Extracting Rigorous Conclusions from Model-Data Comparisons

Scott Pratt, Michigan State University

MADAI Collaboration

Models and Data Analysis Initiative

http://madai.us

Cyber-enabled Discovery and Innovation

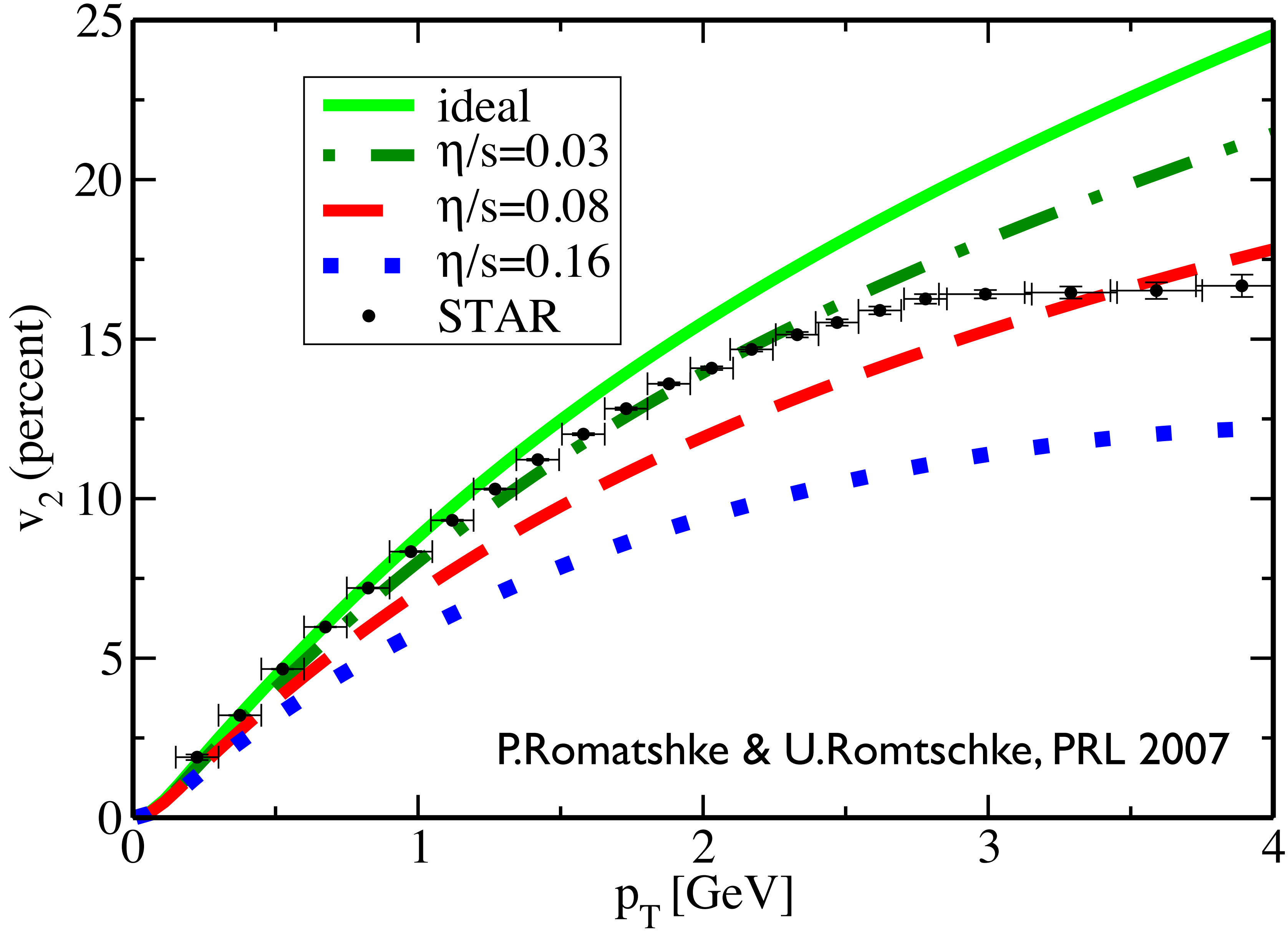
MICHIGAN STATE UNIVERSITY Duke UNIVERSITY THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL renci

