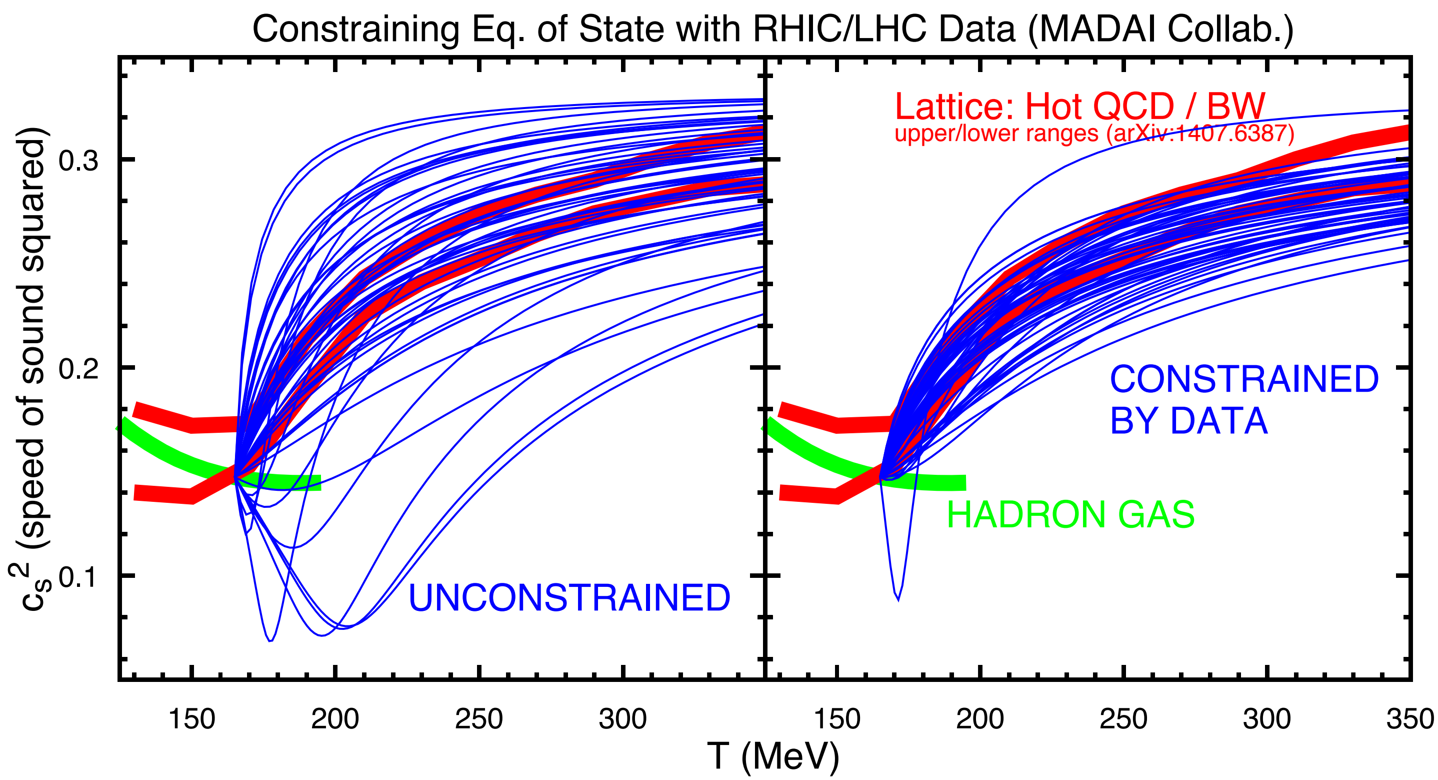
Correlated Errors



Scott Pratt, Michigan State University

Goals & Challenges

Parameter Determination:

- **Determine Full Likelihood**
- **Many (dozen(s)) parameters**
- **Computationally expensive models**
- **Heterogenous data sets**

