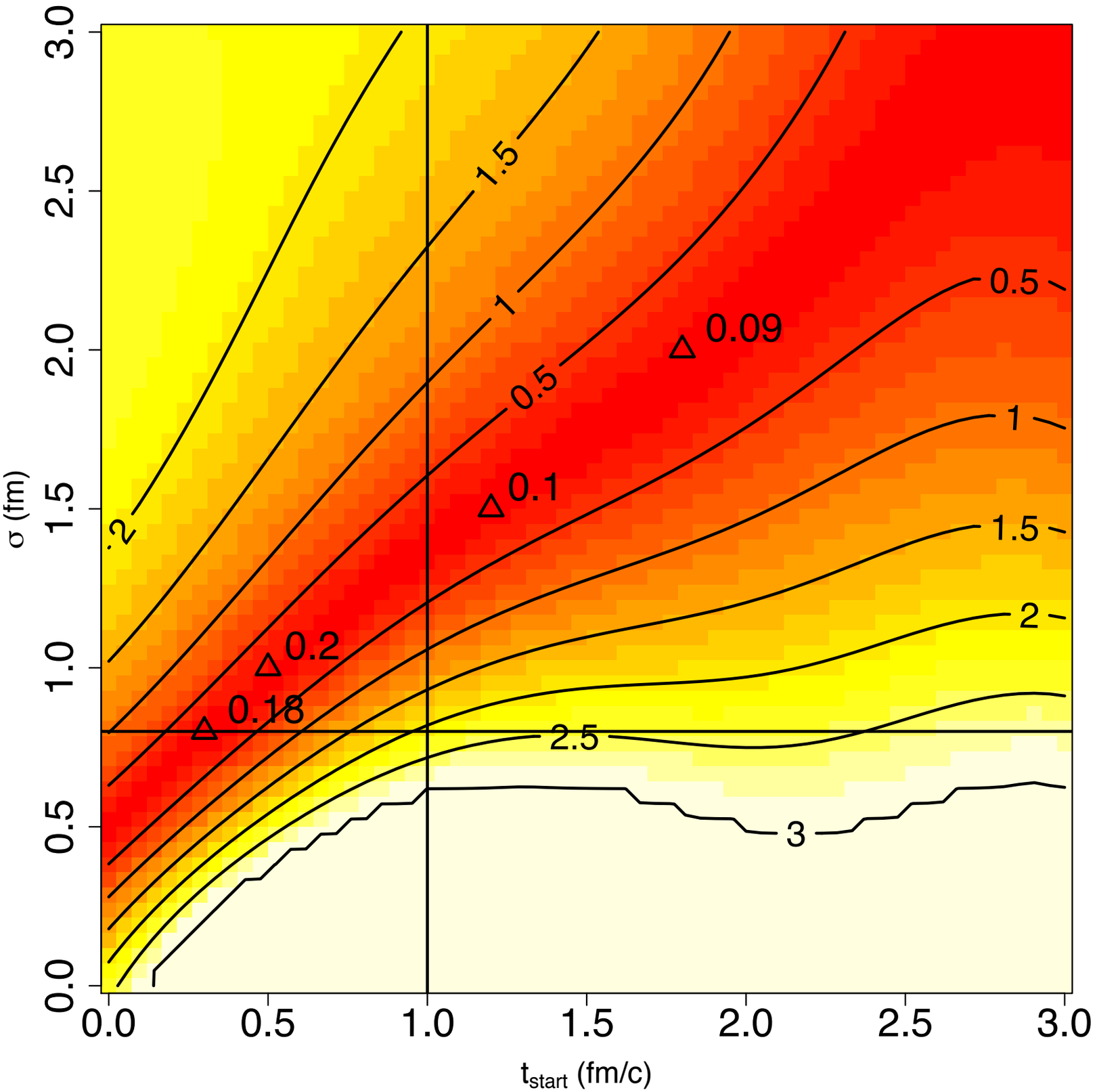


- We actually have information about the uncertainty and we should use it!
  - Systematic and Stat errors in reported data.
  - Systematic and Stat errors from the code.
  - The variance from the emulator

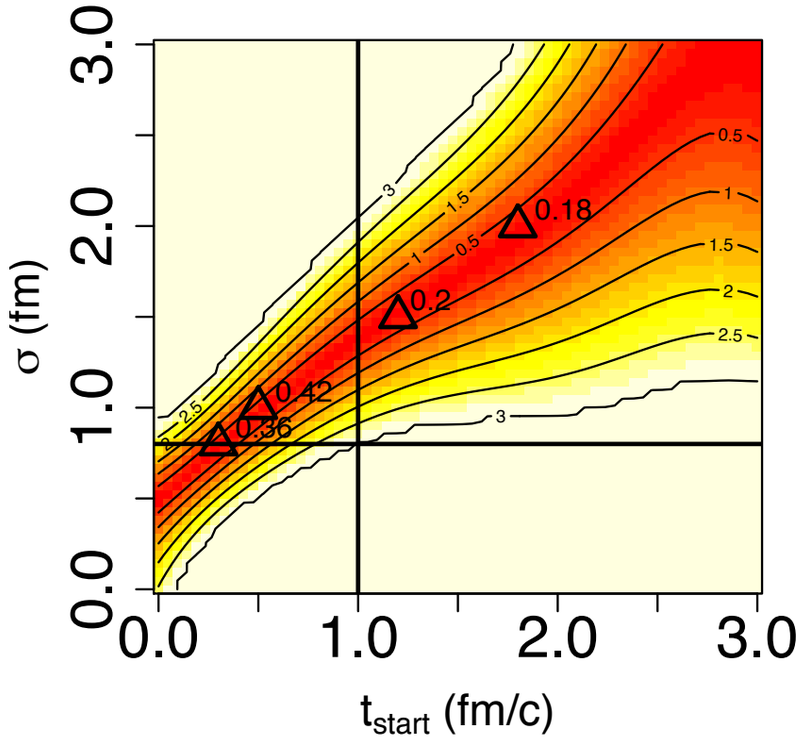
$$I^2 = \frac{\left(E[y_{emu}] - Z\right)^2}{\mathrm{Var(model)} + \mathrm{Var(data)} + \mathrm{Var(emu)}}$$

- Use all this information to slice up the parameter space into “possible” and “not-so-possible” regions

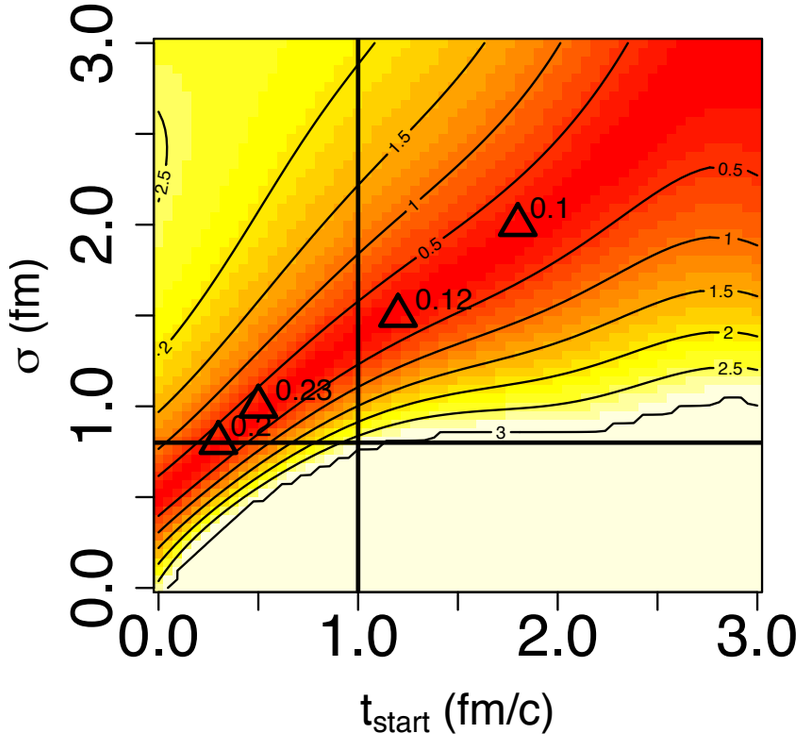
Feasibility, including all known error sources



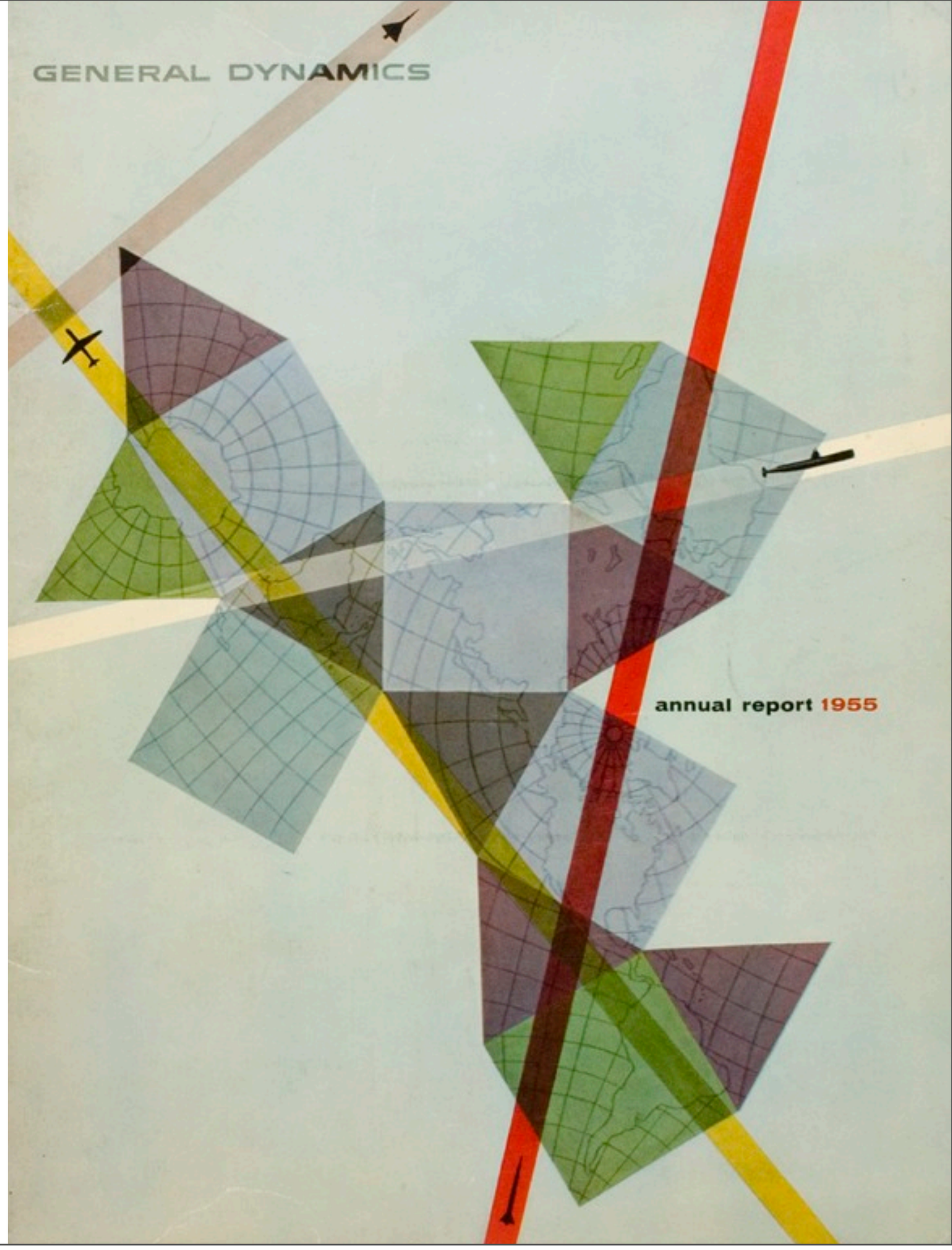
Emulator + Data Errors



Emulator + Urqmd Errors

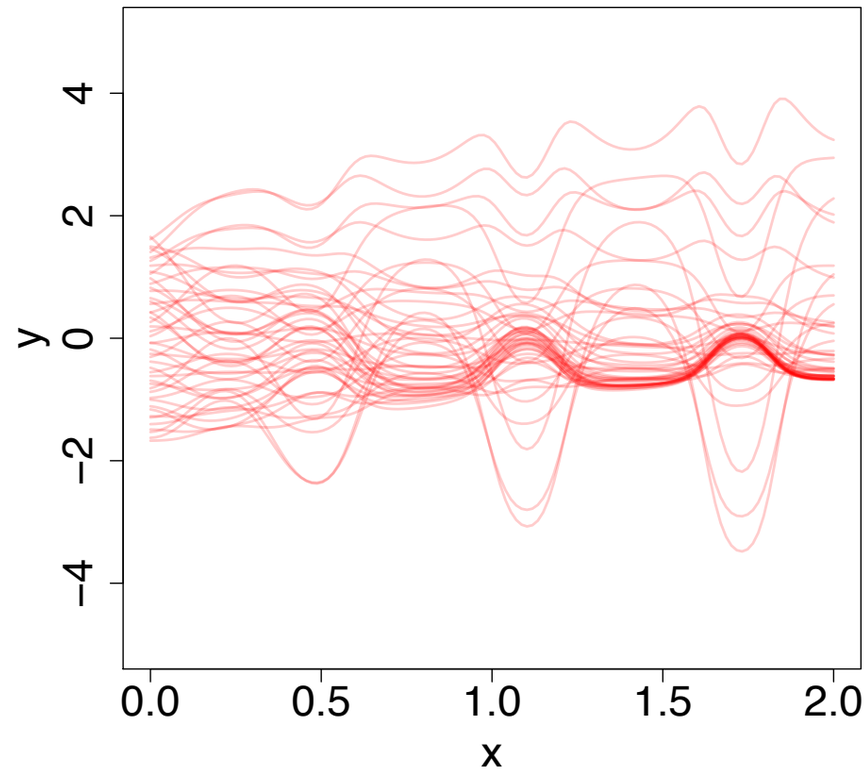


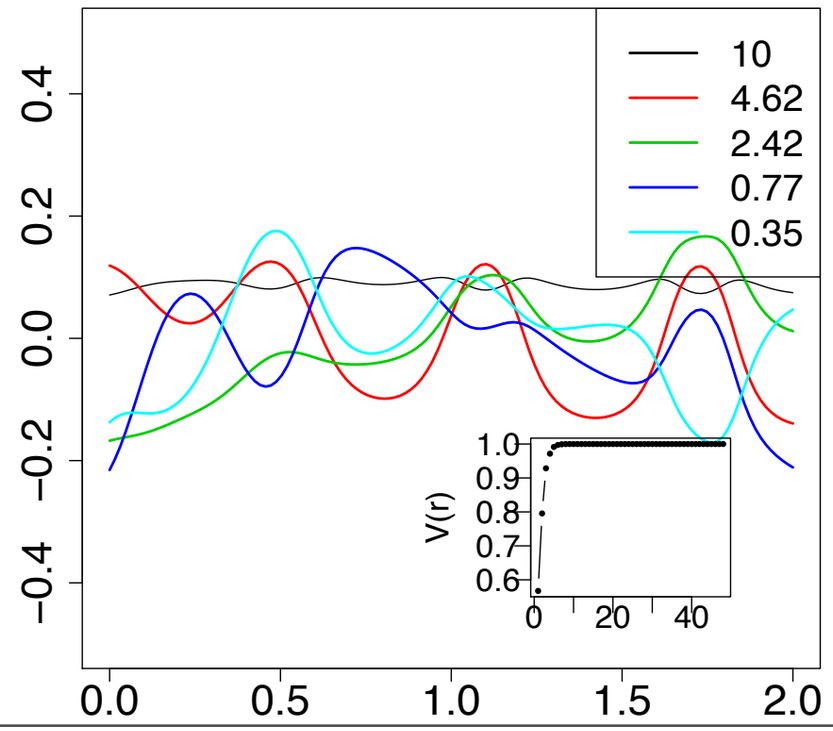
Functional Data



Functional data, multivariate output.

- What if model output is a vector?
- Emulate each component separately.
  - Not ideal for large dimension vectors
  - May lose correlation between components.
- Principle Components
- Transform data to an orthogonal basis which maximizes variance in each direction.
- Create orthogonal emulators of descending importance.





Principle Components (PCA) - An Aside

- Construct an orthogonal & independent basis from our observations.
- Suppose cpts of vectorial observations are sampled from an MVN

$$Y^T = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\} \quad \mathbf{y}_i \sim \mathrm{MVN}(\hat{\mu}, \hat{\Sigma}) \quad (\hat{\mu})_j = \frac{1}{n}\sum_i (y_i)_j$$

$$\hat{\Sigma} = E[(\mathbf{y}_i - \mu_i)^T(\mathbf{y}_i - \mu_i)]$$

  - An eigendecomposition of the sample-covariance matrix, provides this basis.

$$\hat{\Sigma} = U\Lambda U^T$$

- Eigenvectors for largest eigenvalue provides direction with largest variance.
  - Successive eigenvalues correspond to less “various” directions.
  - Only need to keep the top r eigenvalues.
- Project observations onto this basis $Z_r = \frac{1}{\sqrt{\Lambda_r}} U_r^T (Y - \hat{\mu})$
- Now construct emulators for each cpt of Z.

$$Y^\star \approx \hat{\mu} + U_r\sqrt{\Lambda_r} Z_r$$

Watch out for Outliers

- Eigendecomposition is very sensitive to outlying points
- perturbation due to outlying points

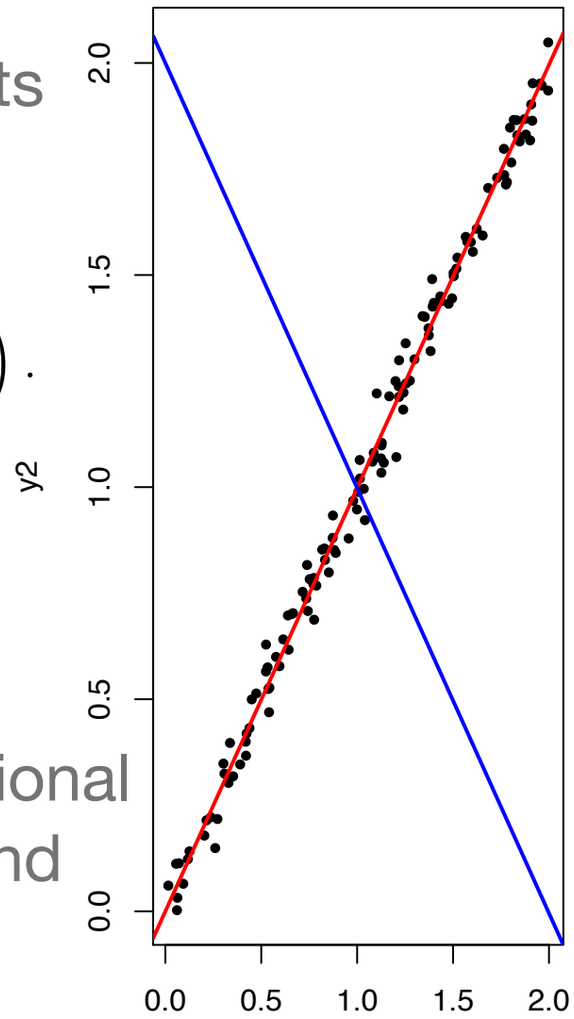
$$\hat{\Sigma}' \approx \hat{\Sigma} + \epsilon V + \mathcal{O}(\epsilon^3)$$

$$V_{\alpha,\beta} = \left(-2\hat{\mu}_\beta Y^\alpha_{\text{out}} - 2\hat{\mu}_\alpha Y^\beta_{\text{out}} + Y^\alpha_{\text{out}} Y^\beta_{\text{out}} + 3\hat{\mu}_\alpha\hat{\mu}_\beta\right).$$

$$V_{\alpha,\beta} = \left\{2\hat{\mu}_\beta\left(\hat{\mu}_\alpha - Y^\alpha_{\text{out}}\right) + 2\hat{\mu}_\alpha\left(\hat{\mu}_\beta - Y^\beta_{\text{out}}\right) + \left(Y^\alpha_{\text{out}} Y^\beta_{\text{out}} - \hat{\mu}_\alpha\hat{\mu}_\beta\right)\right\}.$$

