Phenomenological constraints on transport properties of QCD matter with quantified theoretical uncertainties

Sunil Jaiswal

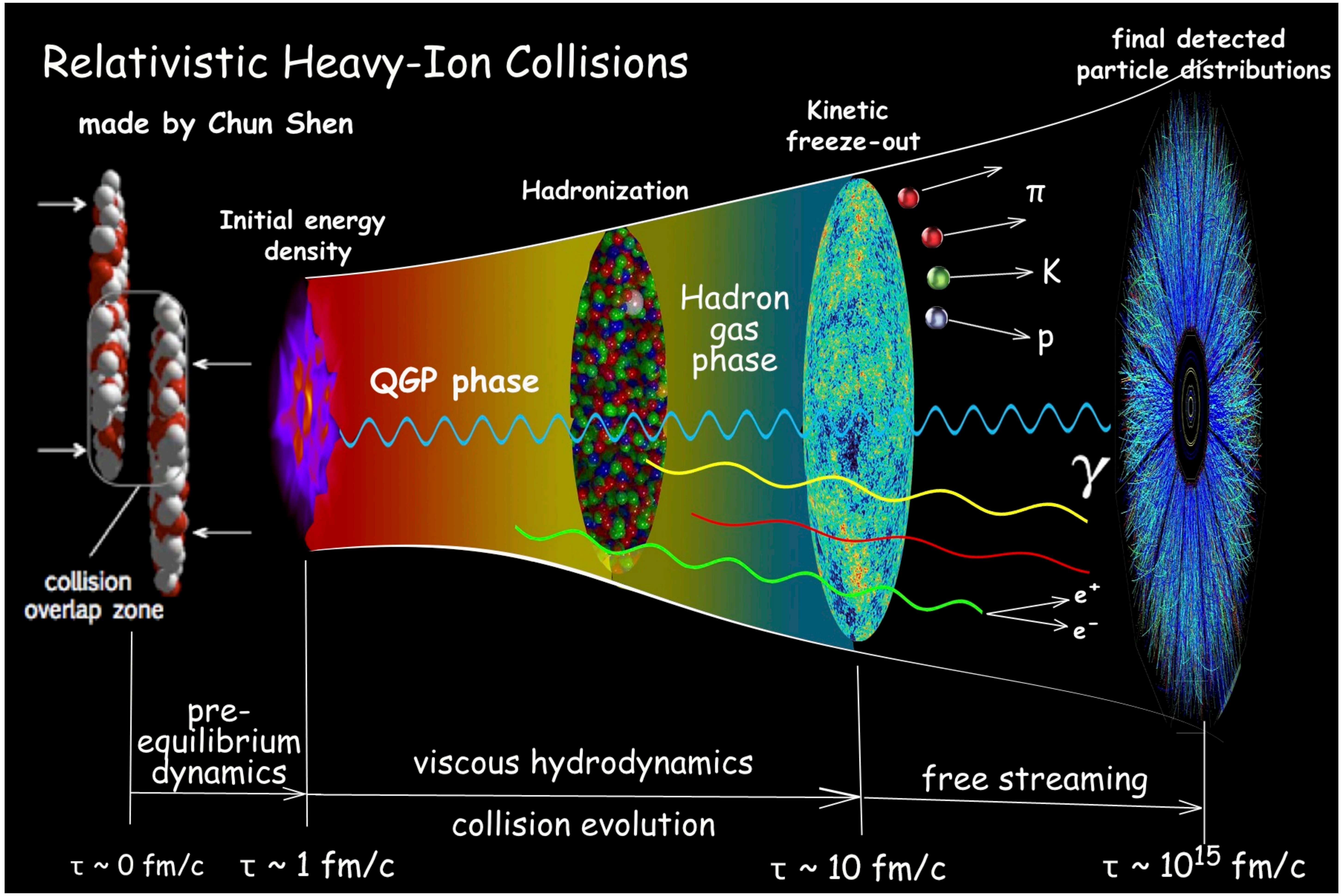
Ohio State University, Columbus, US

India-JINR Workshop

Based on arXiv: 2504.13144, 2509.19759
Collaborators: Richard Furnstahl, Ulrich Heinz, Matthew Pratola, Chun Shen

November 12, 2025

Heavy-ion collisions: Many stages of evolution



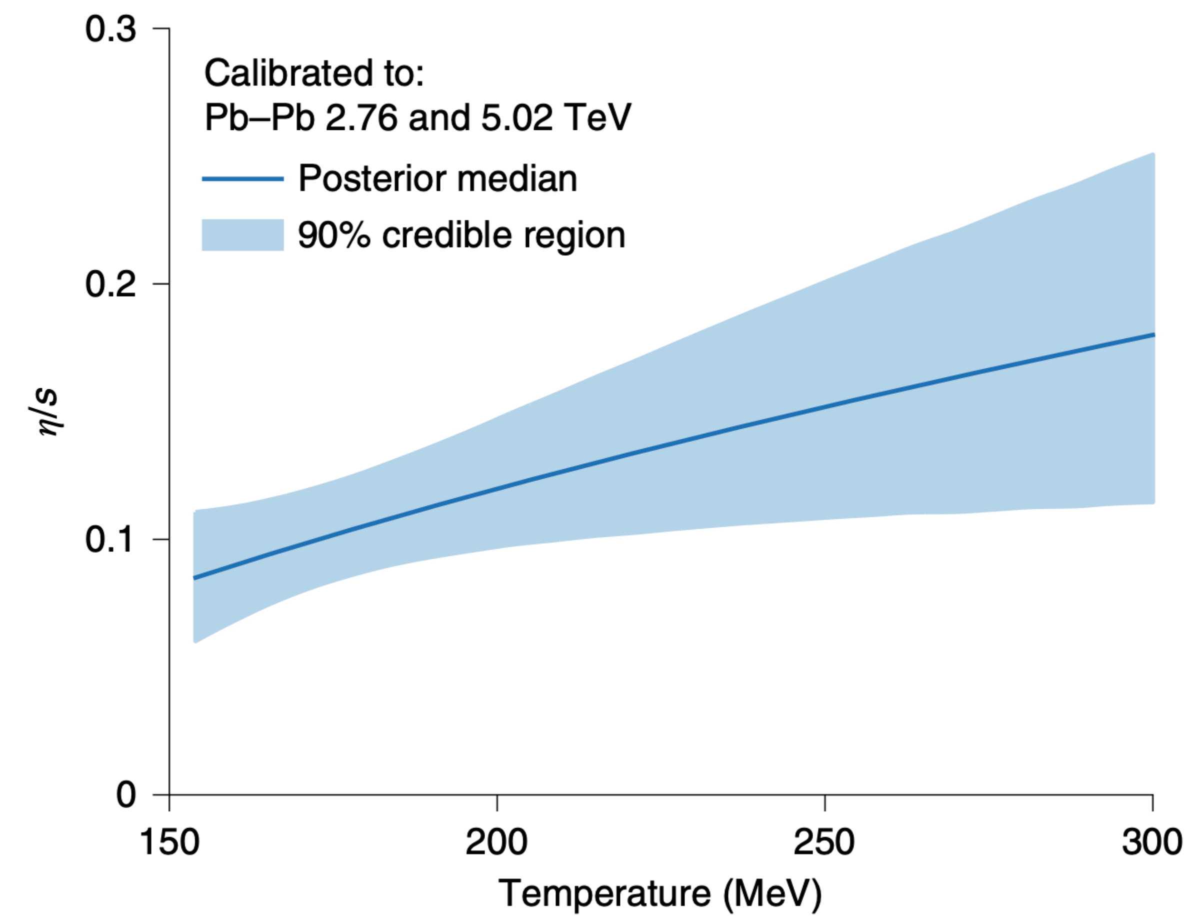
- Many stages of evolution involving many theoretical models.
- Physical parameters of model (like specific viscosities) needs to be inferred from data.

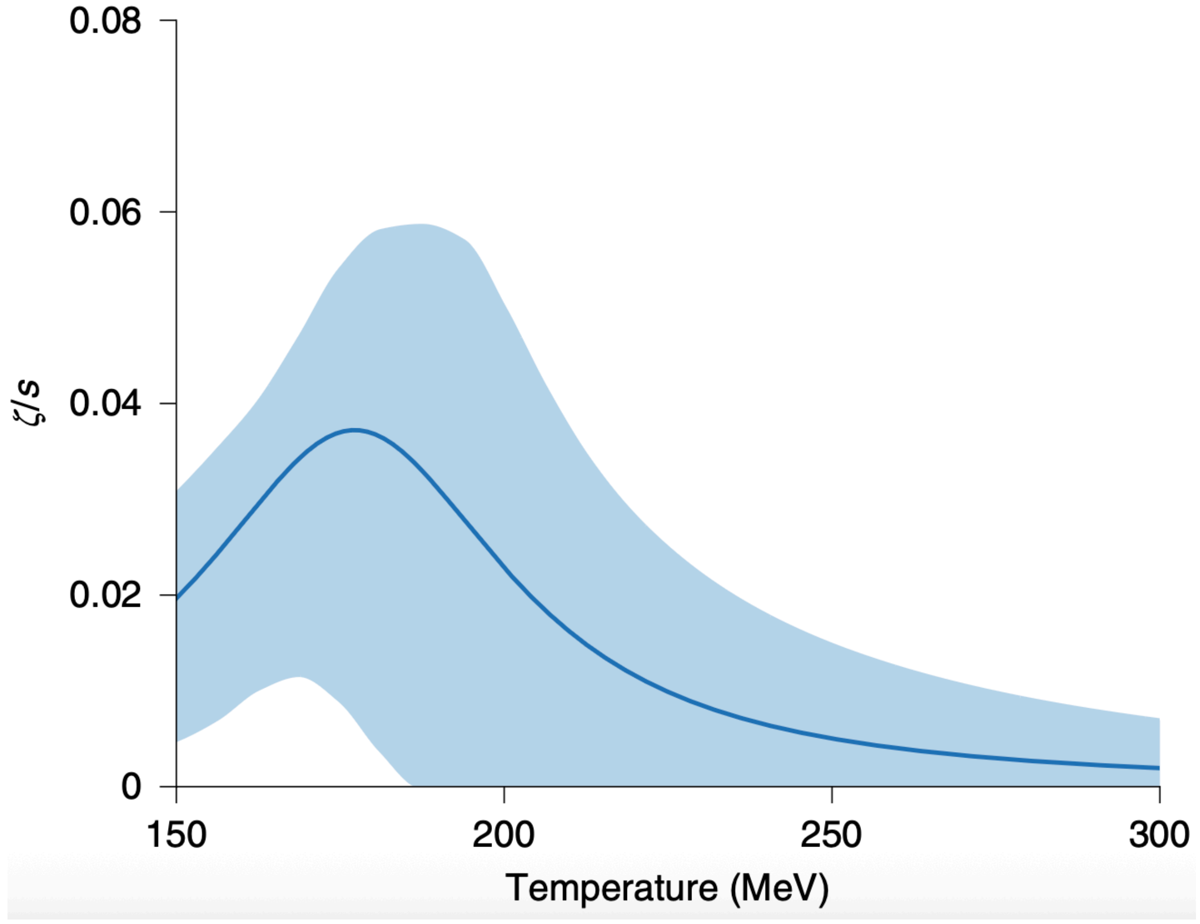
Inconsistent QGP viscosity estimates from different models

