Bayesian model comparison for one-dimensional azimuthal correlations

Hans C Eggers, Michiel B de Kock
Stellenbosch University

and

Thomas A Trainor
University of Washington

XLV International Symposium on Multiparticle Dynamics
Wildbad Kreuth, 4–9 October 2015

Phys. Rev. C **92**, 034908 (2015) arxiv 1502.04475

STAR 2D data: 200 GeV AuAu autocorrelations

1.2 million events, 11 centrality classes

Variables

$$\phi_\Delta = \phi_1 - \phi_2$$

$$\eta_\Delta = \eta_1 - \eta_2$$

Pair densities (autocorrelations)

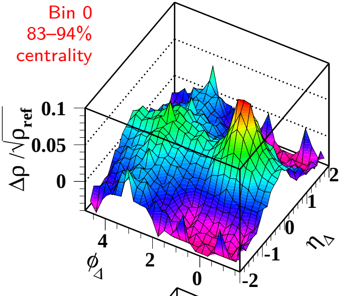
$$\rho(\phi_\Delta, \eta_\Delta) \qquad \rho_{\rm ref}(\phi_\Delta, \eta_\Delta)$$

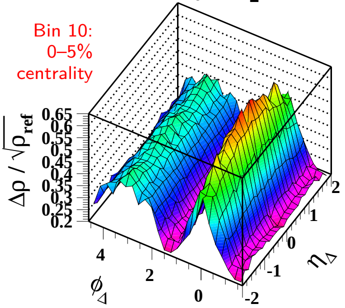
1-particle differential cross section

$$\rho_0 = d^2 n_{ch} / d\eta_\Delta \phi_\Delta$$

per-particle autocorrelation data

$$A(\phi_\Delta, \eta_\Delta) = \frac{\Delta\rho}{\sqrt{\rho_{\rm ref}}} = \rho_0 \left[\frac{\rho}{\rho_{\rm ref}} - 1 \right]$$





2D Parametrization $f(\phi_\Delta, \eta_\Delta | \boldsymbol{w})$

Simplify notation: $\phi_\Delta \equiv \phi \qquad \eta_\Delta \equiv \eta$

Parametrization used for 2D STAR data has **eleven parameters**

$$\boldsymbol{w} = (A_{2D}, \sigma_\phi, \sigma_\eta, A_0, A_\mathrm{D}, A_\mathrm{Q}, A_\mathrm{soft}, \sigma_0, A_\mathrm{BE}, w_\phi, w_\eta)$$

$$f(\phi, \eta | \boldsymbol{w}) = A_{2D} \exp\left\{ -\frac{1}{2}\left[\left(\frac{\phi}{\sigma_\phi}\right)^2 + \left(\frac{\eta}{\sigma_\eta}\right)^2 \right] \right\} \qquad \text{SS peak}$$

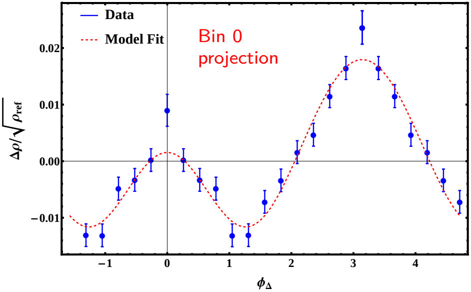
$$+ A_0 + A_\mathrm{D}[\cos(\phi - \pi) + 1]/2 \qquad \text{constant, AS dipole}$$

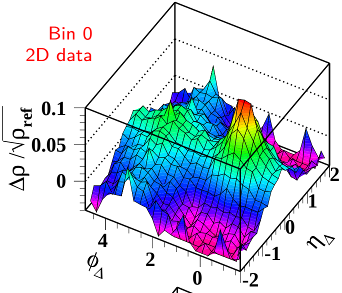
$$+ A_\mathrm{Q}\, 2\cos(2\phi) + A_\mathrm{soft} \exp\left\{ -\frac{1}{2}\left(\frac{\eta}{\sigma_0}\right)^2 \right\} \qquad \text{quadrupole, soft}$$