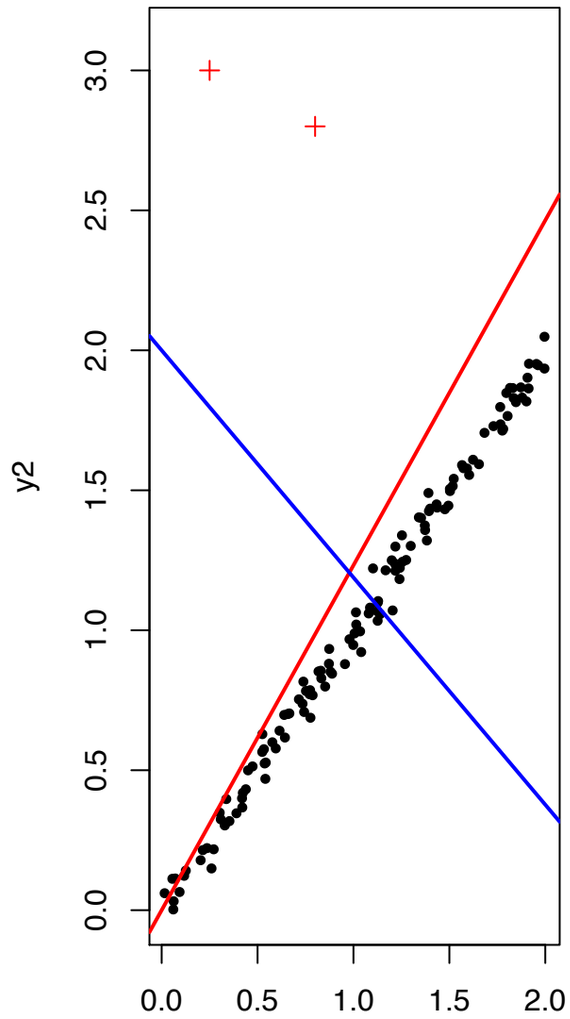
- New matrix elements are proportional to extent that points outly the trend

$$\lambda^1_j = u^{0\intercal}_j V u^0_j, \quad u^1_j = \sum_{\ell=1,\ell\neq j}^{k} \frac{u^{0\intercal}_\ell V u^0_j}{\lambda^0_j - \lambda^0_\ell},$$



$$\lambda_1 = 0.998$$

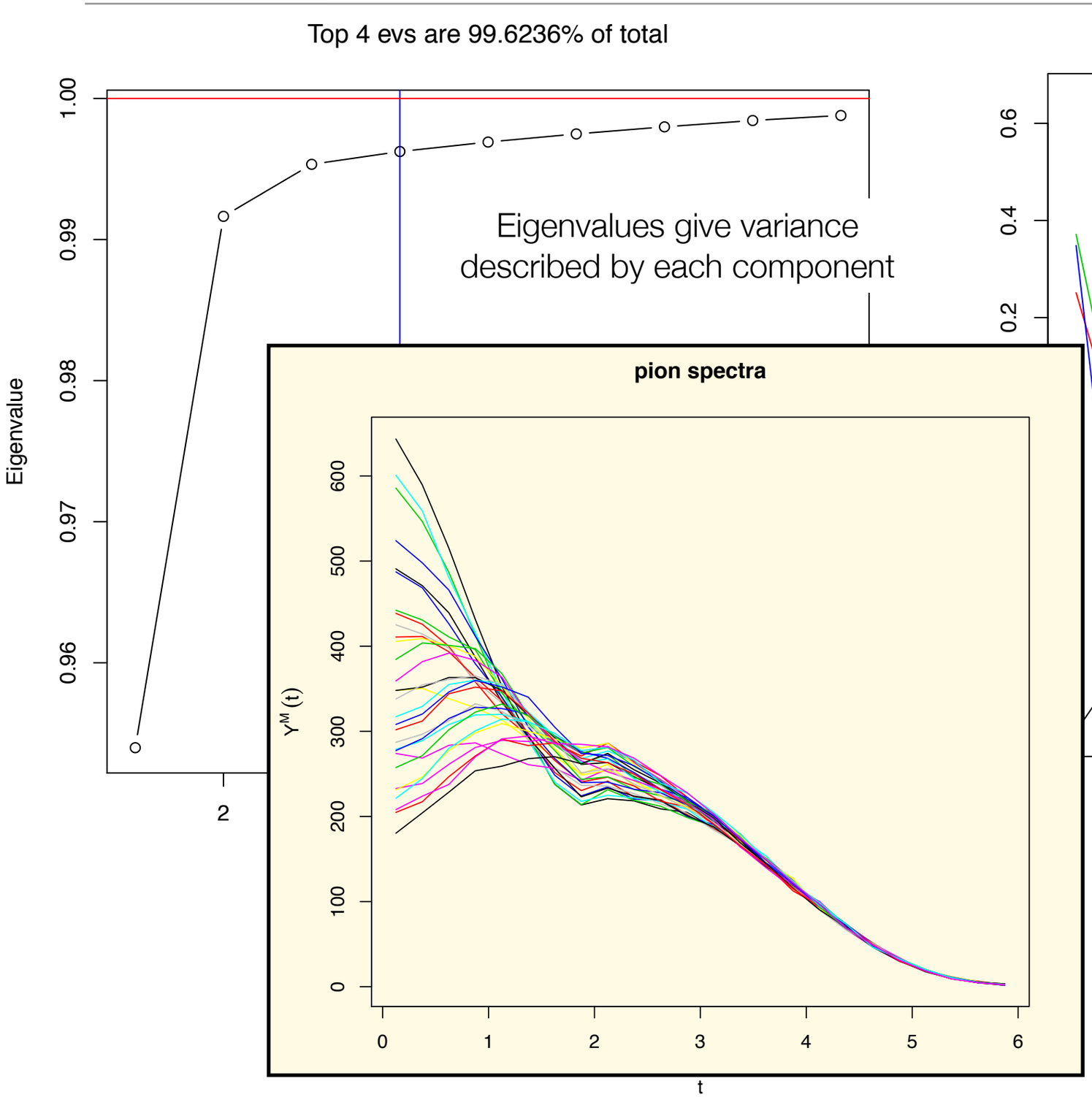
$$\lambda_2 = 0.0018$$

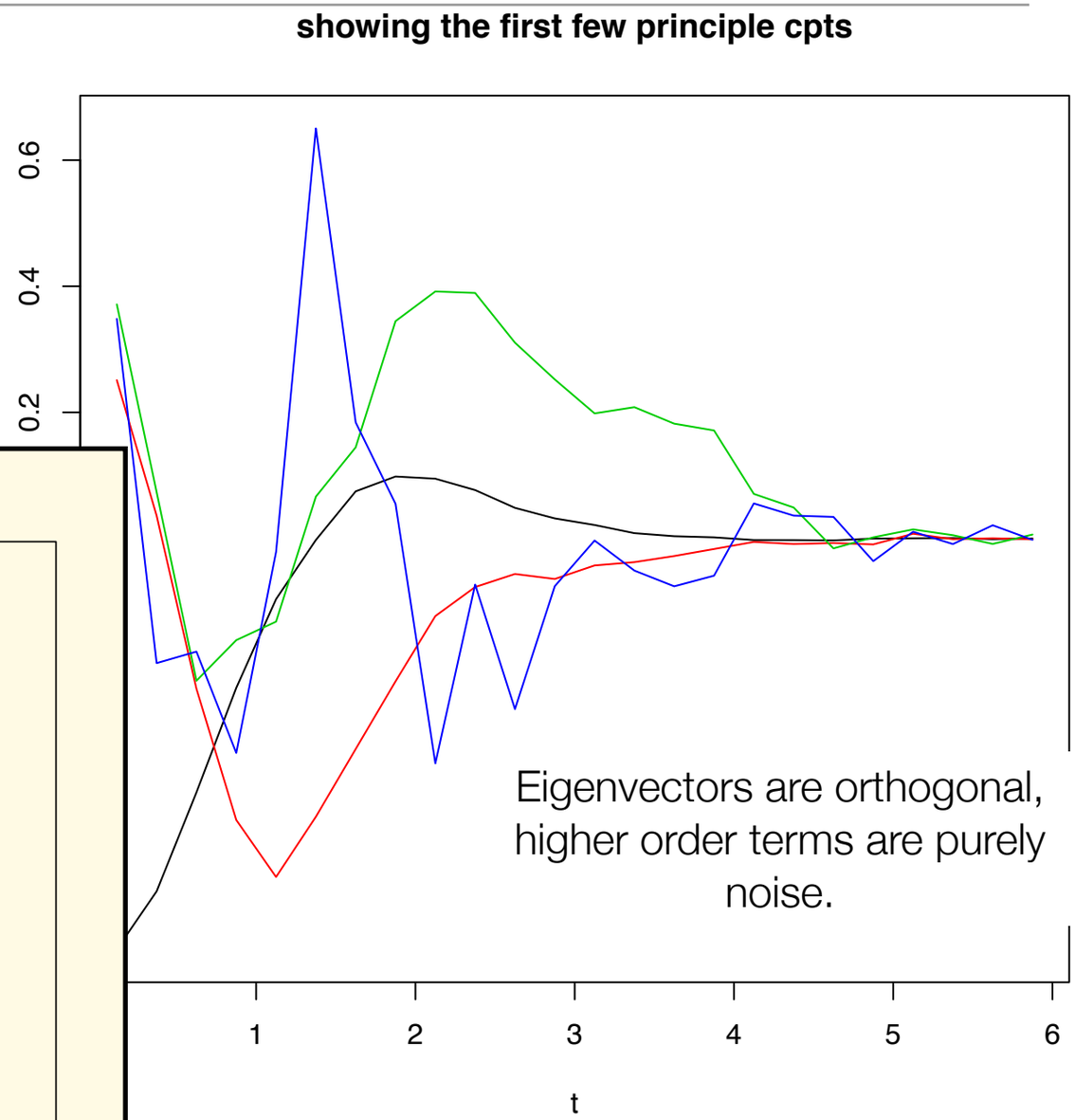


$$\lambda_1 = 0.904$$

$$\lambda_2 = 0.095$$

Pion Spectra - PCA





Eigenvectors are orthogonal, higher order terms are purely noise.

Centering and scaling the data improves emulator and PCA performance.

Principle Components - Joint Implausibility

- Emulators are constructed P.C space, independent.

- Mutual covariance matrix is therefore diagonal. Can predict variance at any new location (diagonal matrix).

- Rotate this cov-matrix back to ‘observable space’. We can predict the full covariance matrix at any location in the parameter space.

$$\mathcal{K}_{lj}(\boldsymbol{x}_i) = \mathrm{Cov}[y_l(\boldsymbol{x}_i), y_j(\boldsymbol{x}_i)] \approx \sum_{\alpha,\beta,\gamma=1}^{t} U_{l\alpha}\Lambda^{1/2}_{\alpha\beta}U_{j\gamma}\Lambda^{1/2}_{\gamma\beta}\mathrm{Var}[Z_\beta(\boldsymbol{x}_i)],$$

- Use this to construct the joint implausibility.

$$J^2(\mathbf{x}) = \left(E[Y_f] - \hat{m}(\mathbf{x})\right)\left(\mathcal{K}_{ij}(\mathbf{x}) + V[Y_f]\right)^{-1}\left(E[Y_f] - \hat{m}(\mathbf{x})\right)$$

- This gives a *just-smart-enough* measure of how the model deviates from the field data

Galaxy Formation ChemTreeN

With F.Gomez (MSU), B.O'Shea (MSU), R.Wolpert

ChemtreeN

- A galaxy formation model, captures the interplay of gravitational (darkmatter) and nuclear physics (star formation life-cycle etc)
- A hybrid code:
  - First: darkmatter only simulations model bulk gravitational structure. This is an N-body tree code. Very Slow. Supercomputer Slow.
    - A discrete set of differential equations models the baryonic evolution. Fastish (relative to N-body)
- Goal is to understand the relative importance of the baryonic and darkmatter evolutions.
- Attempt this by exploring the influence of varying the parameters in the baryonic sector combined with using a small set of N-body histories.

z = 48.4

T = 0.05 Gyr

The darkmatter component

Merger trees

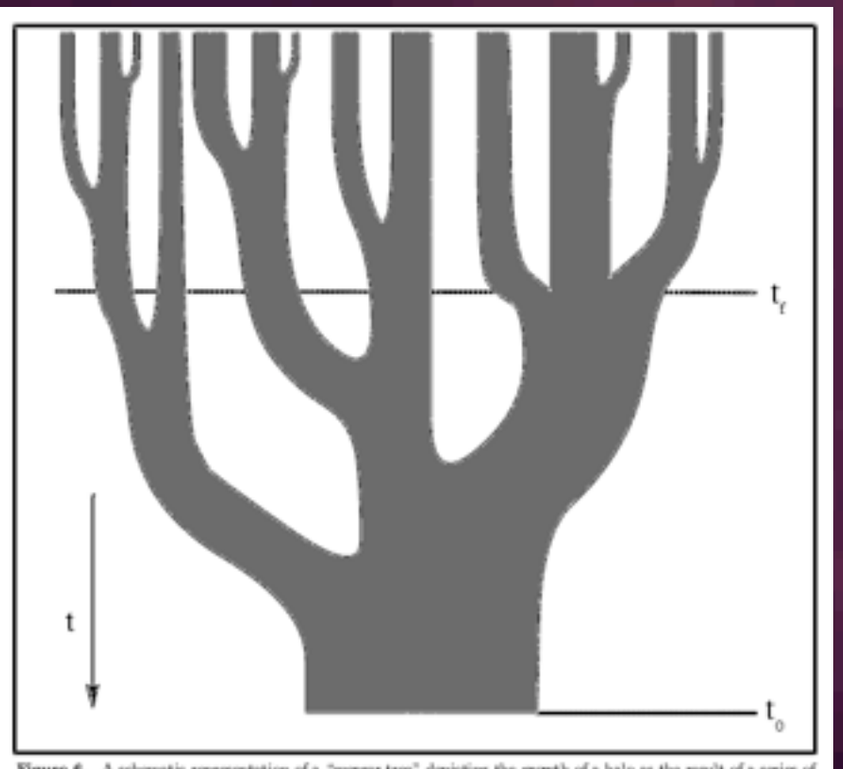


Figure 6. A schematic representation of a "merger tree" depicting the growth of a halo as the result of a series of mergers. Time increases from top to bottom in this figure and the widths of the branches of the tree represent the masses of the individual parent halos. Slicing through the tree horizontally gives the distribution of masses in the parent halos at a given time. The present time t_0 and the formation time t_f are marked by horizontal lines, where the formation time is defined as the time at which a parent halo containing in excess of half of the mass of the final halo was first created.

500 kpc

z = 36.0

T = 0.08 Gyr

The darkmatter component

Merger trees

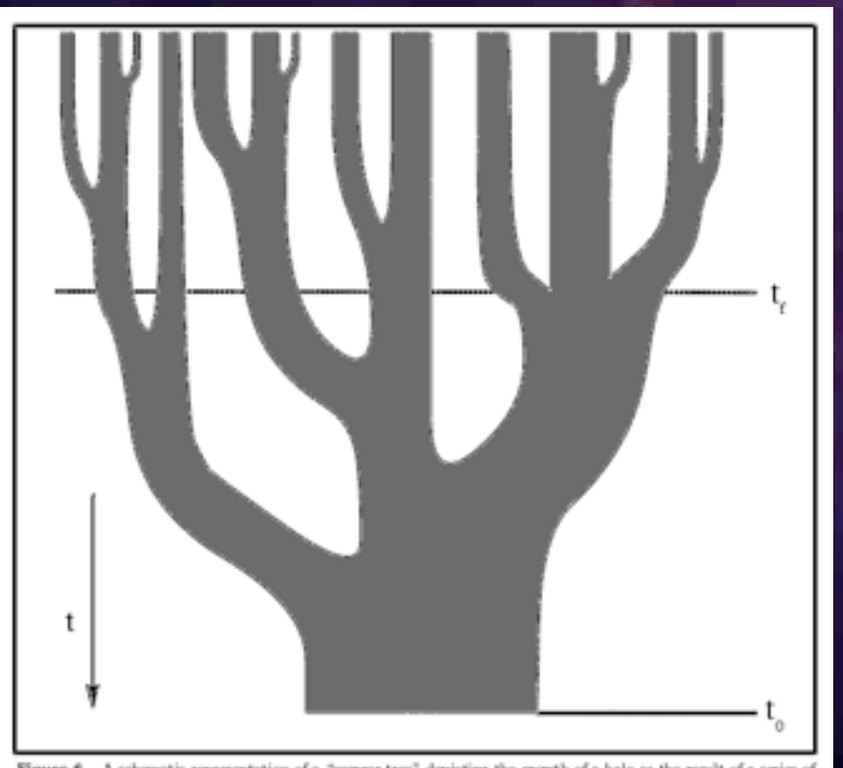


Figure 6. A schematic representation of a "merger tree" depicting the growth of a halo as the result of a series of mergers. Time increases from top to bottom in this figure and the widths of the branches of the tree represent the masses of the individual parent halos. Slicing through the tree horizontally gives the distribution of masses in the parent halos at a given time. The present time t_0 and the formation time t_f are marked by horizontal lines, where the formation time is defined as the time at which a parent halo containing in excess of half of the mass of the final halo was first created.

500 kpc

z = 30.8

T = 0.10 Gyr

The darkmatter component

Merger trees

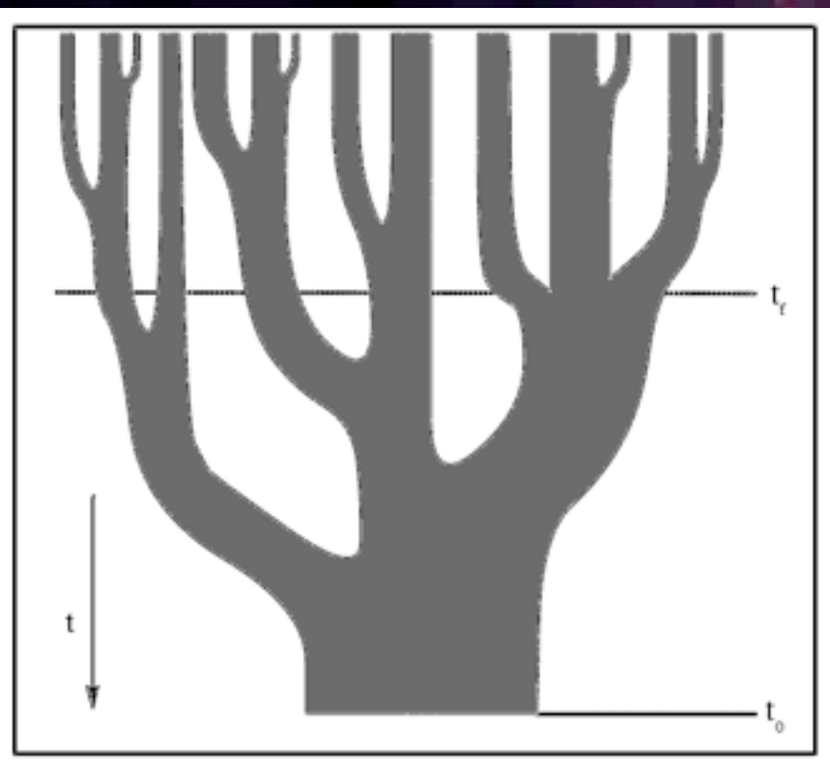


Figure 6. A schematic representation of a "merger tree" depicting the growth of a halo as the result of a series of mergers. Time increases from top to bottom in this figure and the widths of the branches of the tree represent the masses of the individual parent halos. Slicing through the tree horizontally gives the distribution of masses in the parent halos at a given time. The present time t_0 and the formation time t_f are marked by horizontal lines, where the formation time is defined as the time at which a parent halo containing in excess of half of the mass of the final halo was first created.

500 kpc

z = 24.3

T = 0.14 Gyr

The darkmatter component

Merger trees

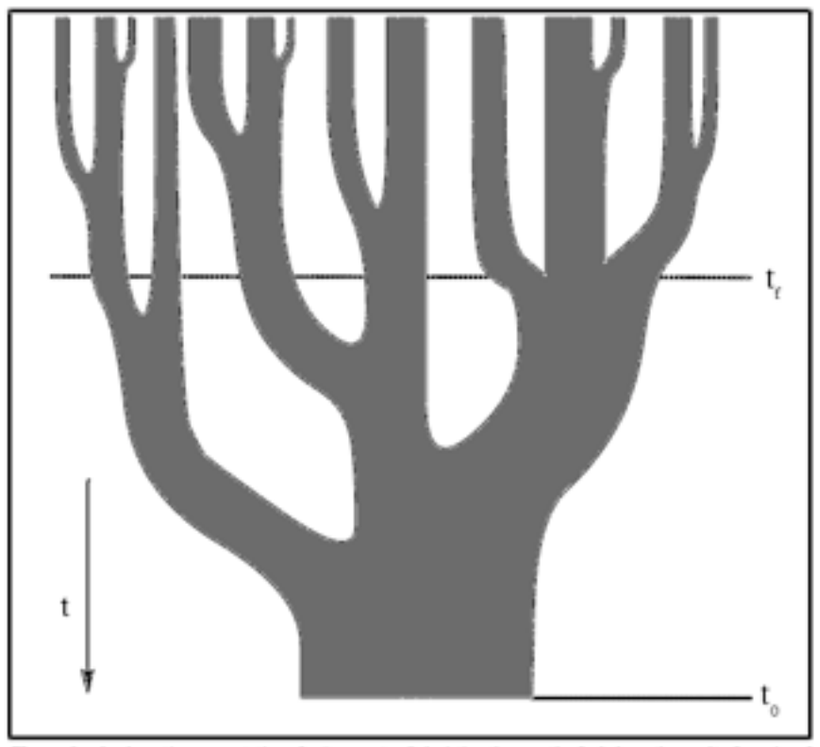


Figure 6. A schematic representation of a "merger tree" depicting the growth of a halo as the result of a series of mergers. Time increases from top to bottom in this figure and the widths of the branches of the tree represent the masses of the individual parent halos. Slicing through the tree horizontally gives the distribution of masses in the parent halos at a given time. The present time t_0 and the formation time t_f are marked by horizontal lines, where the formation time is defined as the time at which a parent halo containing in excess of half of the mass of the final halo was first created.

500 kpc

z = 20.3

T = 0.18 Gyr

The darkmatter component

Merger trees

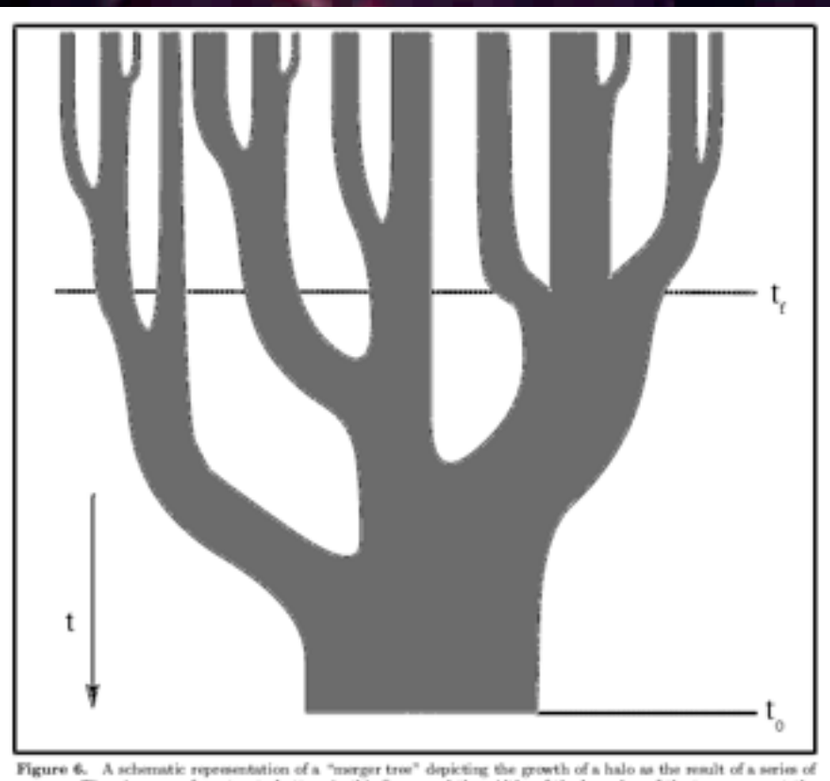


Figure 6. A schematic representation of a "merger tree" depicting the growth of a halo as the result of a series of mergers. Time increases from top to bottom in this figure and the widths of the branches of the tree represent the masses of the individual parent halos. Slicing through the tree horizontally gives the distribution of masses in the parent halos at a given time. The present time t_0 and the formation time t_f are marked by horizontal lines, where the formation time is defined as the time at which a parent halo containing in excess of half of the mass of the final halo was first created.