2019

DUKE group

Bernhard *et al.* *Nat. Phys.* **15**, 1113–1117

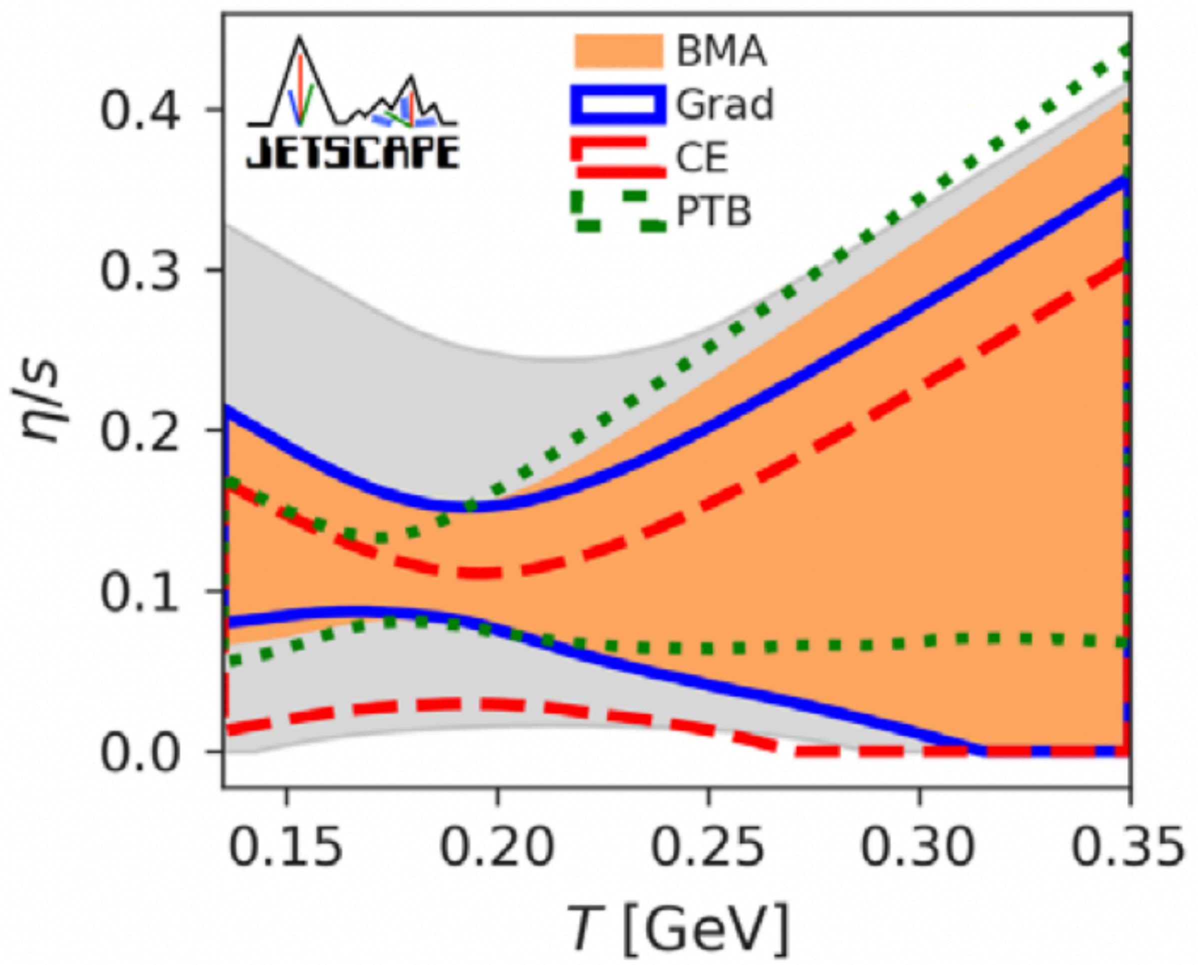


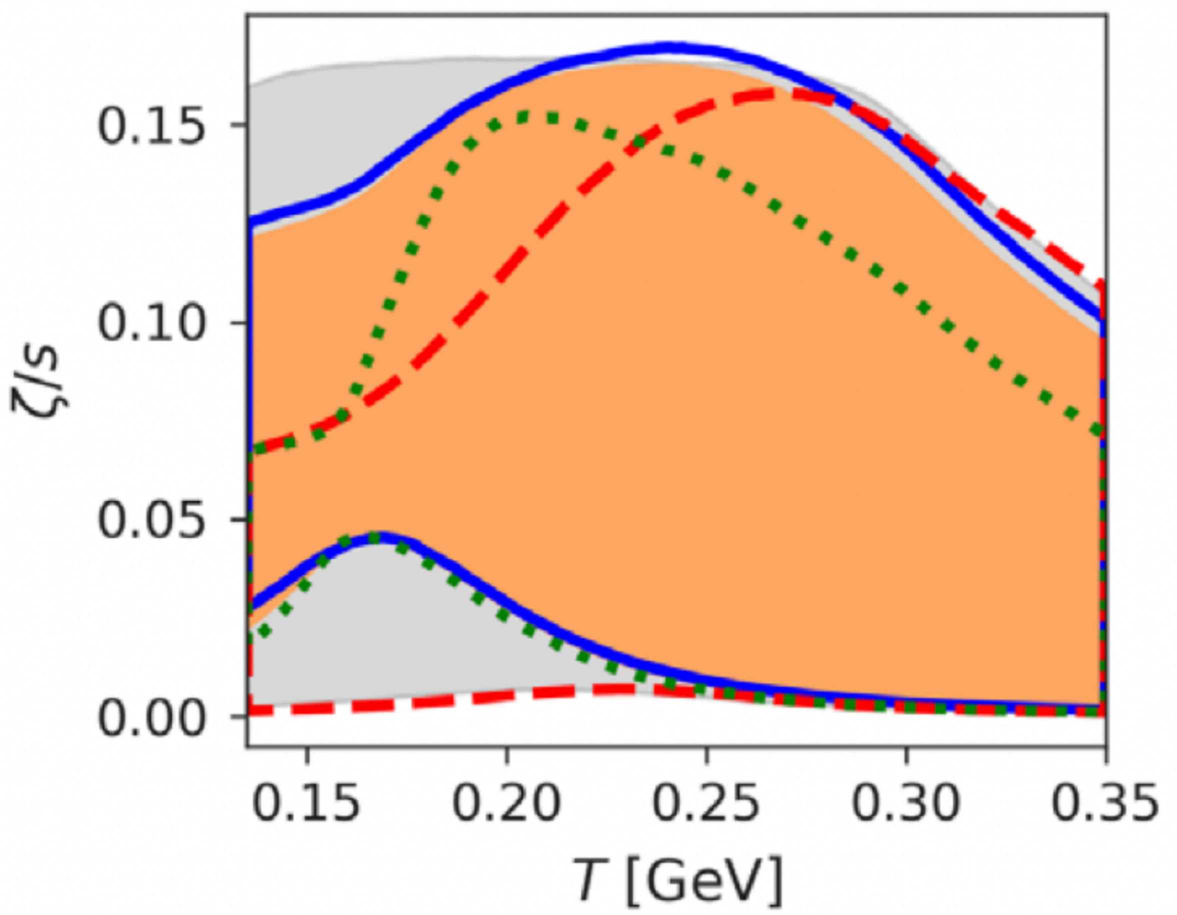


2020

JETSCAPE

D. Everett *et al.* 2010.03928, 2011.01430

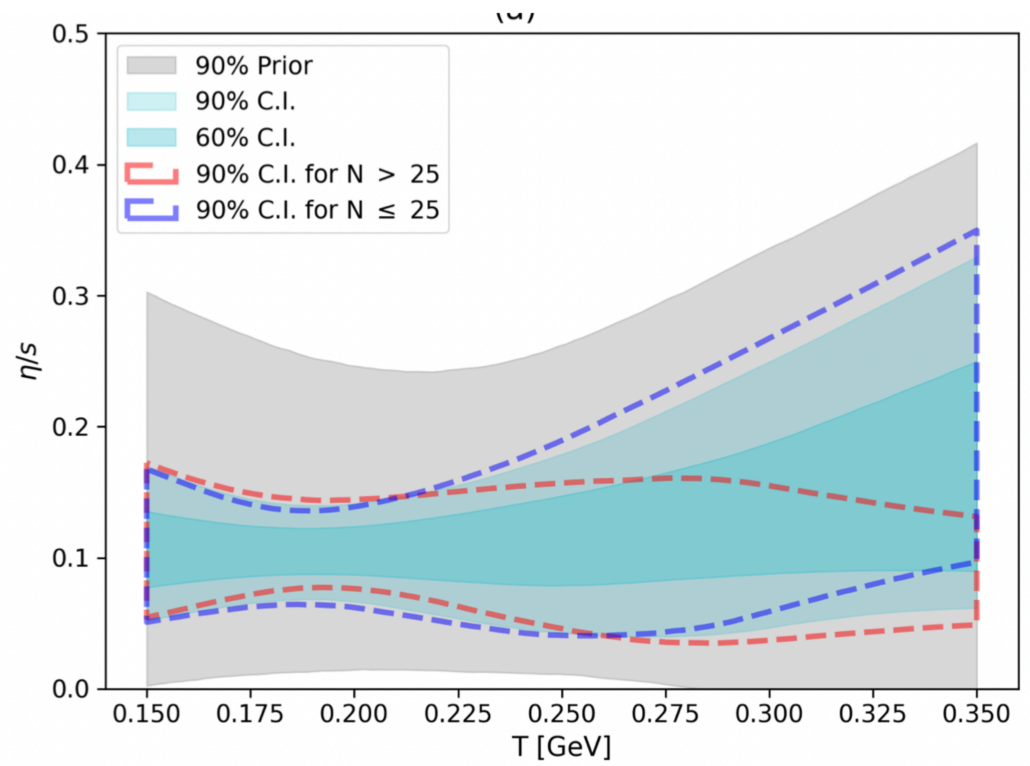


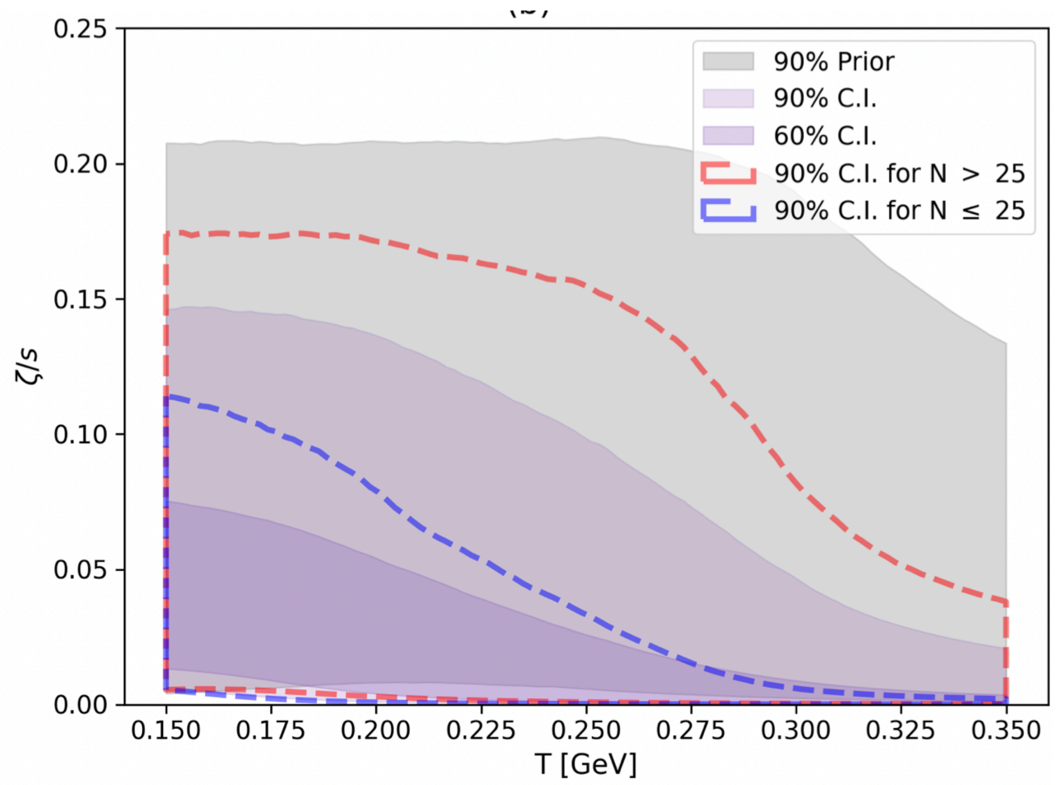


2023

OSU-VAH model

D. Liyanage, O. Surer, *et al.* 2302.14184

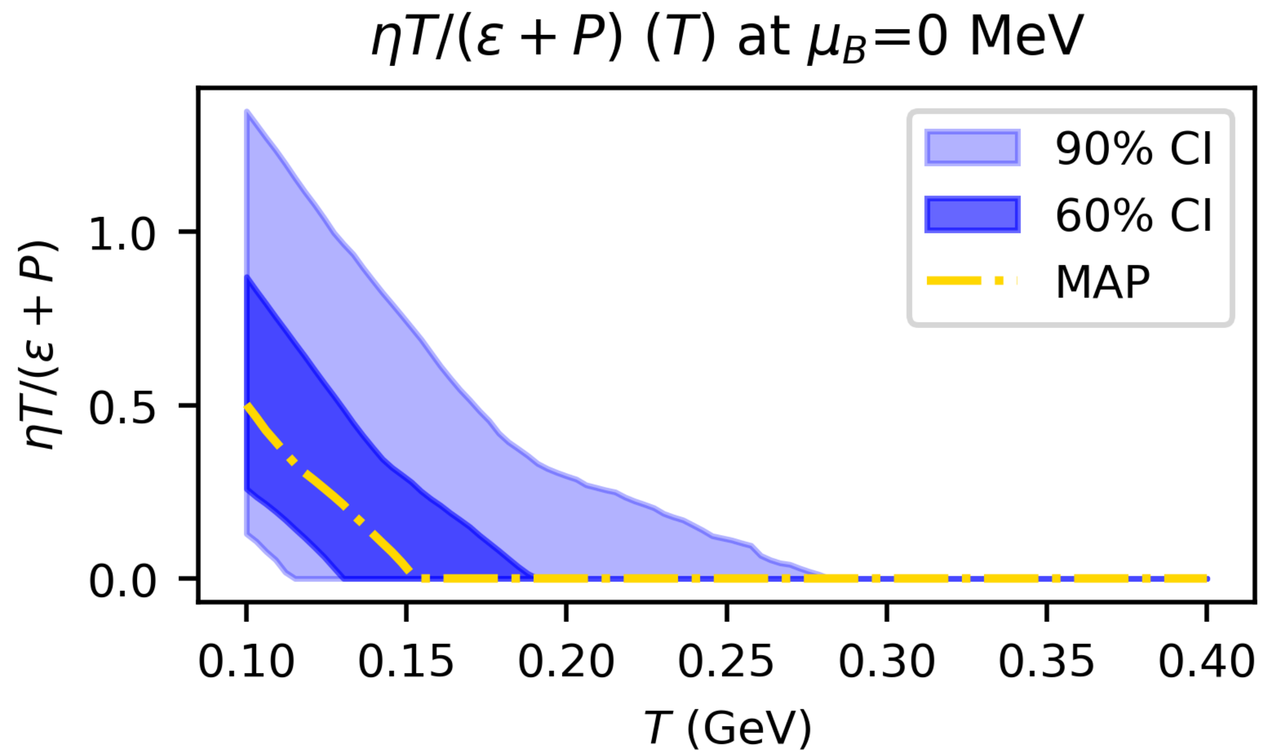


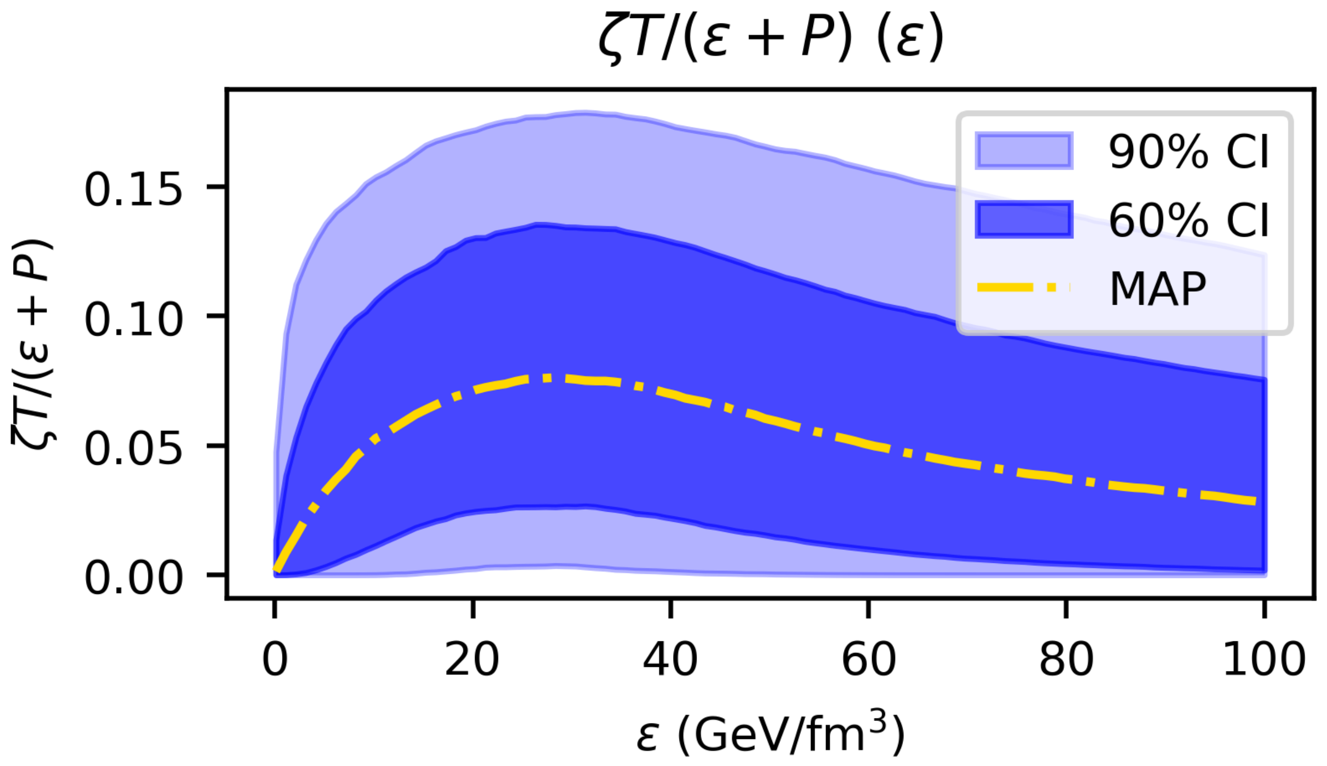


2025

SMASH-vHLLE-hybrid model

Niklas Gotz, *et al. 2503.10181*





Multi-stage physics model

Initial conditions
- Trento
- IP-Glasma

Pre-Equilibrium
- Free Stream
- KoMPoST
- KTiSO

Viscous Hydrodynamics
- 2nd order viscous hydro
- Viscous anisotropic hydro

Particlization
- Grad
- CE
- PTB
- PTMA

After Burner
- SMASH
- URQMD

Slide adapted from D. Liyanage

Multi-stage physics model

Initial conditions
- Trento
- IP-Glasma

Pre-Equilibrium
- Free Stream
- KoMPoST
- KTiSO

Viscous Hydrodynamics
- 2nd order viscous hydro
- Viscous anisotropic hydro

Particlization
- Grad
- CE
- PTB
- PTMA

After Burner
- SMASH
- URQMD

T$_R$ENTo
- N : normalization
- p : generalized mean power
- σ_k : multiplicity fluct. variance
- w : nucleon width
- d_{min} : min. dist. betw. nucleons

Free-stream
- τ_R : free-stream time scale
- α : energy density scaling of τ_R

MUSIC
- $(\eta/s)(T)$: shear viscosity
- $(\zeta/s)(T)$: bulk viscosity
- b_π: Shear relax time factor

GRAD | **CE**
- T_{sw}: switching temperature

SMASH

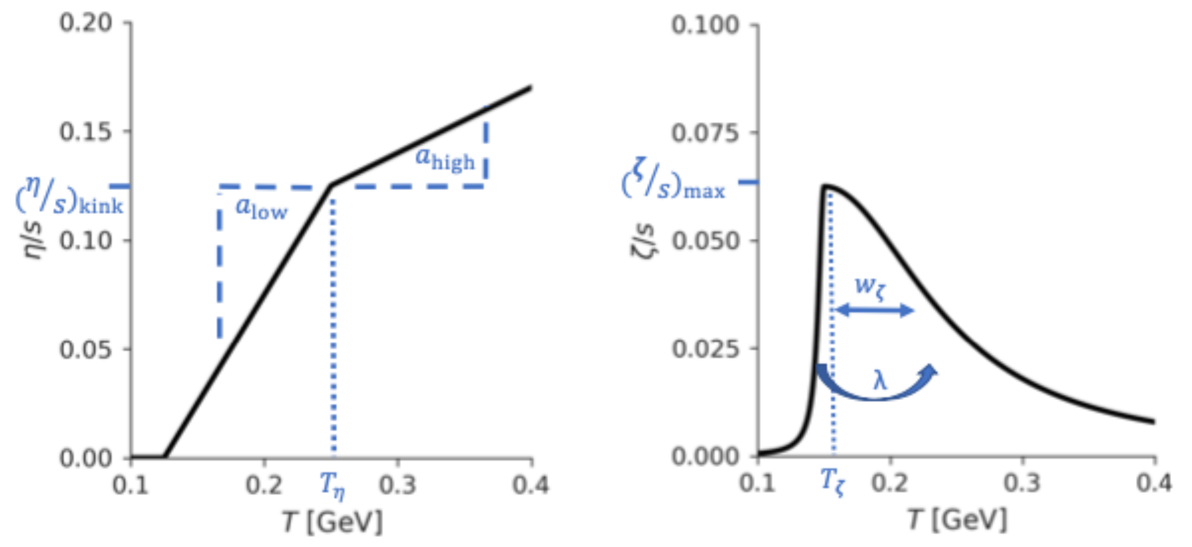
Pb - Pb at 2.76 TeV
Model predictions
for observables

JETSCAPE SIMS calibration

D. Everett *et al.* 2010.03928, 2011.01430

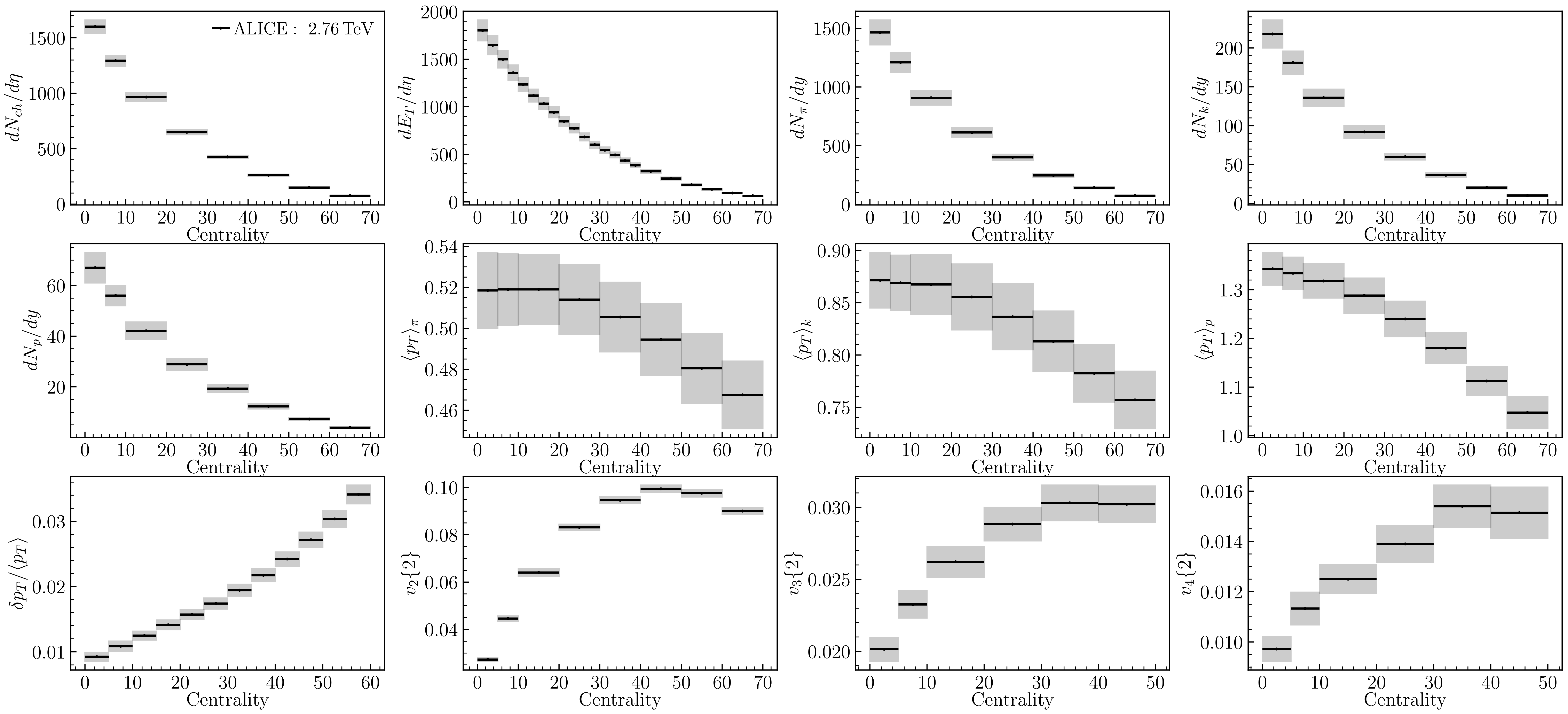
17 model parameters

Parameter name	Symbol	
Normalization	N	
Generalized mean	p	
Multiplicity fluctuation	σ_k	
Nucleon width	w	
Min. dist. btw. nucleons	$d^3_{\min}$	
Free-streaming time scale	τ_R	
Free-streaming energy dep.	α	
Temperature of (η/s) kink	$T_{\rm kink}$	$(\eta/s)(T)$
Low temp. slope of (η/s)	$a_{\rm low}$	
High temp. slope of (η/s)	$a_{\rm high}$	
(η/s) at kink	$(\eta/s)_{\rm kink}$	
Maximum of (ζ/s)	$(\zeta/s)_{\max}$	$(\zeta/s)(T)$
Temperature of (ζ/s) peak	T_ζ	
Width of (ζ/s) peak	w_ζ	
Asymmetry of (ζ/s) peak	λ_ζ	
Shear relaxation time factor	b_π	
Particlization temperature	$T_{\rm sw}$	



Slide adapted from D. Liyanage

Experimental data: Pb+Pb at 2.76 TeV



SJ, arXiv: 2509.19759

(i) the number of charged hadrons per unit pseudorapidity $dN_{ch}/d\eta$ [28]; (ii) the transverse energy per unit pseudorapidity $dE_T/d\eta$ [29]; (iii) the number of identified charged hadrons per unit rapidity dN_i/dy, $i \in \{\pi, k, p\}$ [30]; (iv) the mean transverse momenta of identified hadrons $\langle p_T \rangle_i$, $i \in \{\pi, k, p\}$ [30]; (v) the two-particle cumulant flow coefficients $v_n\{2\}$ for $n = 2, 3, 4$ [31]; (vi) the fluctuation in the mean transverse momentum $\delta p_T/\langle p_T \rangle$ [32].

[28] K. Aamodt, et al., Centrality dependence of the charged-particle multiplicity density at mid-rapidity in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, Phys. Rev. Lett. 106 (2011) 032301. arXiv:1012.1657, doi:10.1103/PhysRevLett.106.032301.

[29] J. Adam, et al., Measurement of transverse energy at midrapidity in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, Phys. Rev. C94 (3) (2016) 034903. arXiv:1603.04775, doi:10.1103/PhysRevC.94.034903.

[30] B. Abelev, et al., Centrality dependence of π, K, p production in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, Phys. Rev. C88 (2013) 044910. arXiv:1303.0737, doi:10.1103/PhysRevC.88.044910.

[31] K. Aamodt, et al., Higher harmonic anisotropic flow measurements of charged particles in Pb-Pb collisions at $\sqrt{s_{NN}}$=2.76 TeV, Phys. Rev. Lett. 107 (2011) 032301. arXiv:1105.3865, doi:10.1103/PhysRevLett.107.032301.

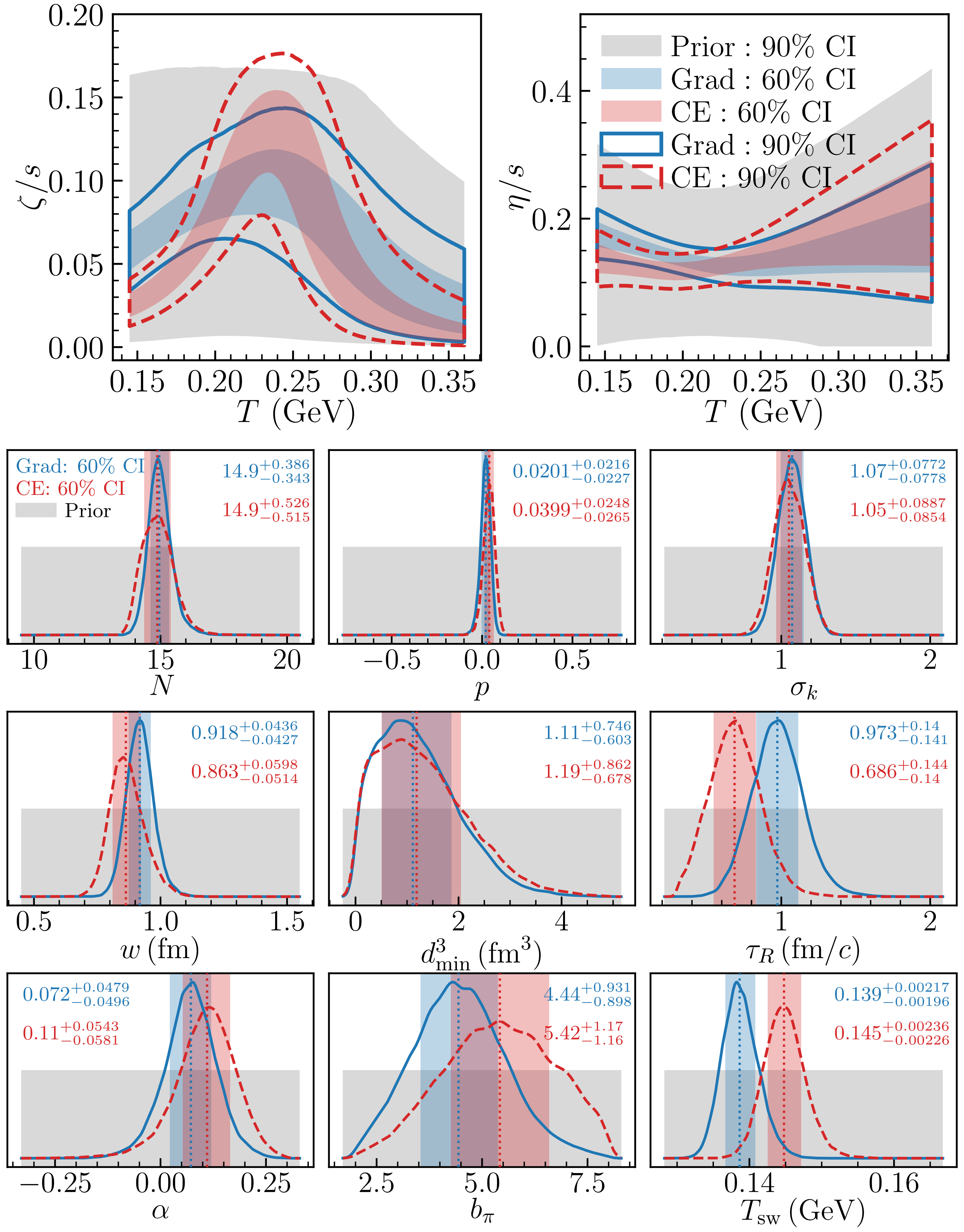
[32] B. B. Abelev, et al., Event-by-event mean p_T fluctuations in pp and Pb-Pb collisions at the LHC, Eur. Phys. J. C74 (10) (2014) 3077. arXiv:1407.5530, doi:10.1140/epjc/s10052-014-3077-y.