Notation

x_i : **model parameters**

$y_a^{(\mathrm{exp})}$: **observables**

$y_a^{(\mathrm{mod})}(\vec{x})$: **model predictions**

$y_a^{(\mathrm{emu})}(\vec{x})$: **emulator**

z_a : **principal components**

S.P, E.Sangaline, P.Sorenson & H.Wang, PRL 2014

MADAI Analysis of Relativistic Heavy Ion Collisions

Choose observables y_a from RHIC/LHC

Assign uncertainties

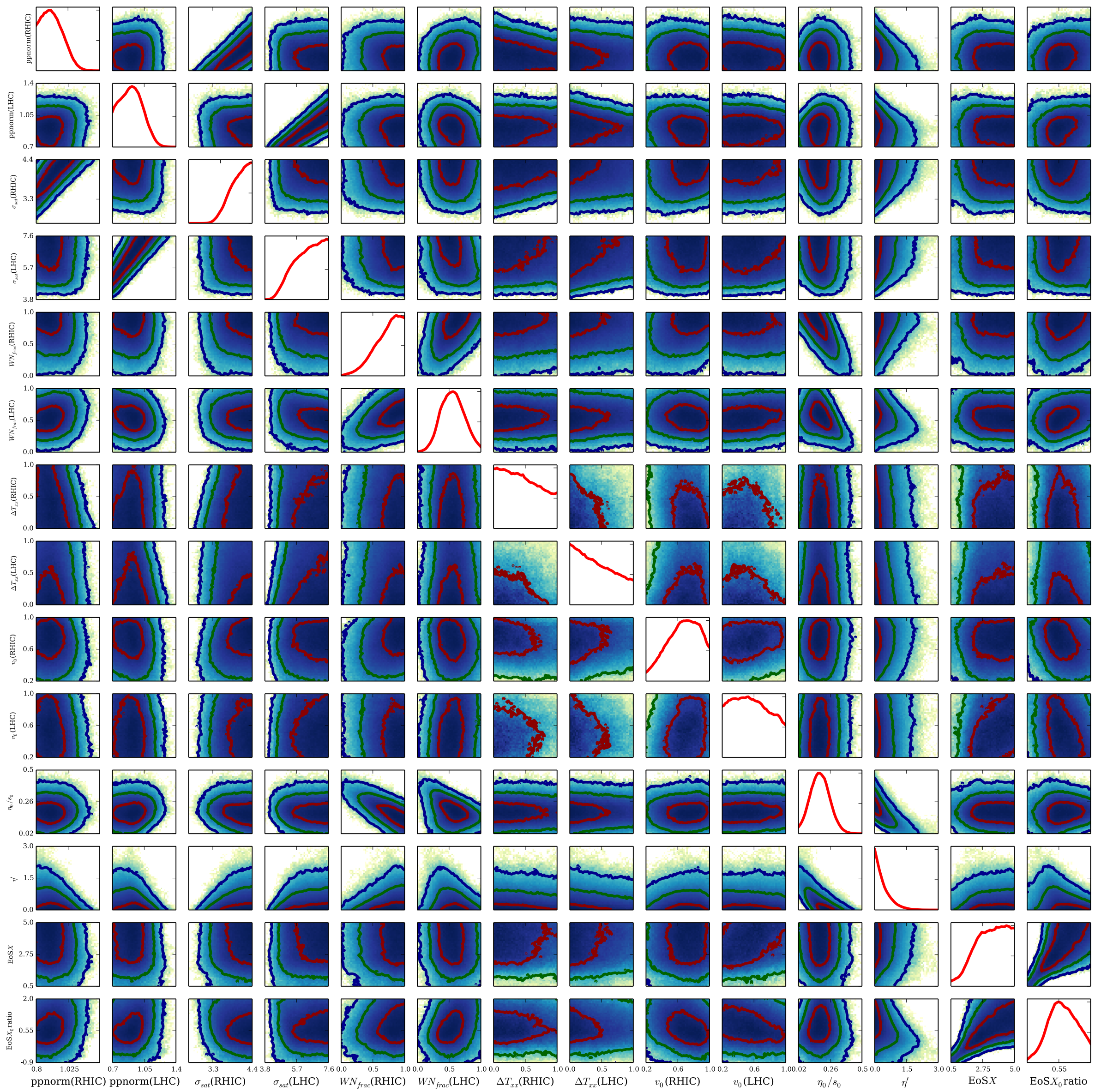
PCA: $y_a \rightarrow z_a$

Construct Emulator

$$z_a^{(\mathrm{emu})}(\vec{x}) \approx z_a^{(\mathrm{mod})}(\vec{x})$$

Perform MCMC:

$$\mathcal{L} \sim \exp\left\{ -\frac{1}{2} \sum_a (z_a^{(\mathrm{mod})} - z_a^{(\mathrm{exp})})^2 \right\}$$