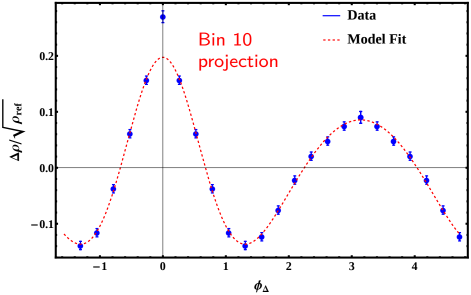
$$+ A_\mathrm{BE} \exp\left\{ -\frac{1}{2}\left[\left(\frac{\phi}{w_\phi}\right)^2 + \left(\frac{\eta}{w_\eta}\right)^2 \right]^{1/2} \right\} \qquad \text{Bose-Einstein}$$

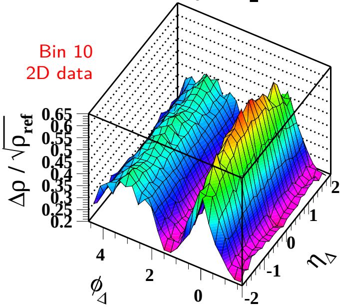
STAR, PRC **86**, 064902 (2012)

Projection onto azimuth ϕ_Δ









1D Parametrizations: Gauss-plus-multipoles

Parametrization 1D projection with 4 parameters

$$f(\phi|\boldsymbol{w}) = A_0 \; + \; A_{1\mathrm{D}} \exp\left[-\frac{1}{2}\left(\frac{\phi}{\sigma_\phi}\right)^2\right] \; - \; A'_{\mathrm{D}}\, 2\cos(\phi)$$

Basic Model: Gaussian plus dipole

Data must integrate to zero, hence reduce to 3 parameters:

BM $$f(\phi|\boldsymbol{w}) = A_{1\mathrm{D}} \left\{ \exp\left[-\frac{1}{2}\left(\frac{\phi}{\sigma_\phi}\right)^2\right] - \frac{\sigma_\phi}{\sqrt{2\pi}} \right\} - A'_{\mathrm{D}}\, 2\cos(\phi)$$

Augment Basic Model by Quadrupole, Sextupole, Octupole

BM+ $$f(\phi|\boldsymbol{w}) = A_{1\mathrm{D}} \left\{ \exp\left[-\frac{1}{2}\left(\frac{\phi}{\sigma_\phi}\right)^2\right] - \frac{\sigma_\phi}{\sqrt{2\pi}} \right\} - A'_{\mathrm{D}}\, 2\cos(\phi)$$

$$- A'_{\mathrm{Q}}\, 2\cos(2\phi) - A'_{\mathrm{S}}\, 2\cos(3\phi) - A'_{\mathrm{O}}\, 2\cos(4\phi)$$

1D Parametrizations: Fourier Series

Motivated by collective-flow paradigm v_2 v_4 $\ldots$,
Many papers eg
STAR, PRC**72**, 014904 (2005)
Poskanzer and Voloshin, PRC **58**, 1671 (1998)

Different **Fourier Series models** for $K = 1, 2, \ldots, 11$

$$\textbf{FS} \qquad f(\phi|\mathbf{w}, K) = 2\sum_{j=1}^{K} w_j \cos(j\phi)$$

sampled at ϕ_n i.e. series of models of different length up to K=11

$$f(\phi_n|\mathbf{w}, 2) = 2w_1\cos(\phi_n) + 2w_2\cos(2\phi_n)$$
$$f(\phi_n|\mathbf{w}, 3) = 2w_1\cos(\phi_n) + 2w_2\cos(2\phi_n) + 2w_3\cos(3\phi_n)$$

successively incorporate dipole, quadrupole, sextupole, . . .

Basics of fitting, common to all approaches

Definitions

$\boldsymbol{\phi} = (\phi_1 \ \ldots \ \phi_N)$	x-coordinates of data (angles)
$\boldsymbol{D} = (A_1 \ \ldots \ A_N)$	set of N data points (autocorrelations) taken at $\boldsymbol{\phi}$
$\boldsymbol{\sigma} = (\sigma_1 \ \ldots \ \sigma_N)$	experimental uncertainties
$f(\phi\|\boldsymbol{w})$	**parametrization** (model function) sampled at $(\phi_1 \ \ldots \ \phi_N)$
$\boldsymbol{w} = (w_1 \ \ldots \ w_K)$	set of K parameters