12 observables. 110 observations

Bayesian inference

SJ, arXiv: 2509.19759

D. Everett *et al.* 2010.03928, 2011.01430

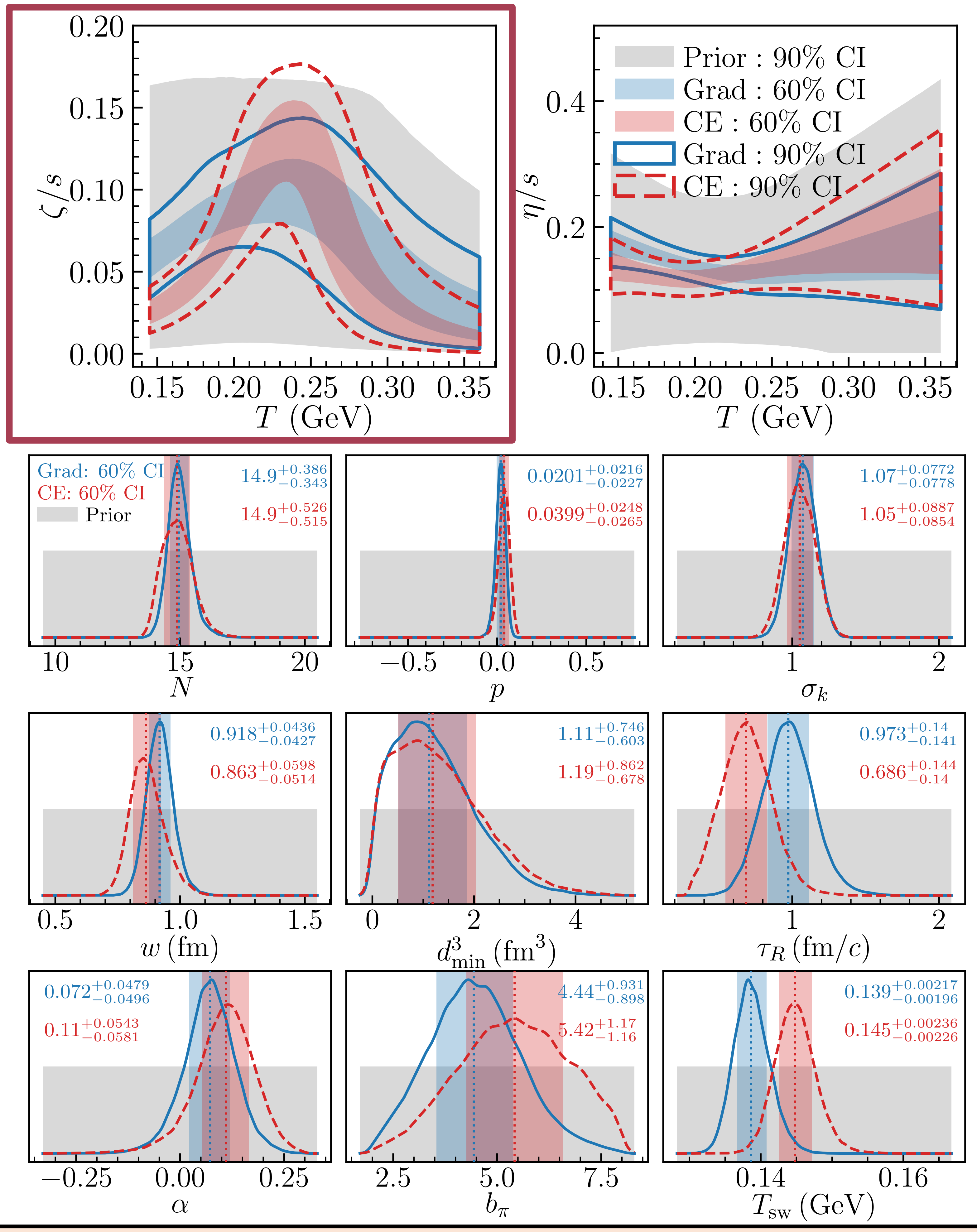


17 parameter Bayesian inference

Bayesian inference

SJ, arXiv: 2509.19759

D. Everett *et al.* 2010.03928, 2011.01430



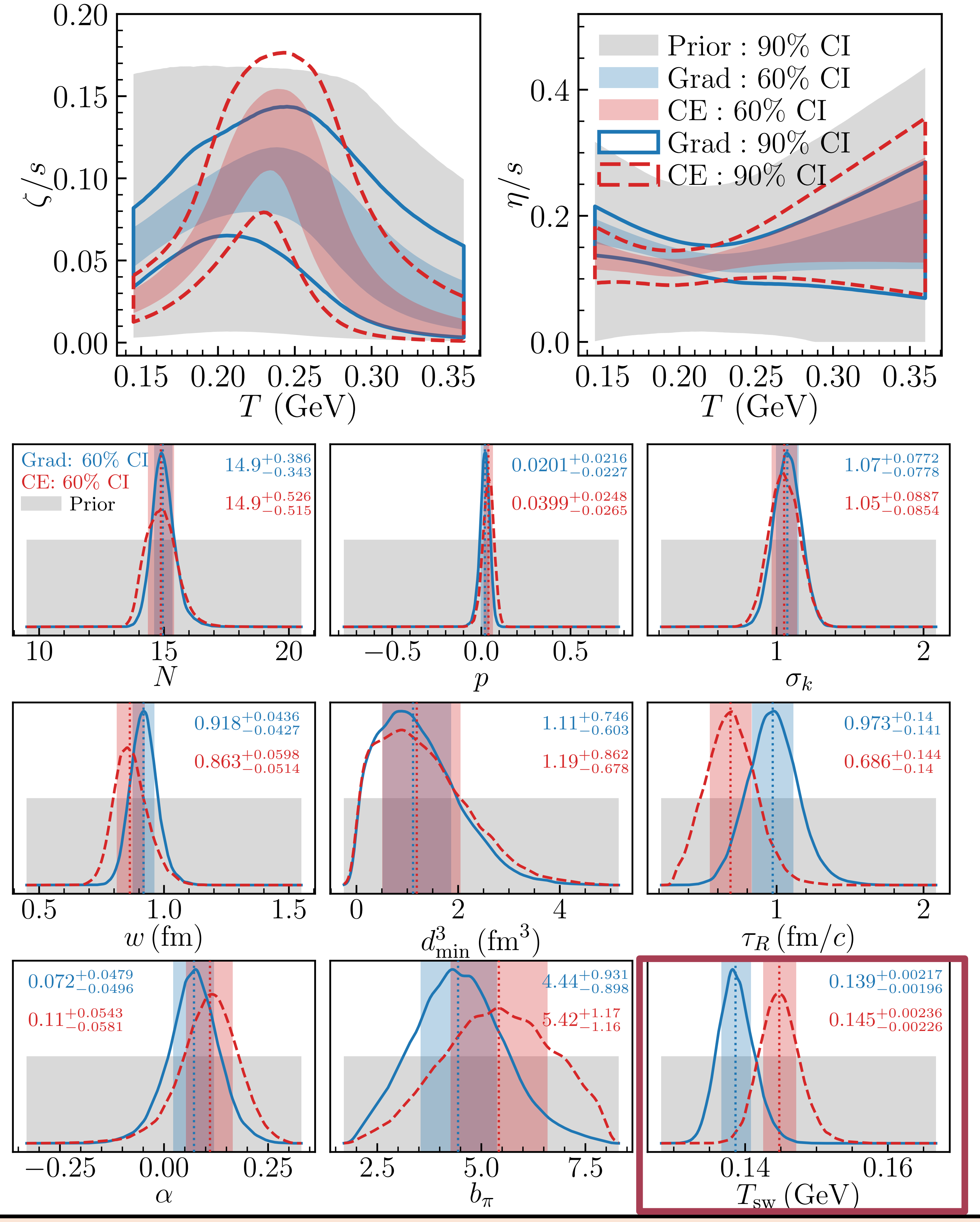
17 parameter Bayesian inference

- Tension observed between the specific bulk viscosity posteriors from Grad (blue) and CE (red).

Bayesian inference

SJ, arXiv: 2509.19759

D. Everett *et al.* 2010.03928, 2011.01430



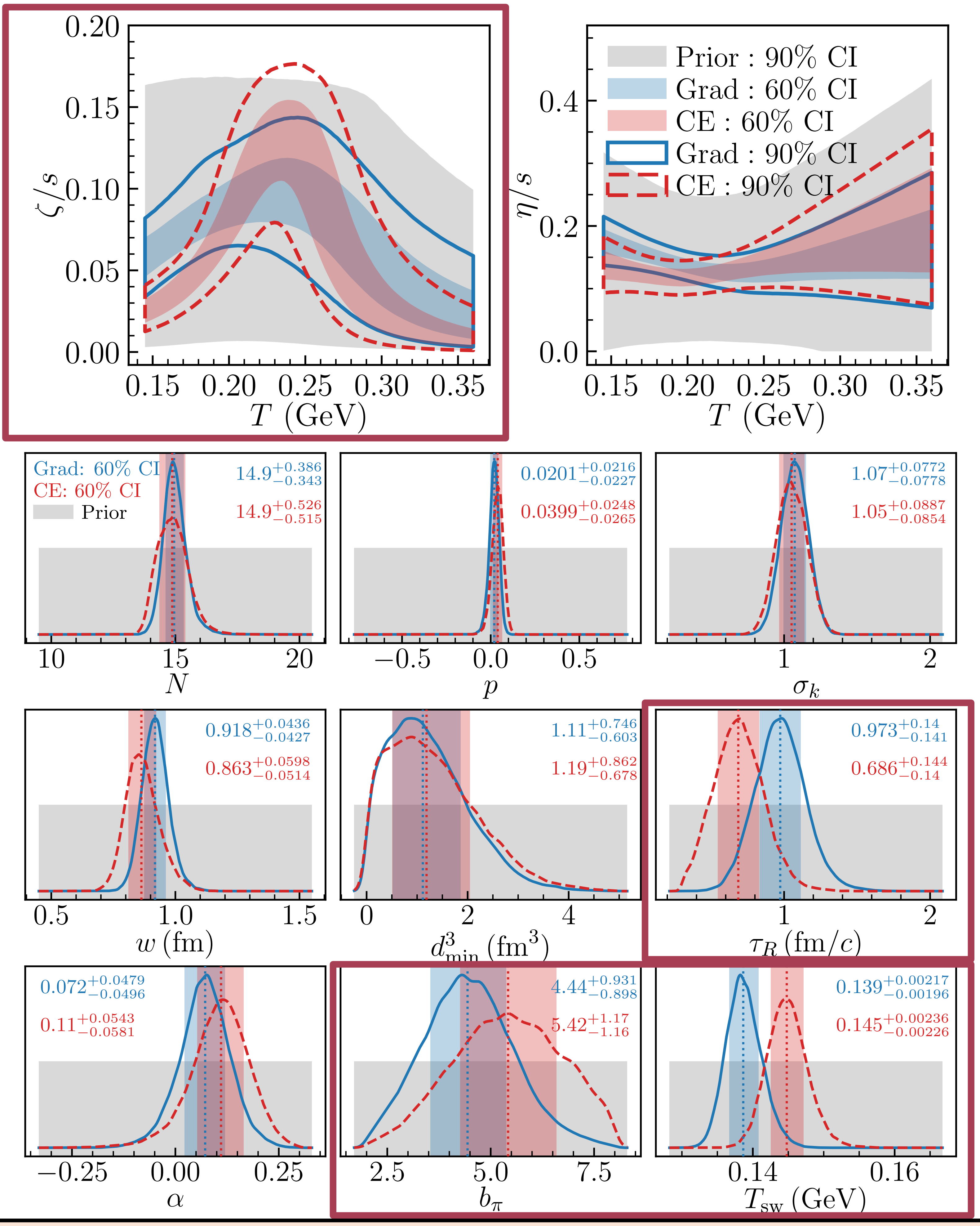
17 parameter Bayesian inference

- Tension observed between the specific bulk viscosity posteriors from Grad (blue) and CE (red). Differences also visible for other parameters.
- Particlization temperature lower for Grad: $T_{sw} = 138.65^{+2.17}_{-1.96}$ MeV, than for CE: $144.79^{+2.36}_{-2.26}$ MeV.

Bayesian inference

SJ, arXiv: 2509.19759

D. Everett *et al.* 2010.03928, 2011.01430



17 parameter Bayesian inference

Error of theory propagates across stages. Results in different $T^{\mu\nu}$ evolution. Will effect studies where medium evolution needs to be considered (like jet studies).

- Tension observed between the specific bulk viscosity posteriors from Grad (blue) and CE (red). Differences also visible for other parameters.
- Particlization temperature lower for Grad: $T_{\rm sw} = 138.65^{+2.17}_{-1.96}$ MeV, than for CE: $144.79^{+2.36}_{-2.26}$ MeV.

Model discrepancy framework

- Every theoretical model is an **approximation** of the true underlying physics.
- Usually, the error of theory is not known quantitatively.

$$\underbrace{\zeta(x)}_{\text{Predictions from true theory}} = \underbrace{\eta(x, \theta)}_{\text{Predictions from approximate theory}} + \underbrace{\delta(x)}_{\text{Error of approximate theory predictions}}$$

Model discrepancy framework

- Every theoretical model is an **approximation** of the true underlying physics.
- Usually, the error of theory is not known quantitatively.

$$\underbrace{\zeta(x)}_{\text{Predictions from true theory}} = \underbrace{\eta(x,\theta)}_{\text{Predictions from approximate theory}} + \underbrace{\cancelto{0}{\delta(x)}}_{\text{Error of approximate theory predictions}}$$

- Therefore, conventional model-data comparison ignores theoretical uncertainty, implicitly assuming the model is perfect: $\delta(x) = 0$. Consequently, inferred **parameters can become mere fitting variables**, no longer corresponding to their true physical values.

Model discrepancy framework

- Every theoretical model is an **approximation** of the true underlying physics.

- Usually, the error of theory is not known quantitatively.

$$\underbrace{\zeta(x)}_{\text{Predictions from true theory}} = \underbrace{\eta(x, \theta)}_{\text{Predictions from approximate theory}} + \underbrace{\delta(x)}_{\text{Error of approximate theory predictions}}$$

- Therefore, conventional model-data comparison ignores theoretical uncertainty, implicitly assuming the model is perfect: $\delta(x) = 0$. Consequently, inferred **parameters can become mere fitting variables**, no longer corresponding to their true physical values.

- **Our approach :** SJ, C. Shen, R. J. Furnstahl, U. Heinz, and M. T. Pratola, PLB 870, 139946 (2025), arXiv: 2504.13144
  - ➡ We usually have some qualitative knowledge of the domain of validity of our theories. Example: "*We know the theory is more reliable in this regime than in that one*".
  - ➡ We use such important, qualitative knowledge of theory error to statistically model $\delta(x)$.
  - ➡ Model theory error with a Gaussian Process $\delta(\,\cdot \mid \phi) \sim \mathrm{GP}\left(\mathbf{0}, K(\,\cdot\,,\cdot \mid \phi)\right)$.

    By constructing covariance kernel K to reflect the prior qualitative knowledge of the error.

M. Kennedy, A. O'Hagan
doi:10.1111/1467-9868.00294

J. Brynjarsdóttir and A. O'Hagan,
doi:10.1088/0266-5611/30/11/114007

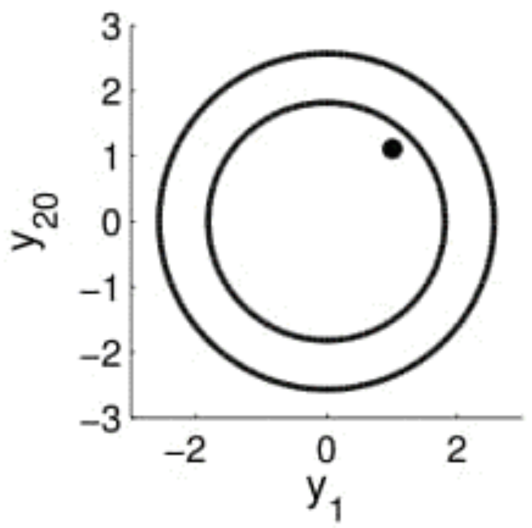
D. Higdon, M. Kennedy, *et. al.*,
doi:10.1137/S1064827503426693

Gaussian Process: Distribution over functions

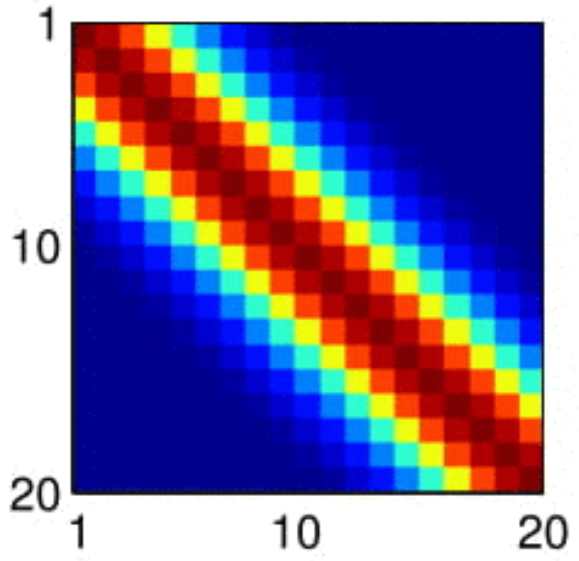
Formal definition: A Gaussian process is a collection of random variables, any finite number of which also have a joint Gaussian distribution.

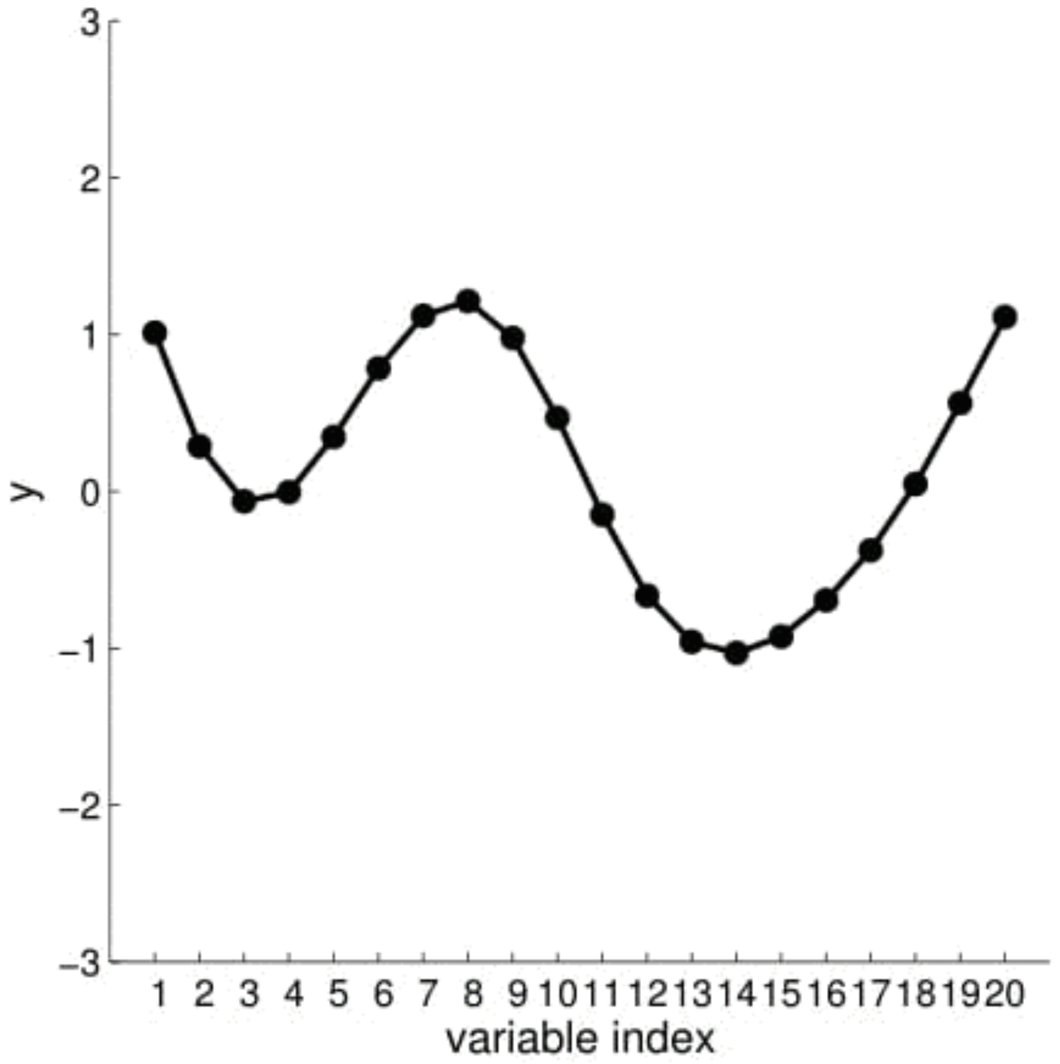
Consider a multivariate Gaussian distribution: $p(\mathbf{y} \mid \Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}\right)$