1st MADAI Collaboration Meeting, SANDIA 2010

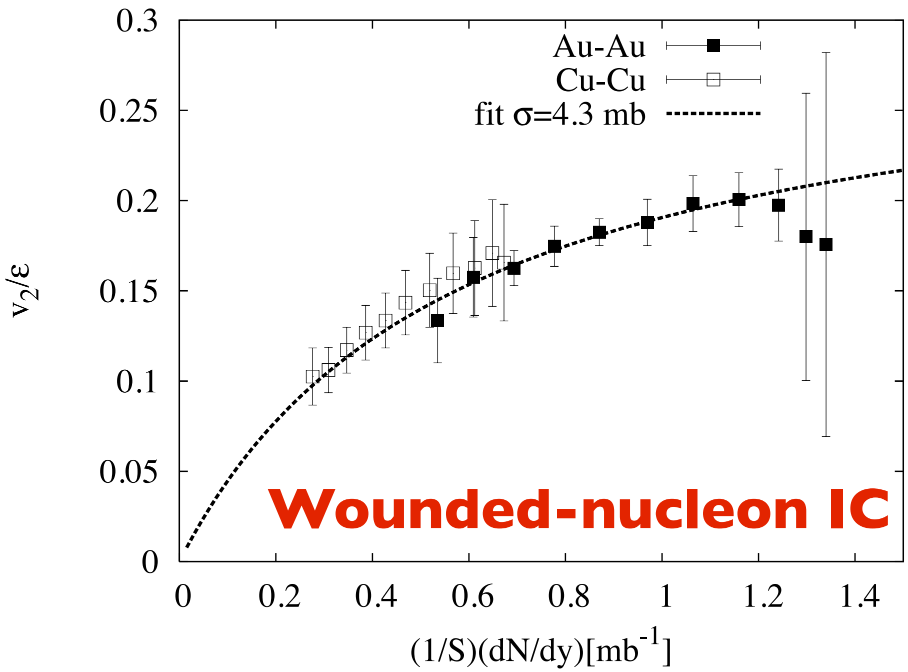
How to write a PRL for RHIC physics

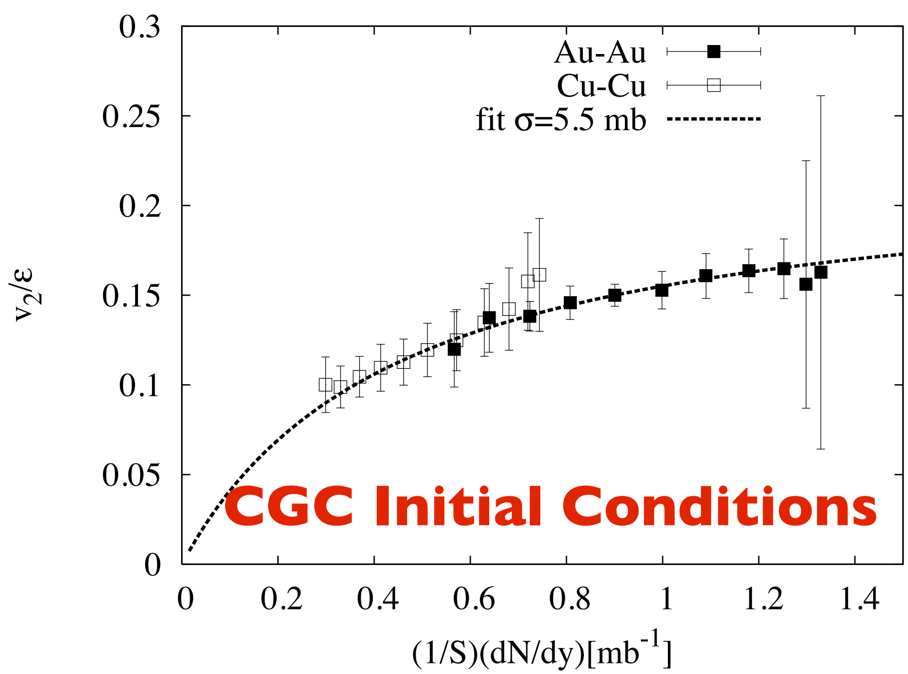
1. **Identify a strong relationship between parameter *x* and observable *y* (e.g., η and v_2)**
2. **Run your model for several values of *x* and show that *y* changes**
3. **Add hand-waving arguments that this is the principal sensitivity**
4. **Wait for someone else to show other parameters affect *y***

Example: v_2 and η/s



Drescher, Dumitru, Gombeaud &Ollitrault PRC 2007





v2 depends on ....

- η/s
- saturation model
- pre-thermal flow

weakly on ....

- T-dependence of η/s

~independent of ....