Gaussian fluctuations in deviations

$$A_n - f(\phi_n|\boldsymbol{w}) \ = \ \varepsilon_n \qquad\qquad p(\varepsilon_n) \ = \ \text{gaussian with std deviation } \sigma_n$$

lead to usual theory with likelihood, χ^2, Hessian and covariance matrix

$$p(\boldsymbol{D} \,|\, \boldsymbol{w}) = \mathrm{const} \cdot e^{-\chi^2/2} \qquad\qquad \chi^2 = \sum_{n=1}^{N} \left(\frac{A_n - f(\phi_n|\boldsymbol{w})}{\sigma_n} \right)^2$$

$$\mathbb{H}_{k\ell} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial w_k \, \partial w_\ell} \qquad\qquad \mathbb{C} = \mathbb{H}^{-1}$$

The Bayesian framework

- Allow probability of data variable $\boldsymbol{y}$, parameters $\boldsymbol{w}$, models $\mathcal{H}$, ...
- **Conditional notation** $p(A|B)$ distinguishes between what is unknown / predicted A and what is known B (hypothesis)
- **Data $\boldsymbol{D}$** is a **specific realization** of data variable $\boldsymbol{y}$. Data $\boldsymbol{D}$ **is fixed**.
  Likelihood $p(\boldsymbol{D}|\boldsymbol{w},\mathcal{H}) \;\equiv\; p(\boldsymbol{y}{=}\boldsymbol{D}|\boldsymbol{w},\mathcal{H})$
- Joint probability for *data variables* $\boldsymbol{y}$ and parameters $\boldsymbol{w}$

$$p(\boldsymbol{y},\boldsymbol{w}|\mathcal{H}) = p(\boldsymbol{y}|\boldsymbol{w},\mathcal{H})\,p(\boldsymbol{w}|\mathcal{H}) \;=\; p(\boldsymbol{w}|\boldsymbol{y},\mathcal{H})\,p(\boldsymbol{y}|\mathcal{H})$$

- **Bayes' Theorem: For any particular model $\mathcal{H}$ consisting of a likelihood (including parametrization) and a prior**

$$p(\boldsymbol{w}|\boldsymbol{D},\mathcal{H}) = \frac{p(\boldsymbol{D}|\boldsymbol{w},\mathcal{H})\,p(\boldsymbol{w}|\mathcal{H})}{p(\boldsymbol{D}|\mathcal{H})}$$

$$(\text{posterior}) = \frac{(\text{likelihood})\,(\text{prior})}{(\text{evidence})}$$

- **Evidence** is an average of the likelihood **for all values of $\boldsymbol{w}$**

$$p(\boldsymbol{D}|\mathcal{H}) = \int d\boldsymbol{w}\,p(\boldsymbol{D}|\boldsymbol{w},\mathcal{H})\,p(\boldsymbol{w}|\mathcal{H})$$

Bayesian model comparison

- **Model Hypotheses:** $\mathcal{H}_1 =$ **BM** or **BM+** $\qquad \mathcal{H}_2 =$ **FS**
- **Ratio of model probabilities, given data D, is**

$$\frac{p(\mathcal{H}_1|\boldsymbol{D})}{p(\mathcal{H}_2|\boldsymbol{D})} = \frac{p(\boldsymbol{D}|\mathcal{H}_1)}{p(\boldsymbol{D}|\mathcal{H}_2)}\,\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_2)} = \frac{p(\boldsymbol{D}|\mathcal{H}_1)}{p(\boldsymbol{D}|\mathcal{H}_2)} = \frac{\text{evidence for } \mathcal{H}_1}{\text{evidence for } \mathcal{H}_2}$$

- (evidence) $= \int d\boldsymbol{w}$ (likelihood) (prior)

$$p(\boldsymbol{D}|\mathcal{H}) = \int d\boldsymbol{w}\, p(\boldsymbol{D}|\boldsymbol{w},\mathcal{H})\; p(\boldsymbol{w}|\mathcal{H})$$

$$= p(\boldsymbol{D}|\tilde{\boldsymbol{w}},\mathcal{H})\; p(\tilde{\boldsymbol{w}}|\mathcal{H})\sqrt{(2\pi)^K \det \mathbb{C}} \quad \text{(Laplace approximation)}$$

$\tilde{\boldsymbol{w}} =$ maximum-likelihood parameter values

- Hence **(-2 log evidence)** is $\boxed{-2\mathrm{LE} = \text{minimum}\,\chi^2 + 2\,\mathcal{I}}$ with

$$-\tfrac{1}{2}\log(\text{max likelihood}) = \min\chi^2 = -2\log p(\tilde{\boldsymbol{w}}|\mathcal{H})$$

$$\text{Information} = \mathcal{I} = -\log\left[p(\tilde{\boldsymbol{w}}|\mathcal{H})\,(2\pi)^{K/2}\sqrt{\det \mathbb{C}}\right]$$