20D Gaussian distribution — A different representation



$\Sigma =$



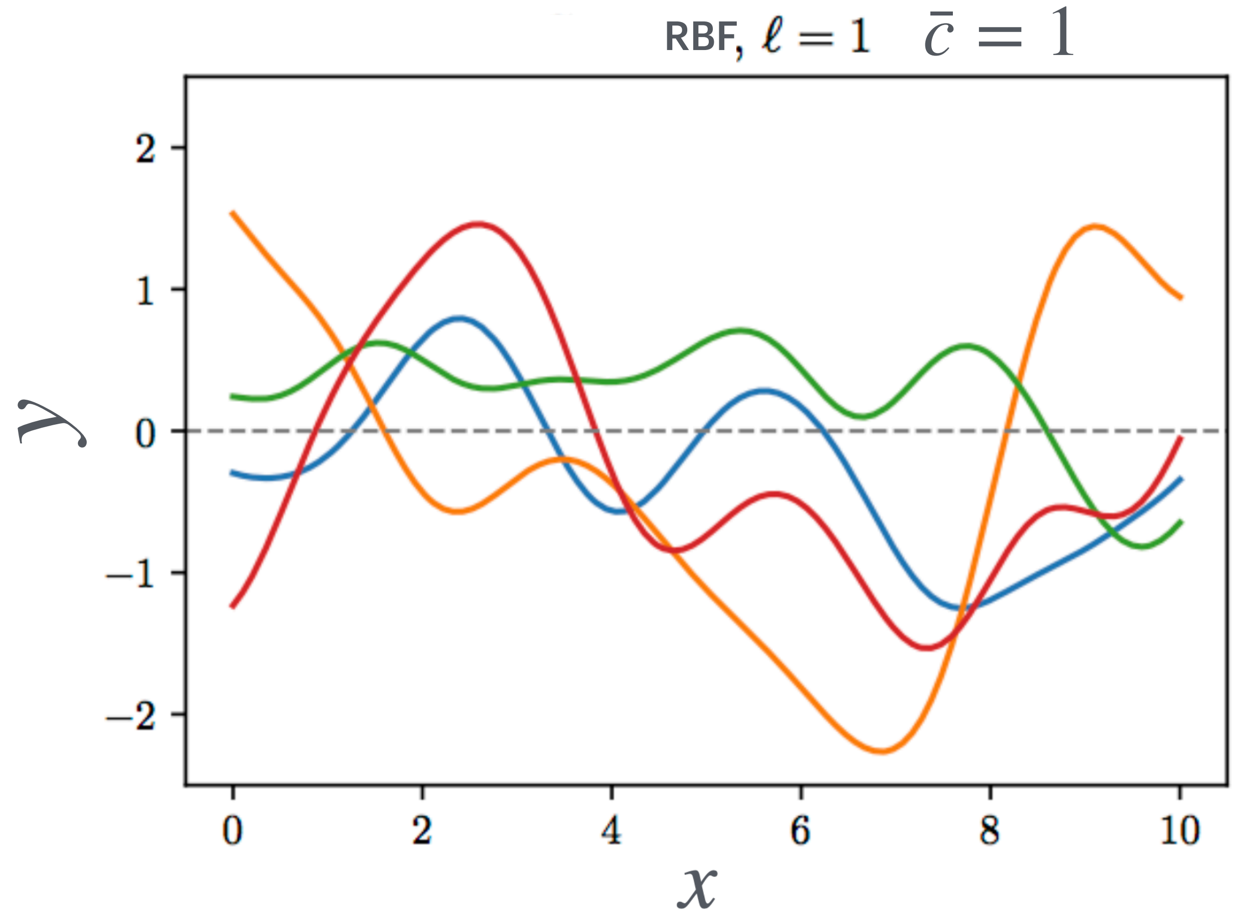


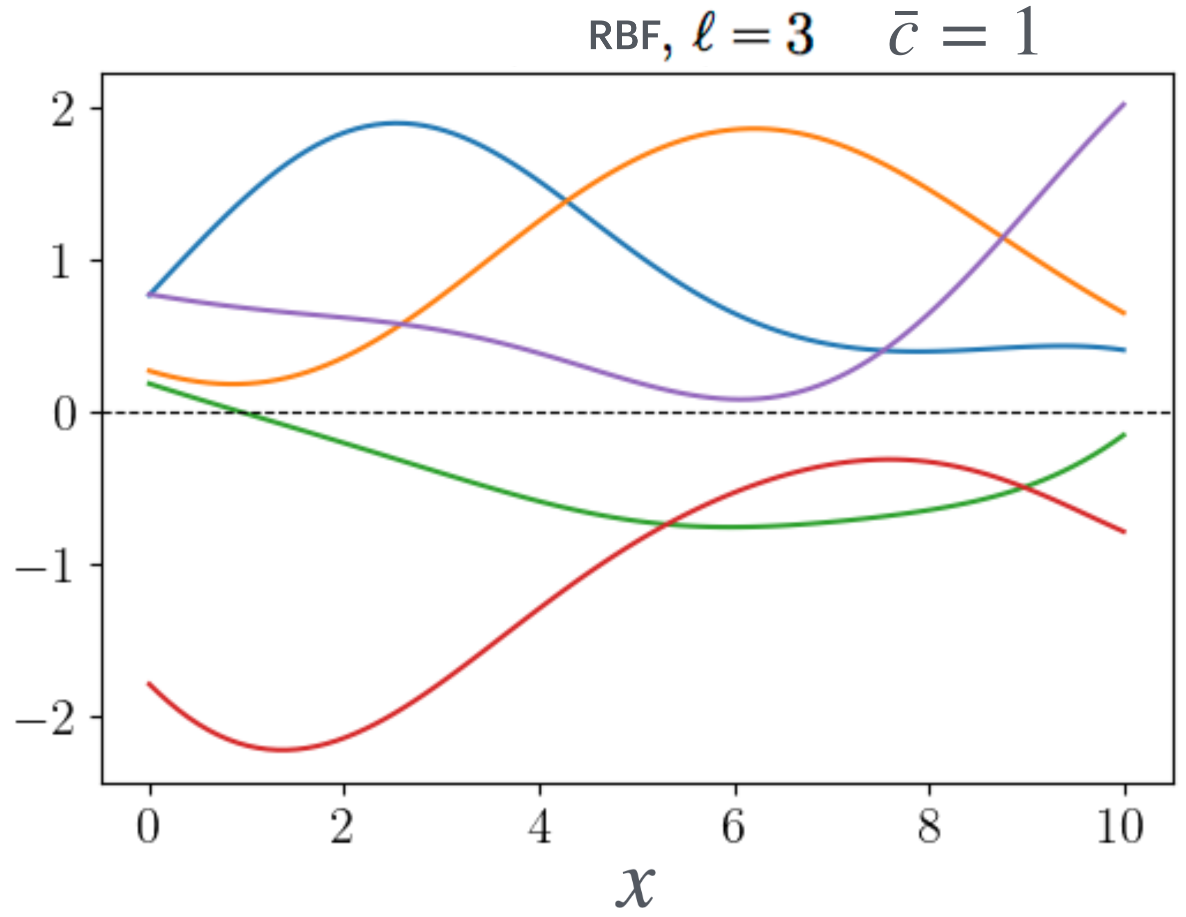
https://thegradient.pub/gaussian-process-not-quite-for-dummies/

Computing Σ from the **covariance kernel**

$$\Sigma \equiv K(x_i, x_j) = \bar{c}^2 \exp\left(-\frac{\|x_i - x_j\|^2}{2\ell^2}\right)$$

$\mathbf{x}$ is a vector of length n ($= 20$ here)





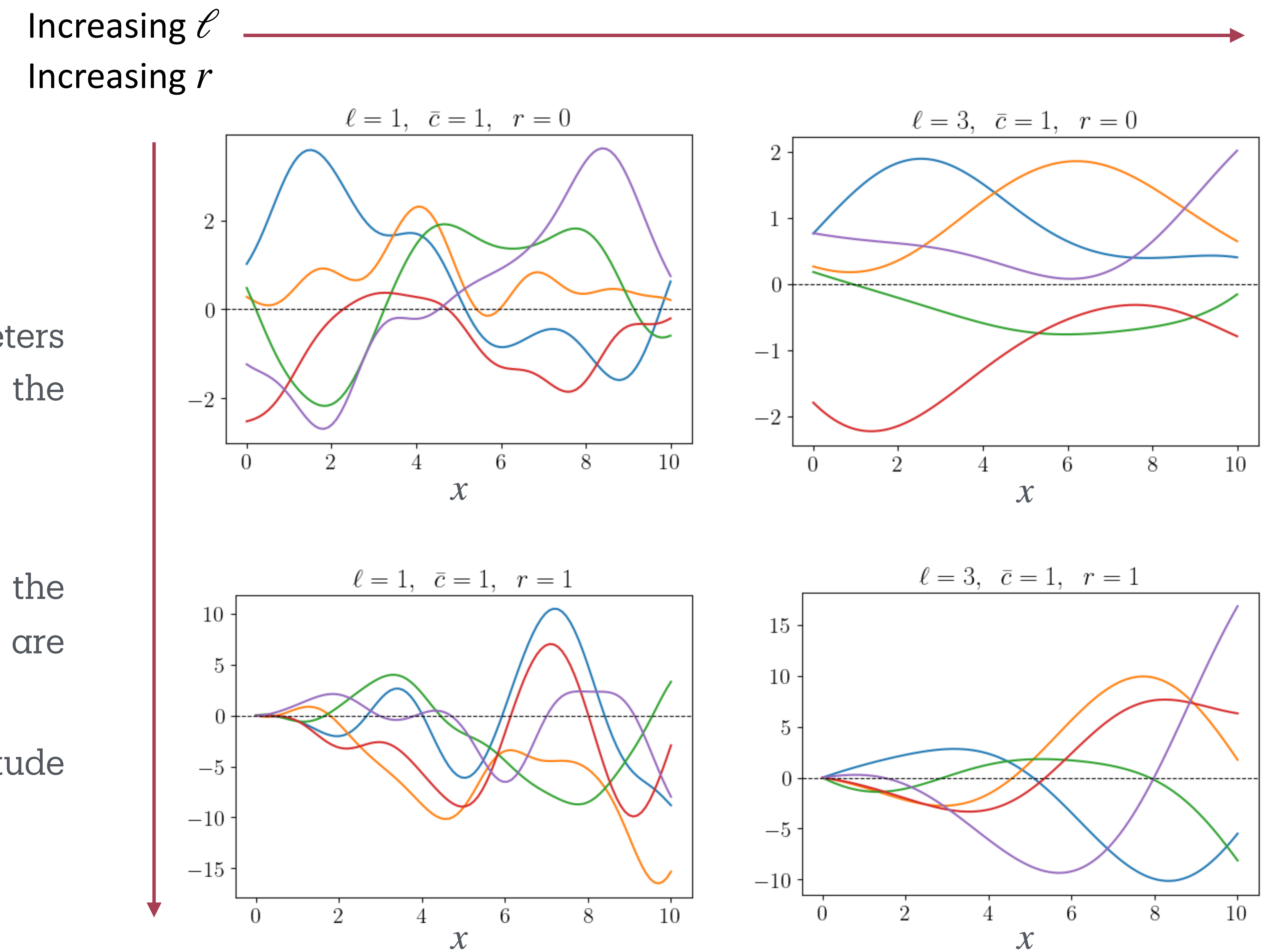
Takeaway: A Gaussian process is a probability distribution over functions. The functions have special properties determined by the covariance kernel.

Covariance kernel: central object of GP

Consider the kernel:

$$K(x_i, x_j) = \bar{c}^2 (x_i x_j)^r \exp\left(-\frac{\|x_i - x_j\|^2}{2\ell^2} \right)$$

- The kernel, for fixed values of the parameters $(\bar{c}, r, \ell)$, generates covariance matrix for the multivariate gaussian distribution.
- ℓ: correlation length between two points.
- $\bar{c}$: controls the overall scale over which the functions vary. As $\bar{c} \to 0$, GP predictions are Dirac-delta distributions.
- r (≥ 0): generates functions whose magnitude increase with x.



Modeling theoretical uncertainty (error) in heavy-ion

- Statistical model:

$$\underbrace{y(x_i)}_{\text{Experimental observation}} = \underbrace{\eta_{\text{mod}}(x_i, \theta)}_{\text{Theoretical prediction}} + \underbrace{\delta_{\text{MD}}(x_i, \phi)}_{\text{Theoretical error}} + \underbrace{\epsilon(x_i)}_{\text{Observation error}}$$

- Model $\delta(\cdot \mid \phi) \sim \text{GP}\left(\mathbf{0}, K(\cdot, \cdot \mid \phi)\right)$

Prior knowledge about theory error: **JETSCAPE model is more reliable at small centralities than at large centralities.**

Reliability of hydrodynamics degrades as one moves from the most central to more peripheral collisions (smaller overlap). Also, Grad and Chapman–Enskog particlization schemes perform better in more central collisions (near-equilibrium corrections).

Modeling theoretical uncertainty (error) in heavy-ion

- Statistical model:

$$\underbrace{y(x_i)}_{\text{Experimental observation}} = \underbrace{\eta_{\mathrm{mod}}(x_i, \theta)}_{\text{Theoretical prediction}} + \underbrace{\delta_{\mathrm{MD}}(x_i, \phi)}_{\text{Theoretical error}} + \underbrace{\epsilon(x_i)}_{\text{Observation error}}$$

- Model $\delta(\cdot \mid \phi) \sim \mathrm{GP}\left(\mathbf{0}, K(\cdot, \cdot \mid \phi)\right)$

Prior knowledge about theory error: **JETSCAPE model is more reliable at small centralities than at large centralities.**

Reliability of hydrodynamics degrades as one moves from the most central to more peripheral collisions (smaller overlap). Also, Grad and Chapman–Enskog particlization schemes perform better in more central collisions (near-equilibrium corrections).

- Consider the kernel: $K(x_i, x_j \mid \phi) \equiv s^2 + \bar{c}^2 (x_i x_j)^r \exp\left(-\frac{\|x_i - x_j\|^2}{2\ell^2}\right), \quad (r \geq 0).$

➡ Allows theory errors to be constant or increase with centrality (but not decrease).

➡ Non-vanishing theory error allowed even for ultra-central collisions ($s \neq 0$).

➡ $s = \bar{c} = 0 \implies \delta_{\mathrm{MD}} = 0.$

Modeling theoretical uncertainty (error) in heavy-ion

- Statistical model:

$$\underbrace{y(x_i)}_{\text{Experimental observation}} = \underbrace{\eta_{\text{mod}}(x_i, \theta)}_{\text{Theoretical prediction}} + \underbrace{\delta_{\text{MD}}(x_i, \phi)}_{\text{Theoretical error}} + \underbrace{\epsilon(x_i)}_{\text{Observation error}}$$

- Model $\delta(\,\cdot \mid \phi) \sim \text{GP}\left(\mathbf{0}, K(\,\cdot\,, \cdot \mid \phi)\right)$

Prior knowledge about theory error: **JETSCAPE model is more reliable at small centralities than at large centralities.**

Reliability of hydrodynamics degrades as one moves from the most central to more peripheral collisions (smaller overlap). Also, Grad and Chapman–Enskog particlization schemes perform better in more central collisions (near-equilibrium corrections).

- Consider the kernel: $K(x_i, x_j \mid \phi) \equiv s^2 + \bar{c}^2 (x_i x_j)^r \exp\left(-\frac{\|x_i - x_j\|^2}{2\ell^2}\right), \quad (r \geq 0).$

➡ Allows theory errors to be constant or increase with centrality (but not decrease).

➡ Non-vanishing theory error allowed even for ultra-central collisions ($s \neq 0$).

➡ $s = \bar{c} = 0 \implies \delta_{\text{MD}} = 0.$

- Fit $\zeta(x; \theta, \phi) \equiv \eta_{\text{mod}}(x, \theta) + \delta_{\text{MD}}(x, \phi)$ to data and estimate θ, ϕ simultaneously.

Bayesian updating of knowledge

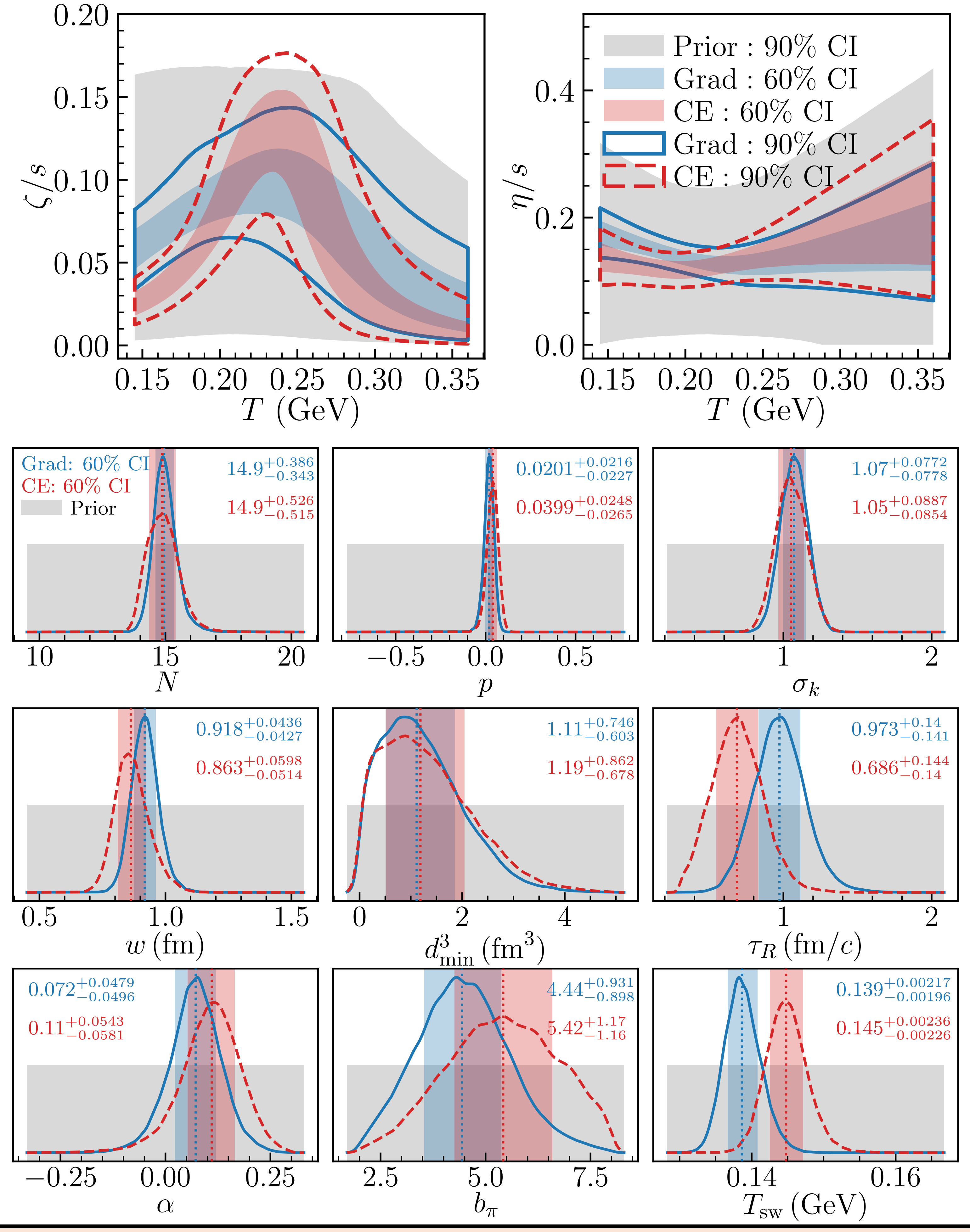
$$\text{pr}(A|B, I) = \frac{\text{pr}(B|A, I)\text{pr}(A|I)}{\text{pr}(B|I)} \implies \underbrace{\text{pr}(\alpha|\mathbf{y}_{\text{exp}}, I)}_{\text{posterior}} \propto \underbrace{\text{pr}(\mathbf{y}_{\text{exp}}|\alpha, I)}_{\text{likelihood}} \times \underbrace{\text{pr}(\alpha|I)}_{\text{prior}}$$

Bayesian inference

SJ, arXiv: 2509.19759

Not accounting for theory uncertainties (w/o MD)

D. Everett *et al.* 2010.03928, 2011.01430



17 parameter Bayesian inference

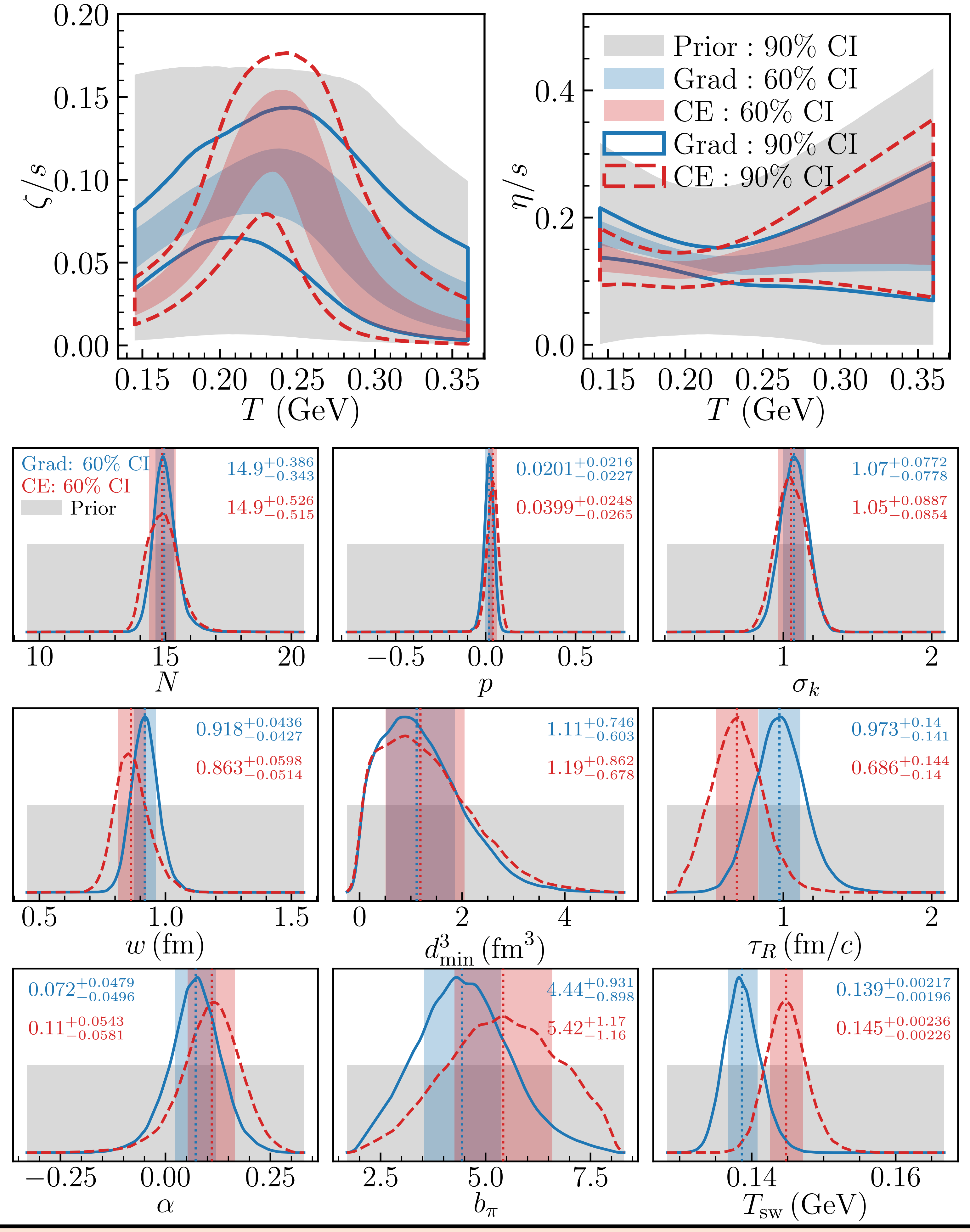
- Tension observed between the specific bulk viscosity posteriors from Grad (blue) and CE (red). Differences also visible for other parameters.
- Particlization temperature lower for Grad: $T_{sw} = 138.65^{+2.17}_{-1.96}$ MeV, than for CE: $144.79^{+2.36}_{-2.26}$ MeV.