S.P, E.Sangaline, P.Sorenson & H.Wang, PRL 2014

MADAI Analysis of Relativistic Heavy Ion Collisions

Choose observables y_a from RHIC/LHC

Assign uncertainties

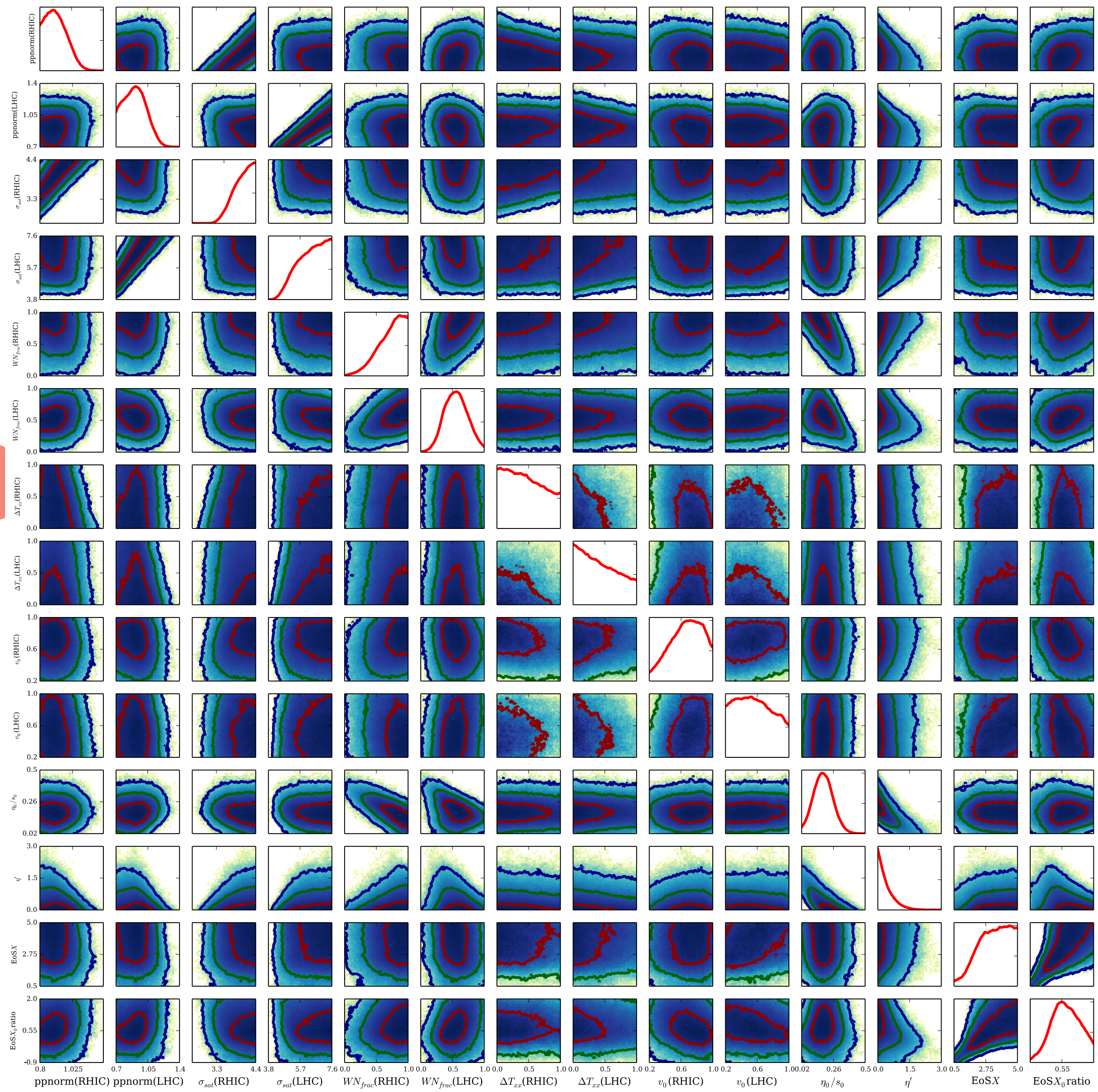
PCA: $y_a \rightarrow z_a$

Construct Emulator

$$z_a^{(\mathrm{emu})}(\vec{x}) \approx z_a^{(\mathrm{mod})}(\vec{x})$$

Perform MCMC:

$$\mathcal{L} \sim \exp\left\{ -\frac{1}{2} \sum_a (z_a^{(\mathrm{mod})} - z_a^{(\mathrm{exp})})^2 \right\}$$



$$\mathcal{L} \sim \exp\left\{-\frac{1}{2}(y_a^{(\mathrm{exp})} - y_a^{(\mathrm{emu})})\Sigma_{ab}^{-1}(y_b^{(\mathrm{exp})} - y_b^{(\mathrm{emu})})\right\}$$

Σ_{ab} combines uncertainties:

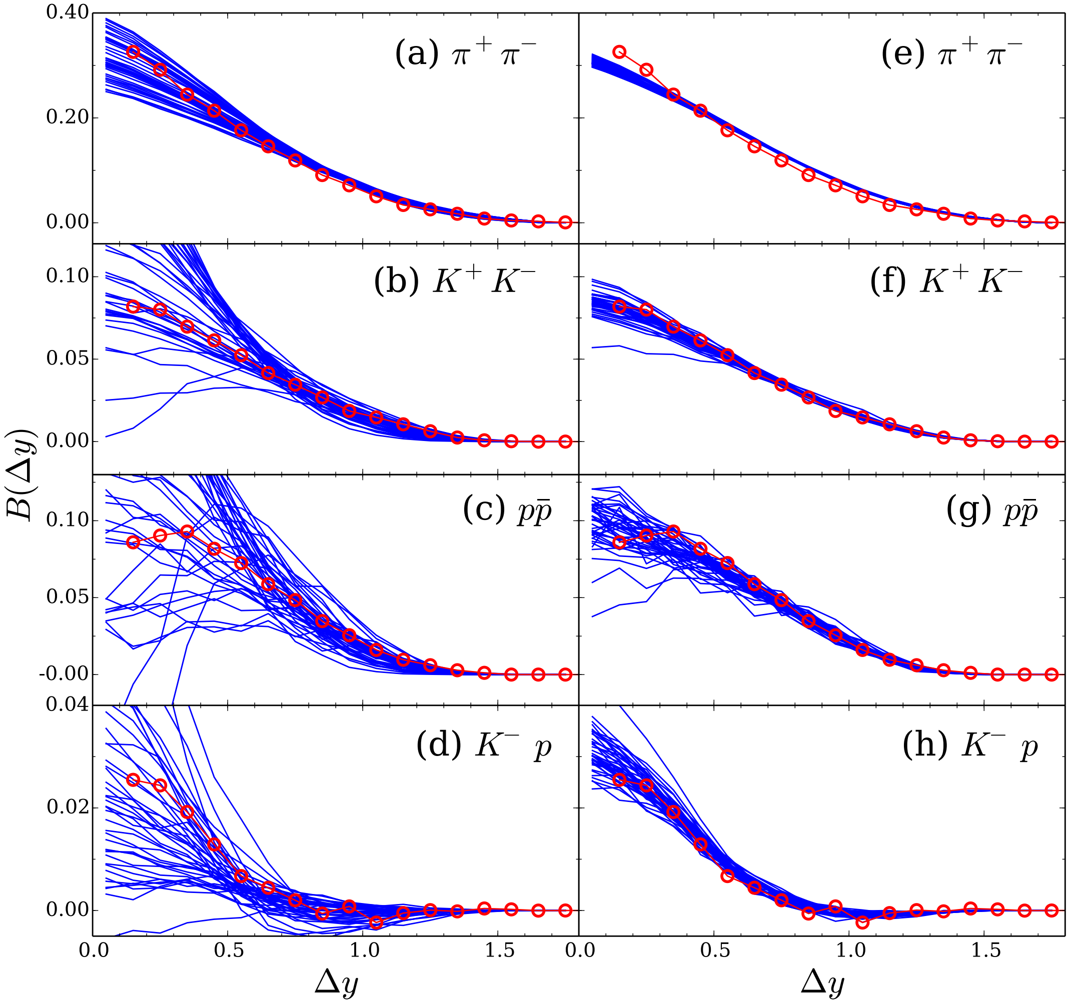
- **aleotoric, both theory and experiment**
- **experimental systematic**
- **model systematic (missing physics)**
- **emulator accuracy**
- **correlated errors (off-diagonal elements)**