Priors

- We use **uniform priors**

$$p(\boldsymbol{w}|\mathcal{H}) = \prod_{k=1}^{K} \frac{1}{\Delta_k}$$

We assign same widths across different models

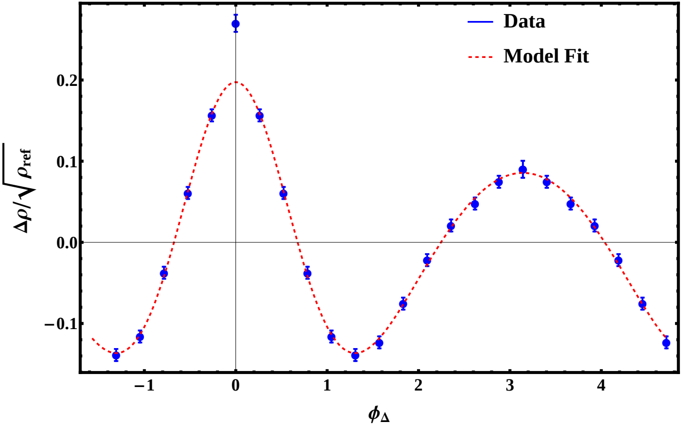
$\sigma_{\phi_\Delta} \in [0, \pi/2]$ (Same-side peak width)

$\Delta_k = 1$ to Gaussian amplitude parameters

$\Delta_k = 1/3$ to cosine coefficients
include uncertainty bands with $\max \Delta_k = 1$ $\min \Delta_k = 1/5$

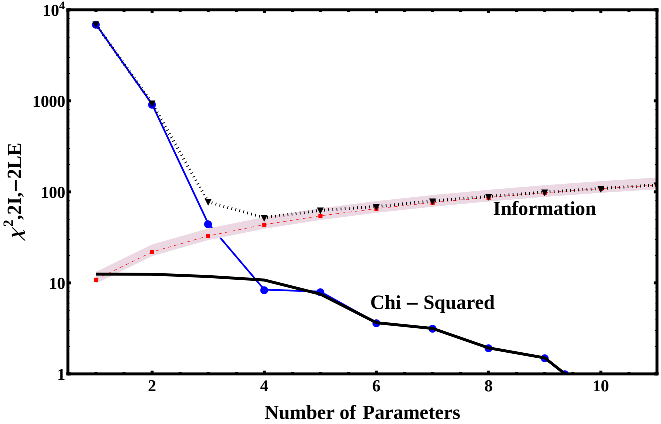
Central collisions: 1D data on azimuth

- Symmetrized, so only **13 unique points**
- Bin ϕ=0 excluded from fits (BE peak)
- $\sum_\phi A(\phi) = 0$, hence $N = \mathbf{11}$ **is small**
- Curve: Basic Model **and** Fourier Series



Central collisions: Fourier series models

χ^2	log-likelihood	Blue solid line **keeps dropping**
		keeps decreasing as you add more parameters
	Residuals	**Solid black line** χ^2 of data minus FS model
$2\mathcal{I}$	Information	Red dashed curve **rises with** K
$-2\mathrm{LE}$	log-evidence	**Black dashed** sum of χ^2 and $2I$
		reaches minimum at $K = 4$



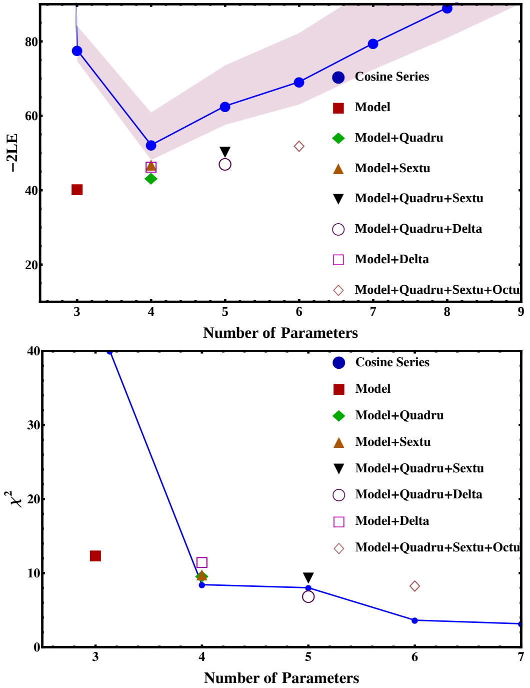
$-2\mathrm{LE}$ for different models

- Picture changes in $-2\mathrm{LE}$
- Basic Model best by far
- Larger Occam factor for FS than for BM models
- **Odds**