Bayesian inference

SJ, arXiv: 2509.19759

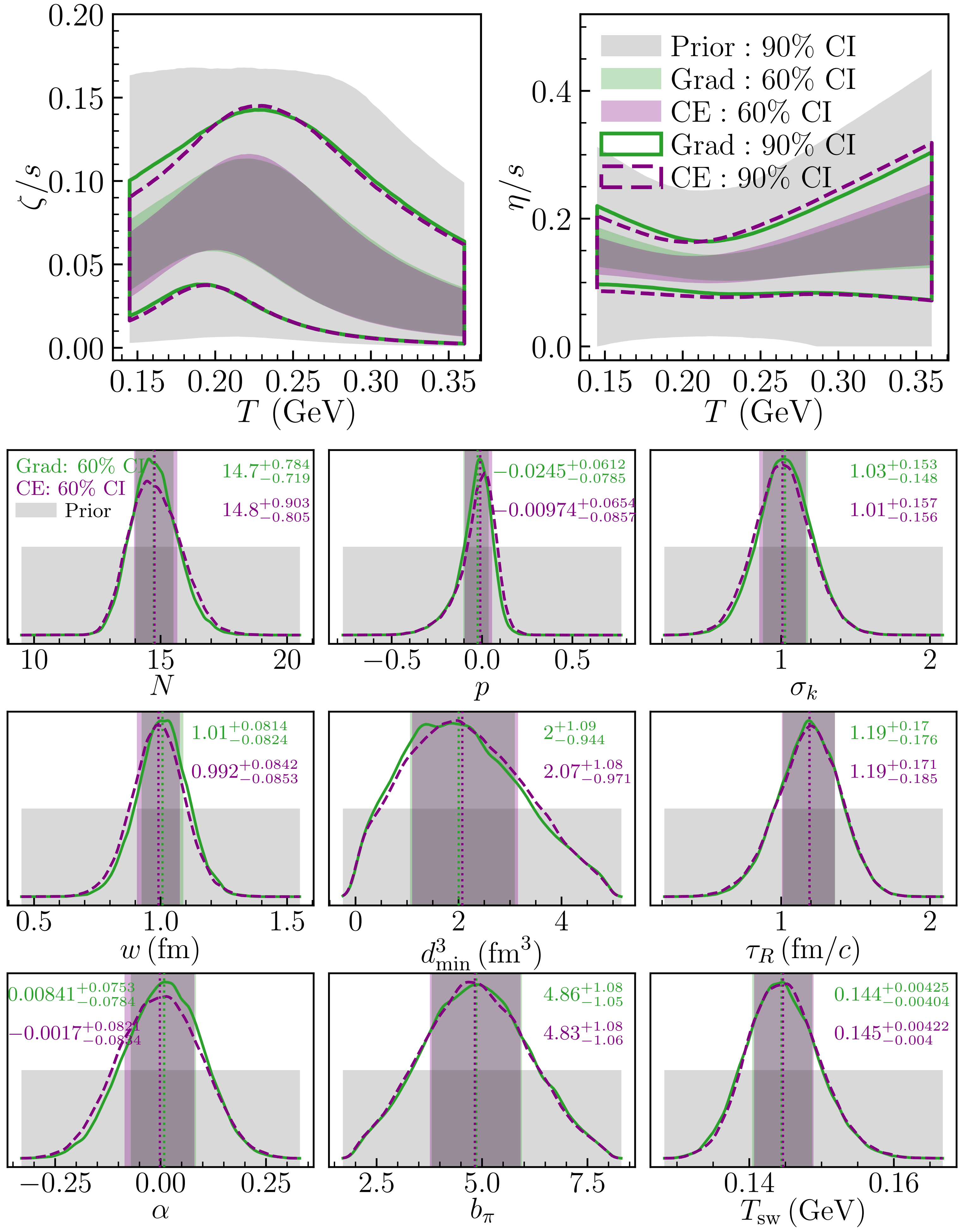
D. Everett *et al.* 2010.03928, 2011.01430

Not accounting for theory uncertainties (w/o MD)



17 parameter Bayesian inference

Accounting for theory uncertainties (w/ MD)



65 parameter Bayesian inference: 17 model parameters + 4 GP param per observable (12 observables ⇒ 48 GP param).

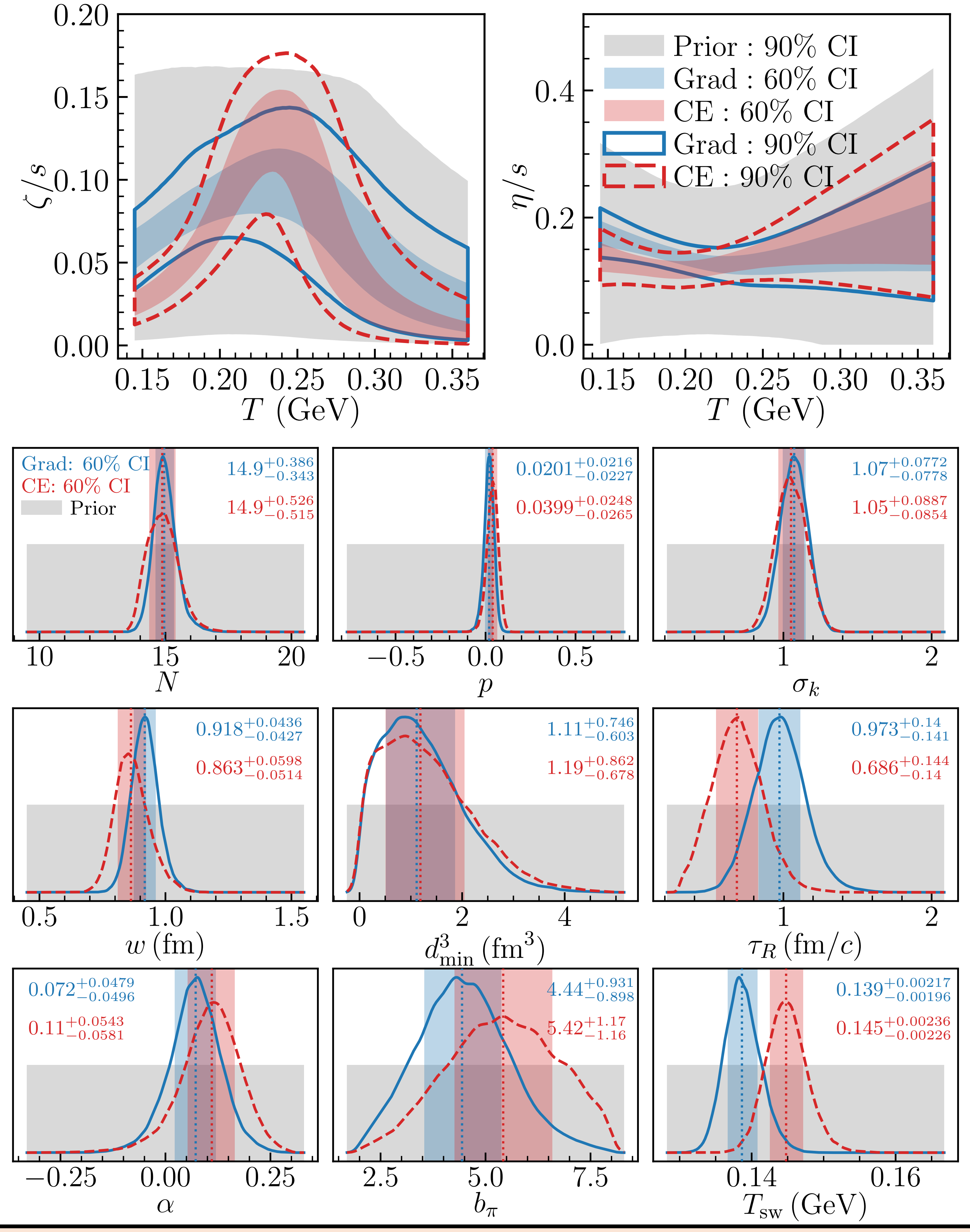
- Tension observed between the specific bulk viscosity posteriors from Grad (blue) and CE (red). Differences also visible for other parameters.
- Particlization temperature lower for Grad: $T_{\rm sw} = 138.65^{+2.17}_{-1.96}$ MeV, than for CE: $144.79^{+2.36}_{-2.26}$ MeV.

Bayesian inference

SJ, arXiv: 2509.19759

D. Everett *et al.* 2010.03928, 2011.01430

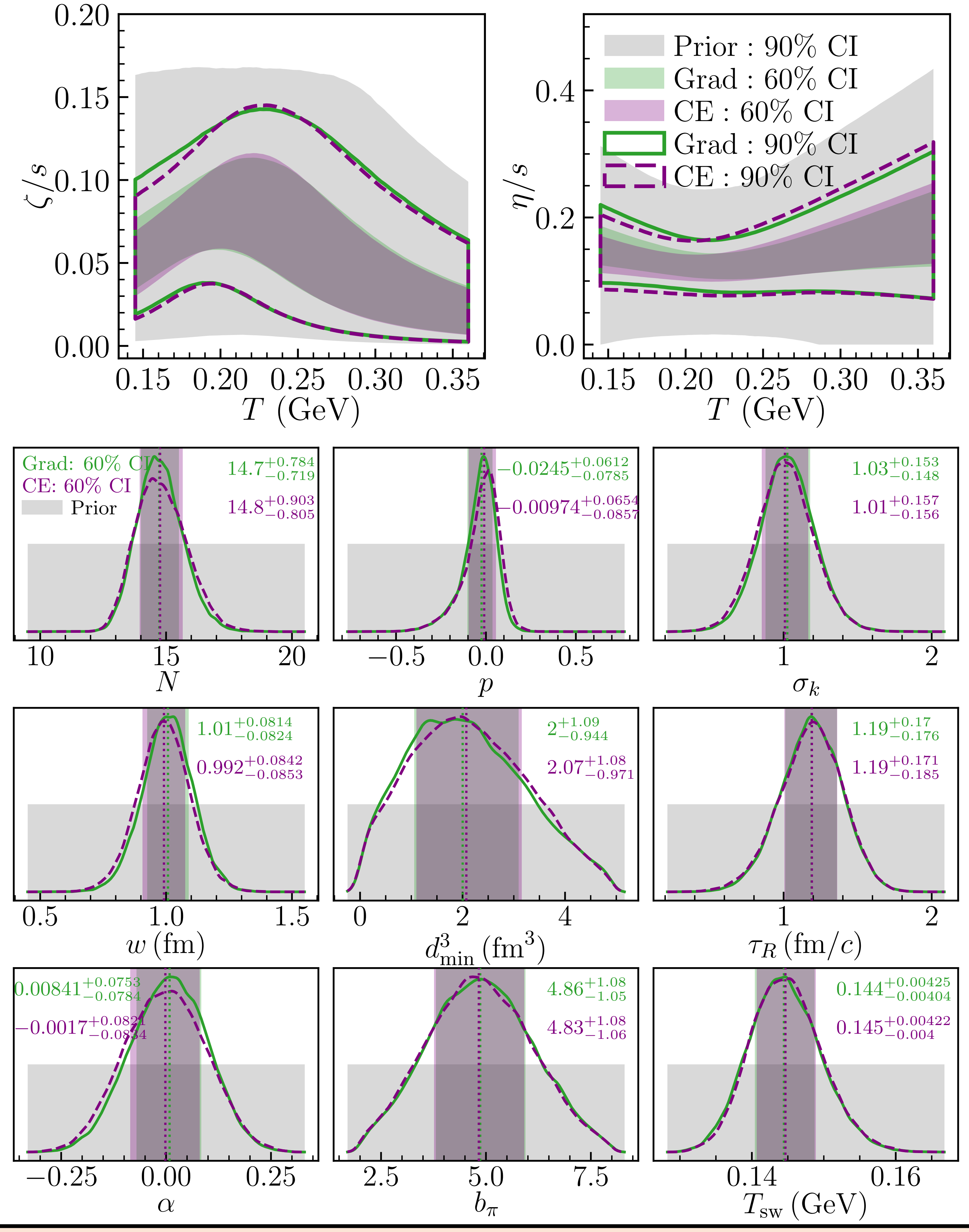
Not accounting for theory uncertainties (w/o MD)



17 parameter Bayesian inference

- Tension observed between the specific bulk viscosity posteriors from Grad (blue) and CE (red). Differences also visible for other parameters.
- Particlization temperature lower for Grad: $T_{\rm sw} = 138.65^{+2.17}_{-1.96}$ MeV, than for CE: $144.79^{+2.36}_{-2.26}$ MeV.

Accounting for theory uncertainties (w/ MD)



65 parameter Bayesian inference: 17 model parameters + 4 GP param per observable (12 observables ⇒ 48 GP param).

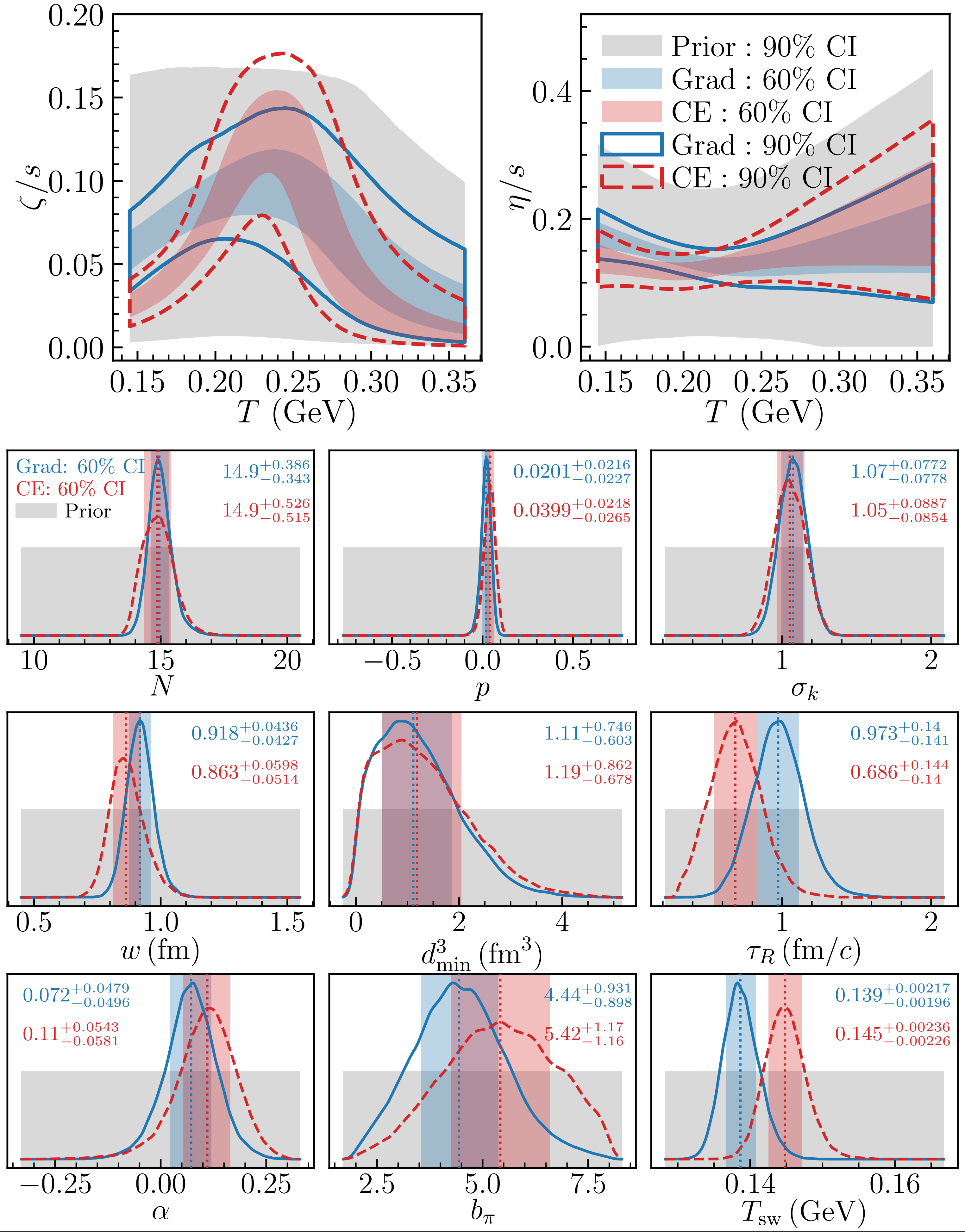
- Posteriors are almost indistinguishable for all parameters for Grad (green) and CE (purple).

Bayesian inference

SJ, arXiv: 2509.19759

D. Everett *et al.* 2010.03928, 2011.01430

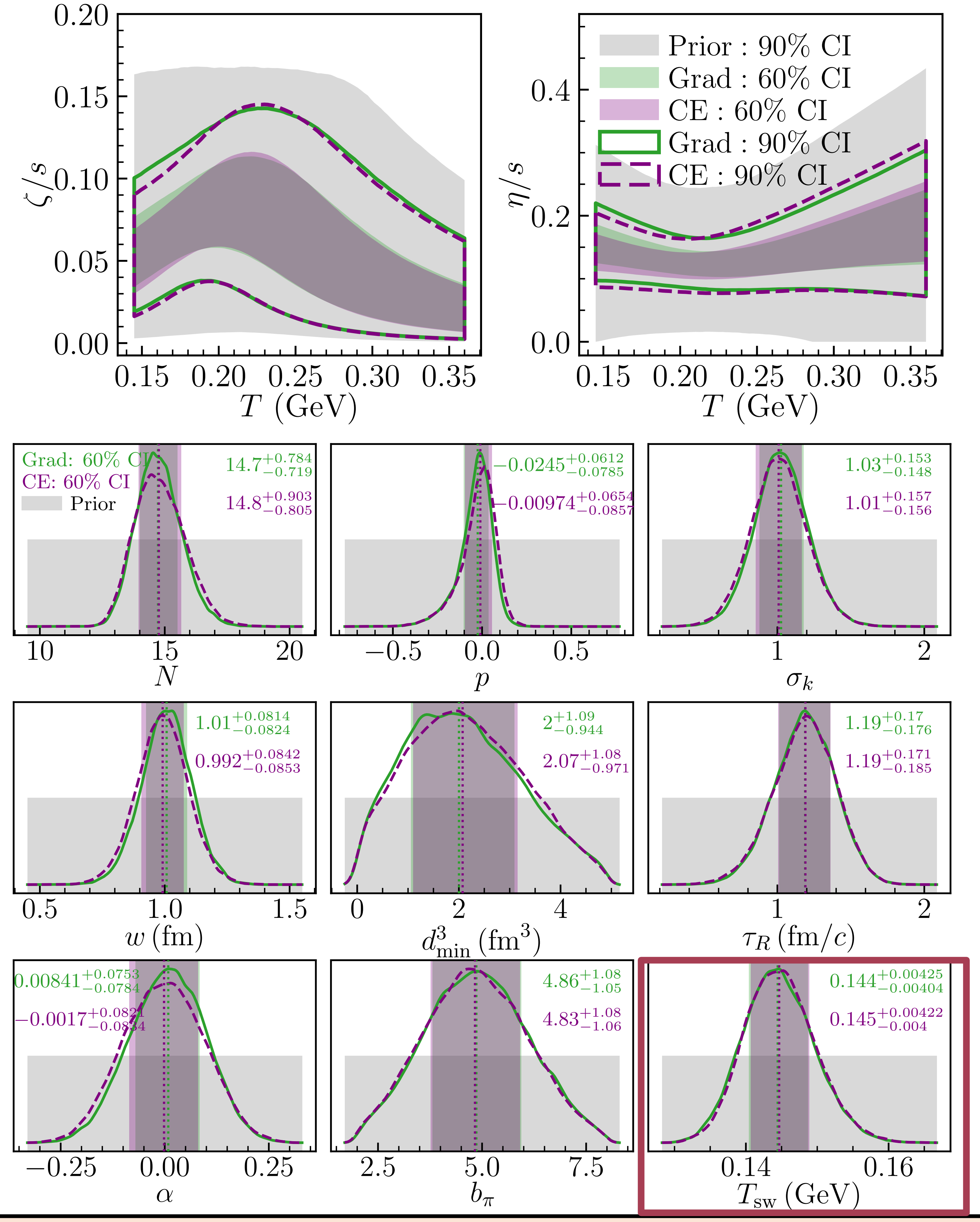
Not accounting for theory uncertainties (w/o MD)



17 parameter Bayesian inference

- Tension observed between the specific bulk viscosity posteriors from Grad (blue) and CE (red). Differences also visible for other parameters.
- Particlization temperature lower for Grad: $T_{\rm sw} = 138.65^{+2.17}_{-1.96}$ MeV, than for CE: $144.79^{+2.36}_{-2.26}$ MeV.