500 kpc

z = 18.2

T = 0.21 Gyr

The darkmatter component

Merger trees

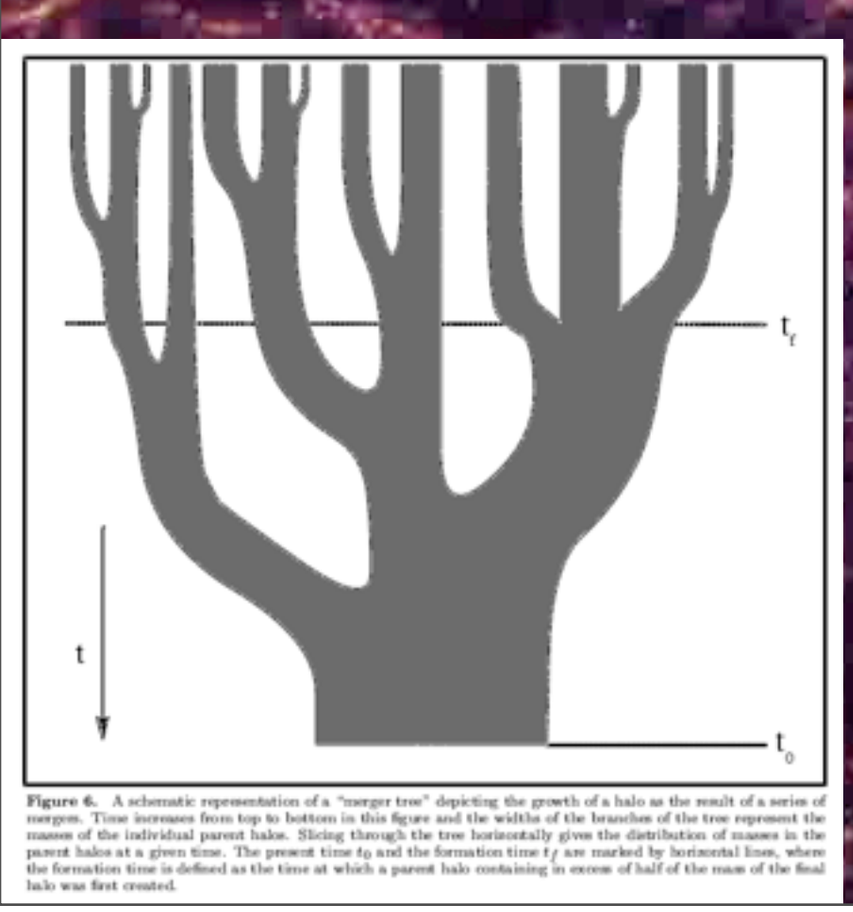


Figure 6. A schematic representation of a "merger tree" depicting the growth of a halo as the result of a series of mergers. Time increases from top to bottom in this figure and the widths of the branches of the tree represent the masses of the individual parent halos. Slicing through the tree horizontally gives the distribution of masses in the parent halos at a given time. The present time t_0 and the formation time t_f are marked by horizontal lines, where the formation time is defined as the time at which a parent halo containing in excess of half of the mass of the final halo was first created.

500 kpc

z = 15.0

T = 0.28 Gyr

The darkmatter component

Merger trees

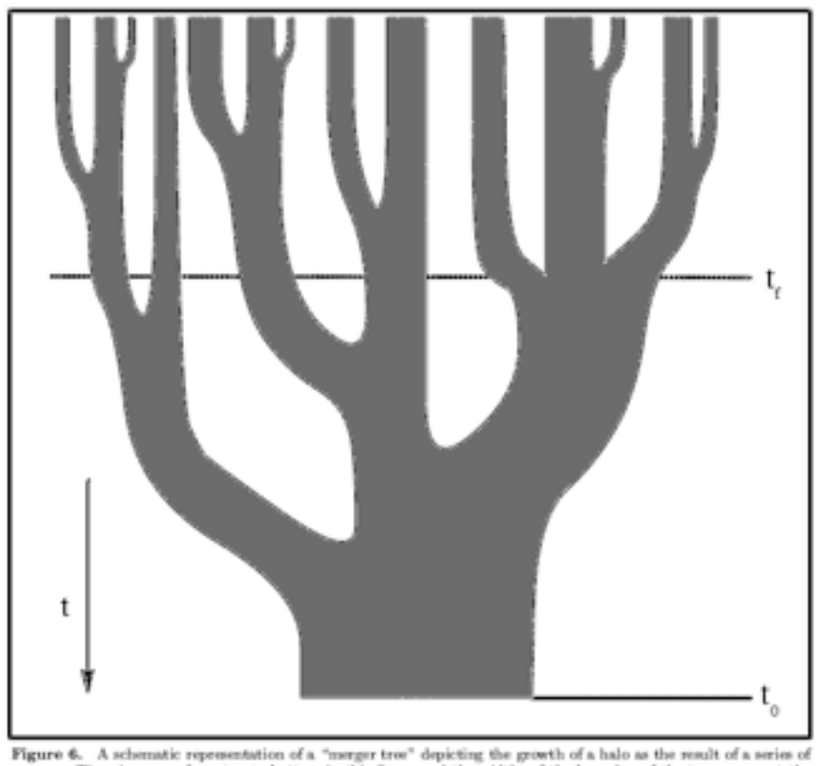
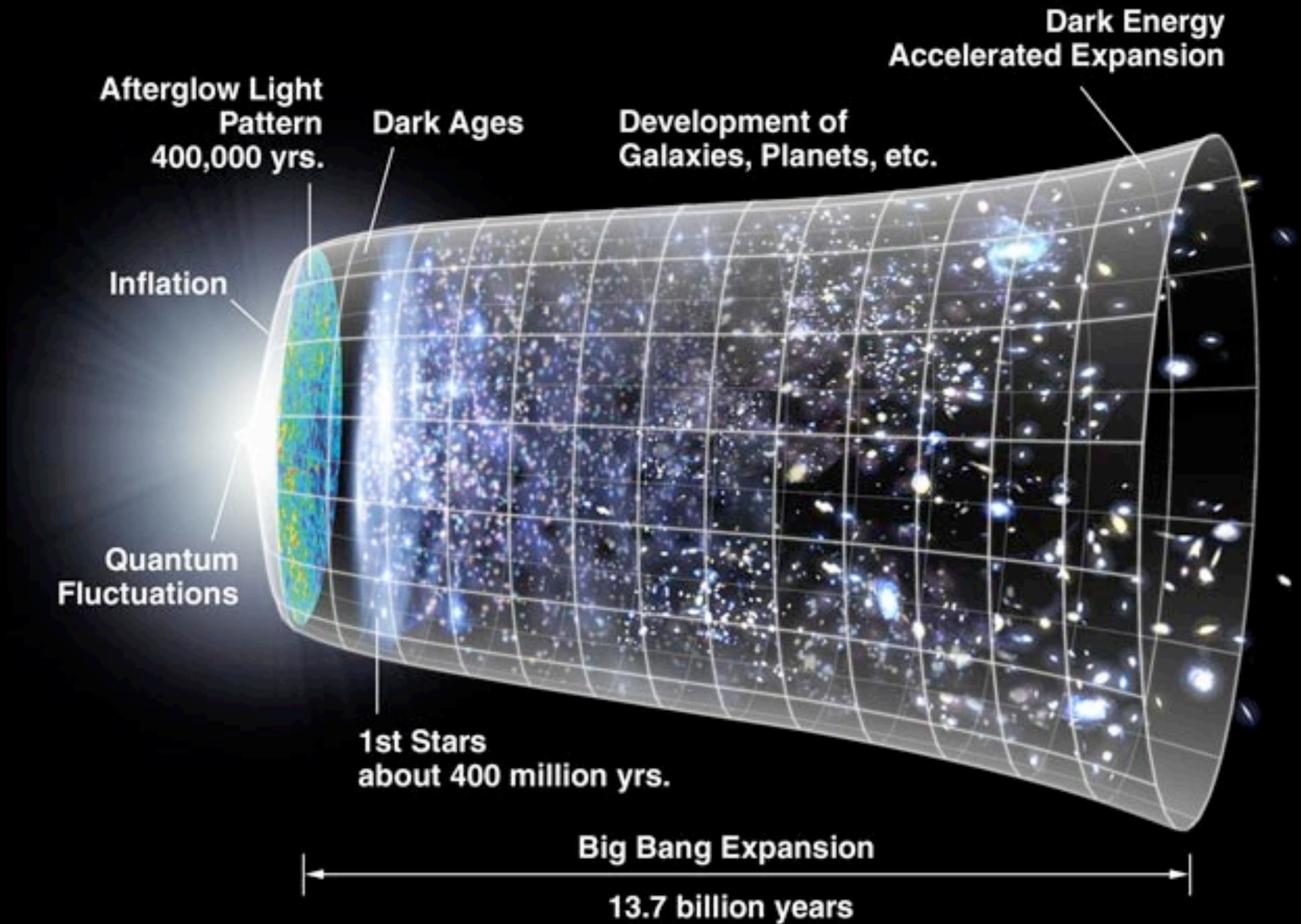


Figure 6. A schematic representation of a "merger tree" depicting the growth of a halo as the result of a series of mergers. Time increases from top to bottom in this figure and the widths of the branches of the tree represent the masses of the individual parent halos. Slicing through the tree horizontally gives the distribution of masses in the parent halos at a given time. The present time t_0 and the formation time t_f are marked by horizontal lines, where the formation time is defined as the time at which a parent halo containing in excess of half of the mass of the final halo was first created.

500 kpc

The baryonic component



Astrophysical processes modeled:
ChemTreeN (Tumlinson 2010)

- Shock heating & radiative cooling
- Photoionization squelching
- Star formation (quiescent & burst)
- SN heating & SN-driven winds
- Chemical evolution
- Stellar populations

Particle tagging
to distribute stars:
e.g. 5% most bound

TABLE 2
MODEL PARAMETERS.

Parameter	Fiducial Value	Range	Description	Explored
$z_{\rm r}$	10	5 - 19	Epoch of re-ionization	Yes
$f_{\rm bary}$	0.05	0 - 0.2	Baryonic mass fraction	Yes
$f_{\rm esc}$	50	0 - 110	Escape factor of metals	Yes
ϵ_*	1×10^{-10}	0.2 - 1.8	Star formation efficiency (10^{-10} yr^{-1})	Yes
$m_{\rm Fe}^{\rm II}$	0.07	0.04 - 0.2	SN II iron yield ($M_\odot$)	Yes
f_{Ia}	0.015	···	SN Ia probability	No
$\epsilon_{\rm SN}$	0.0015	···	SNe energy coupling	No
$m_{\rm Fe}^{\rm Ia}$	0.5	···	SN Ia iron yield ($M_\odot$)	No

ChemTreeN Analysis

CHARACTERIZING THE FORMATION HISTORY OF MILKY WAY-LIKE STELLAR HALOES WITH MODEL EMULATORS

FACUNDO A. GÓMEZ
Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA and
Institute for Cyber-Enabled Research, Michigan State University, East Lansing, MI 48824, USA

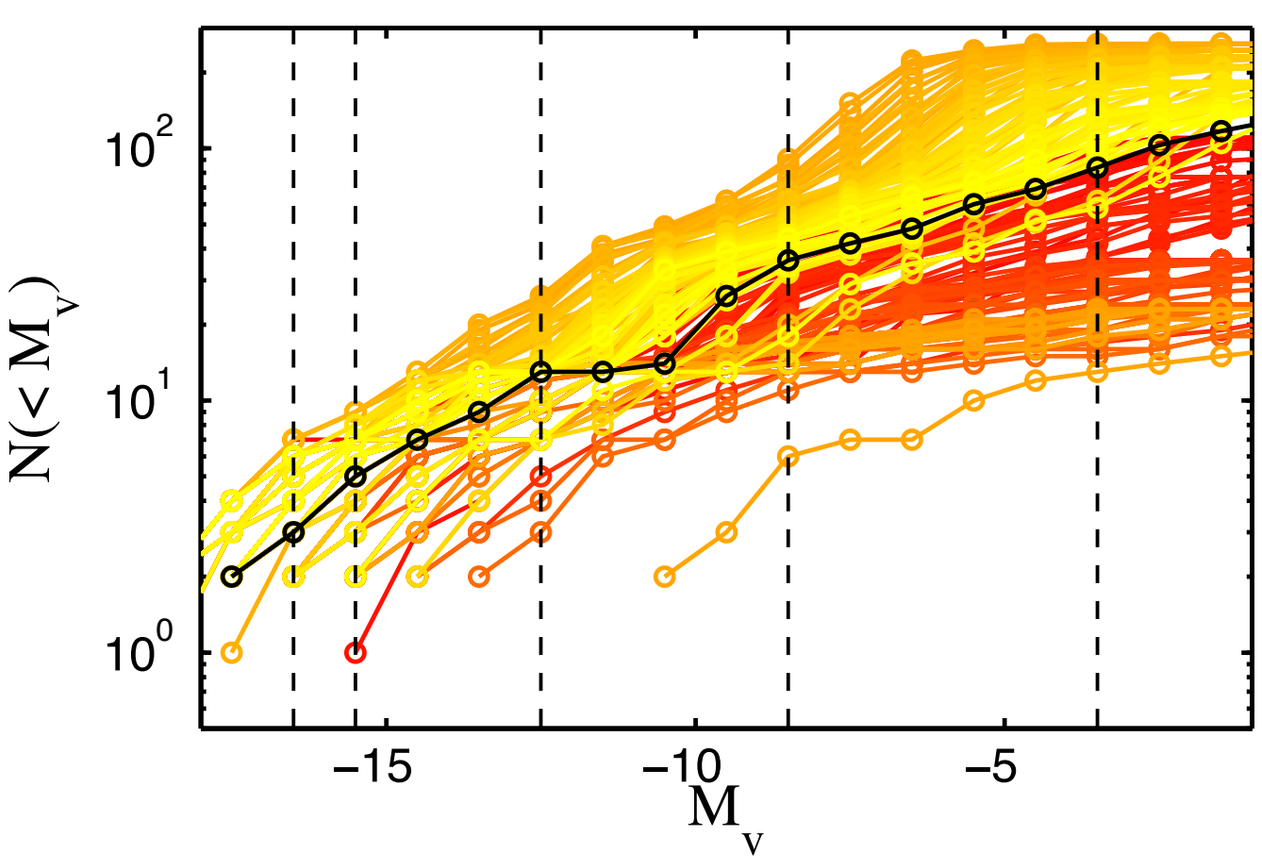
CHRISTOPHER E. COLEMAN-SMITH
Department of Physics, Duke University, Durham, NC, 27708, USA

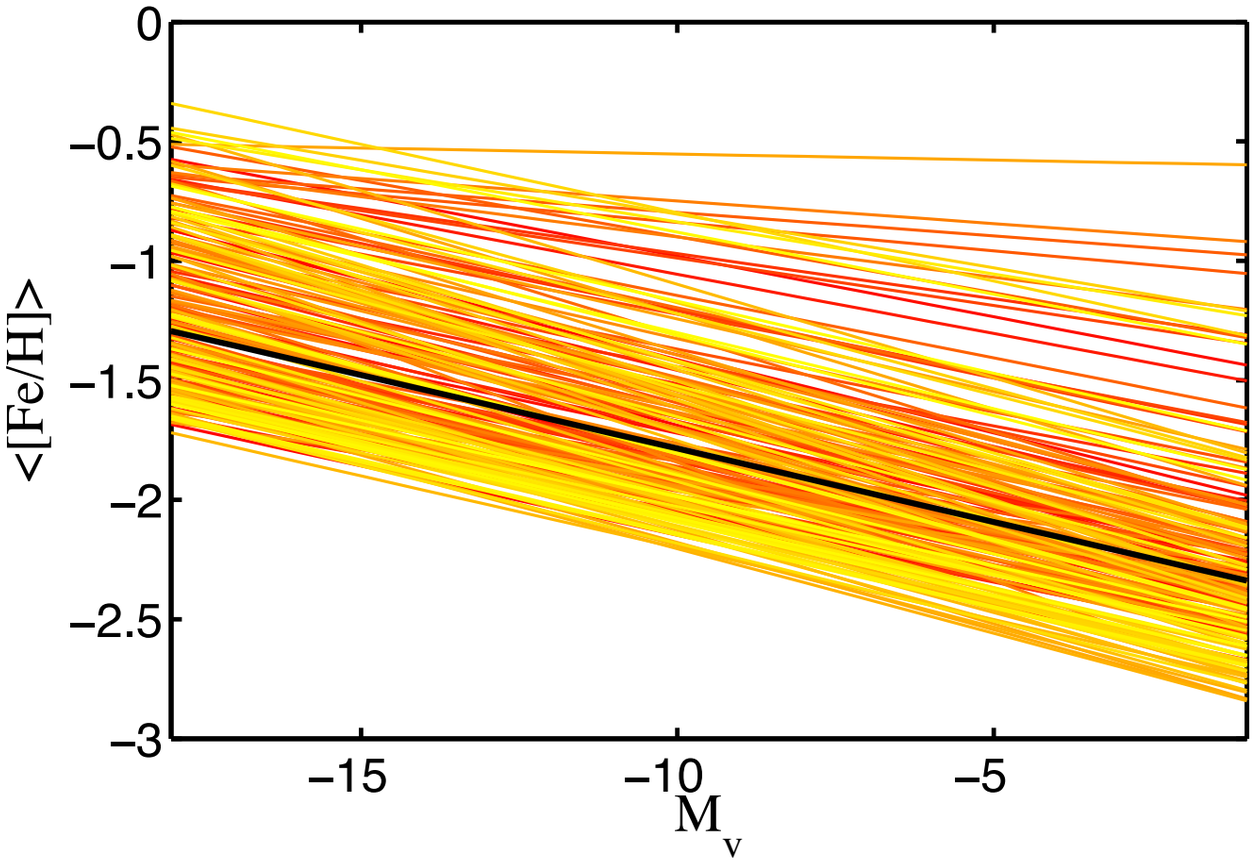
- Controlled experiments: Take true "parameters" from model itself. 3 parameter design

$$\boldsymbol{x}_{\rm obs} = (z_{\rm r},\ f_{\rm esc},\ f_{\rm bary}) = (10,\ 50,\ 0.05).$$

- 5 values from the Luminosity Function (LF)
- Intercept and slope of the Luminosity-Metallicity relation