$$\frac{p(BM|\boldsymbol{D})}{p(FS|\boldsymbol{D})} = \frac{360}{1}$$

$$\frac{p(BM|\boldsymbol{D})}{p(BM+Q|\boldsymbol{D})} = \frac{4.6}{1}$$

- Basic Model has largest χ^2



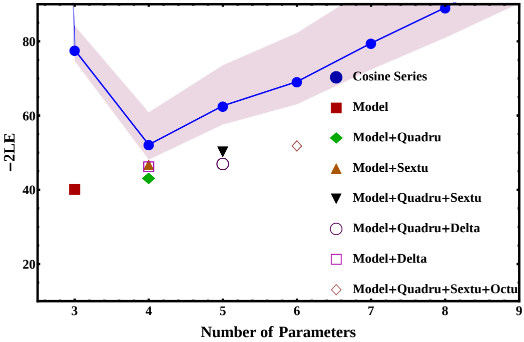
$-2\mathrm{LE}$ **for different models**

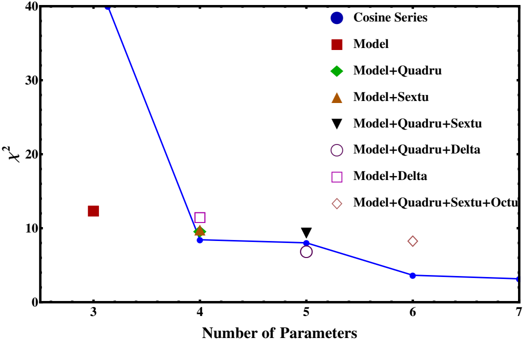
- Picture changes in $-2\mathrm{LE}$
- Basic Model best by far
- Larger Occam factor for FS than for BM models
- **Odds**

$$\frac{p(BM|\boldsymbol{D})}{p(FS|\boldsymbol{D})} = \frac{360}{1}$$

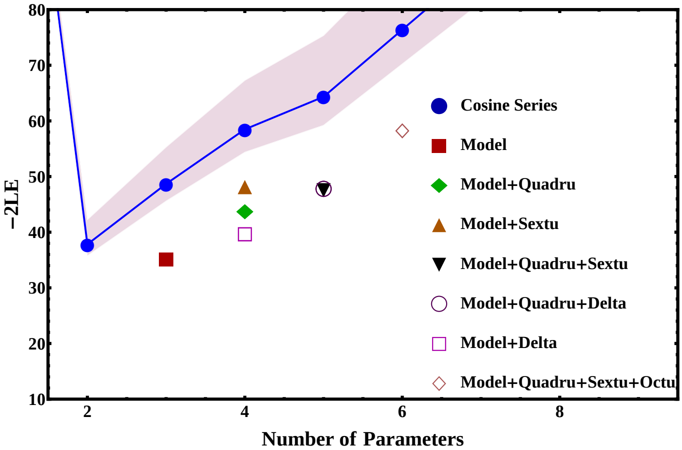
$$\frac{p(BM|\boldsymbol{D})}{p(BM{+}\mathrm{Q}|\boldsymbol{D})} = \frac{4.6}{1}$$

- Basic Model has largest χ^2





Peripheral collisions



Comparable evidence for BM ($K=3$) and FS ($K=2$)