Accounting for theory uncertainties (w/ MD)



65 parameter Bayesian inference: 17 model parameters + 4 GP param per observable (12 observables ⇒ 48 GP param).

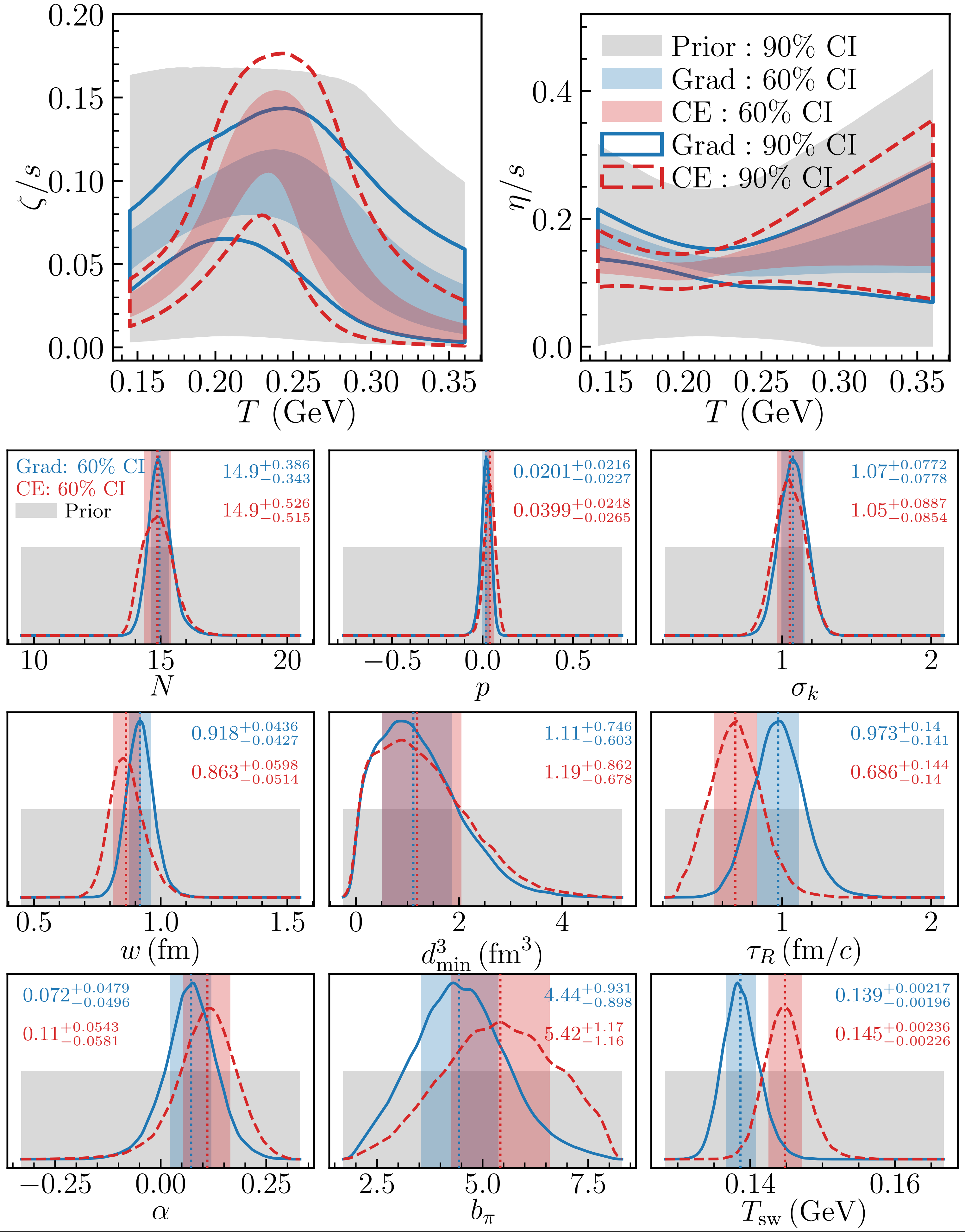
- Posteriors are almost indistinguishable for all parameters for Grad (green) and CE (purple).
- Particlization temperature consistent: $T_{\rm sw} = 144.46^{+4.25}_{-4.04}$ MeV (Grad) and $T_{\rm sw} = 144.66^{+4.22}_{-4.00}$ MeV (CE).

Bayesian inference

SJ, arXiv: 2509.19759

D. Everett *et al.* 2010.03928, 2011.01430

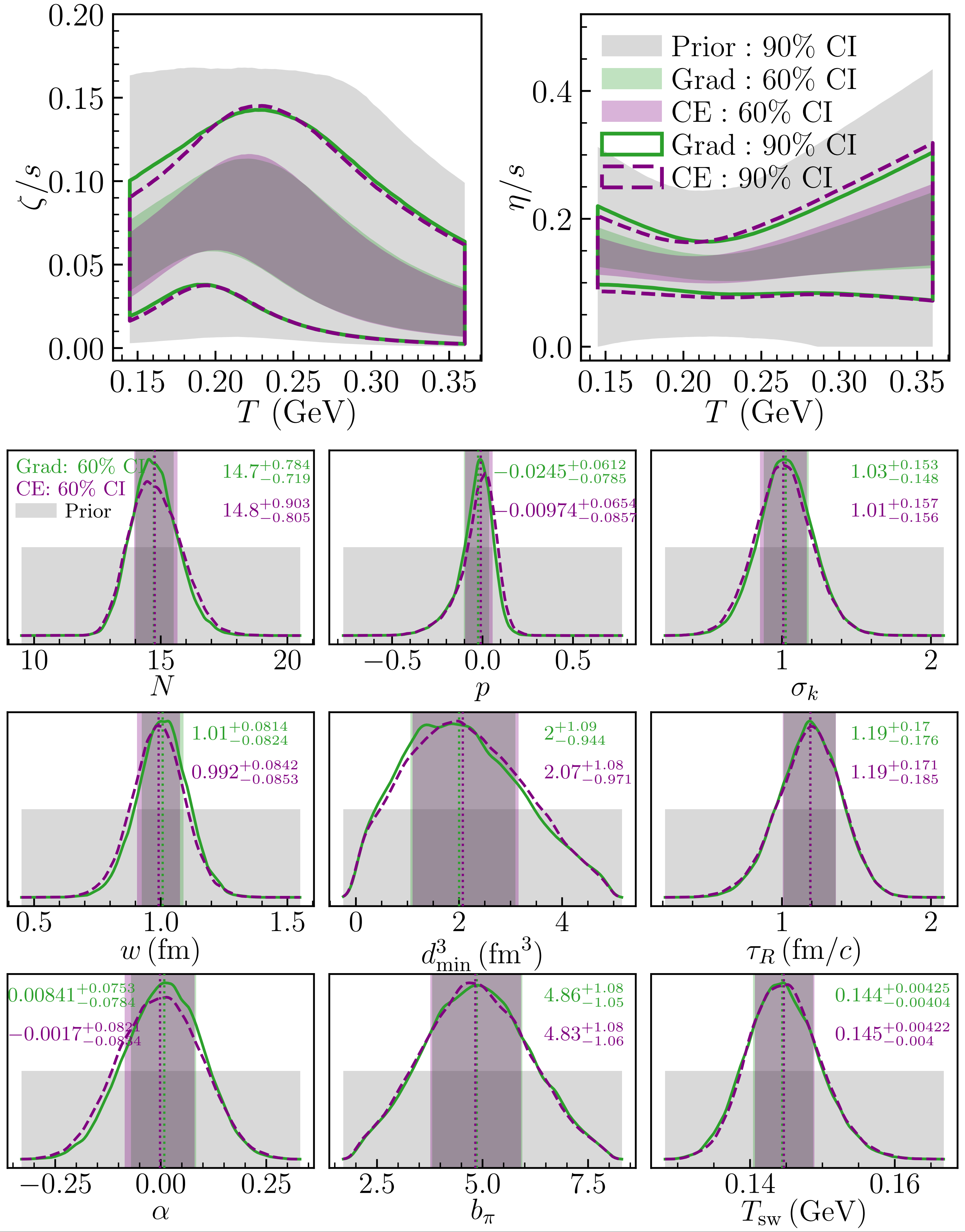
Not accounting for theory uncertainties (w/o MD)



17 parameter Bayesian inference

Error of theory is quantified and does not propagate across stages => same $T^{\mu\nu}$ evolution for Grad and CE.

Accounting for theory uncertainties (w/ MD)



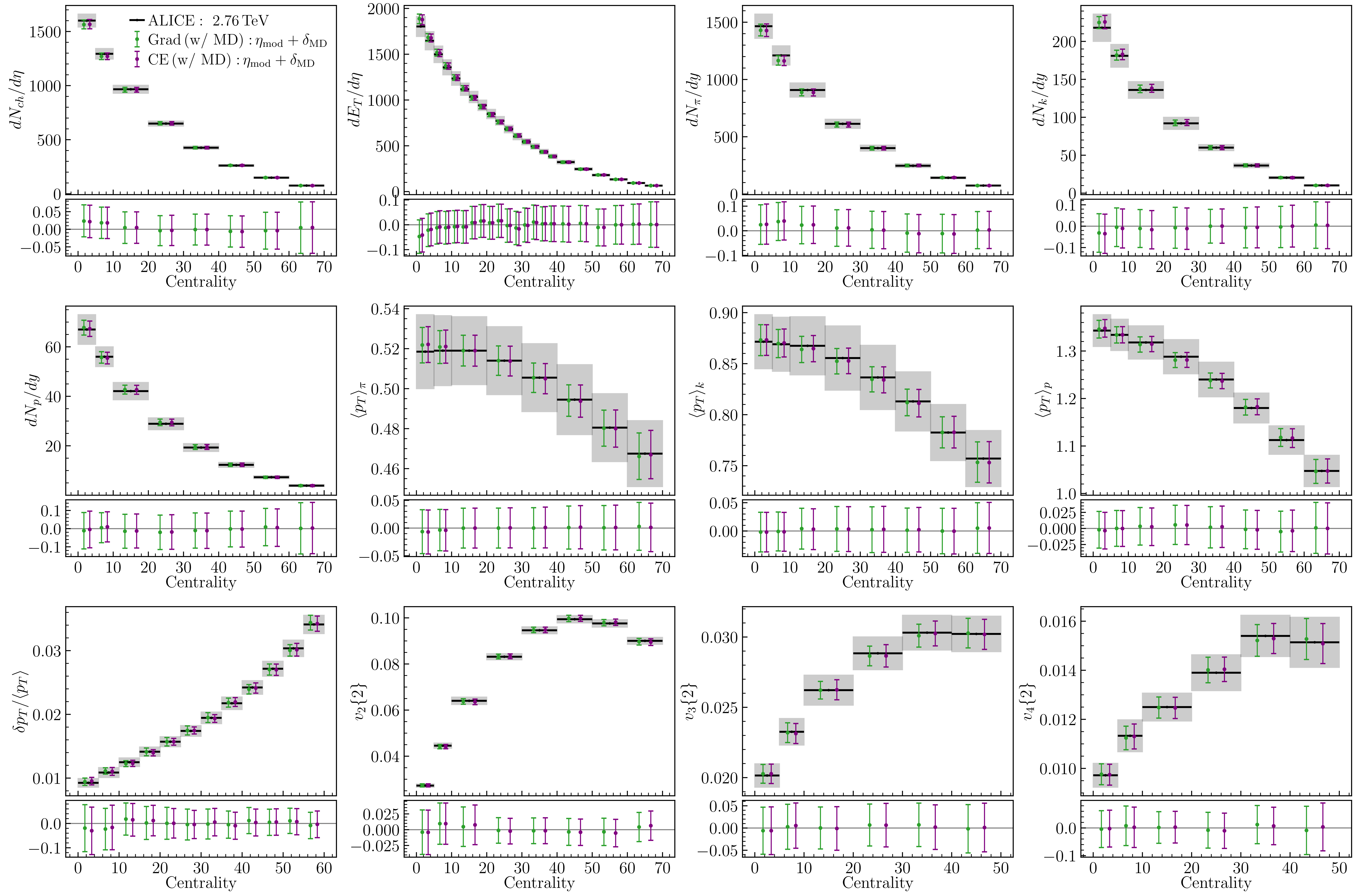
65 parameter Bayesian inference: 17 model parameters + 4 GP param per observable (12 observables ⇒ 48 GP param).

- Tension observed between the specific bulk viscosity posteriors from Grad (blue) and CE (red). Differences also visible for other parameters.
- Particlization temperature lower for Grad: $T_{\rm sw} = 138.65^{+2.17}_{-1.96}$ MeV, than for CE: $144.79^{+2.36}_{-2.26}$ MeV.

- Posteriors are almost indistinguishable for all parameters for Grad (green) and CE (purple).
- Particlization temperature consistent: $T_{\rm sw} = 144.46^{+4.25}_{-4.04}$ MeV (Grad) and $T_{\rm sw} = 144.66^{+4.22}_{-4.00}$ MeV (CE).

Model + discrepancy predictions

SJ, arXiv: 2509.19759



normalized error $(y_{\rm exp} - \eta_{\rm mod} - \delta_{\rm MD})/|\langle y_{\rm exp} \rangle|$

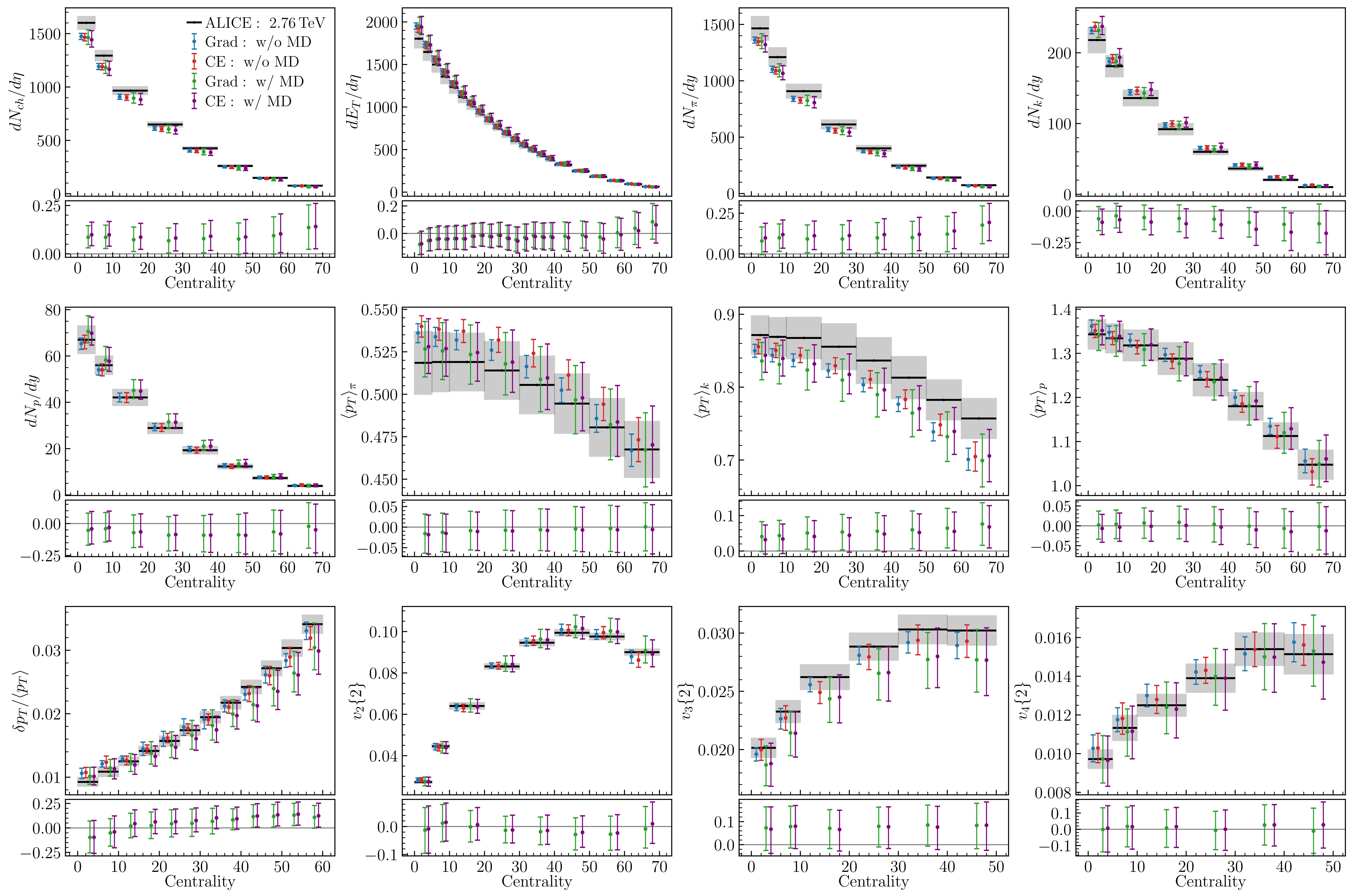
Predictions from

$\zeta(x; \theta, \phi) = \eta_{\rm mod}(x, \theta) + \delta_{\rm MD}(x, \phi)$

Excellent agreement with data.
Expected.

Model predictions

SJ, arXiv: 2509.19759



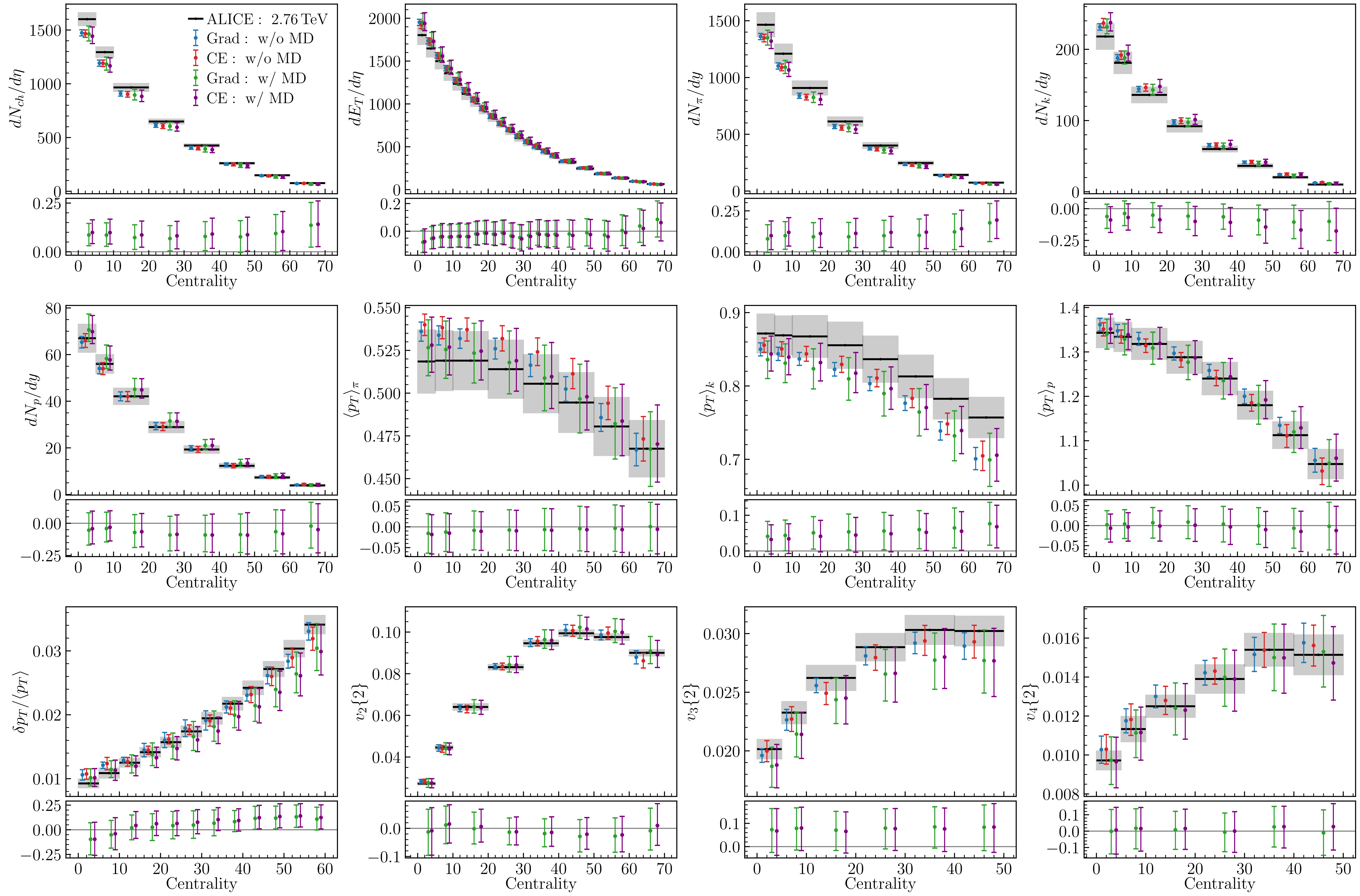
Predictions from $\eta_{\text{mod}}(x, \theta)$.

normalized discrepancies $(y_{\text{exp}} - \eta_{\text{mod}})/|\langle y_{\text{exp}} \rangle|$