Bigger uncertainties ➜ less constrained parameter space

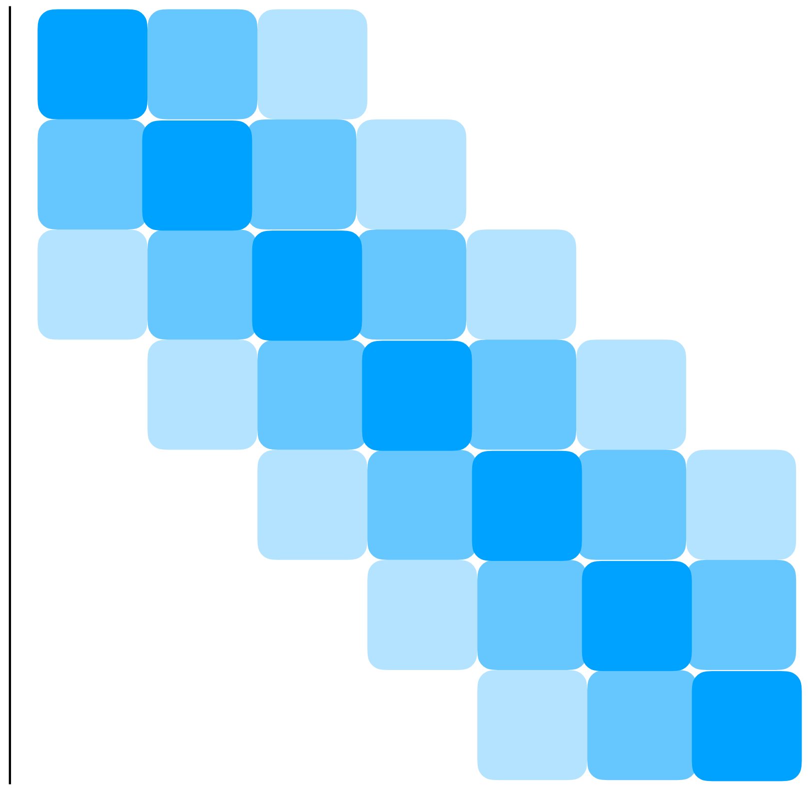
Four Strategies for Correlated Errors

1. Keep full error matrix
2. Exaggerate errors
3. Distillation
4. Nuisance parameters

I. Keep Full Error Matrix



Neighbors correlated



Gaussian process? (Furnstahl)

I. Keep Full Matrix - PCA modified

Transform y $\leftrightarrow$ z

$$\Sigma'_{ad} = U_{ab}\Sigma_{bc}U^{\dagger}_{cd} = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix},$$

$$y'_a = U_{ab}y_b$$

$$\tilde{y}_a = y'_a/\sigma_a$$

average over prior

$$\Lambda_{ad} = \tilde{U}_{ab}\langle\delta\tilde{y}_b\ \delta\tilde{y}_c\rangle\tilde{U}^{\dagger}_{cd} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} = \langle z_a z_b\rangle$$

$$z_a = \tilde{U}_{ab}\tilde{y}_b$$

Only emulate z_a observables with $\lambda_a \gtrsim 1$

$$\mathcal{L} \sim \exp\left\{-\frac{1}{2}\sum_a (z_a^{(\mathrm{mod})} - z_a^{(\mathrm{exp})})^2\right\}$$

II. Exaggerate Errors

If $y_1, y_2, y_3, \ldots$ are redundant:

$$\Sigma = \Sigma_{aa} \left(\begin{array}{ccc} 1 & \cdots & 1 \\ 1 & \cdots & 1 \\ 1 & \cdots & 1 \end{array} \right)$$

$$y = y_1 = y_2 = y_3 \ldots = y_n$$
$$P(y) \sim \exp\left\{-y^2/2\sigma^2 - y^2/2\sigma^2 \ldots - y^2/2\sigma^2\right\}$$
$$\langle \delta y^2 \rangle = \sigma^2/n,$$
$$\sigma^2 = n\Sigma_{aa}$$

You can ignore off-diagonal elements, but set

$$\Sigma_{ab} \rightarrow n\Sigma_{aa}\delta_{ab}$$

If you think strong correlation extends ~ *n* points, just increase σ by $n^{1/2}$