- initial T_{xx}/T_{yy}

J.Vredevoogd, ArXiv 1307.7677

Shortcomings of approach

- **Many parameters, *x*, affect any given observable, y E.g. v_2 affected by viscosity, saturation parameters, initial flow...**
- **Each parameter affects several observables**
  $$\vec{y}(\vec{x})$$
- **Can't quantitatively determine parameters**

Some parameters (bulk physics)

- **Shear Viscosity (*T*,ρ dependent)**
- **Bulk Viscosity (*T*,ρ dependent)**
- **Eq. of State (*T*,ρ dependent)**
- **Initial Conditions:**
  choice of parameterization/model, initial flow, saturation scale, energy normalization, rapidity width, baryon stopping
- **Hadronization parameters: T_{had}, fugacity**
- **In-medium screening of σ**
- **Could easily exceed a dozen**

Some observables (soft)

- Yields and Spectra
- v_2, v_3, v_4...
- Femtoscopic Radii
- Can reduce functions to a few numbers. E.g. spectra → yield, $\langle p_T \rangle$, $\langle p_T^2 \rangle$
- long-range correlations, e.g. charge balance
- All depend on species, p_T , y, centrality, beam particles, beam energy
- Perhaps ~100 observables

Markov Chain Monte Carlo (MCMC)

- **Find sample "posterior" distribution: collection of *x* weighted by likelihood**

$$\vec{L}(\vec{x}) \propto \exp\left\{-\sum_a \frac{(y_a(\vec{x}) - y_a^{\exp})^2}{2\sigma_a^2}\right\}$$

- **σ_a incorporates both experimental & model uncertainties**

- **Typical algorithm Metropolis:**
  1. Calculate *L(x)*
  2. Calculate *L(x+δx)*
  3. If *L(x+δx)*> *L(x)* or *L(x+δx)/L(x)*<*Random*: keep
  otherwise: reject
  Repeat .... perhaps millions of times

Markov Chain Monte Carlo (MCMC)

- **Posterior distribution = sampling of *x* consistent with *L*(*x*)**
  From sampling -> $\langle x \rangle$, covariances ..
- **Each point might requires running model with sufficient stats for each impact parameter, beam energy, ...**
- **Could take multiple CPU days for each point**
- **NOT TRACTABLE!!!**

Emulator / Surrogate Model / Metamodel / Interpolator

- **Run full model ~1000 times semi-randomly throughout prior (Latin hyper-cube sampling)**
- **Build emulator:** $y^{\mathrm{emu}}(\vec{x}) \approx y^{\mathrm{mod}}(\vec{x})$
- **Examples emulator schemes:**
  Gaussian process, linear/quadratic fit...

Building emulator

- **Determine principal components (PCA)**

$$y_a(\vec{x}) \to \tilde{y}_a(x) \equiv \frac{y_a(\vec{x}) - \langle y_a \rangle}{\sigma_a}$$

Rotate y to diagonalize covariance

$$\langle \tilde{y}_a \tilde{y}_b \rangle \to \langle z_a z_b \rangle = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{nn} \end{pmatrix}$$