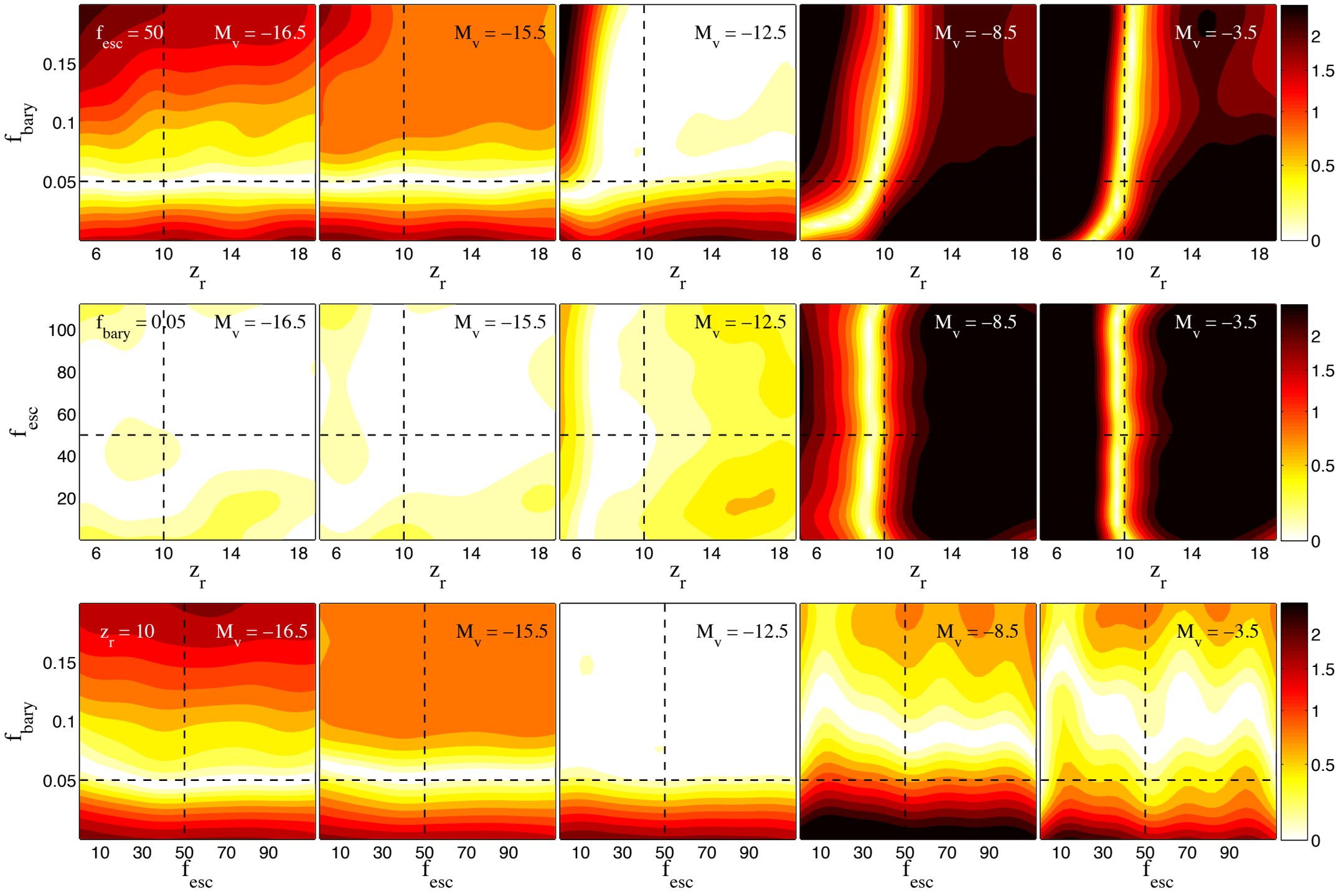
Training set: 200 models
7 observables





ChemTreeN - Implausibility

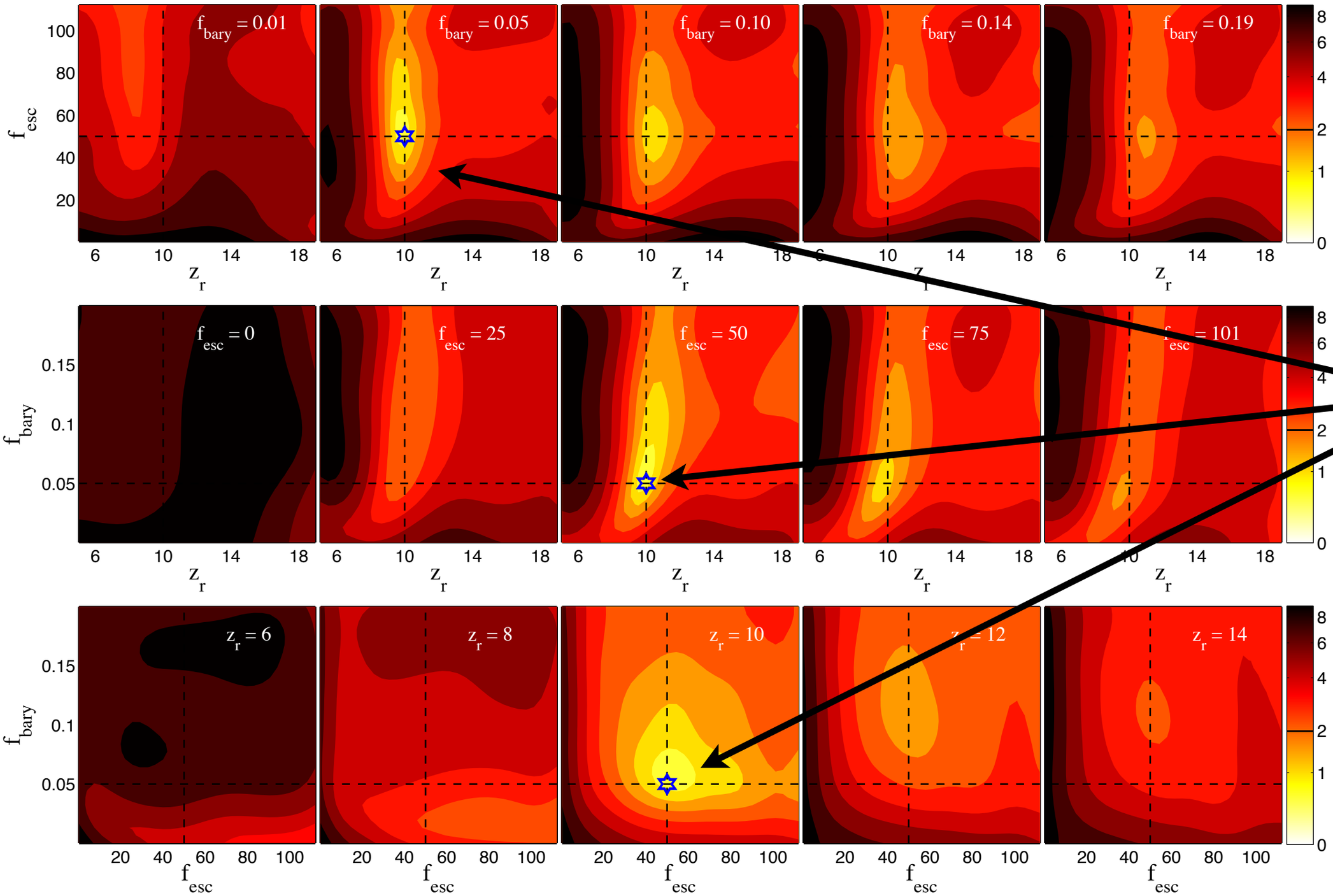
Columns are Luminosity Bins
(separate emulators),
Rows are projections



Difficult to constrain parameters with individual emulators

ChemTreeN - Joint Implausibility

Columns are slices in parameter space
Rows are projections



Joint Implausibility gives a scalar field over parameter space

Strongly constrains parameters

Calibration

Inference about calibration parameters $u_\star$

$$Y_f(x, u_\star) = Y_r(x, u_\star) + \epsilon_f(x)$$
$$Y_r(x, u_\star) = Y_m(x, u_\star) + b(x, u_\star),$$

- Combine field observations Yf and model observations Ym to infer *something* about the calibration parameters

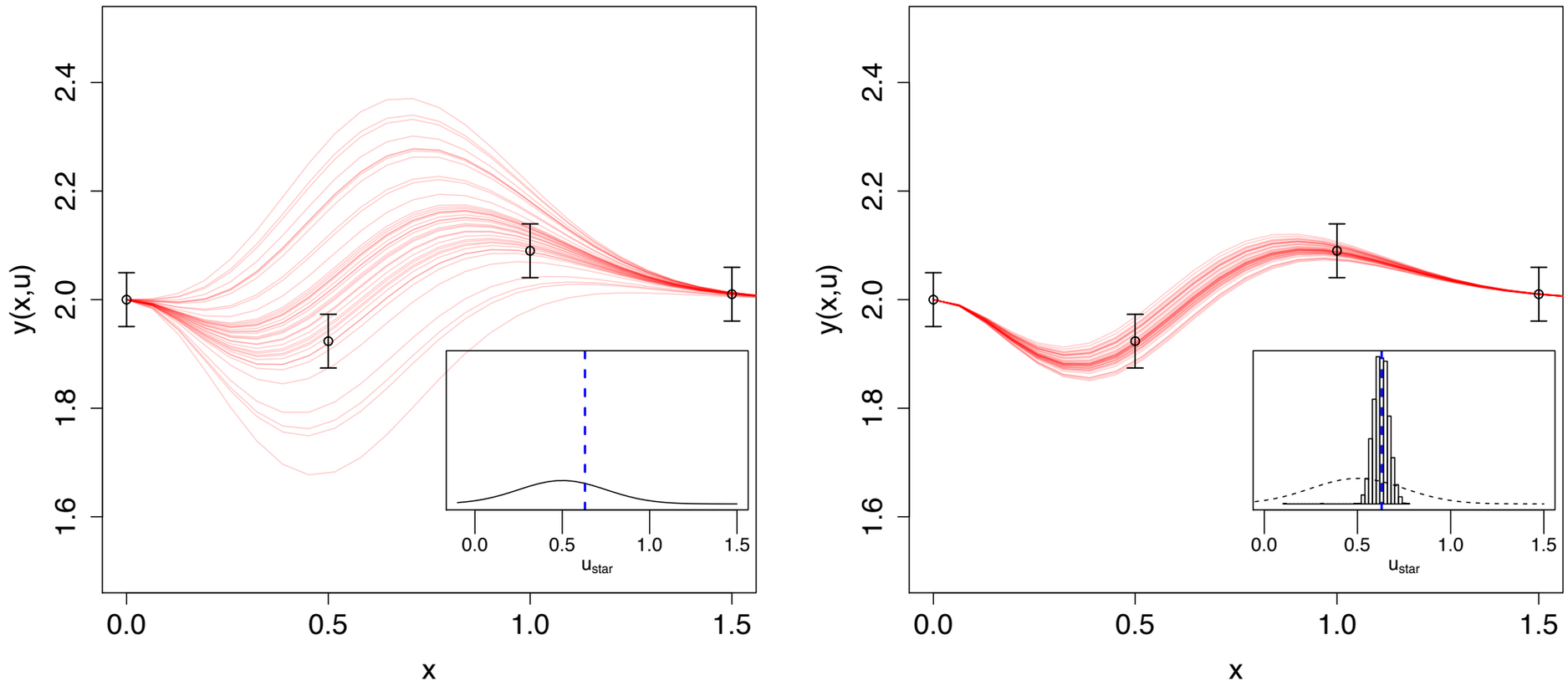
- Field observations
$$y_f = (Y_f(x_1), \ldots, Y_f(x_n)), \quad x_i \in \mathcal{D}_f$$
- Model observations
$$\mathcal{Y} = (Y_m(x_1, u_1), \ldots, Y_f(x_m, u_m)), \quad x_i \in \mathcal{D}$$

- Do **not** have to be made at the same design points

D. Higdon, M. Kennedy, J. Cavendish, J. Cafeo, and R. Ryne, SIAM Journal on Scientific Computing **26**, 448 (2004).

Fast Faithful Model

	HPD lower (95%)	mean	HPD upper (95%)
prior	0.089	0.5	0.911
posterior	0.556	0.629	0.707



- Model can be evaluated for no cost & accurately reproduces reality

$$Y_f(x_i) = Y_m(x_i, u_\star) + \epsilon(x_i), \quad i = 1 \ldots n,$$

- Likelihood for field observations

$$L\left(y_f \mid Y_m(u_\star)\right) \propto \exp\left\{-\frac{1}{2}(y_f - Y_m(u_\star))^\intercal \Sigma_f^{-1}(y_f - Y_m(u_\star))\right\},$$

- Picking some sensible prior

$$\pi\left(u_\star \mid y_f\right) \propto L\big(y_f \mid Y_m(u_\star)\big)\pi(u_\star).$$

Slow Faithful Model

- Simulator is too slow, we can only observe it at some finite set of locations

$$\mathcal{Y} = (Y_m(x_1, u_1), \ldots, Y_f(x_m, u_m)), \quad x_i \in \mathcal{D}$$

- Introduce augmented field and model vector

$$z = (y_f^\intercal, \mathcal{Y}^\intercal)^\intercal$$

- observation likelihood is now

$$L(z \mid u_\star, \mu, \lambda_m, \beta^m, \Sigma_f) \propto |\Sigma_z|^{-1/2} \exp\left\{ -\frac{1}{2} (z - \mu \mathbb{I}_{n+d})^\intercal \Sigma_z^{-1} (z - \mu \mathbb{I}_{n+d}) \right\},$$

$$\Sigma_z = \Sigma_m + \begin{pmatrix} \Sigma_f & 0 \\ 0 & 0 \end{pmatrix}, \quad \Sigma_m = \begin{pmatrix} \Sigma_{y_f y_f} & \Sigma_{y_f y} \\ \Sigma_{y_f y}^\intercal & \Sigma_{yy} \end{pmatrix}$$

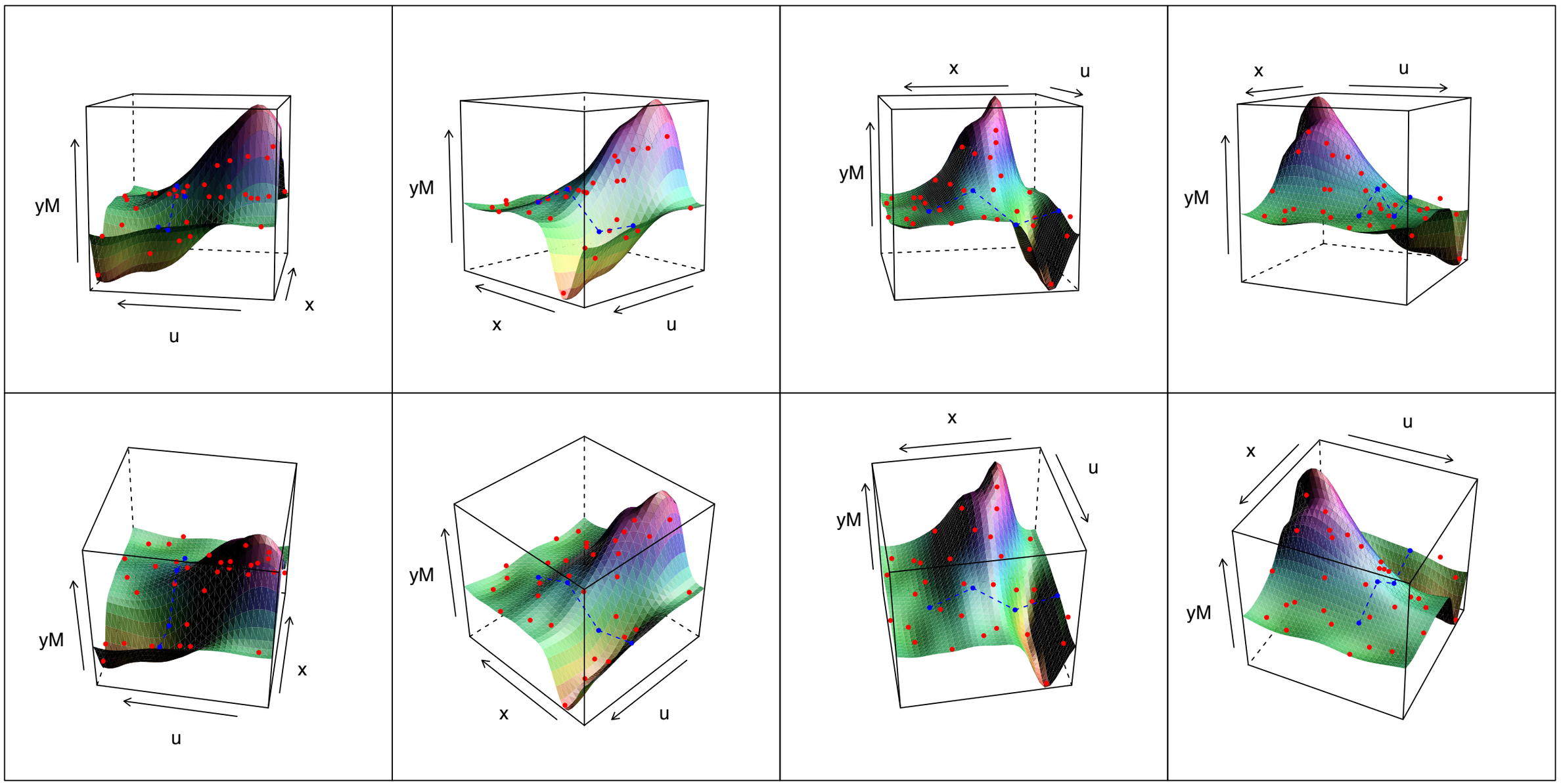
- We have placed a GP prior on the simulator output

$$\mathcal{C}((x,u),(x',u')) = \frac{1}{\lambda_m} \exp\left\{ -\sum_{k=1}^{p_x} \frac{(x_k - x'_k)^\alpha}{(\beta_k^m)^\alpha} - \sum_{k=1}^{p_u} \frac{(u_k - u'_k)^\alpha}{(\beta_{p_x+k}^m)^\alpha} \right\}$$

$$\pi(\lambda_m) \propto \lambda_m^{a_m - 1} e^{-b_m \lambda_m}$$

$$\pi(\beta_k^m) \propto (\beta_k^m)^{a_\beta - 1} (1 - \rho_k)^{b_\beta - 1}, k = 1, \ldots, p_x + p_u.$$