$\approx \delta_{\text{MD}}(x, \phi)$

Model predictions

SJ, arXiv: 2509.19759



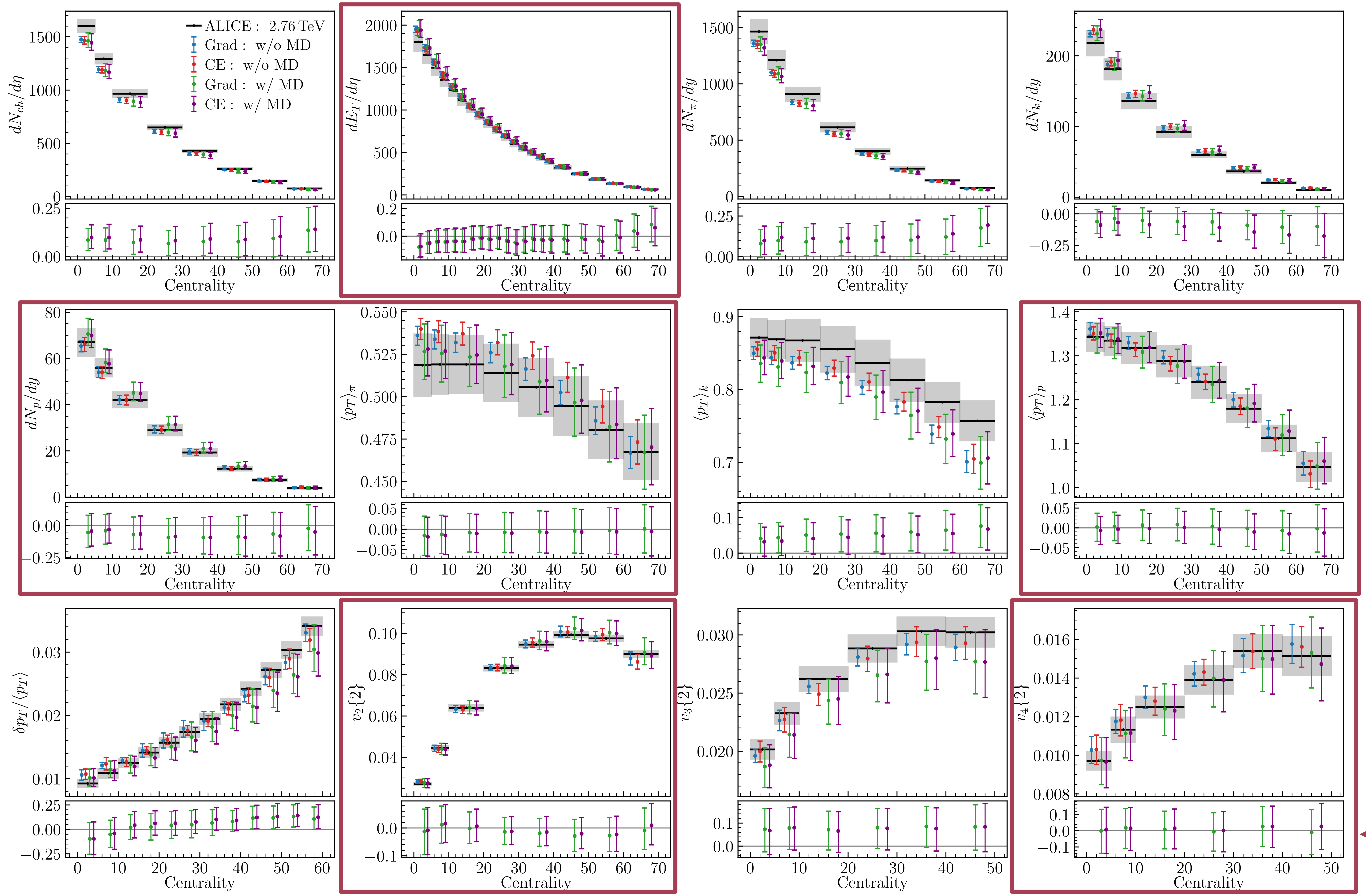
Predictions from $\eta_{\mathrm{mod}}(x, \theta)$.

- No clear preference for Grad or CE is observed.

normalized discrepancies $(y_{\mathrm{exp}} - \eta_{\mathrm{mod}})/|\langle y_{\mathrm{exp}} \rangle|$ $\approx \delta_{\mathrm{MD}}(x, \phi)$

Model predictions

SJ, arXiv: 2509.19759



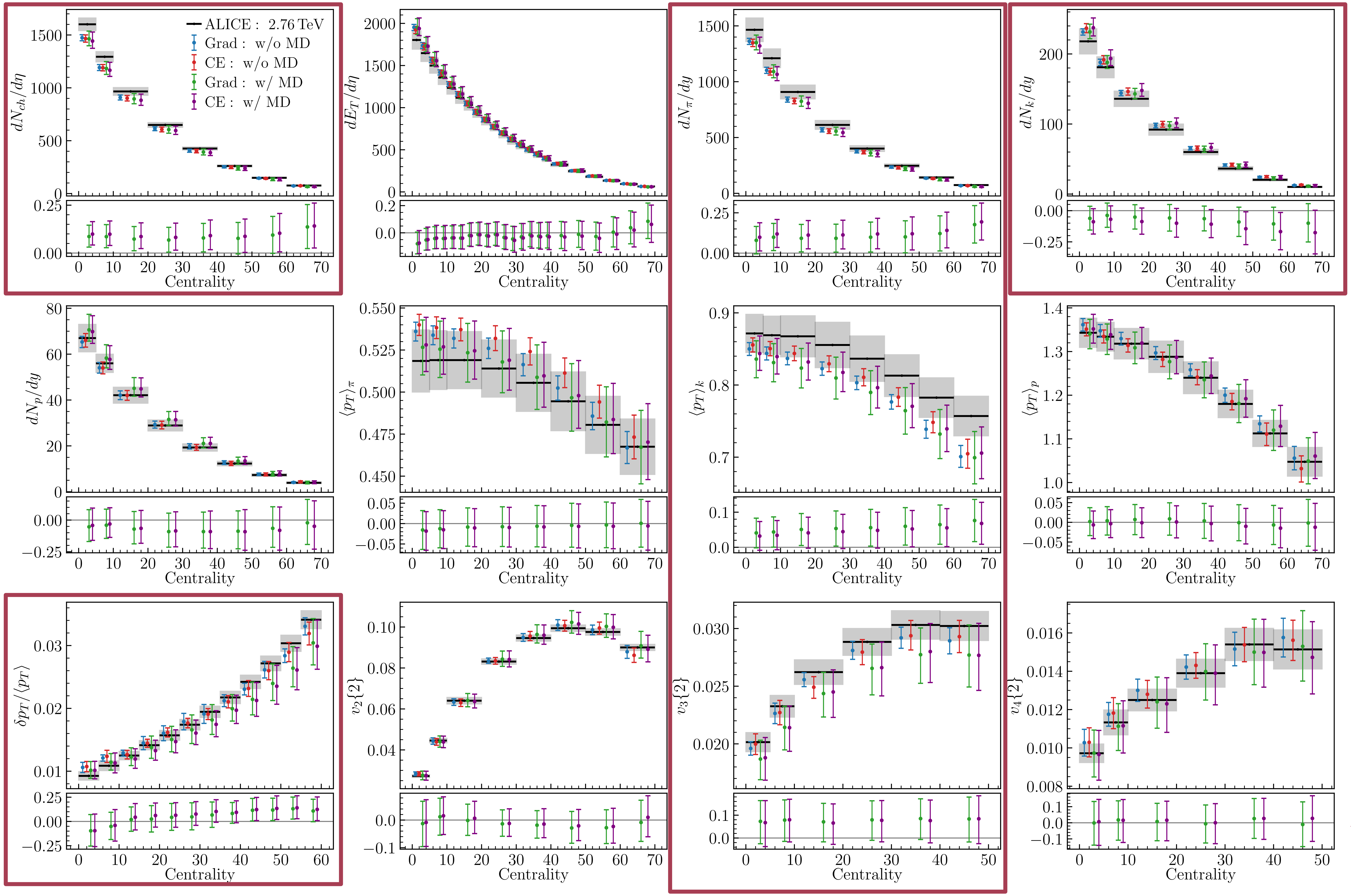
normalized discrepancies $(y_{\text{exp}} - \eta_{\text{mod}})/|\langle y_{\text{exp}} \rangle|$ $\approx \delta_{\text{MD}}(x, \phi)$

Predictions from $\eta_{\text{mod}}(x, \theta)$.

- No clear preference for Grad or CE is observed.
- Theory predicts many observables correctly => learned error are small for these.

Model predictions

SJ, arXiv: 2509.19759



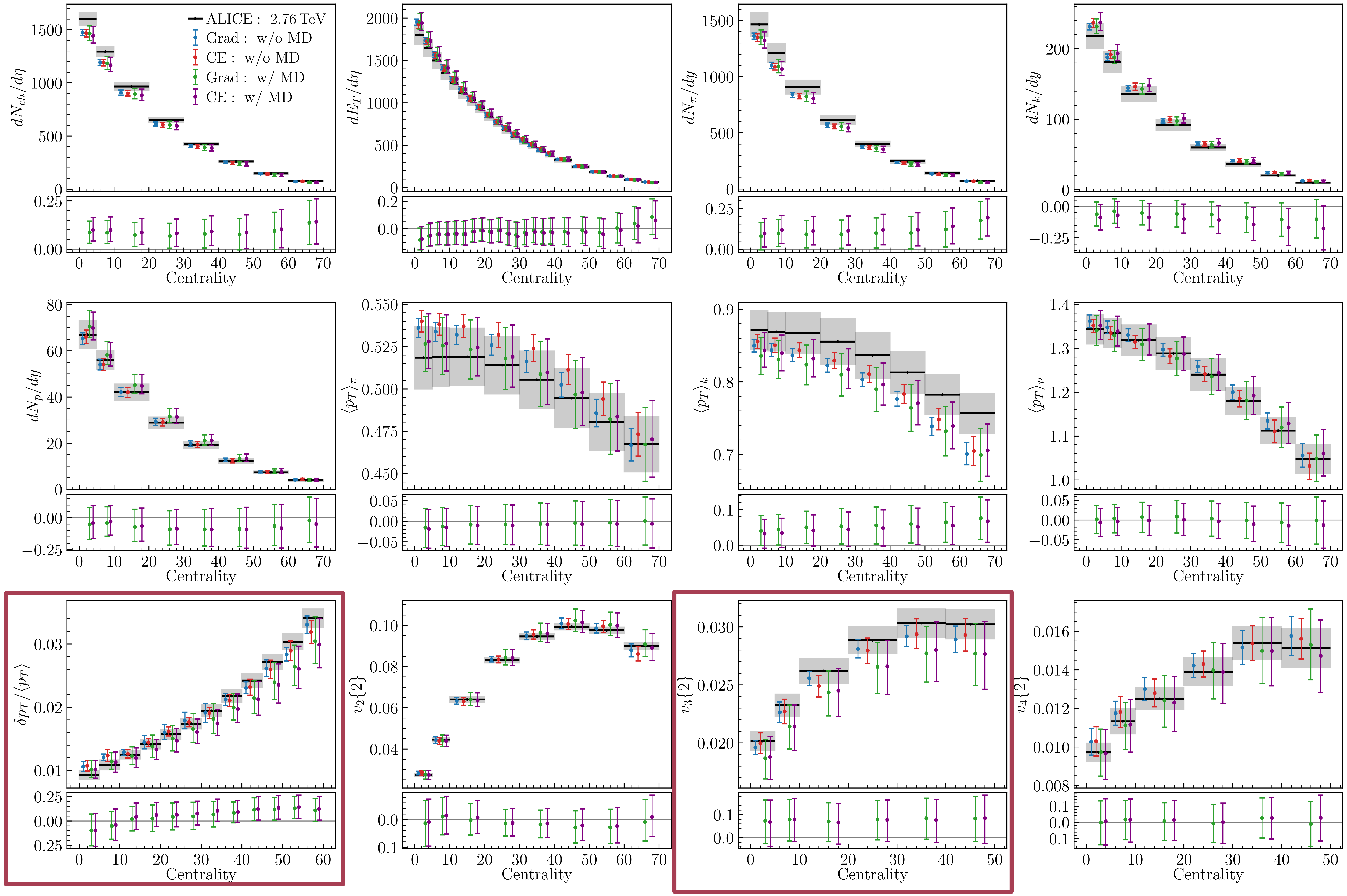
Predictions from $\eta_{\text{mod}}(x, \theta)$.

- No clear preference for Grad or CE is observed.
- Theory predicts many observables correctly => learned error are small for these.
- Deviations observed in w/ MD predictions => learned theory error are large.

normalized discrepancies $(y_{\text{exp}} - \eta_{\text{mod}})/|\langle y_{\text{exp}} \rangle|$ $\approx \delta_{\text{MD}}(x, \phi)$

Model predictions

SJ, arXiv: 2509.19759



Predictions from $\eta_{\rm mod}(x, \theta)$.

- No clear preference for Grad or CE is observed.
- Theory predicts many observables correctly => learned error are small for these.
- Deviations observed in w/ MD predictions => learned theory error are large.
- Particularly large deviations in $v_3\{2\}$ and $\delta p_T/\langle p_T\rangle$ suggest deficiencies in the initial-condition model.

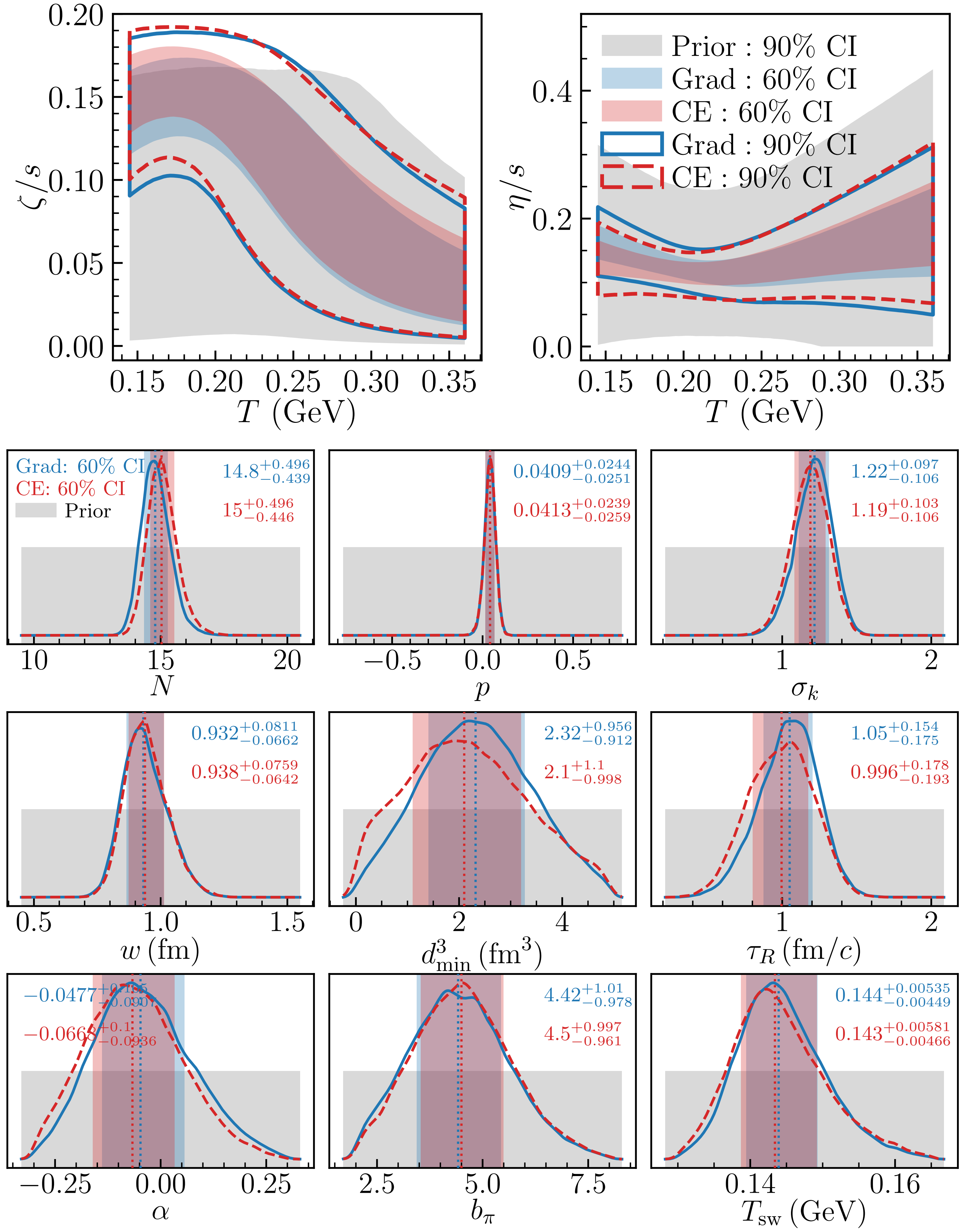
Guides understanding for targeted improvement in models.

normalized discrepancies $(y_{\rm exp} - \eta_{\rm mod})/|\langle y_{\rm exp}\rangle|$ $\approx \delta_{\rm MD}(x, \phi)$

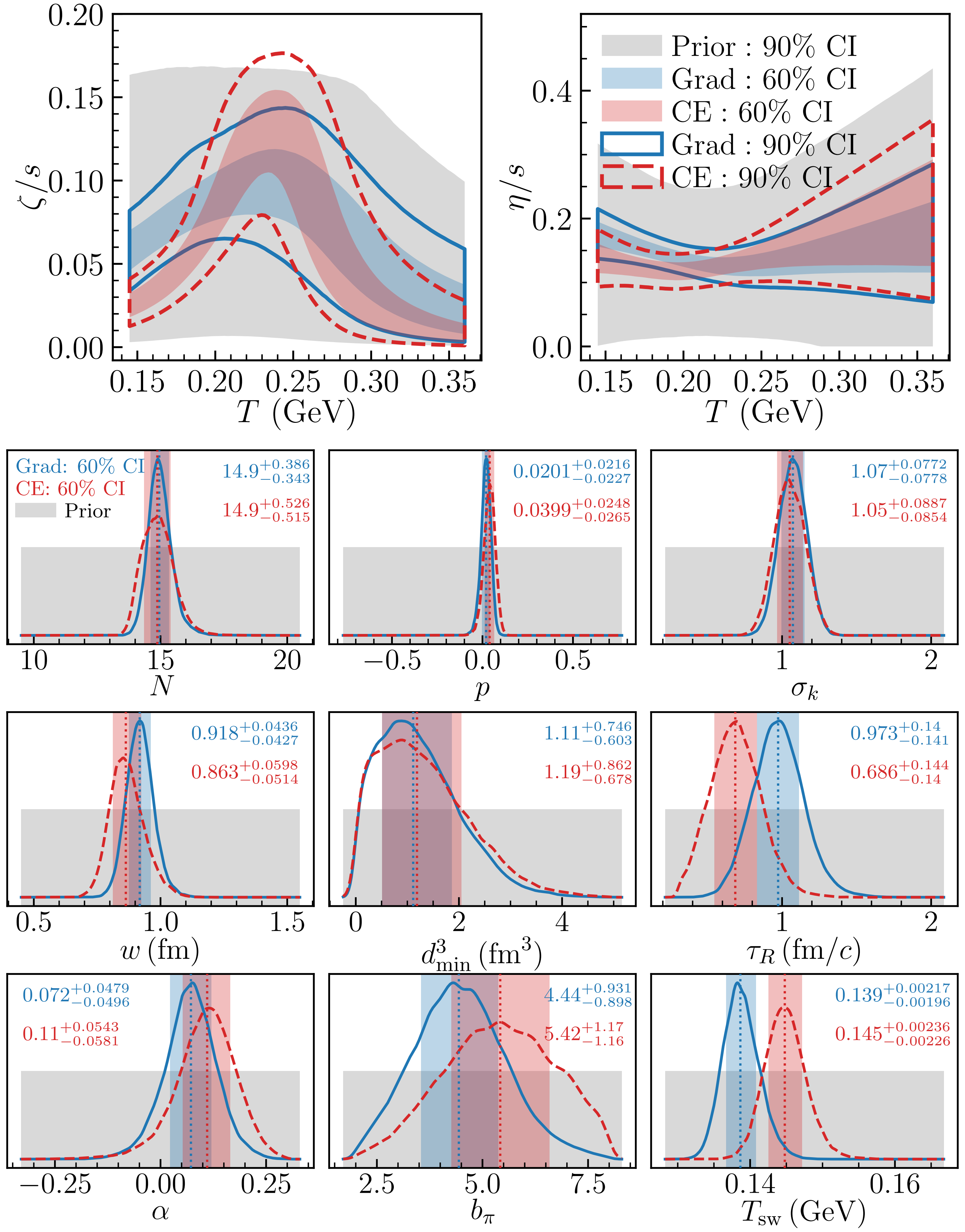
Check (w/o MD): stability of posteriors as more observables are included

SJ, arXiv: 2509.19759

Calibration with subset of observables:
$dN_{\rm ch}/d\eta$, $dE_T/d\eta$, and $v_n\{2\}$ $(n = 2,3,4)$



Calibration with all observables



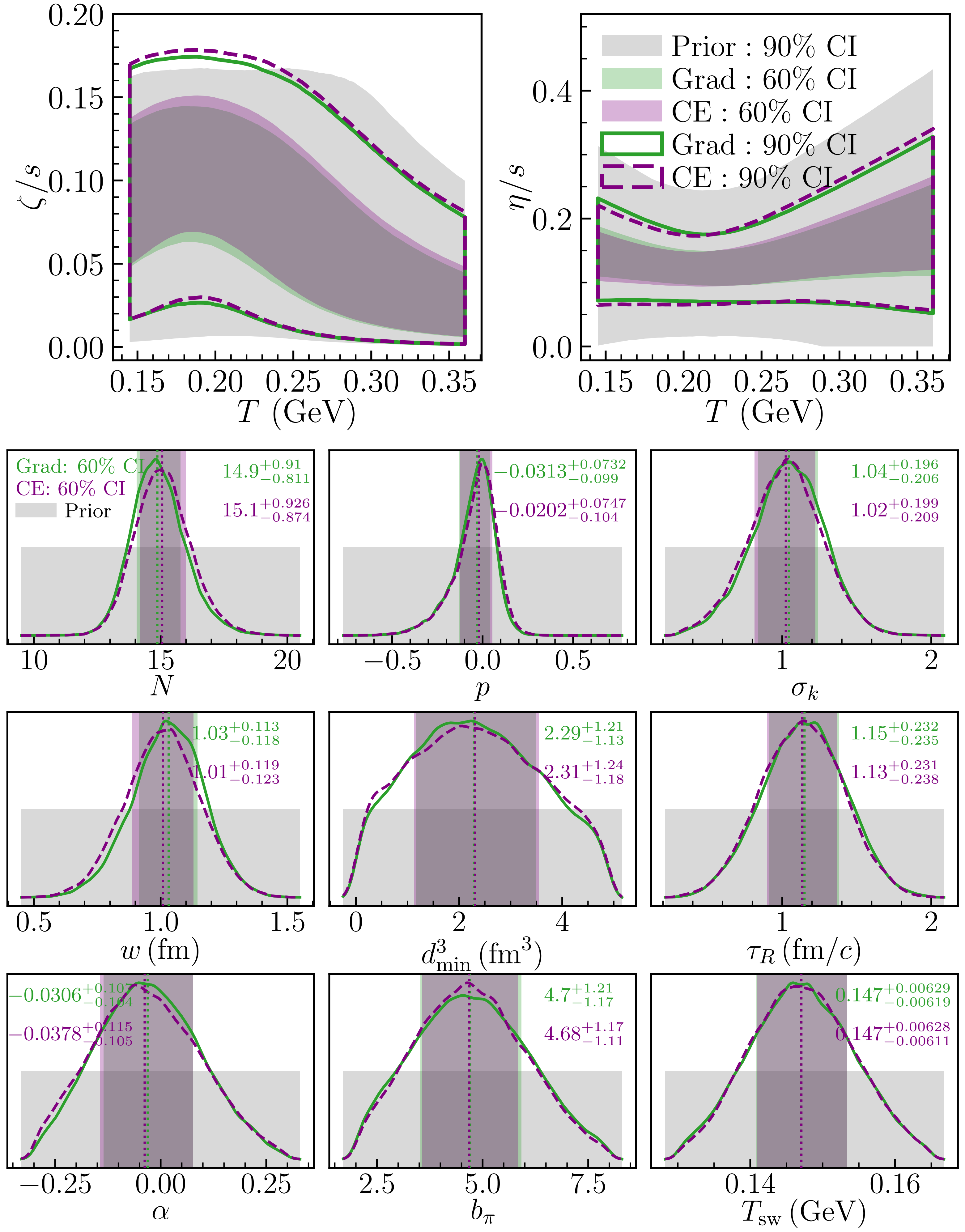
w/o MD

Posteriors for many parameters shift significantly when all measurements are included in the inference, indicating lack of robustness of the parameter estimate.