Only emulate components with λ≳1

- **Determine hyper-parameters: (e.g. linear fit parameters) for each z_a**

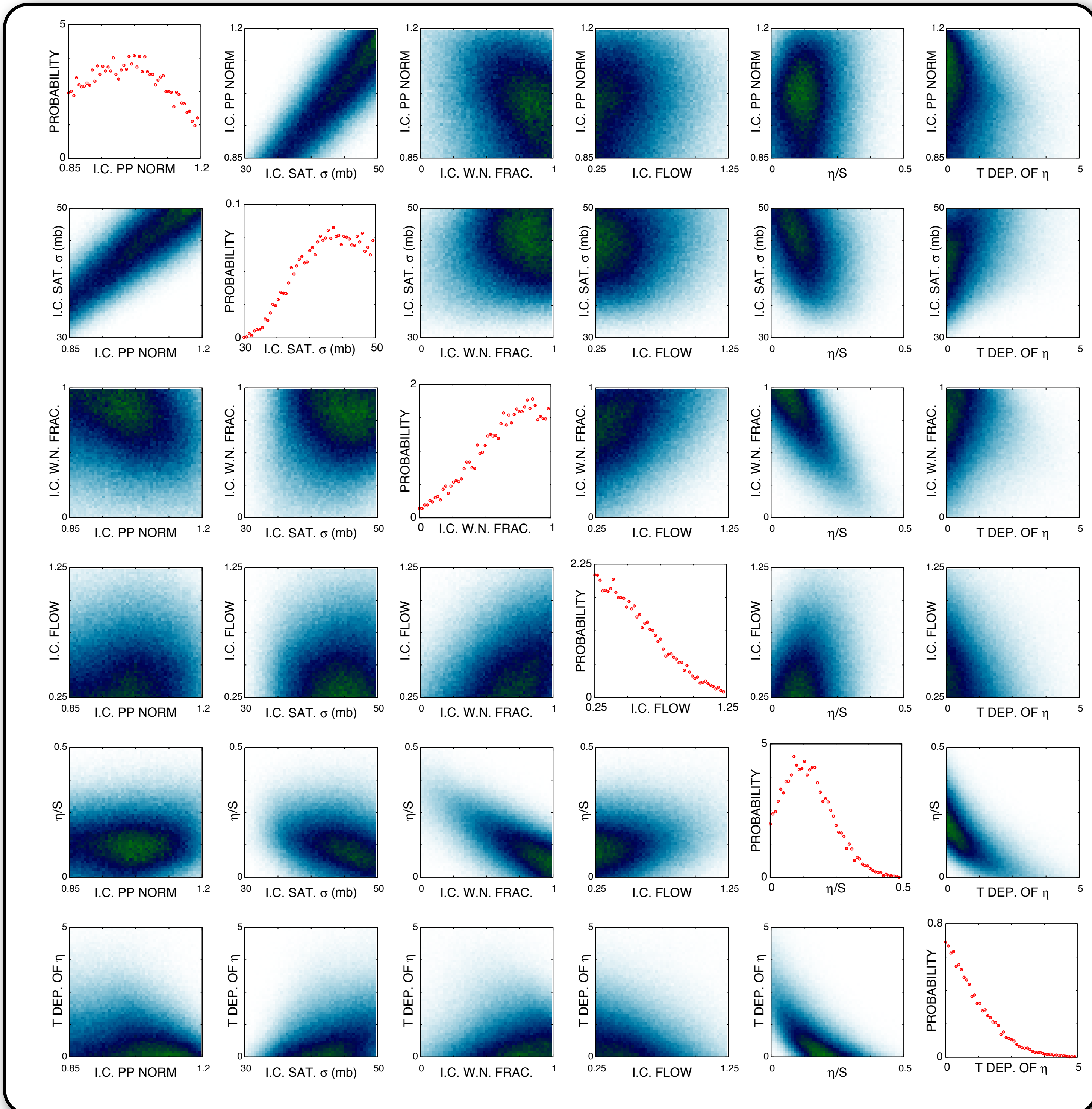
$$\vec{L}(\vec{x}) \approx \exp\left\{ \frac{1}{2} \sum_a (z_a^{\mathrm{emu}}(\vec{x}) - z_a^{\exp})^2 \right\}$$

Test Calculation