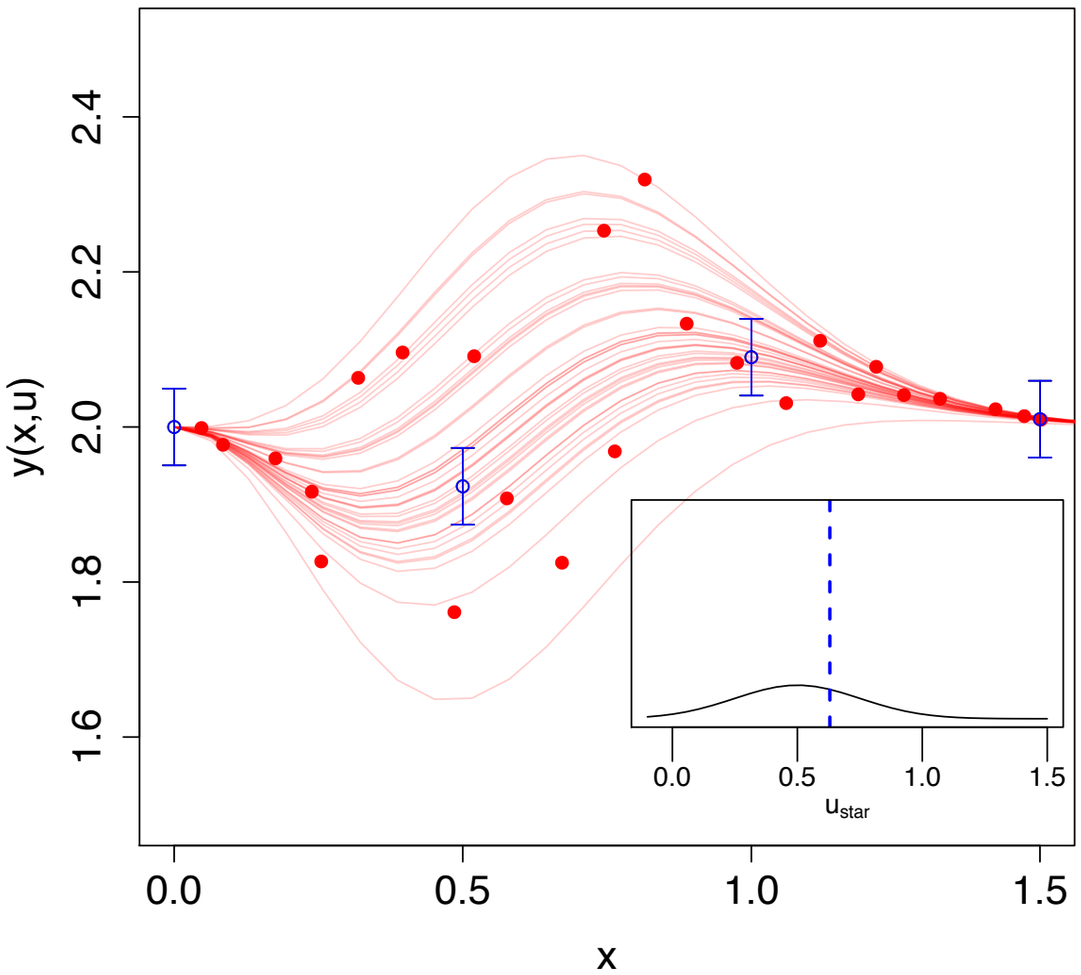
Slow Faithful Model

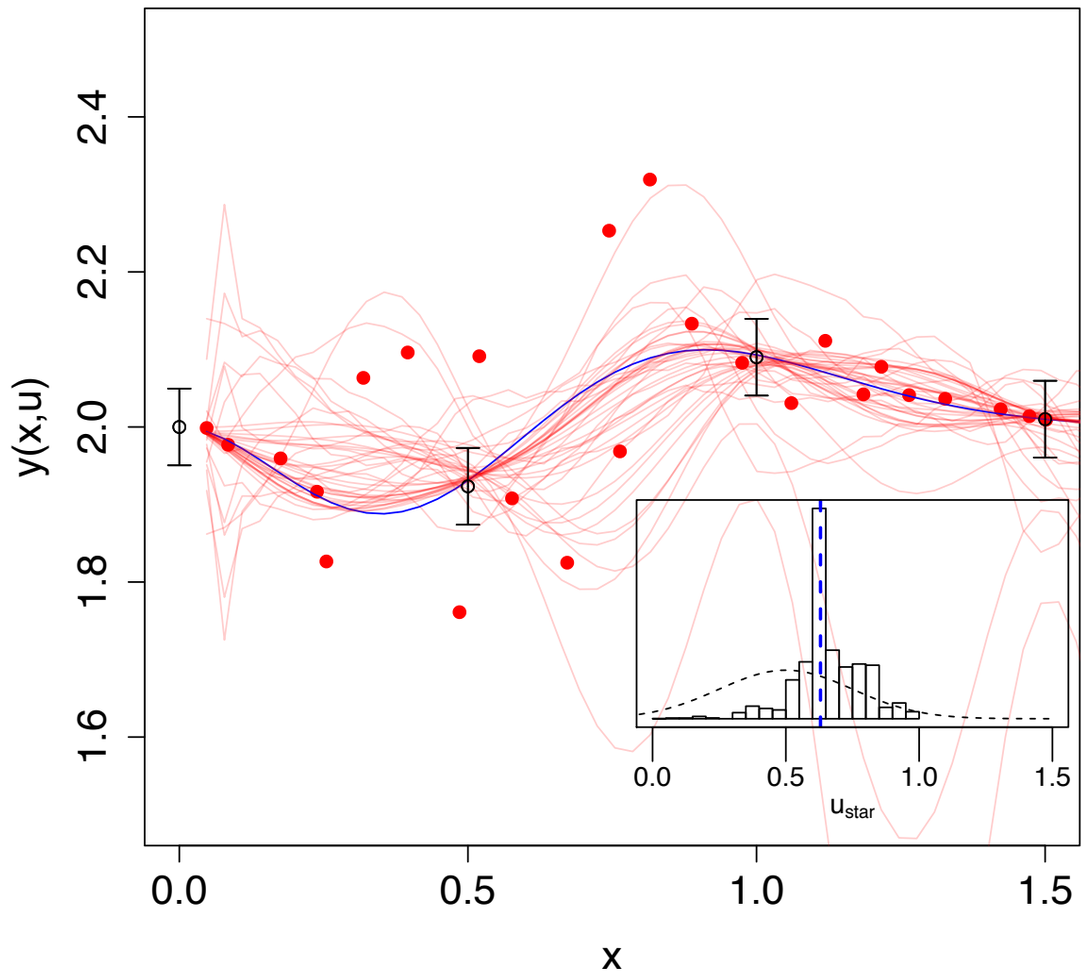


Posterior mean

Slow Faithful Model

$$Y_m(x,u) = 5x^2\exp(-3x^2)\sin(x-u) + 2, \quad x \in [0,2]$$



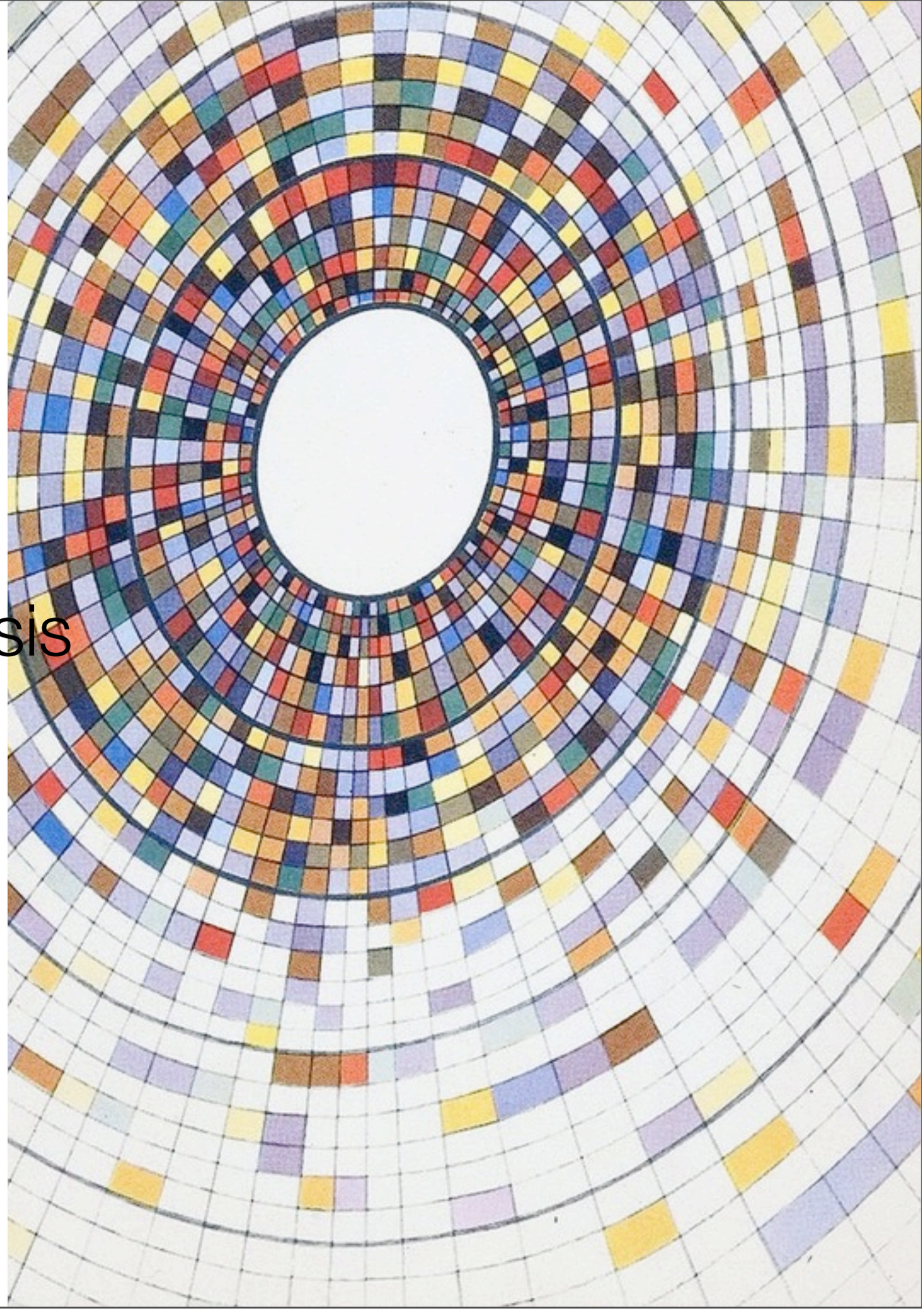


32 training points

	HPD lower (95%)	mean	HPD upper (95%)
prior	0.089	0.5	0.911
posterior	0.462	0.629	0.868

MSU 3+1d Hydro Analysis

With S.Pratt (MSU)



Calibration of MSU 3+1d viscous hydro

- A 3+1d event averaged viscous hydrodynamics code with a hadronic afterburner.
- Design includes parameters which control the initial state energy and flow distributions and the viscosity during the hydro evolution.
- Extensive RHIC experimental data set was collected, these data were empirically reduced to a set of 15 summary variables.
- GP emulator trained on a PCA decomposition of the 15 model output summary variables
- Calibration was carried out using MCMC with a very simple likelihood model

$$\mathcal{L}(\mathbf{x}) \propto \exp\left\{-\frac{1}{2}\sum_i \frac{(z_i^{(\mathrm{emu})}(\mathbf{x}) - z_i^{(\mathrm{exp})})^2}{1+\sigma_e^2}\right\}.$$

6 dimensional design

parameter	description	range
$(dE/dy)_{pp}$	The initial energy per rapidity in the diffuse limit compared to measured value in pp collision	0.85–1.2
$\sigma_{\rm sat}$	This controls how saturation sets in as function of areal density of the target or projectile. In the wounded nucleon model it is assumed to be the free nucleon-nucleon cross section of 42 mb	30 mb–50 mb
f_{wn}	Determines the relative weight of the wounded-nucleon and saturation formulas for the initial energy density described in (3, 4)	0–1
$F_{\rm flow}$	Describes the strength of the initial flow as a fraction of the amount described in (7)	0.25–1.25
$\eta/s\|_{T_c}$	Viscosity to entropy ratio for $T = 170$ MeV	0 – 0.5
α	Temperature dependence of η/s for temperatures above 170 MeV/c, i.e., $\eta/s = \eta/s\|_{T_c} + \alpha \ln(T/T_c)$	0 - 5

TABLE I. Summary of model parameters. Six model parameters were varied. The first four describe the initial state being fed into the hydrodynamic module, and the last two describe the viscosity and its energy dependence.

Analyzed 0-5% and 20-30% centrality classes

Comprehensive RHIC data set

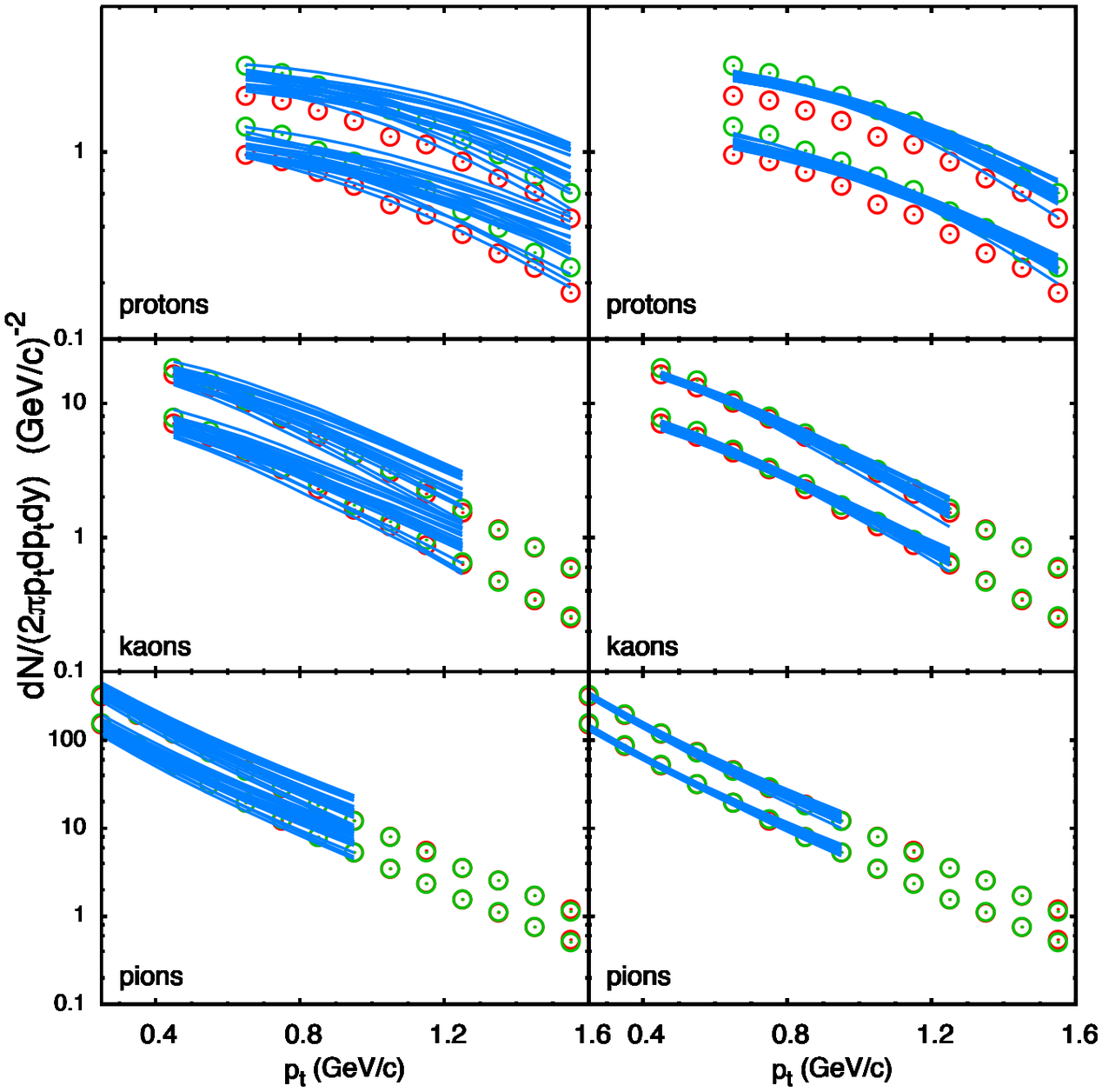
observable	p_t weighting	centrality	collaboration	uncertainty
$v_{2,\pi^+\pi^-}$	ave. over 11 p_t bins from 160 MeV/c to 1 GeV/c	20-30%	STAR* [45]	12%
$R_{\rm out}$	ave. over 4 p_t bins from 150-500 MeV/c	0-5%	STAR [46]	6%
$R_{\rm side}$	ave. over 4 p_t bins from 150-500 MeV/c	0-5%	STAR [46]	6%
$R_{\rm long}$	ave. over 4 p_t bins from 150-500 MeV/c	0-5%	STAR [46]	6%
$R_{\rm out}$	ave. over 4 p_t bins from 150-500 MeV/c	20-30%	STAR [46]	6%
$R_{\rm side}$	ave. over 4 p_t bins from 150-500 MeV/c	20-30%	STAR [46]	6%
$R_{\rm long}$	ave. over 4 p_t bins from 150-500 MeV/c	20-30%	STAR [46]	6%
$\langle p_t \rangle_{\pi^+\pi^-}$	200 MeV/c $< p_t <$ 1.0 GeV/c	0-5%	PHENIX [47]	3%
$\langle p_t \rangle_{K^+K^-}$	400 MeV/c $< p_t <$ 1.3 GeV/c	0-5%	PHENIX [47]	3%
$\langle p_t \rangle_{p\bar{p}}$	600 MeV/c $< p_t <$ 1.6 GeV/c	0-5%	PHENIX [47]	3%
$\langle p_t \rangle_{\pi^+\pi^-}$	200 MeV/c $< p_t <$ 1.0 GeV/c	20-30%	PHENIX [47]	3%
$\langle p_t \rangle_{K^+K^-}$	400 MeV/c $< p_t <$ 1.3 GeV/c	20-30%	PHENIX [47]	3%
$\langle p_t \rangle_{p\bar{p}}$	600 MeV/c $< p_t <$ 1.6 GeV/c	20-30%	PHENIX [47]	3%
$\pi^+\pi^-$ yield	200 MeV/c $< p_t <$ 1.0 GeV/c	0-5%	PHENIX [47]	6%
$\pi^+\pi^-$ yield	200 MeV/c $< p_t <$ 1.0 GeV/c	20-30%	PHENIX [47]	6%

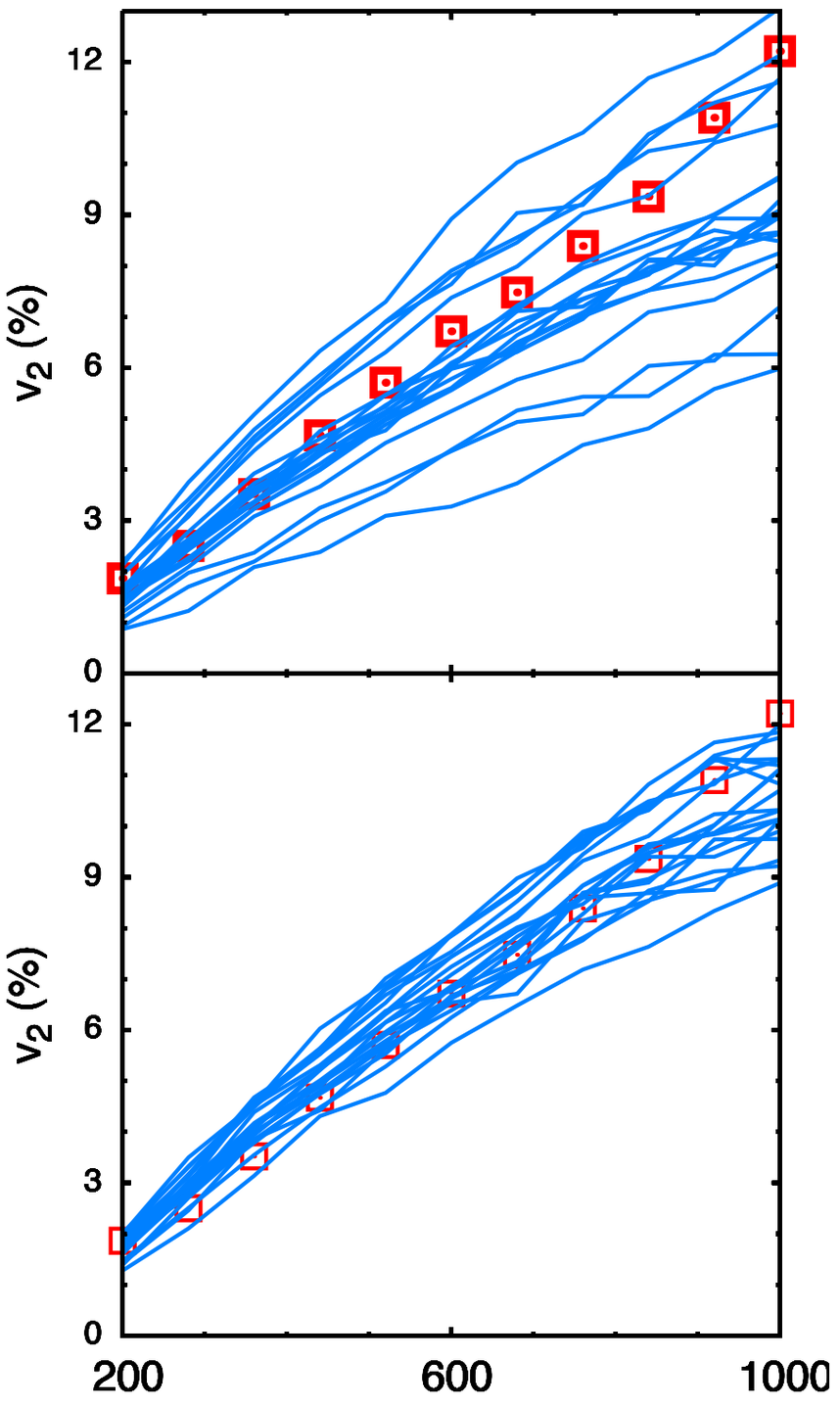
TABLE II. Observables used to compare models to data. *To account for non-flow correlations, the value of v_2 was reduced by 10% from the value reported in [45].

Model was actually calibrated against flow, yields and pt spectra.

Results

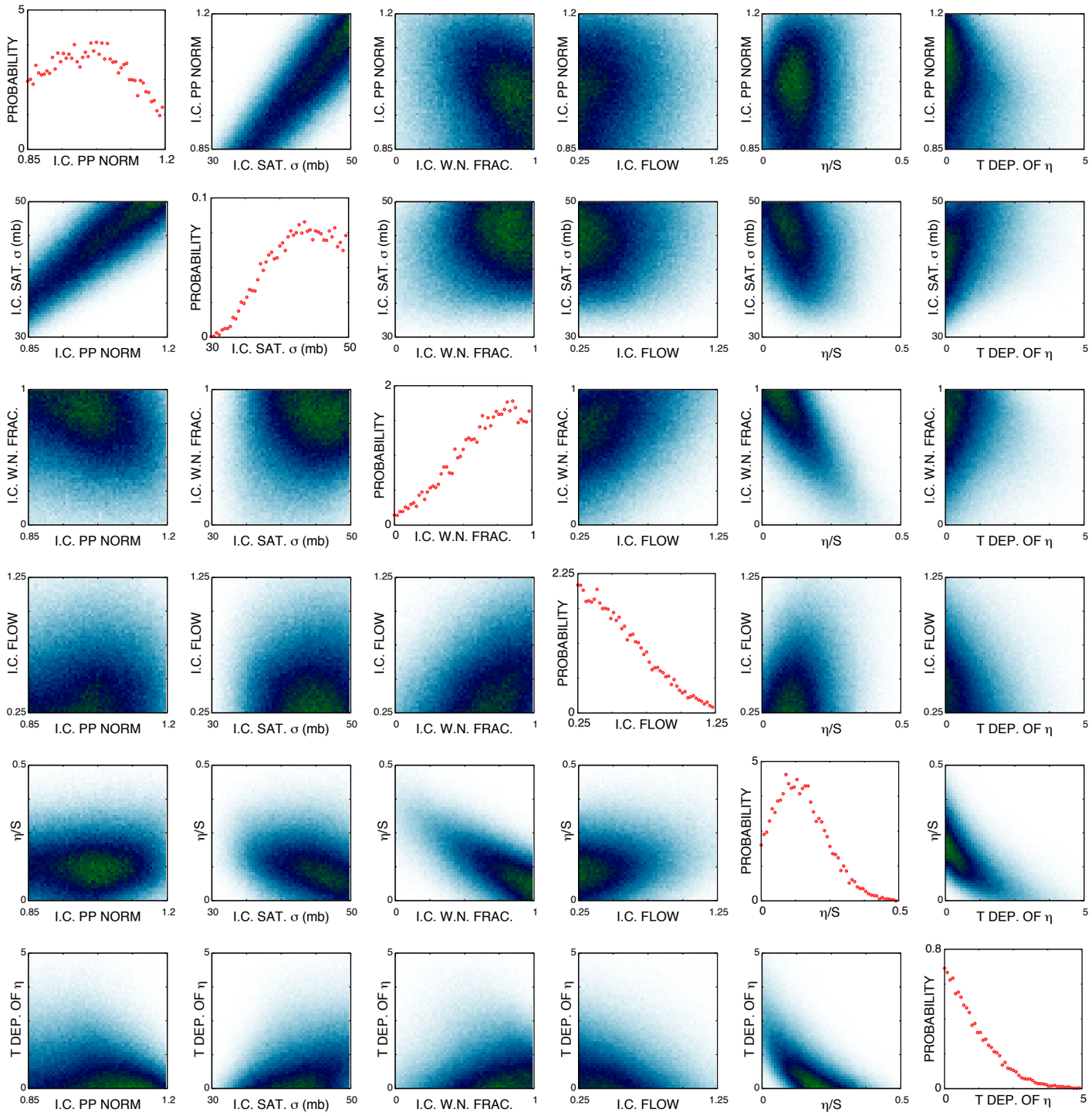
Blue curves are draws from the posterior distribution





Open symbols are data

Joint Posterior



Diagonals show posterior distribution for each of the parameters

off-diagonals show joint-posterior distribution for each pair of parameters

Sensitivity Analysis

With F.Gomez (MSU)

Sensitivity Analysis

- Difficult to visualize output for models in > 2 dimensions
- Calculate how sensitive model output is to varying inputs.
- Decompose output into main-effects and interactions

$$y = \eta(\mathbf{x}) = E(Y) + \sum_{i=1}^{d} z_i(x_i) + \sum_{i<j} z_{i,j}(\mathbf{x}_{i,j}) + \sum_{i<j<k} z_{i,j,k}(\mathbf{x}_{i,j,k}) + \ldots + z_{1,2,\ldots,d}(\mathbf{x}),$$

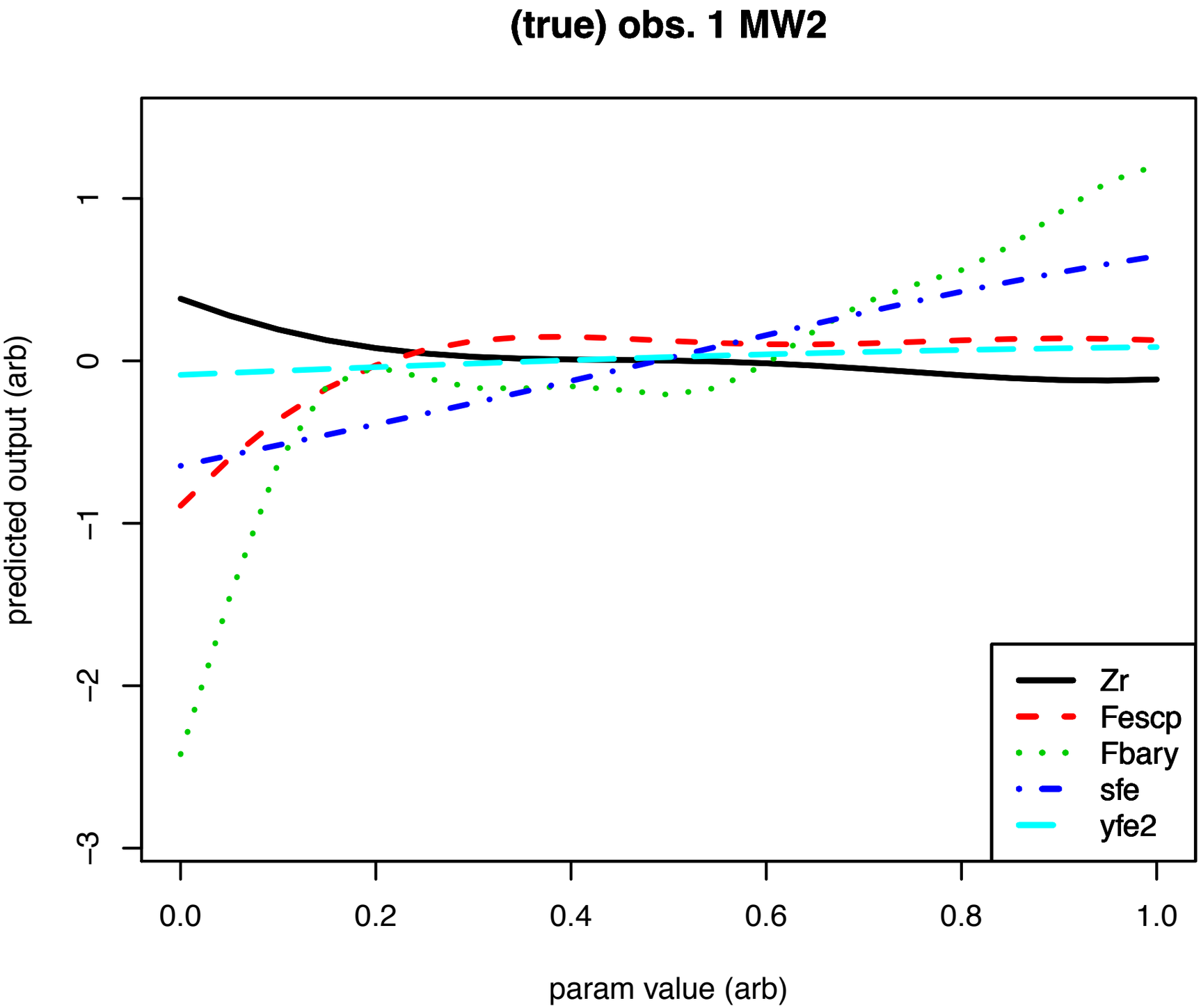
$$z_i(x_i) = E(Y \mid x_i) - E(Y),$$
$$z_{i,j}(\mathbf{x}_{i,j}) = E(Y \mid \mathbf{x}_{i,j}) - z_i(x_i) - z_j(x_j) - E(Y),$$

$$E(Y \mid \mathbf{x}_p) = \int_{\chi_{-p}} \eta(\mathbf{x}) dG_{-p|p}(\mathbf{x}_{-p} \mid \mathbf{x}_p).$$

This is relatively hard to compute with an arbitrary model. Easier for emulator.