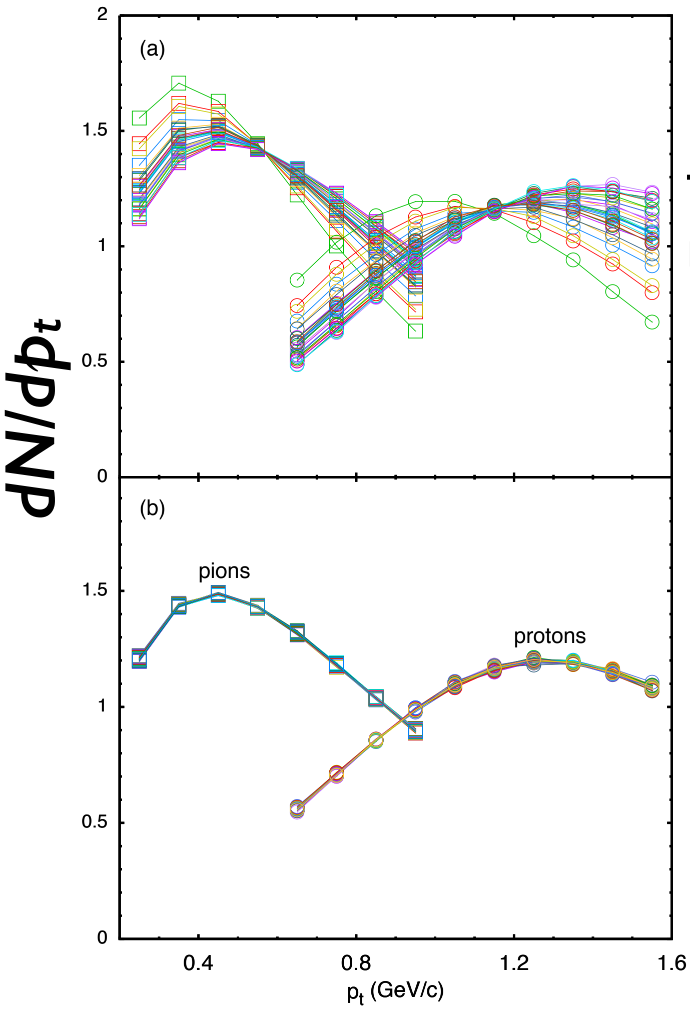
III. Data Distillation

Strategy:
Use minimal number of points

MADAI EoS analysis:
26 plots → 30 observables

III. Data Distillation

π,p spectral *SHAPES*



Example (MADAI EoS):
Reduce spectra to 2 numbers $\langle p_t \rangle$ & yield

from 30 points in parameter space:
randomly from prior

74 pion spectra:
with 573<$\langle p_t \rangle$< 575 MeV

44 proton spectra:
with 1150<$\langle p_t \rangle$< 1152 MeV

III. Data Distillation

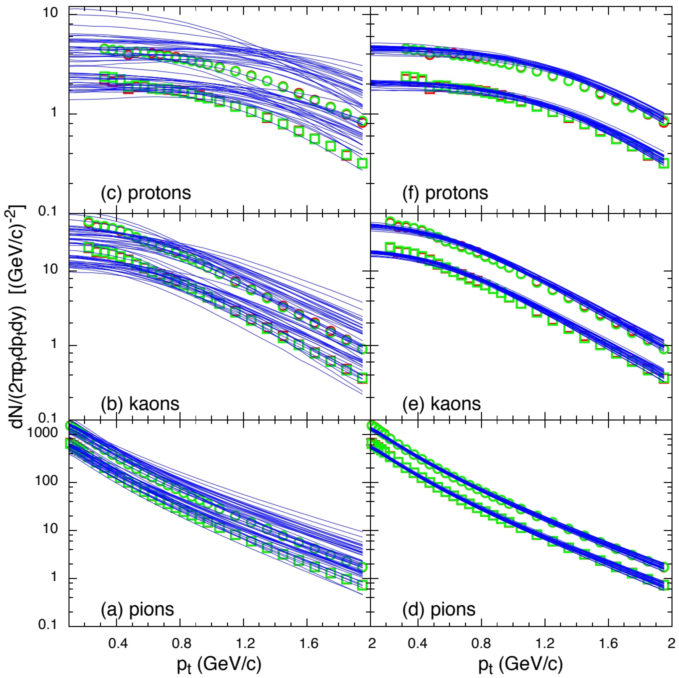
Observables can be chosen by PCA

Consider y_a for one plot

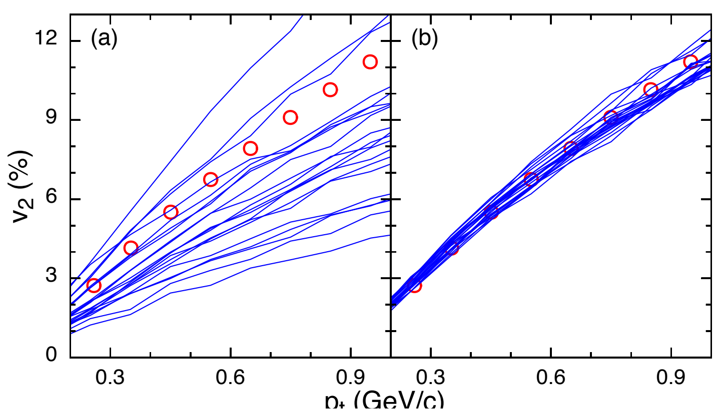
$$\langle \delta y_a \delta y_b \rangle \rightarrow \langle \delta z_a \delta z_b \rangle = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix}$$