See supplementary slides

Why use the Bayesian framework? Part 1

Traditional use of χ^2	Bayesian Inference
Find **maximum likelihoods** $\max p(\boldsymbol{D} \mid \boldsymbol{w}_1, \mathcal{H}_1) \quad \max p(\boldsymbol{D} \mid \boldsymbol{w}_2, \mathcal{H}_2)$	Find **evidences** $p(\boldsymbol{D} \mid \mathcal{H}_1) \qquad p(\boldsymbol{D} \mid \mathcal{H}_2)$
which determines **single best fit parameters** $\tilde{\boldsymbol{w}}_1, \tilde{\boldsymbol{w}}_2$ This is equivalent to	which consider **all possible parameter values** This is equivalent to
$\min \chi_1^2 = -2 \ln p(\boldsymbol{D} \mid \tilde{\boldsymbol{w}}_1, \mathcal{H}_1)$ $\min \chi_2^2 = -2 \ln p(\boldsymbol{D} \mid \tilde{\boldsymbol{w}}_2, \mathcal{H}_2)$	$-2\,\mathrm{LE}_1 = -2 \ln p(\boldsymbol{D} \mid \mathcal{H}_1)$ $-2\,\mathrm{LE}_2 = -2 \ln p(\boldsymbol{D} \mid \mathcal{H}_2)$
Then compare $\dfrac{\chi_1^2}{(N - K_1)} \quad \text{vs} \quad \dfrac{\chi_2^2}{(N - K_2)}$	Then compare $-2\,\mathrm{LE}_1 \quad \text{vs} \quad -2\,\mathrm{LE}_2$
which ignores covariances of $\boldsymbol{w}$ uses only $p(\boldsymbol{D} \mid \tilde{\boldsymbol{w}}, \mathcal{H})$	**which uses covariances of $\boldsymbol{w}$** $p(\boldsymbol{D} \mid \mathcal{H}) =$ $p(\boldsymbol{D} \mid \tilde{\boldsymbol{w}}, \mathcal{H})\, p(\tilde{\boldsymbol{w}} \mid \mathcal{H}) \sqrt{(2\pi)^K \det \mathbb{C}}$

Why use the Bayesian framework? Part 2

Traditional use of χ^2	Bayesian Inference
Look for different likelihoods	**Construct priors and look for different likelihoods**
Estimating uncertainties:	**Estimating uncertainties:**
$\sigma(\boldsymbol{w})$ estimated from parameter covariance matrix $\mathbb{C}$	**Posterior** $p(\boldsymbol{w}\|\boldsymbol{D}, \mathcal{H})$ provides comprensive information on parameters, including $\sigma(\boldsymbol{w})$ from $\mathbb{C}$
Accurate only for large N	**Accurate for all N**

Summary and conclusions

1. **Bayesian methods provide superior approach to model comparison:**
   - **Handles any logical proposition** (parameters, models, the weather . . .)
   - Use of **prior information**
   - **Evidence** takes into account all possible parameter values
   - **Posteriors** provide comprehensive information
   - **Works for small N**
2. **STAR 1D azimuthal correlations**
   - Bayesian model comparison finds that a model based on SS Gaussian plus AS dipole is preferred above a pure Fourier expansion
   - **Flow paradigm is not supported by our results for STAR central (0–5%) and semi-central (9–18%) collisions.**
     Gaussian-plus-multipoles models have larger evidence
   - For peripheral collisions (83–94%) things are less clear.
     Model predictivity becomes important
3. Only **a single** 1D analysis has been done so far;
   Follow-up for other data, other variables and 2D data needed

Acknowledgements

National Institute for Theoretical Physics

Alexander von Humboldt Foundation

Excellence Cluster “Universe”

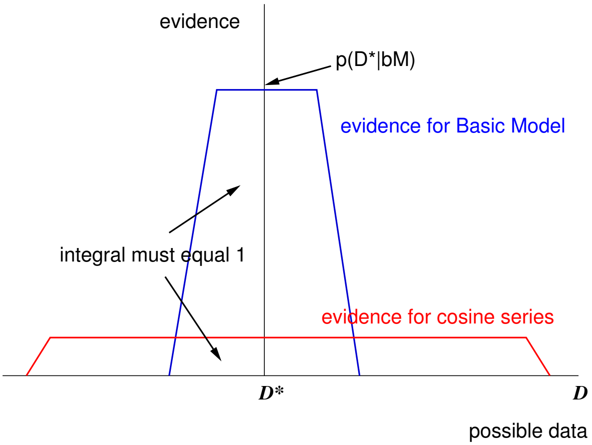
Bayesian model comparison for one-dimensional azimuthal correlations