- A similar decomposition of the model variance is also possible

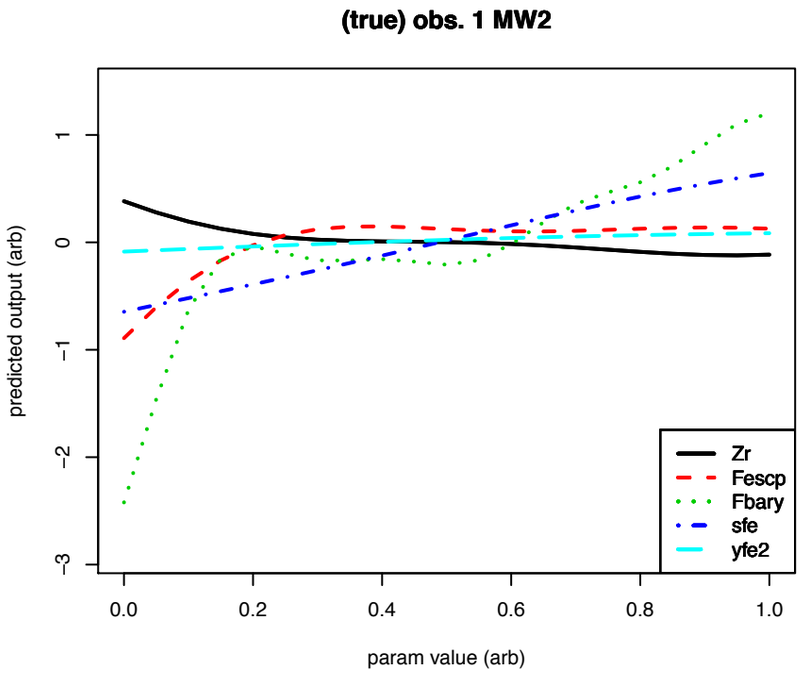
Main Effects

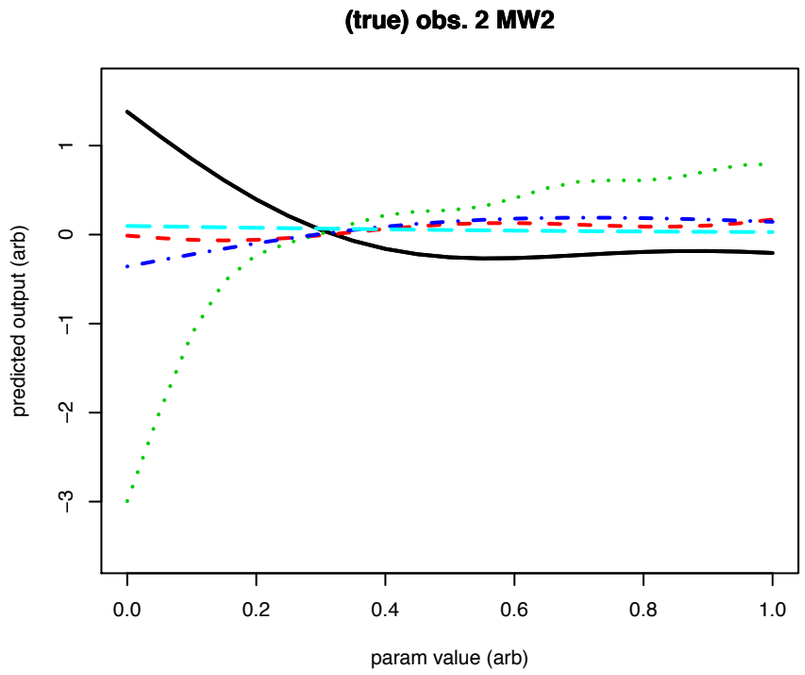


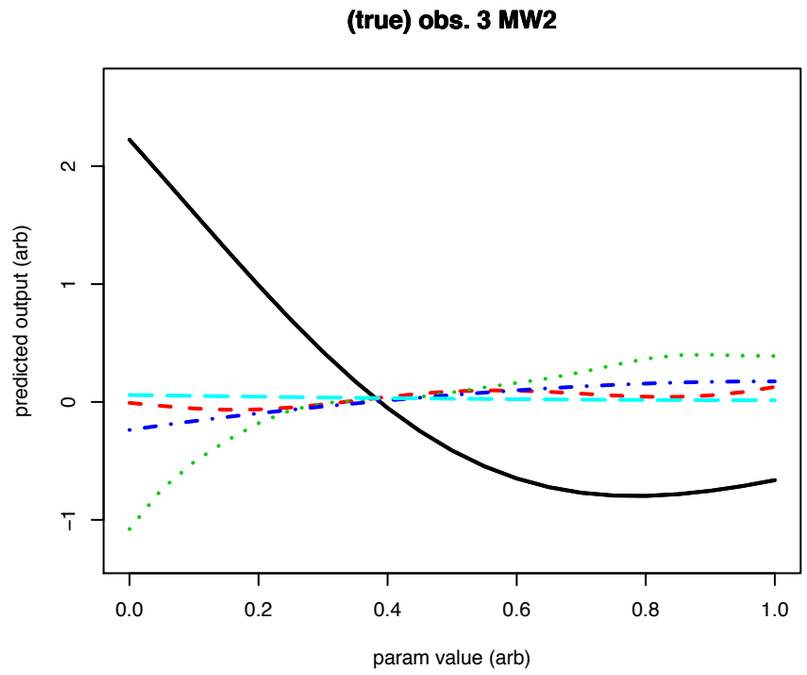
How much does each input parameter contribute to the output?

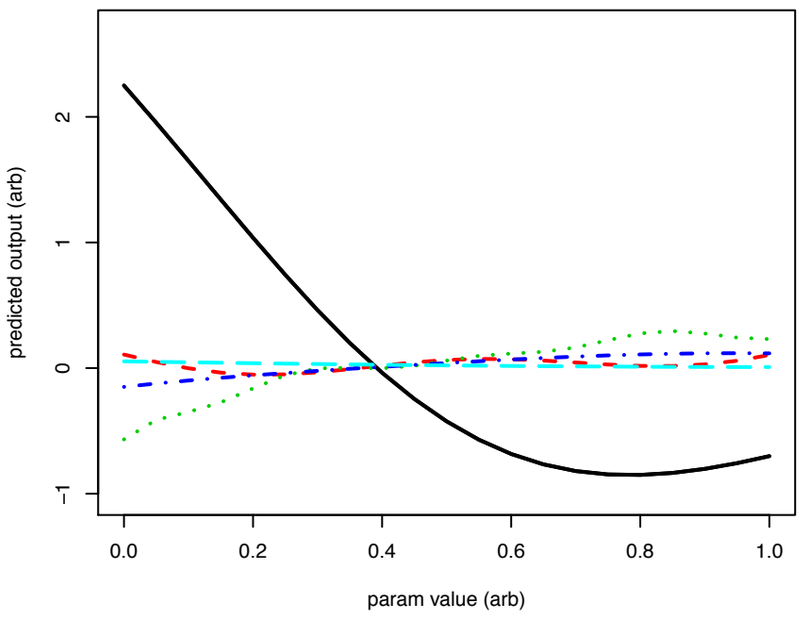
x axis spans entire rescaled range of the model parameters