J.Novak, K. Novak, S.Pratt, C.Coleman-Smith & R.Wolpert, ArXiv:1303.5769

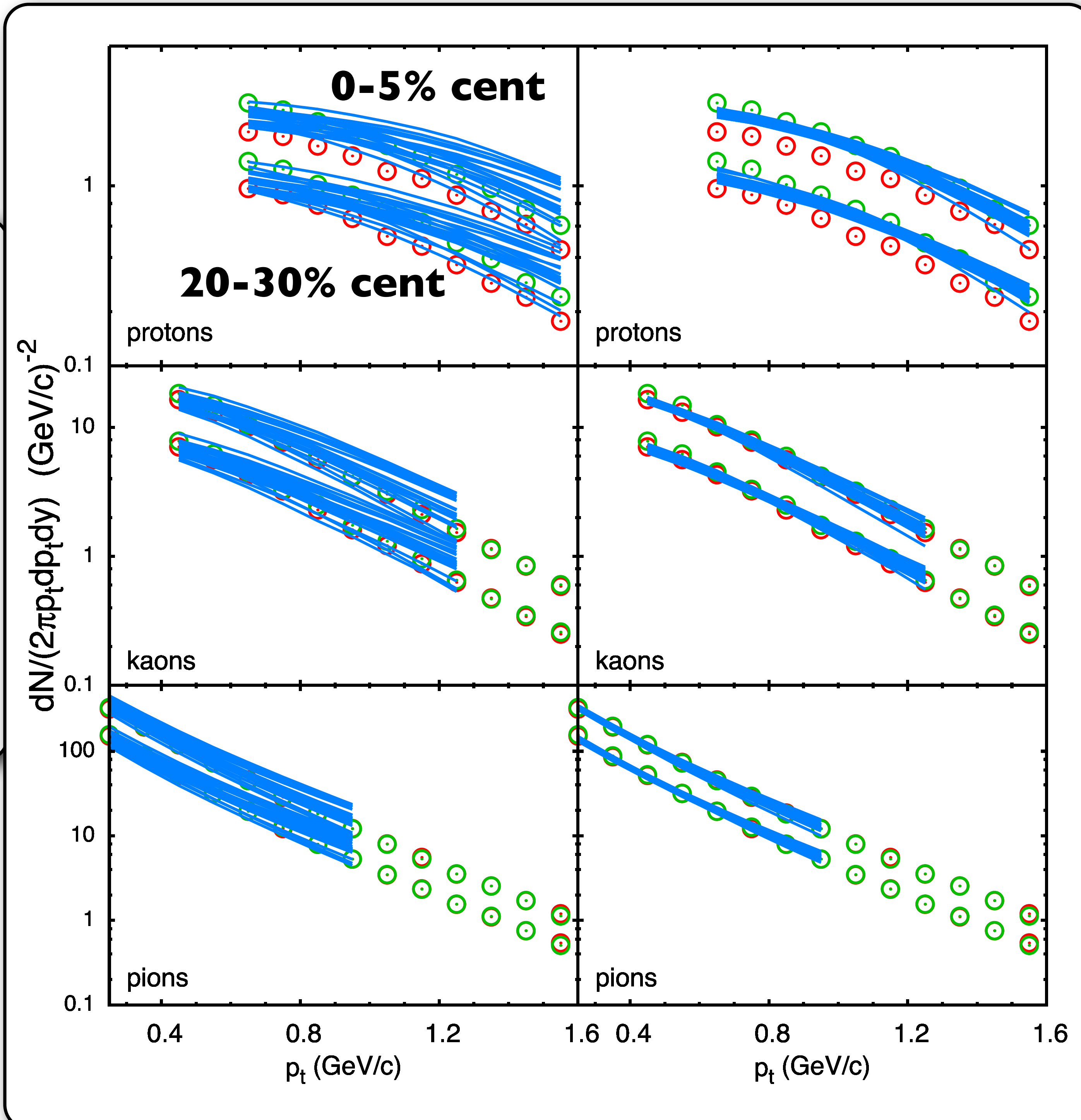
- **Parametrized IC, IS Hydro + Cascade**
- **100 GeV/c + 100 GeV/c Au+Au from RHIC**
- **π, K, p spectra -- π HBT radii -- π v_2**
- **6 parameters: η, *T*-dependence of η, saturation σ, WN vs. CGC weight, ε normalization, initial flow**