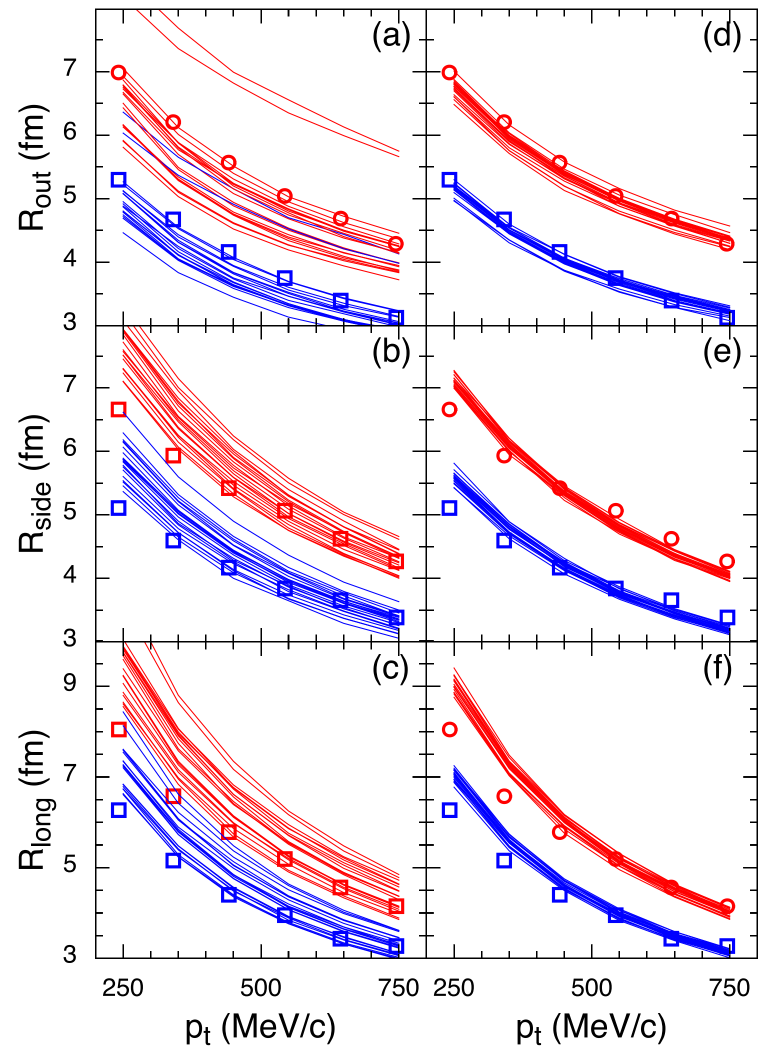
- **Different z_a probably have ~ uncorrelated errors**
- **PCA suggested spectral-shape information carried by one parameter ~$\langle p_t \rangle$**