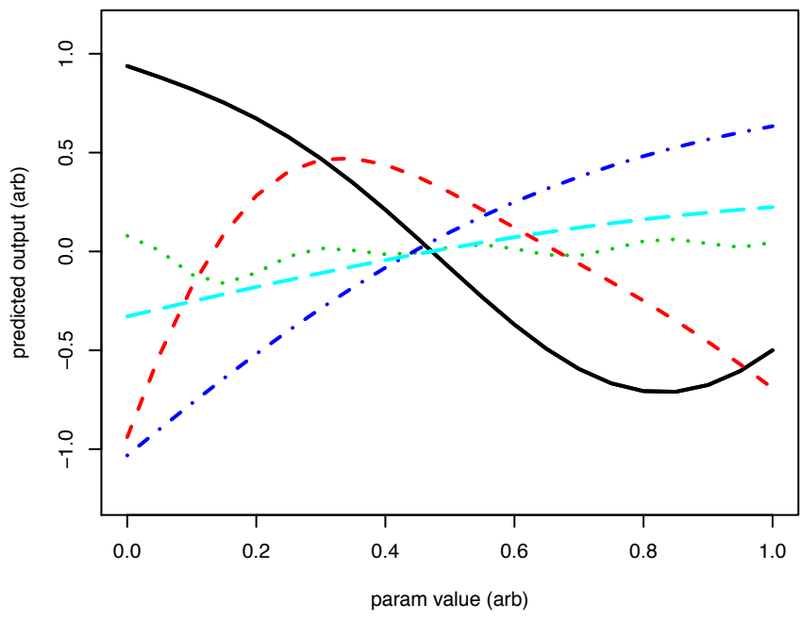
Main Effects Across Observables

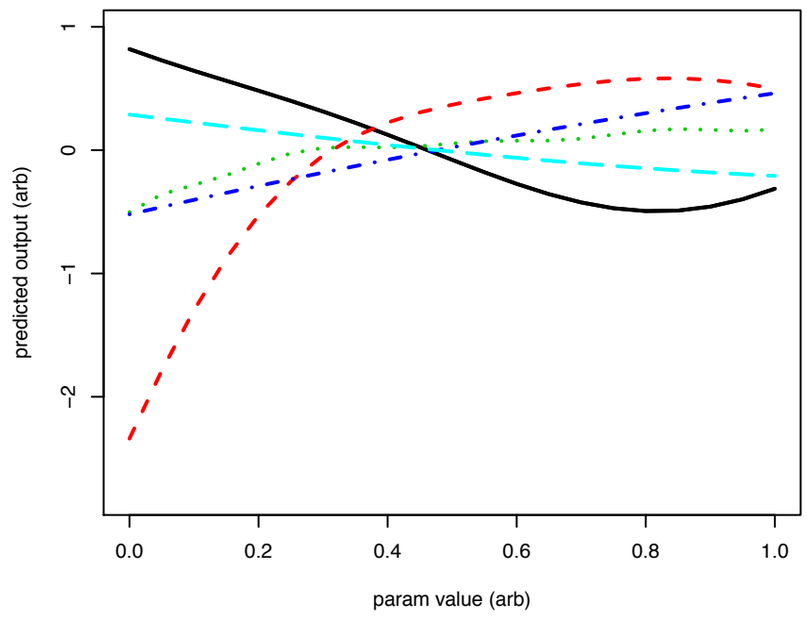








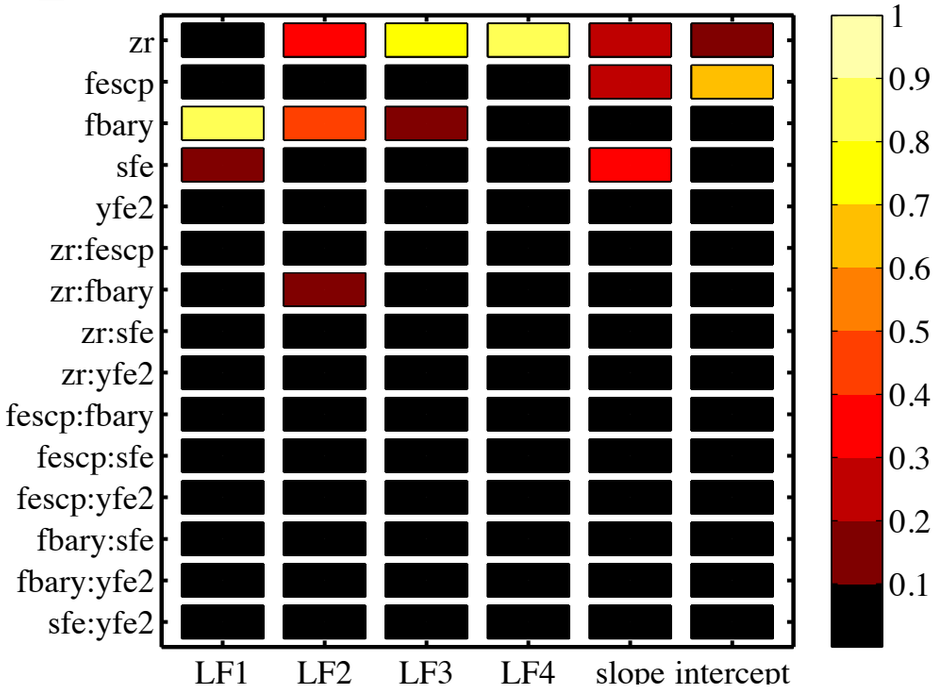




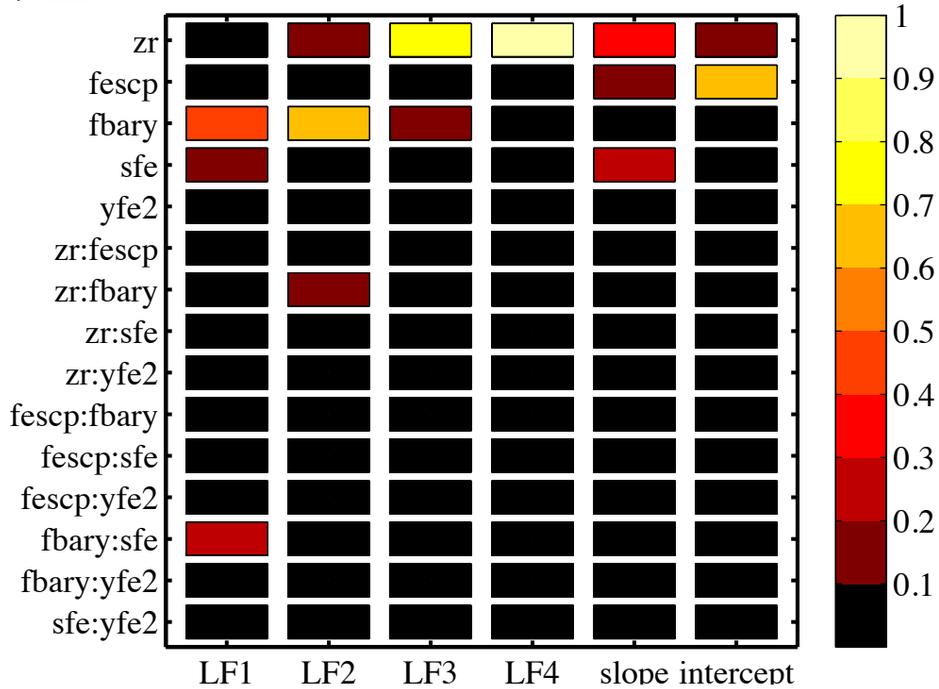
Variance decomposition

Columns are normalized

MW1

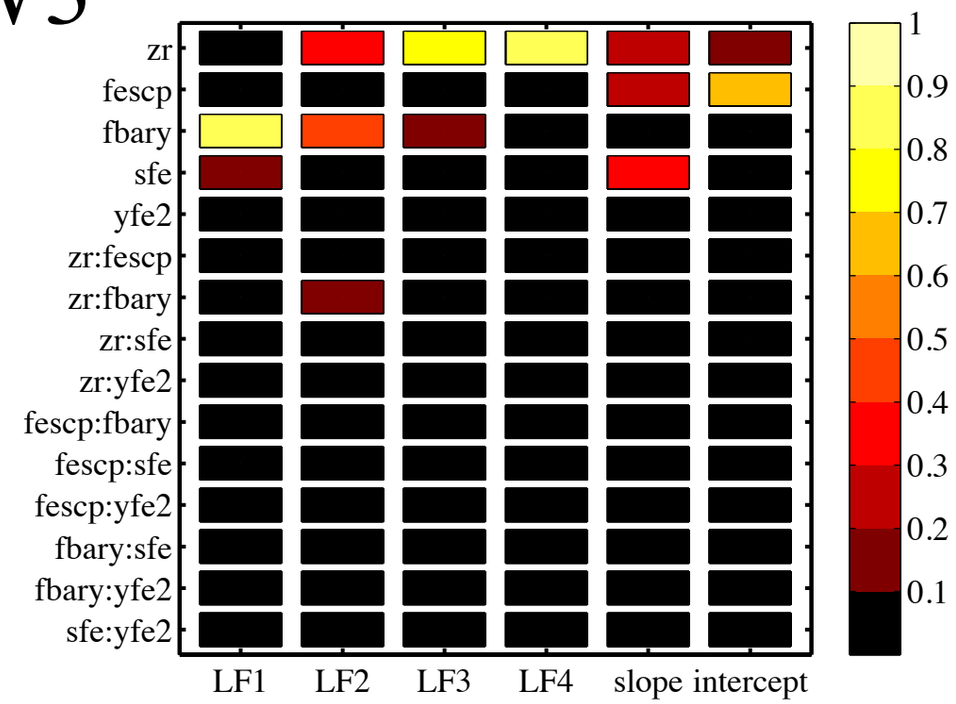


MW2

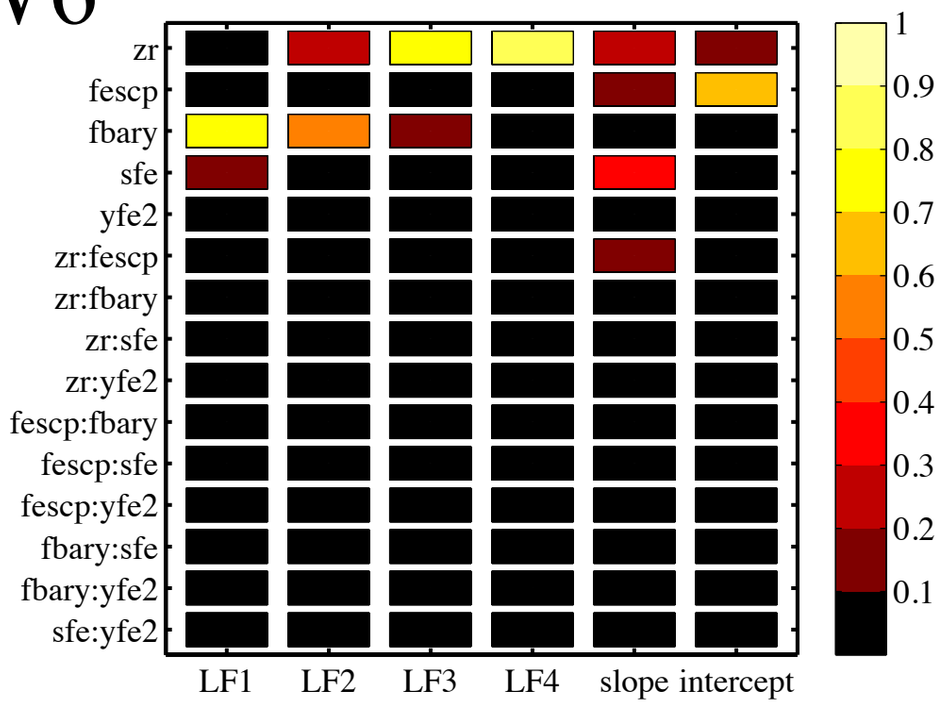


Cells are colored by variance contributions

MW3



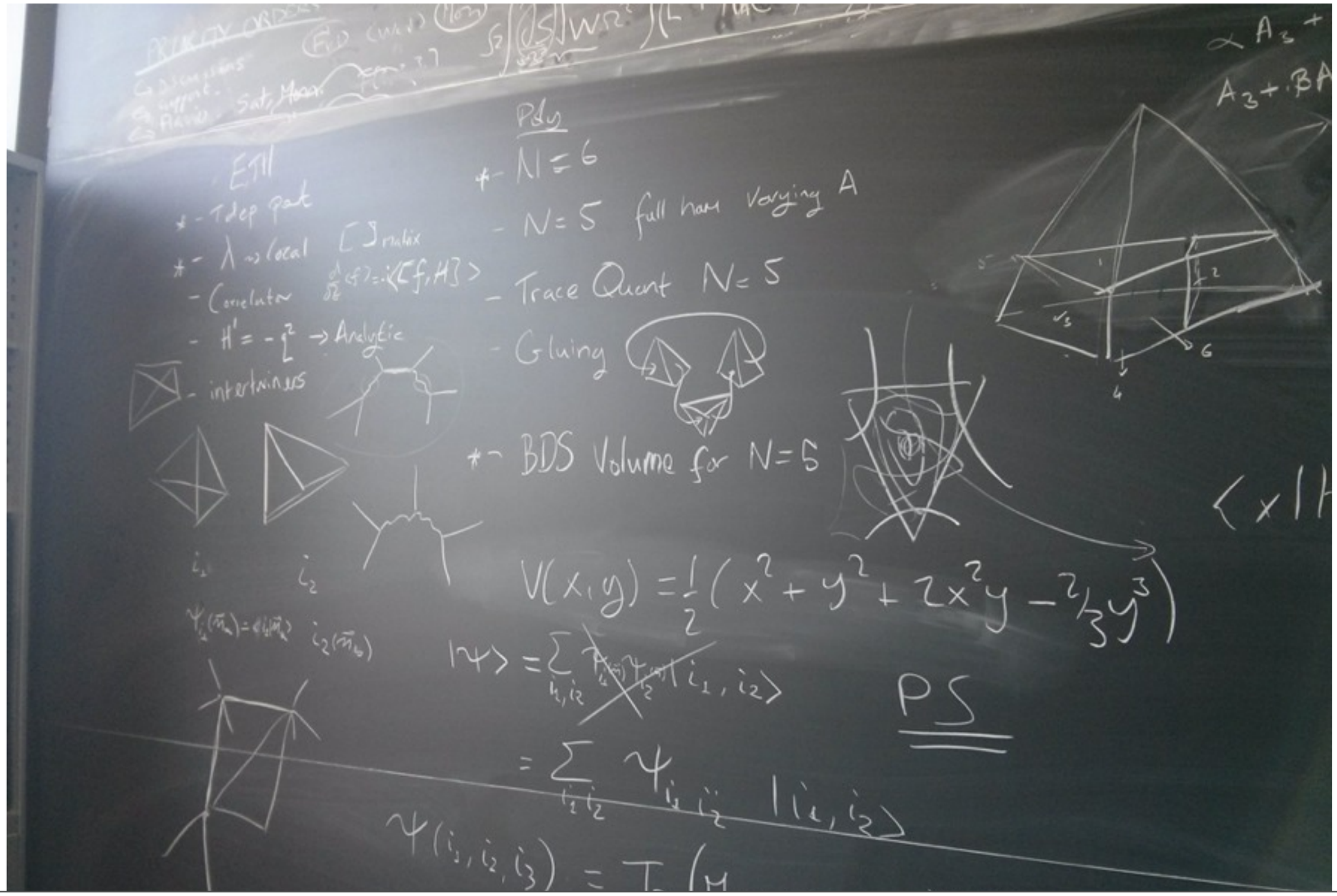
MW6



Summary

- Cutting edge science requires expensive computer experiments
- Computer models should be calibrated to experimental data
- GP Emulators are a good way to interpolate sparse (smoothish) data with input in high dimensions
- Sensitivity Analysis is a useful tool for examining your model
- High dimensional (non convex) joint distributions are difficult to analyze and should be treated with caution.
- This inherently probabilistic approach to interpolation makes inserting an emulator into a larger workflow (calibration, sensitivity analysis, validation) relatively easy

Thanks


Pdy
N=6
N=5 full ham Varying A
Trace Quant N=5
Gluing
BDS Volume for N=5
PS

References / Resources

Santner, Williams, Notz (2003). The Design and Analysis of Computer Experiments. Springer

O'Hagan (2006). Bayesian analysis of computer code outputs: A tutorial. Rel. Eng. Safe. Sys.

Campolongo, Cariboni, Saltelli (2007). An effective screening design for sensitivity analysis of large models. Environ. Model. Soft.

Oakley, O'Hagan (2004). Probabilistic sensitivity analysis of complex models: A Bayesian approach. JRSSB

Oakley, O'Hagan (2002). Bayesian inference for the uncertainty distribution of computer model outputs. Biometrika

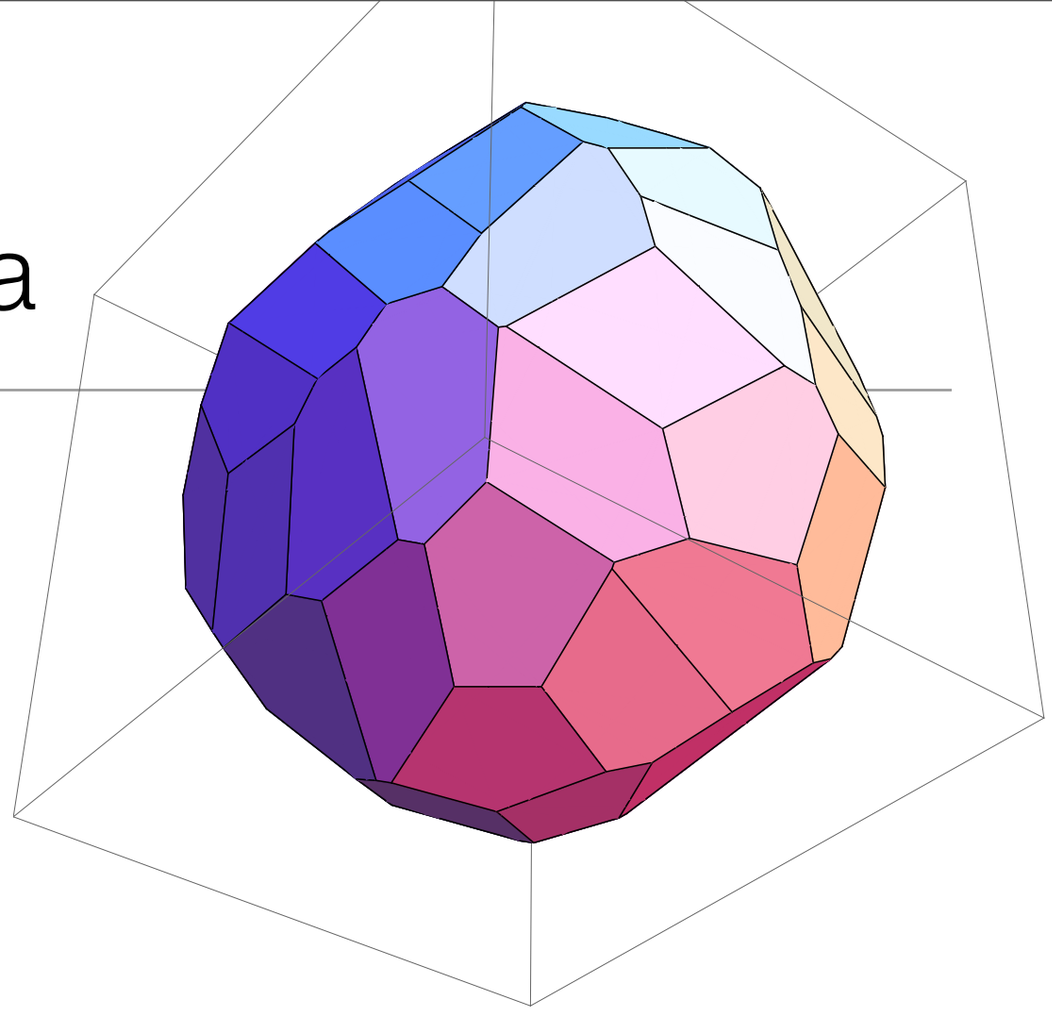
Kennedy and O'Hagan (2001). Bayesian calibration of computer models. JRSSB

Bayarri, Berger, etc (2007). A framework for validation of computer models. Technometrics

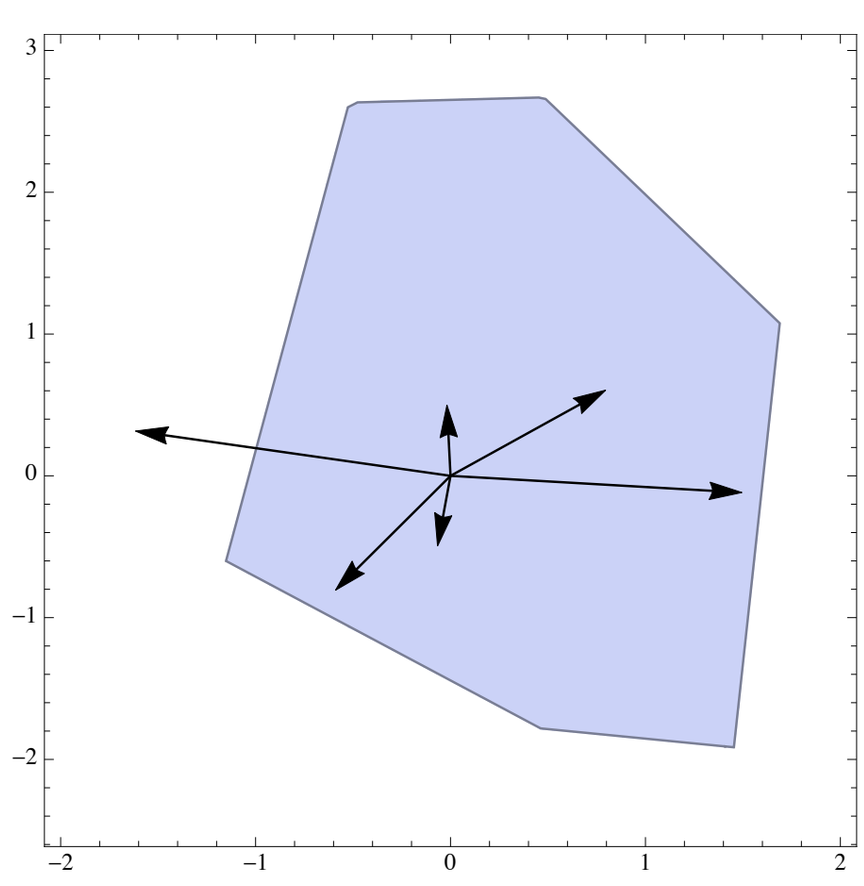
Estimating HPD Regions

- Post model calibration we have a posterior distribution for the model output over the parameters of interest
- This joint distribution embodies all we have learnt about the model
  - can be hard to interpret, may be high dimensional with non trivial relationships between the parameters
- Simplest question we might ask, what is the shape of the “most probable” (Highest Posterior Density) region.
  - A (1-α)100% HPD region is the minimal volume set with posterior probability (1-α)
  - Is it convex? Simply Connected?
- In >1d these are not trivial to find

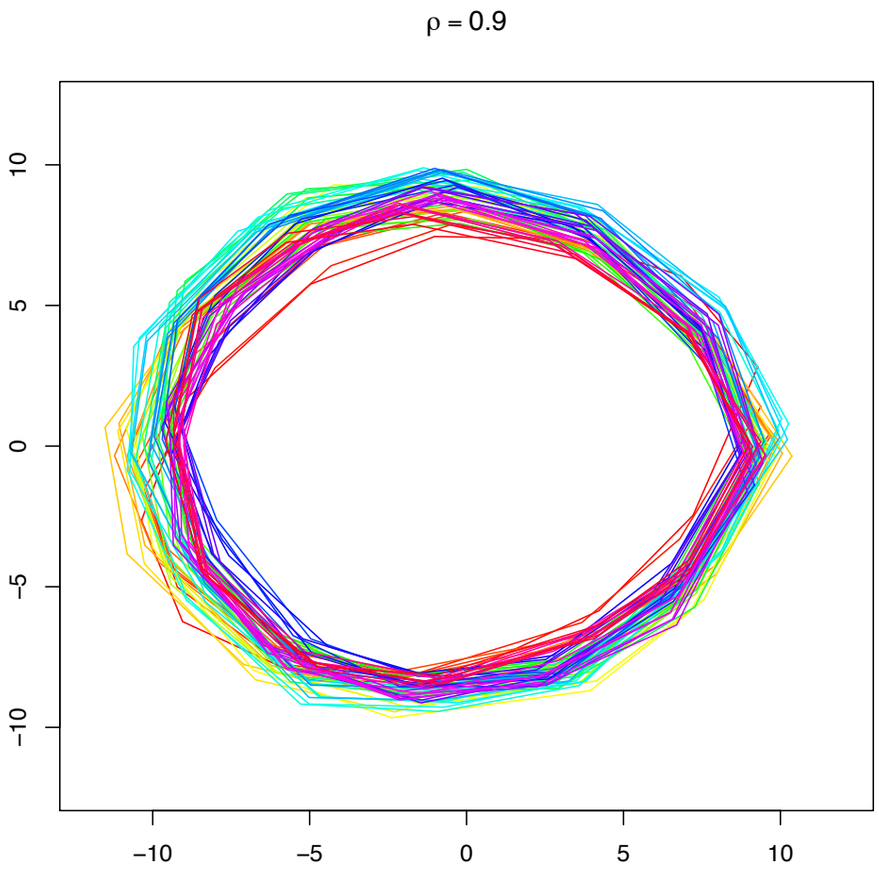
HPD Estimation with polyhedra



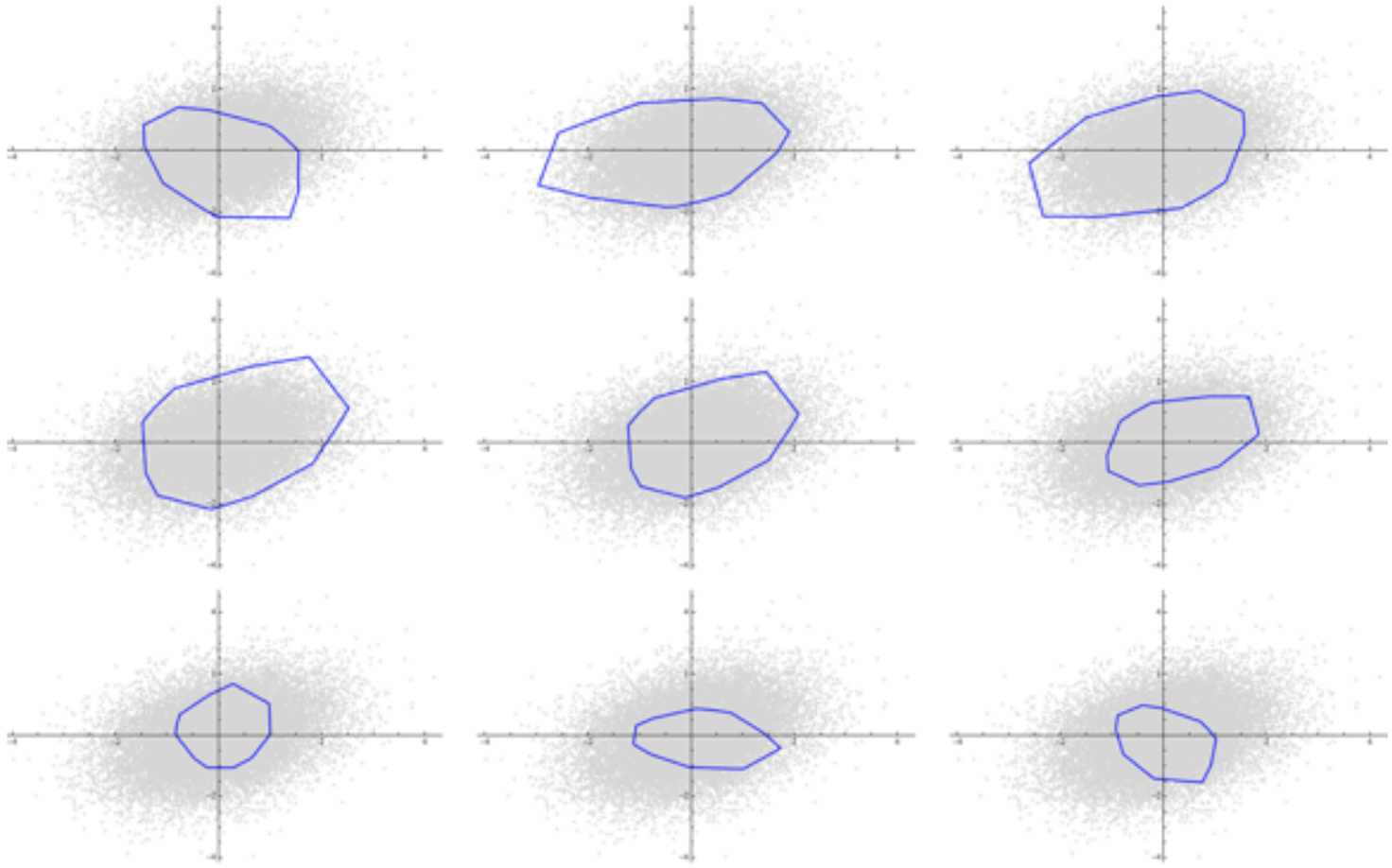
- Polyhedra are a natural way to discretize bounded sets
- Develop a stochastic process of polyhedra to estimate HPD regions in a complex data set
- Respects discrete nature of the data, posterior distribution of HPD regions is given in terms of closed sets
- Easily extended to multi-modal data
- Build a stochastic process of polyhedra and use MHMC to develop this into a process of HPD regions



Preliminary Results

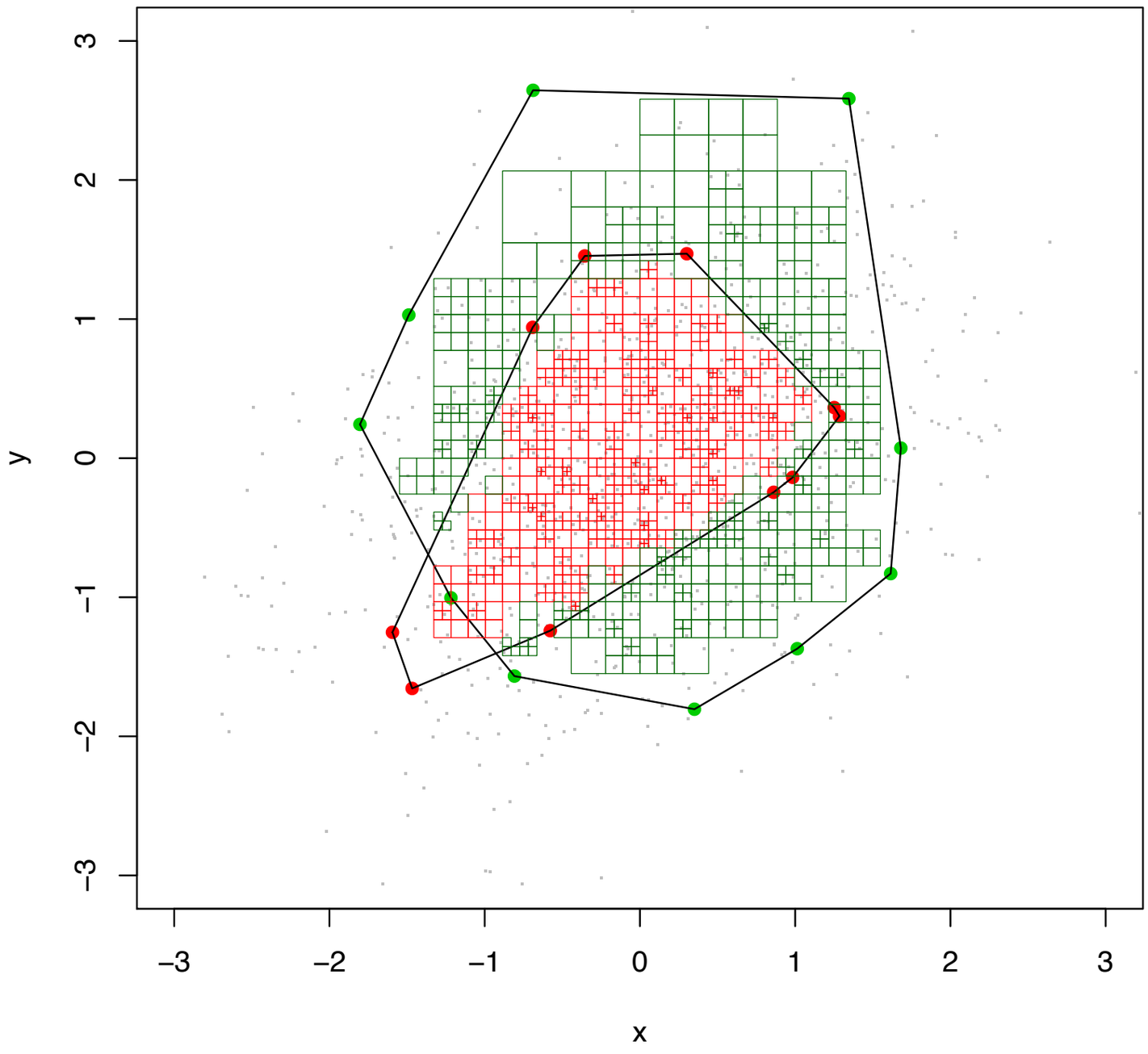


Consecutive samples from Markov Chain of polygons



Random draws from posterior distribution of HPD regions

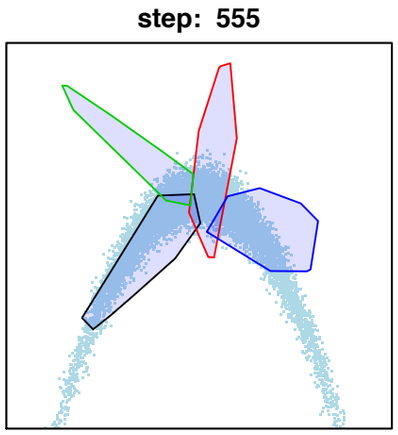
Multiple Polygons

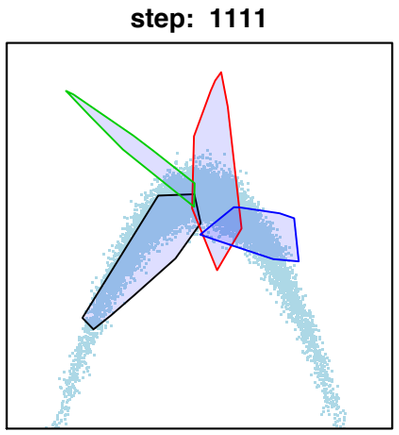


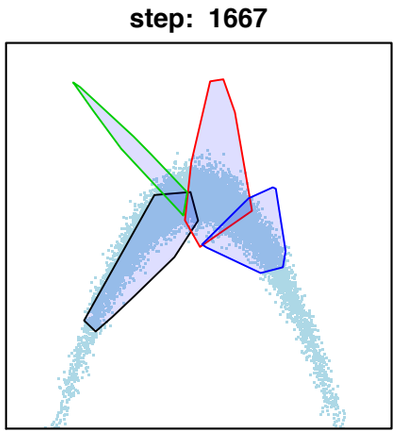
Polygons own the set of points they contain.