Eggers, de Kock and Trainor

Supplementary Slides

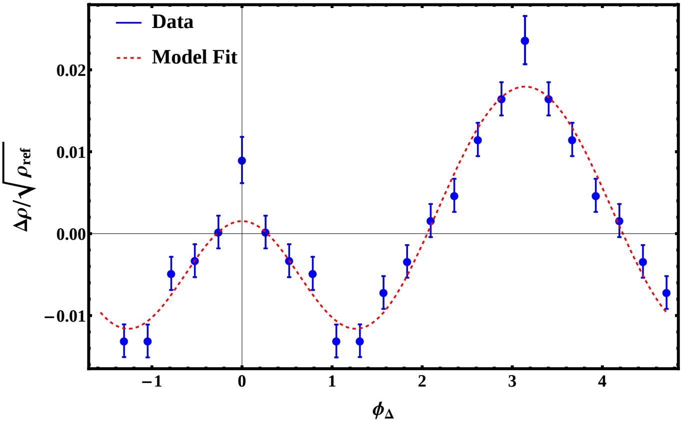
XLV International Symposium on Multiparticle Dynamics
Wildbad Kreuth, 4–9 October 2015

Evidence and model complexity



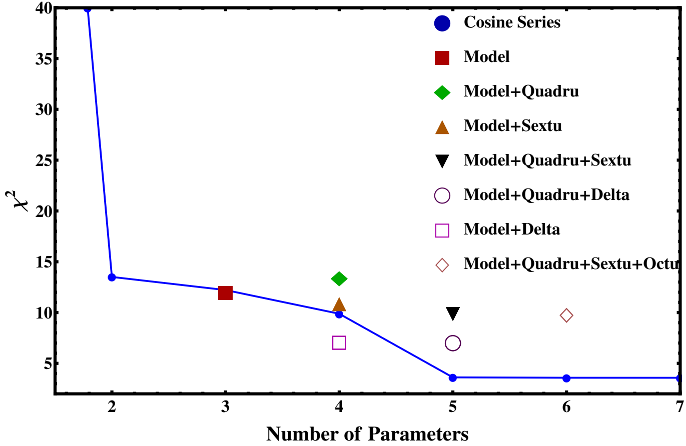
- Complex models such as **FS** can accommodate large set of possible data
- **Normalization $\Longrightarrow$ Smaller FS evidence for the REAL data D^***

Peripheral collisions: the data



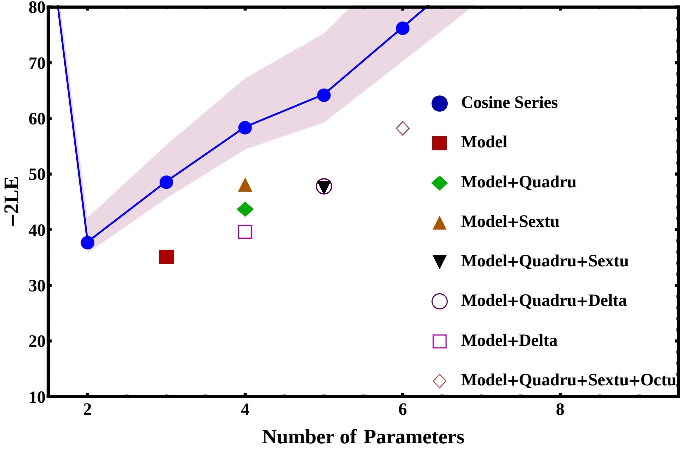
- Essentially $N - N$ collisions
- Much smaller signal (autocorrelations)
- BE correlations are broader, affect the SS peak beyond the ϕ=0 bin
- Broader SS peak

Peripheral collisions: χ^2 only



FS: minimum χ^2 at $K=2$
BM with $K=3$ has similar χ^2
BM plus delta has smaller χ^2 but $K=4$

Peripheral collisions: χ^2, information, evidence



$$\chi^2(\mathbf{BM}) > \chi^2(\mathbf{FS}) \qquad -2\mathrm{LE}(\mathbf{BM}) < -2\mathrm{LE}(\mathbf{FS}) \qquad \frac{p(BM|\boldsymbol{D})}{p(FS|\boldsymbol{D})} = \frac{3.3}{1}$$

Central collisions: 1D data on azimuth

- Symmetrized, so only **13 unique points**
- Bin ϕ=0 excluded from fits (BE peak)
- $\sum_\phi A(\phi) = 0$, hence $N = 11$ **is small**
- Curve: Basic Model **and** Fourier Series

- $\chi^2 = 12.5$
- $A_{1D} = 0.57 \pm 0.007$
- $\sigma_{\phi_\Delta} = 0.635 \pm 0.007$
- $A'_D = 0.115 \pm 0.002$