MADAI EoS Analysis: 26 plots, 30 observables

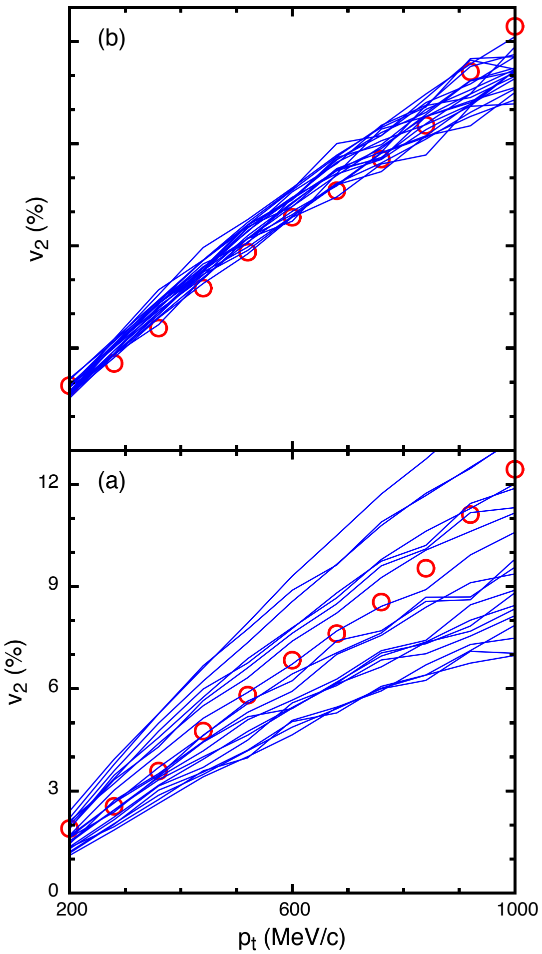


LHC DATA

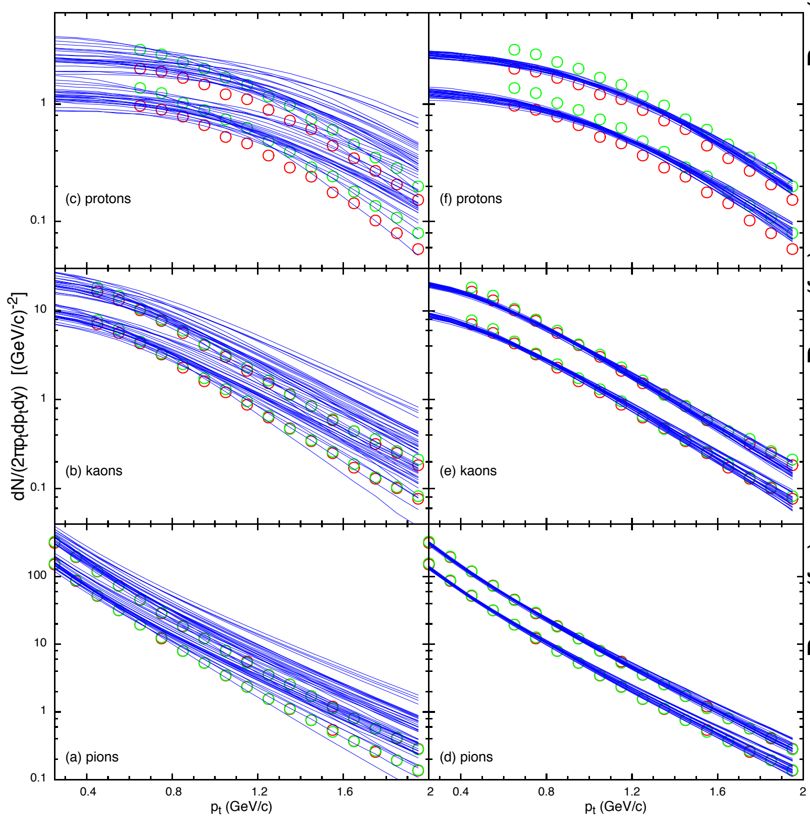


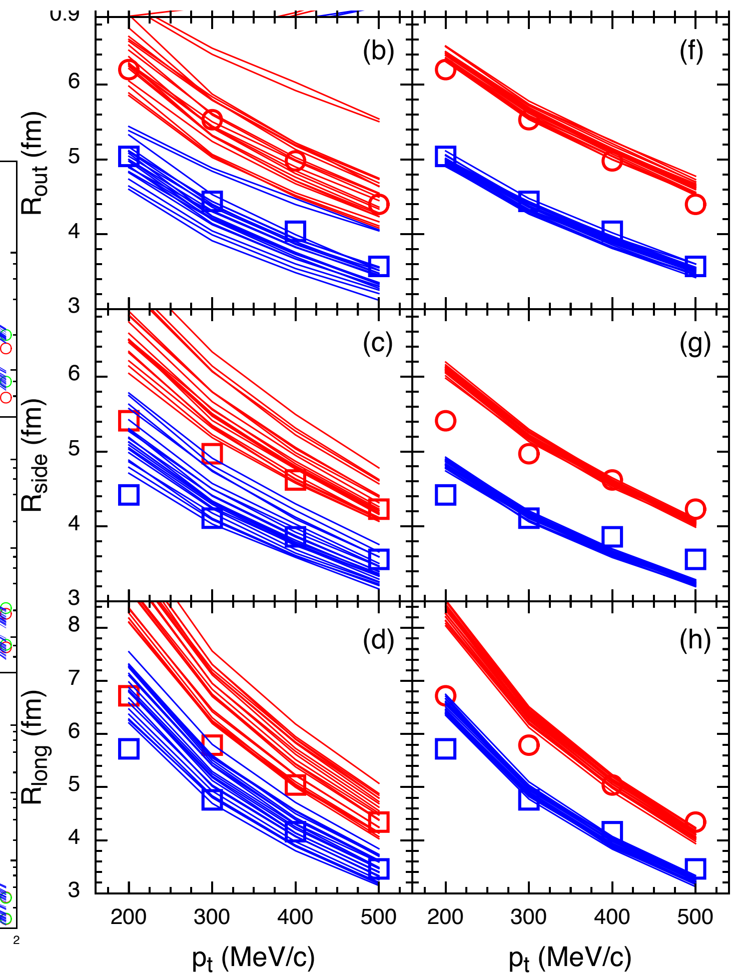


MADAI EoS Analysis: 26 plots, 30 observables



RHIC DATA





IV. Nuisance Parameters

Systematic error has known form:

$$\delta y_a = \boxed{X^{(n)}}\,\boxed{f_a}$$

known form, e.g. exp(-p_t/τ)

Nuisance parameter

$$\Sigma_{ab} = (X^{(\mathrm{n})})^2 f_a f_b$$

- $X^{(n)}$ has prior distribution (Gaussian)
- f_a extends over correlated range of a
- popular in HEP to account for detector response
- could be applied to model mixing

IV. Nuisance Parameters

Example: Uncertain normalization

$$\frac{dN}{dp} = X\,\frac{dN^{(\mathrm{mod})}(x_1, x_2, \cdots x_n)}{dp},$$

$$\mathrm{Prob(X)} \sim e^{-(X-1)^2/2\sigma_X^2}$$

Common-mode errors (Phillips)

Summary

- Several tactics to account for correlated error
- Choice based on specific problem
- All are better than doing nothing!

- Admitting your mistakes is HARD!