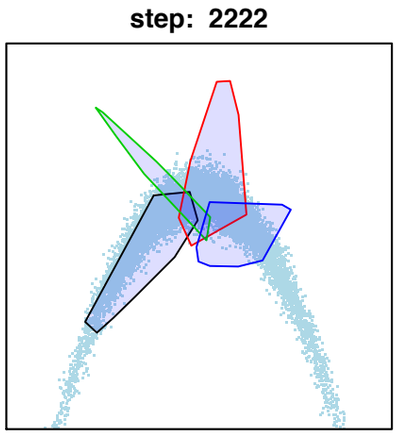
Overlapping point sets are owned on a first-come-first served basis

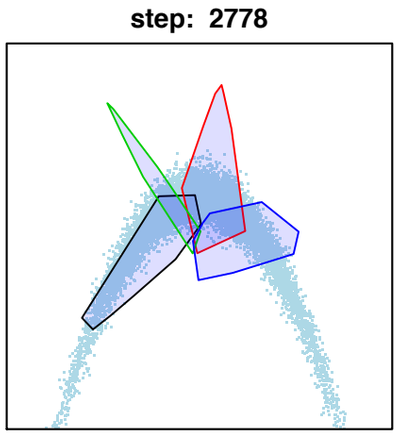
Multiple Polygons

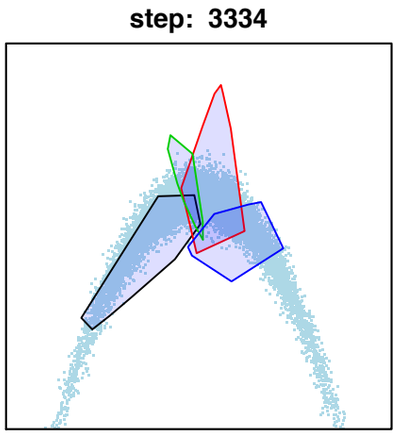


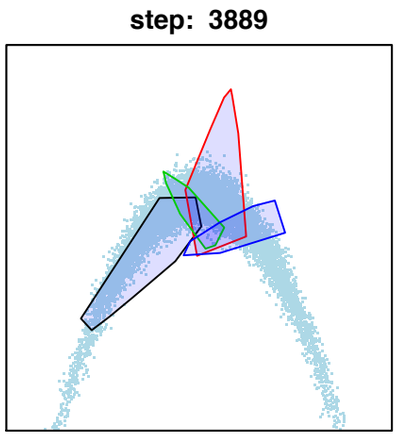


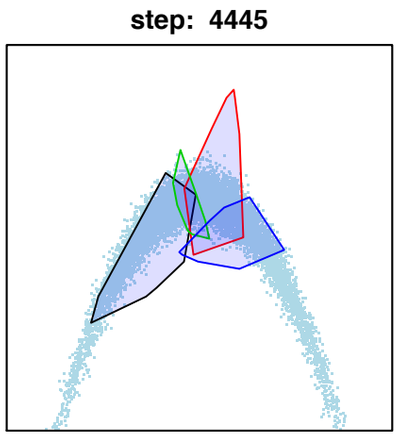


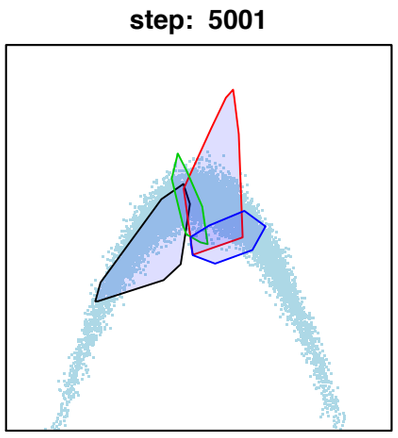












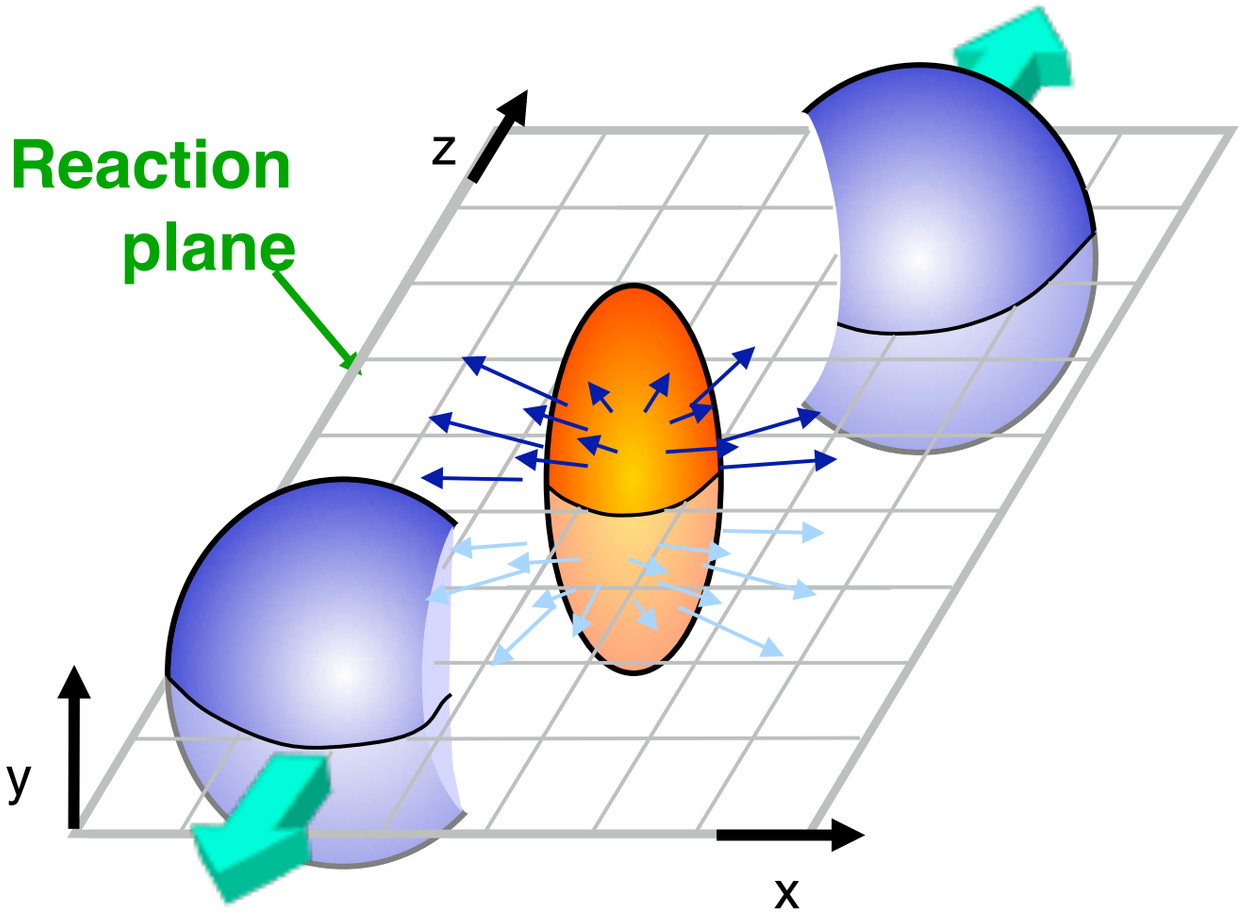
Works as intended on non-convex sets

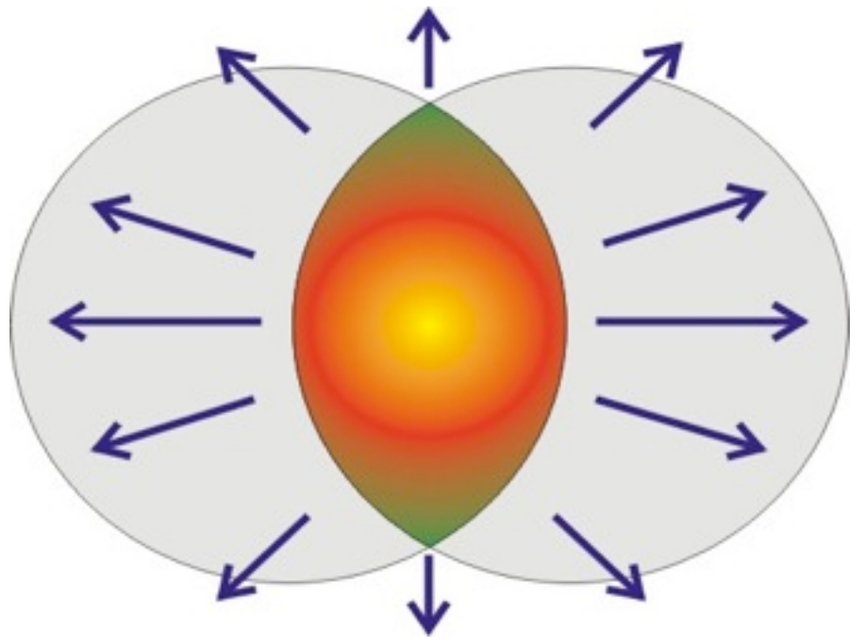
An R package is in preparation

Flow Coefficient Experiment (in progress)

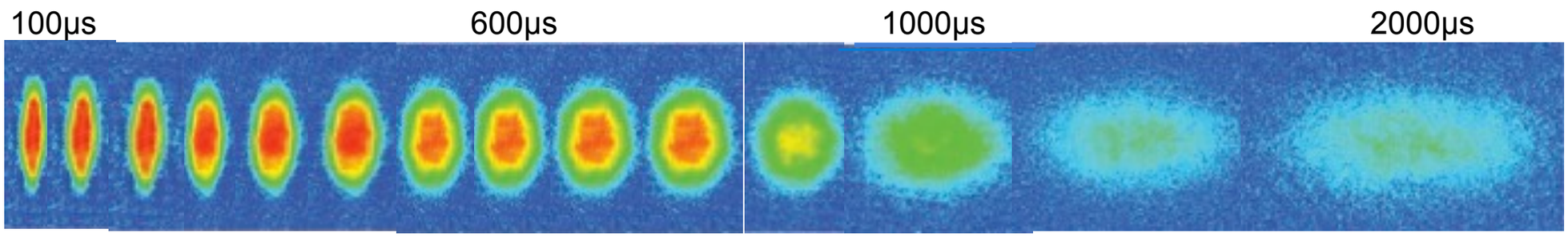
With R.Wolpert (Duke), S.Bass (Duke), U.Heinz (OSU)

Collision Geometry - Elliptic Flow



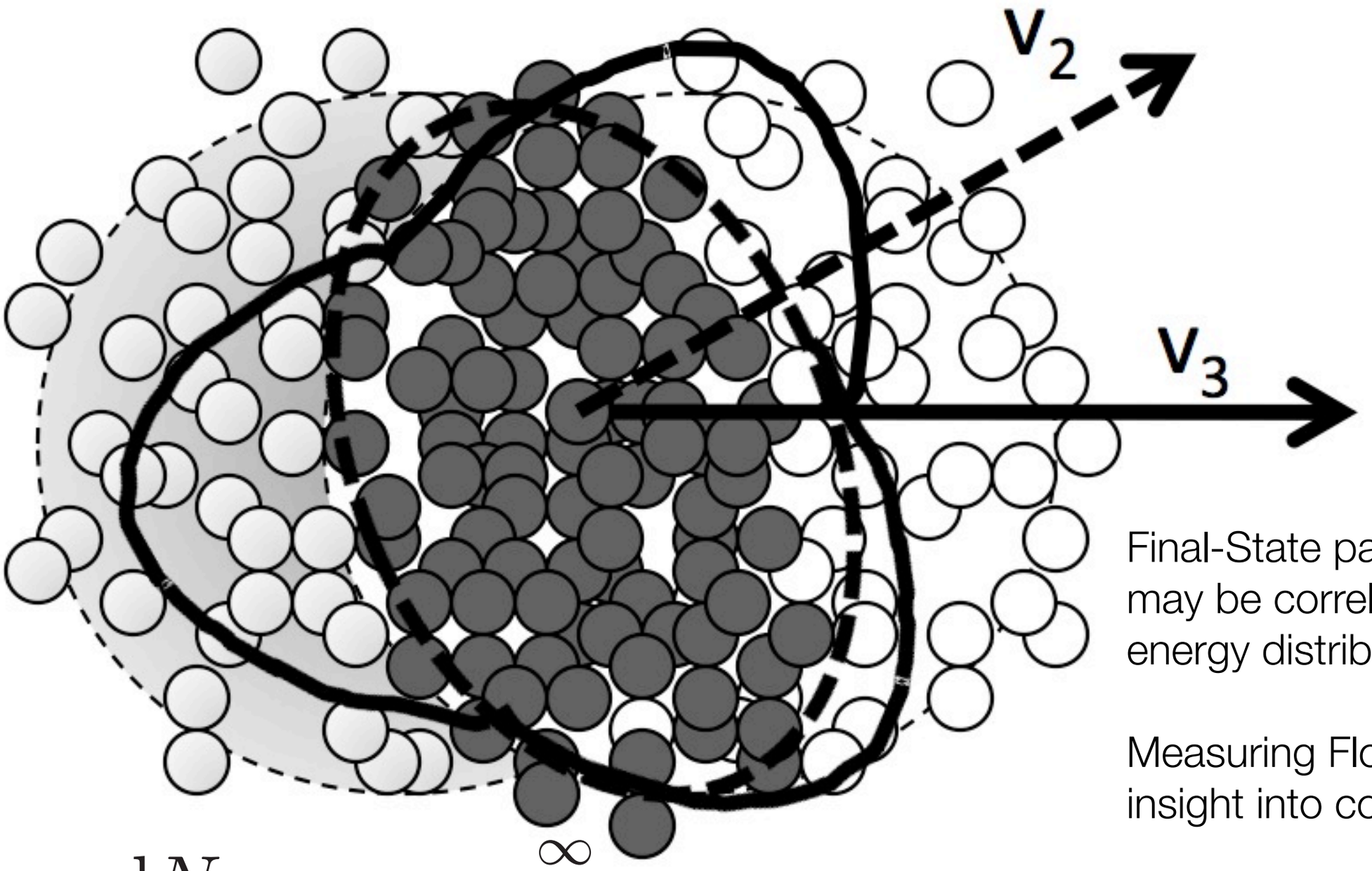


Flow is driven by pressure gradients



–M. Gehm, S. Granade, S. Hemmer, K, O'Hara, J. Thomas - **Science** 298 **2179 (2002)**

Flow Coefficients



Final-State particle distributions may be correlated with initial state energy distributions

Measuring Flow Coefficients gives insight into collision geometry

$$\frac{\mathrm{d}N}{\mathrm{d}\phi} \propto 1 + 2\sum_{n=1}^{\infty} v_n \cos n(\phi - \Phi_n) \ ,$$

Flow Analysis Computer Experiment

- Experimental data now available for Vn distributions, measured on a per event basis, used to be averaged over many collisions)
- Data available for many centrality classes. Try to model the centrality dependence along with the rest of the physics.
- VISHNew fluid dynamics + particle transport state of the art model. End to end simulation, collision -> hydro -> final state particle production
- A large set of calibration parameters, including fundamental properties of the QGP: shear-viscosity to entropy ratio.
- Use this new data set to calibrate the model and infer distributions on the shear-viscosity to entropy ratio. Never been rigorously done before
- Model-Data based Inference is the **only way we can learn about this**.

Transport Model Parameters

Parameter	Values
superMC	
model	Glauber or KLN
normalization factor	20–60 [Glb], 5-11 [KLN]
wounded nucleon / binary collision	0–1 [Glb]
Q_s exponent λ	~ 0.28 [KLN]
VISHNew	
equation of state	discrete choices
shear viscosity η/S	~ 0.08 [Glb], 0.16–0.20 [KLN]
thermalization time τ_0	0.2–1.0 fm
switching energy density $E_{\rm dec}$	~ 0.502 GeV/fm^3
relaxation time VisBeta	0.3–1.0

Table 1: Input parameters for the initialization and hydro models.

Initial Condition & Collision Model Details

Fluid Transport model

9 Important Parameters
2 Discrete

Experimental Data

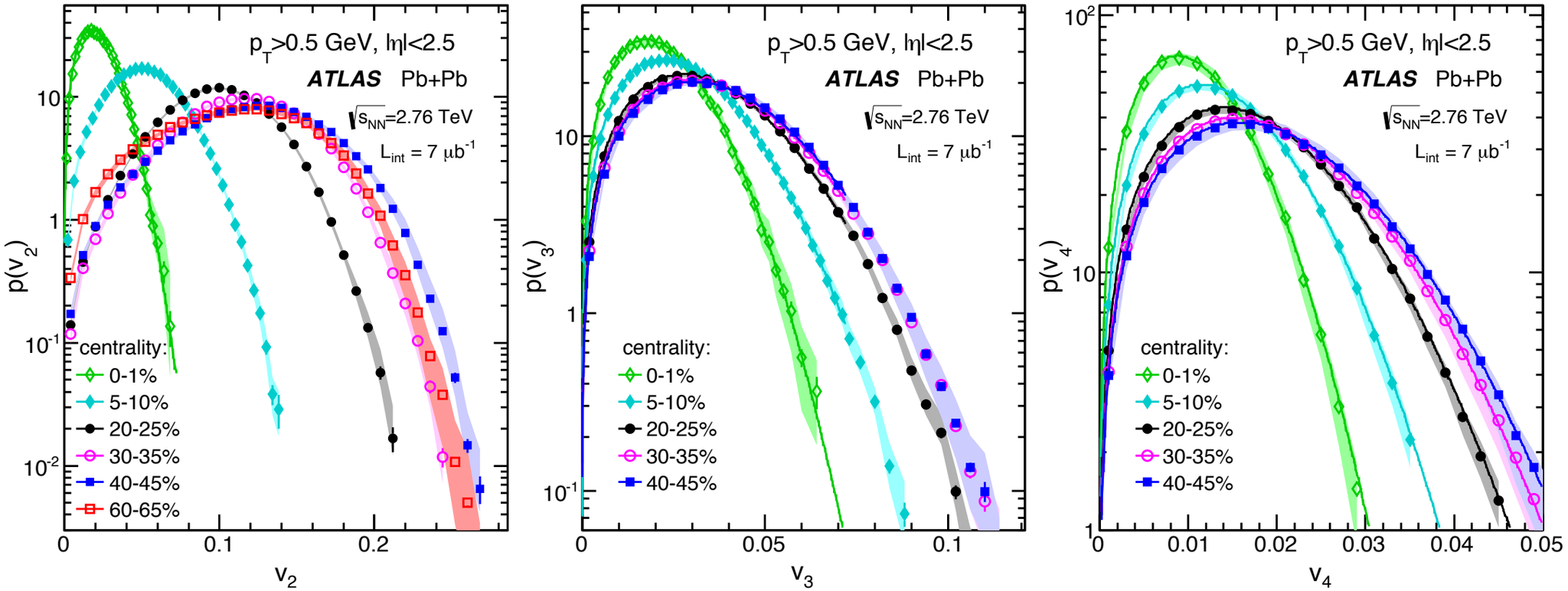


Figure 10. The probability density distributions of the EbyE v_n in several centrality intervals for $n = 2$ (left panel), $n = 3$ (middle panel) and $n = 4$ (right panel). The error bars are statistical uncertainties, and the shaded bands are uncertainties on the v_n-shape. The solid curves are distributions calculated from the measured $\langle v_n \rangle$ according to eq. (1.6). The solid curve is shown only for 0–1% centrality interval for v_2, but for all centrality intervals in case of v_3 and v_4.

Modeling the experimental data

- ATLAS suggests Bessel-Gaussian for Vn distributions (not a very good fit)
- use GRW distribution, fit experimental distributions of flow coefficients

$$f_X(x;\ m, s, \alpha, \gamma) = \frac{\alpha}{|s|\Gamma(\gamma)} \left(\frac{x-m}{s}\right)^{\alpha\gamma-1} \exp\left\{-\left(\frac{x-m}{s}\right)^{\alpha}\right\}$$

- gives a better handle on expt-model comparison
  - fit model output with same process
  - build a likelihood for the model fit-parameters
- Now we can compare estimates of a few numbers instead of entire functional forms