1D and 2D Posterior Projections

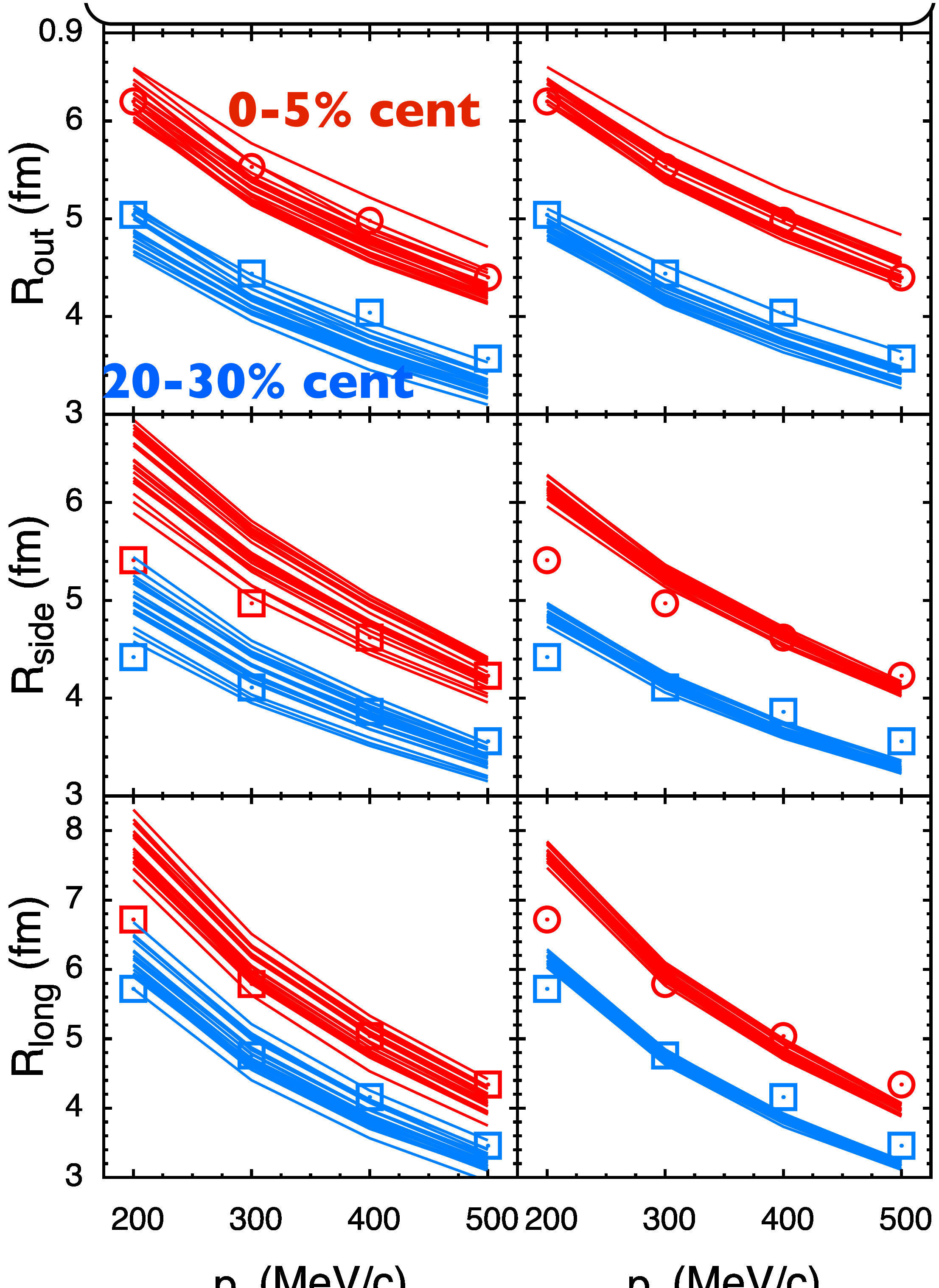


J.Novak, K. Novak, S.Pratt, C.Coleman-Smith & R.Wolpert, ArXiv:1303.5769

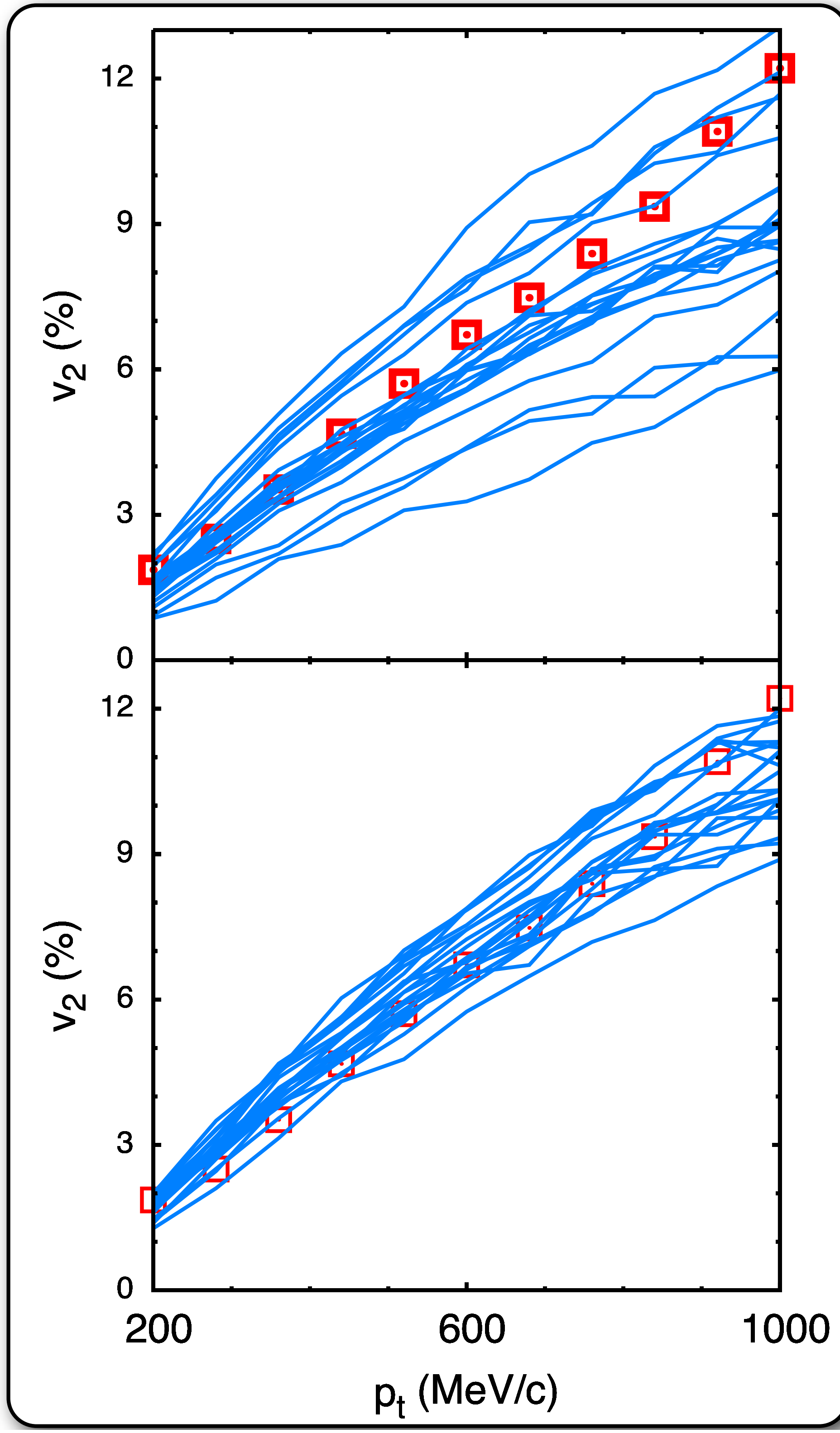
Sample Spectra from Prior and Posterior



Sample HBT from Prior and Posterior

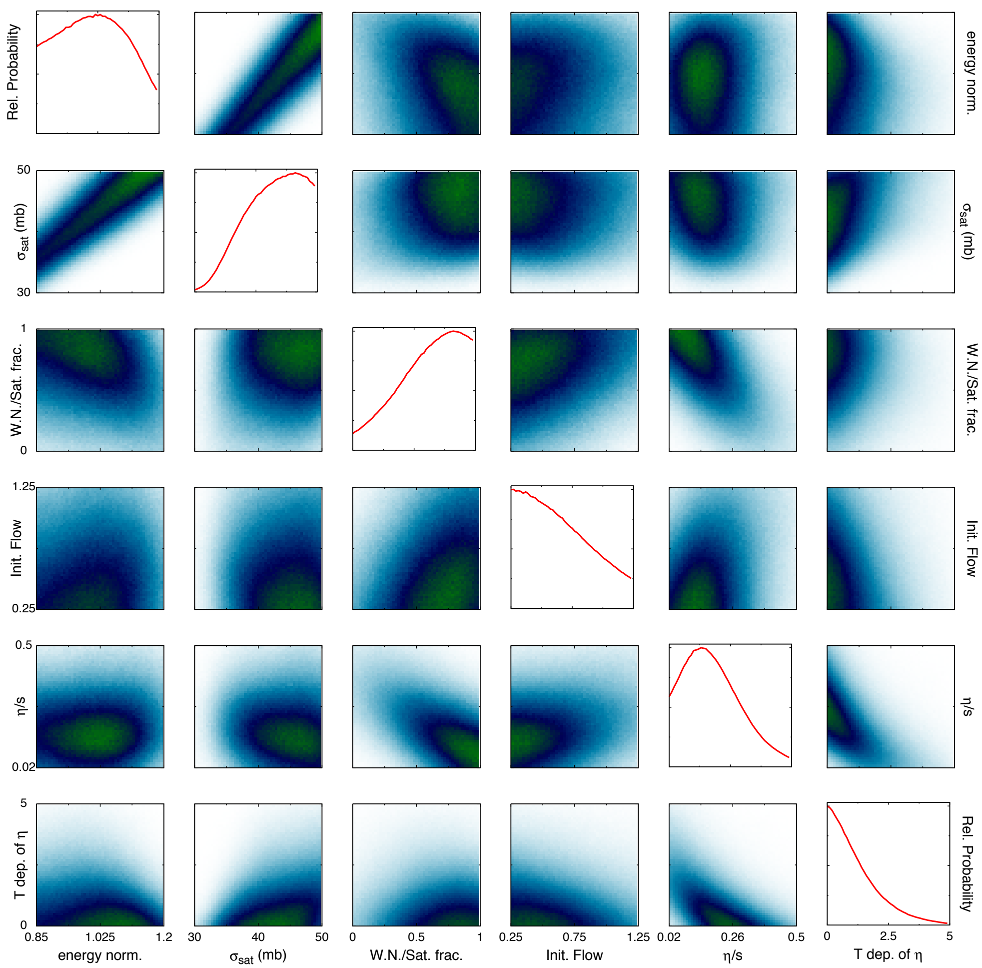


Sample v_2 from Prior and Posterior

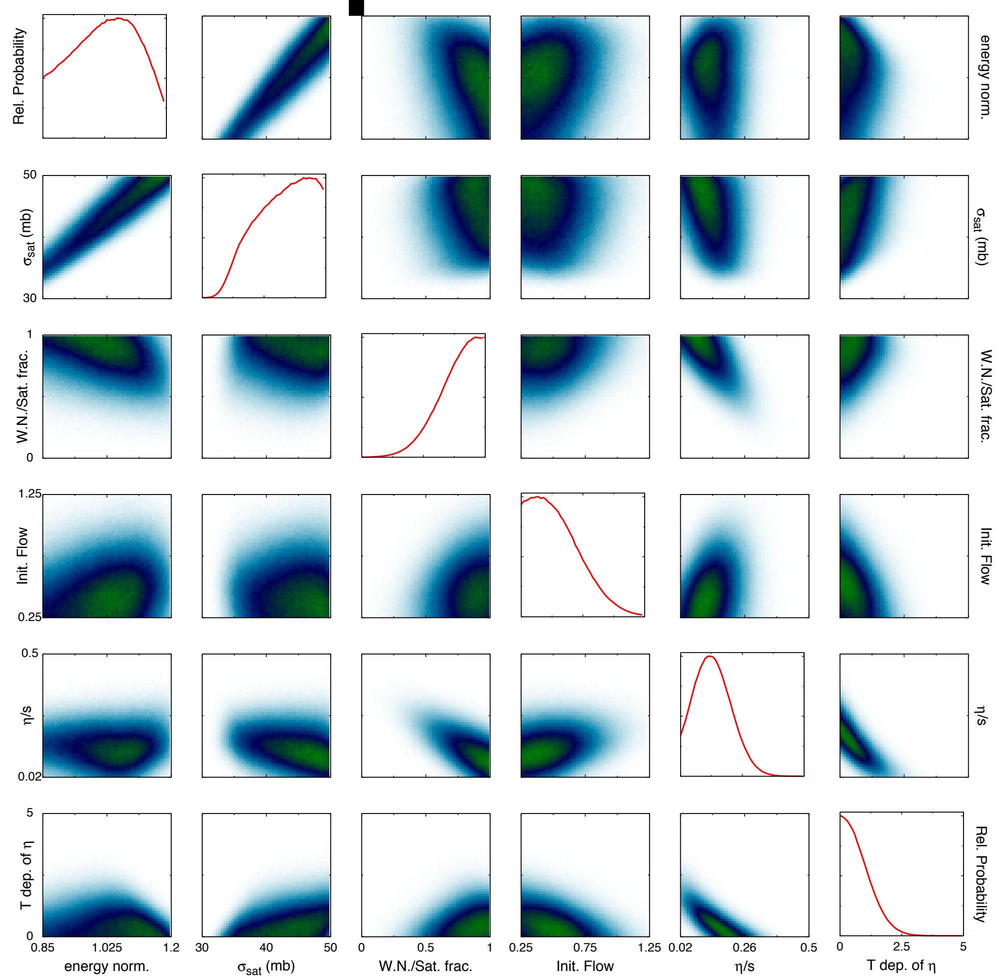


Sensitivity to Uncertainties

Pessimistic



Optimistic



Proof of Principle was successful

- Robust, Repeatable
- Emulation works splendidly
- Method scales well to more parameters & more data

Numerical Requirements for 10-20 parameters

- Run Code for 2000 points in parameter space
- 20 impact parameters / beam energies / A+B
- 400 fluctuations × 100 cascade events
  =40,000 cascade events & 400 hydro runs
  for each x, b, A+B, E/A...
- 1.6×10^{9} cascade runs, 1.6×10^{7} hydro runs

Numerical Requirements

- 10,000 CPU days for cascade if 0.5 seconds/event (perform example run)
- 40,000 CPU days for hydro if 5 minutes/event
- Important to optimize for multi-core / GPU
- Storage ~ 2 MB/event $\Rightarrow$ 5 PB
  Is it cheaper to regenerate cascade data?
- Not trivial, but tractable
- Inefficient codes/coding wastes many $€£¥...

Modular Modeling

- **Initial Condition (parametric)**
  fully 3D, non-zero baryon number
  3D dynamical models still in flux
- **Viscous 3D Hydro**
- **Hadron Cascade**
- **HBT / EM / Jets after-burners**
- **Statistical Analysis**

Physics (Soft) To-Do List

- **I.C. → Much to Do for 3D/non-zero B**
- **Hydro → seamless approach for 3D & B≠0 for 5 GeV < *E/A* < 14 TeV**
- **Cascade → π Bose factors, baryon annihilation, non-equilibrium hadronization, absorption into hydro region, viscous corrections, mean field, consistent time delays**
- **HBT → good shape**
- **EM → ???**
- **Jets → ???**

Justifying the Approach

STATS approach works because:

- Response to parameters is smooth

Parameterization of physics works because

- Israel-Steward Hydro is flexible
  - handles wide variety of IC
  - Eq. of Motion mainly driven by cons. laws
- Only modest sensitivity to IC
- Main physics assumptions:
  - Components flow together in hydro
  - Close enough to Navier-Stokes so that relaxation to NS is justified
  - Hadronization is close to chem. equil. (aside from quark fugacities)