Check (w/ MD): stability of posteriors as more observables are included

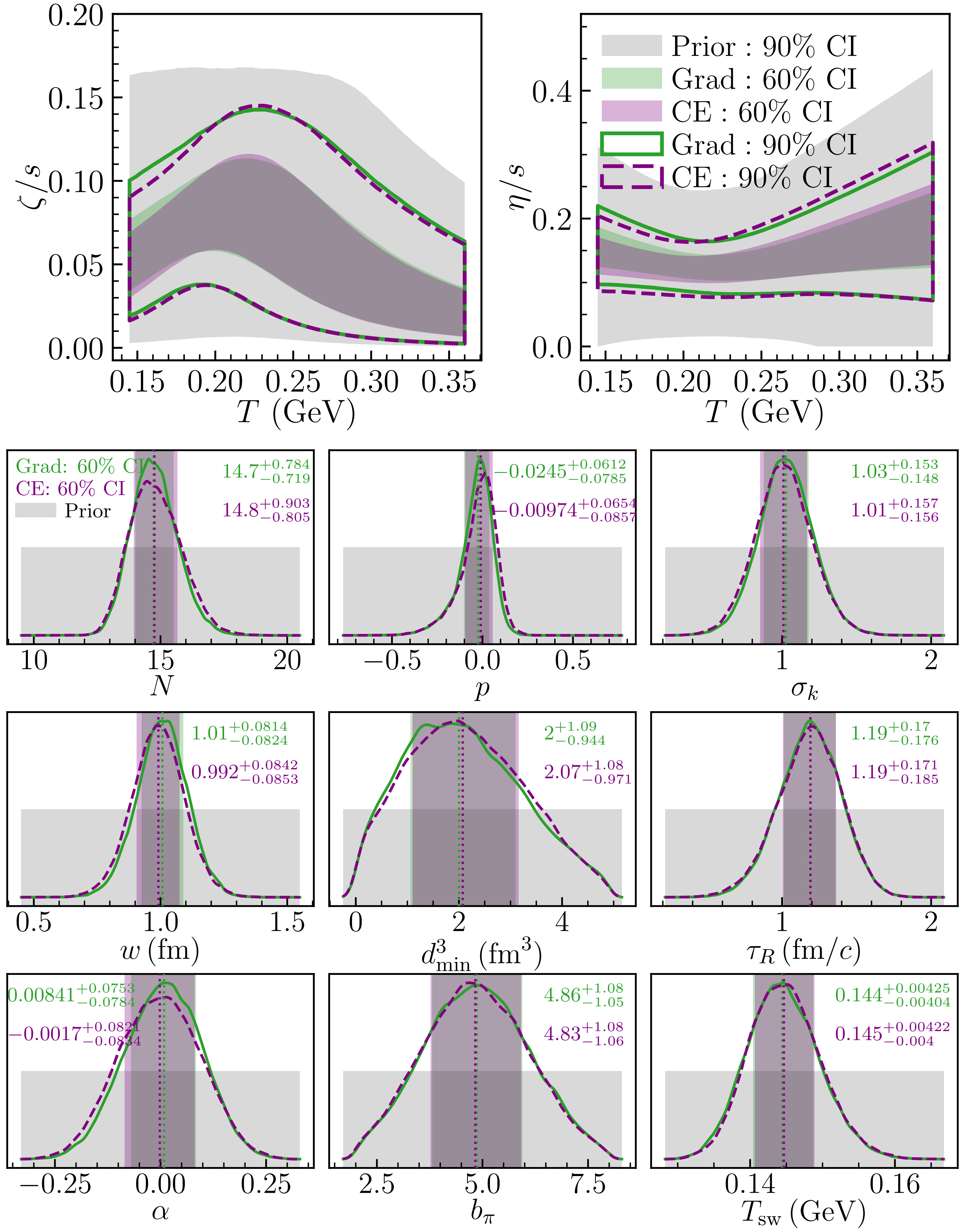
SJ, arXiv: [2509.19759]

Calibration with subset of observables: $dN_{\text{ch}}/d\eta$, $dE_T/d\eta$, and $v_n\{2\}$ $(n = 2,3,4)$



→ Calibration with all observables

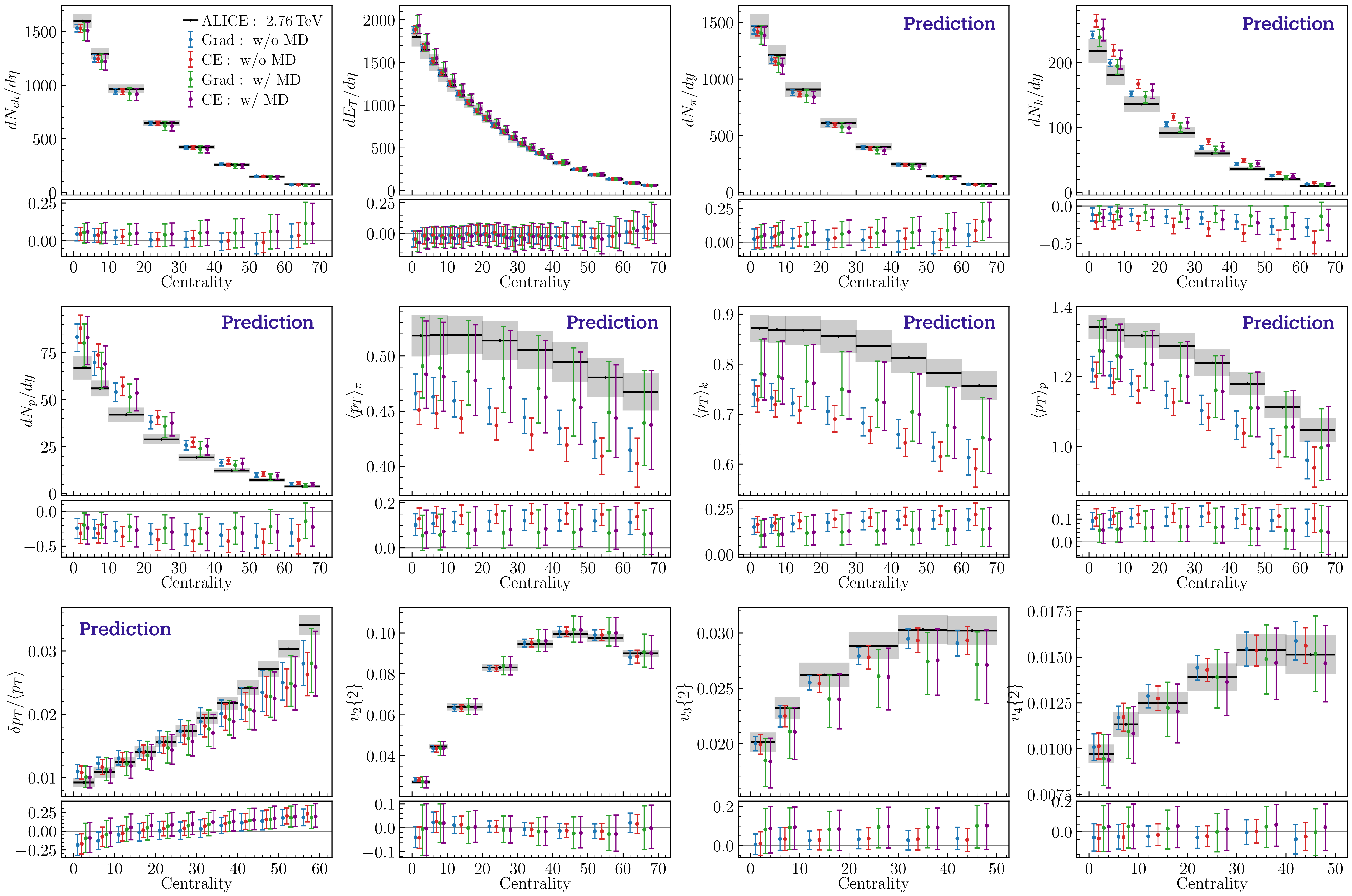
w/ MD



Posteriors tighten without shifting, indicating robust parameter inference.

Model predictions: calibrated on subset of observables

SJ, arXiv: [2509.19759]



Predictions from $\eta_{\text{mod}}(x, \theta)$.

- Observables not used in calibration are labeled "Prediction".
- w/o MD predictions for $\langle p_T \rangle_i,\ i \in \{\pi, k, p\}$ are outside 68 % CI prediction interval, reflecting overfitting.
- w/ MD results show comparatively better agreement with data, albeit with large error bars.

normalized discrepancies $(y_{\text{exp}} - \eta_{\text{mod}})/|\langle y_{\text{exp}} \rangle|$

$\approx \delta_{\text{MD}}(x, \phi)$

Summary

- Often the goal of model-data comparison is to extract physically meaningful model parameters, **not to fit data**.
- Quantifying theoretical uncertainties is **crucial** for correct inference of physical model parameters.

➡ Implicit assumption of the model being universally valid is usually made in model-to-data comparison. This can lead to incorrect and misleading parameter estimates.

- Discussed a Bayesian framework that explicitly quantifies theoretical uncertainties by statistically modeling theory errors, guided by qualitative knowledge of the theory's domain of reliability.
- Derived robust, data-driven constraints on transport properties of QCD with quantified theory uncertainties.
- Framework is general and can be applied to various problems in science.
  Open source code is provided: https://github.com/sjaiswal-tifr/ModelDiscrepancy

Thank you!

Backup slides

Gaussian Process: Distribution over functions

Formal definition: A Gaussian process is a collection of random variables, any finite number of which also have a joint Gaussian distribution.

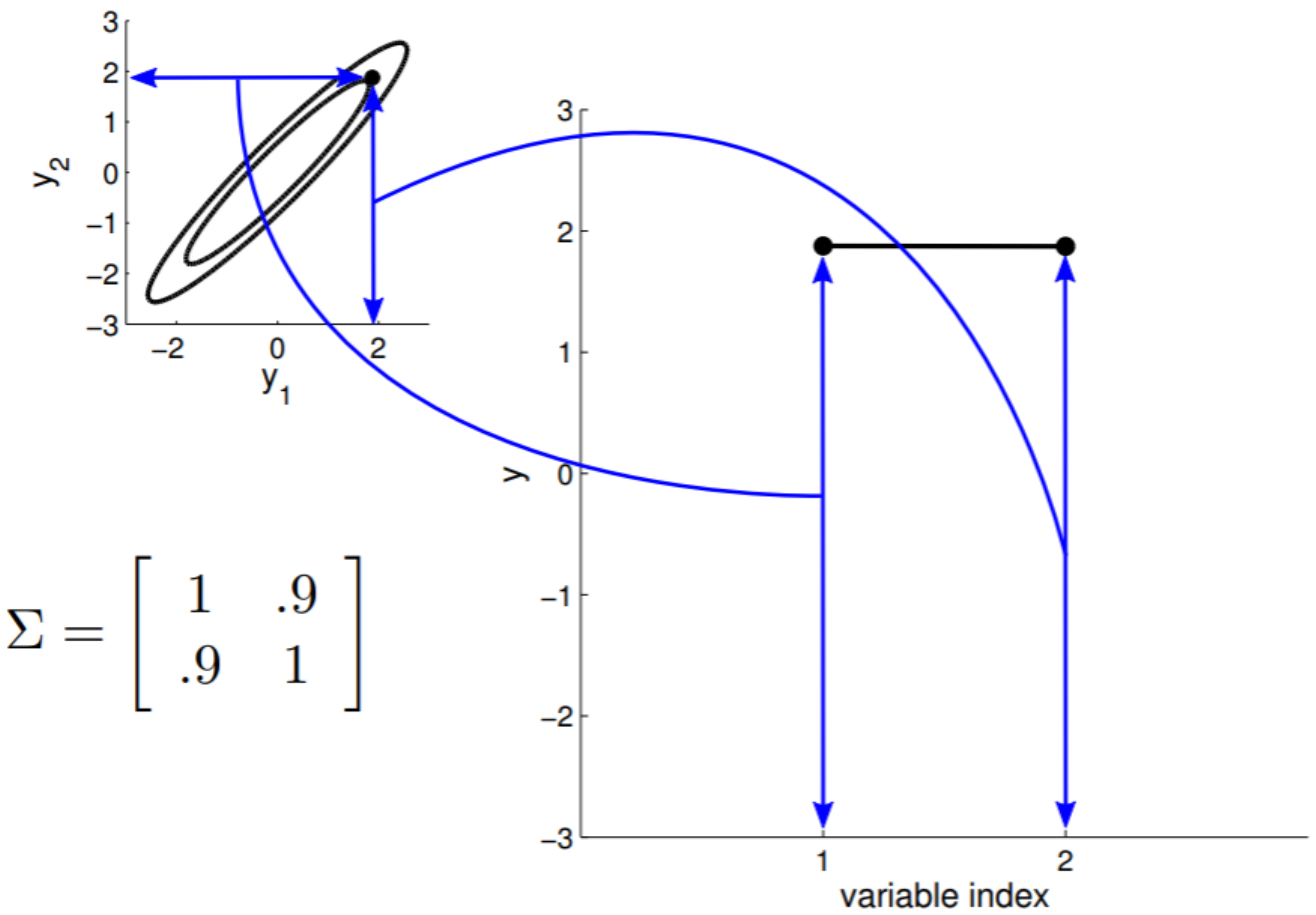
Consider a multivariate Gaussian distribution: $p(\mathbf{y} \mid \Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}\right)$

Gaussian Process: Distribution over functions

Formal definition: A Gaussian process is a collection of random variables, any finite number of which also have a joint Gaussian distribution.

Consider a multivariate Gaussian distribution: $p(\mathbf{y} \mid \Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}\right)$

2D Gaussian distribution — A different representation



$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

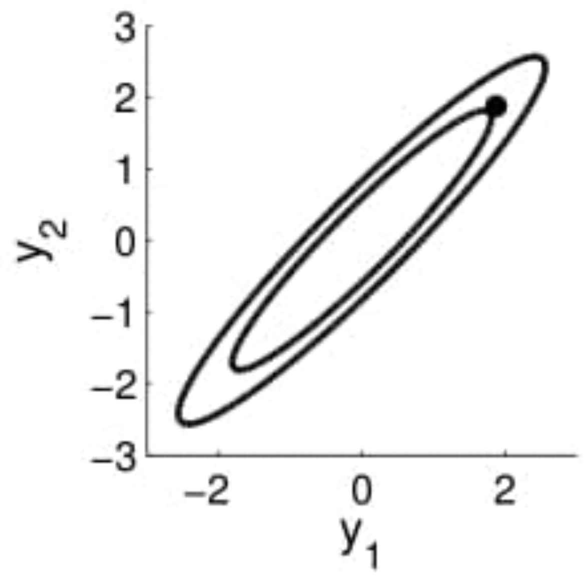
https://thegradient.pub/gaussian-process-not-quite-for-dummies/

Gaussian Process: Distribution over functions

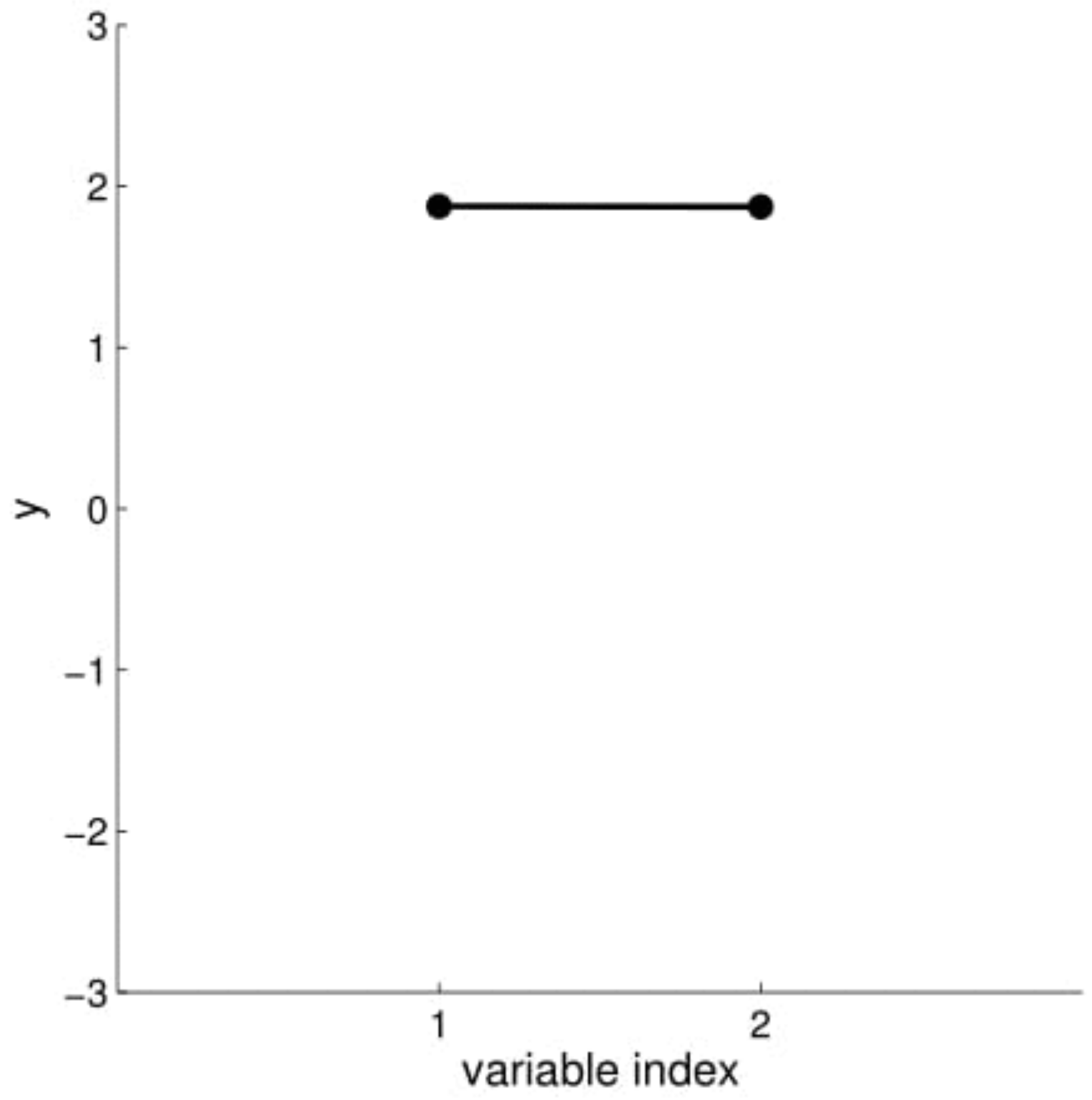
Formal definition: A Gaussian process is a collection of random variables, any finite number of which also have a joint Gaussian distribution.

Consider a multivariate Gaussian distribution: $p(\mathbf{y} \mid \Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}\right)$

2D Gaussian distribution — A different representation



$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



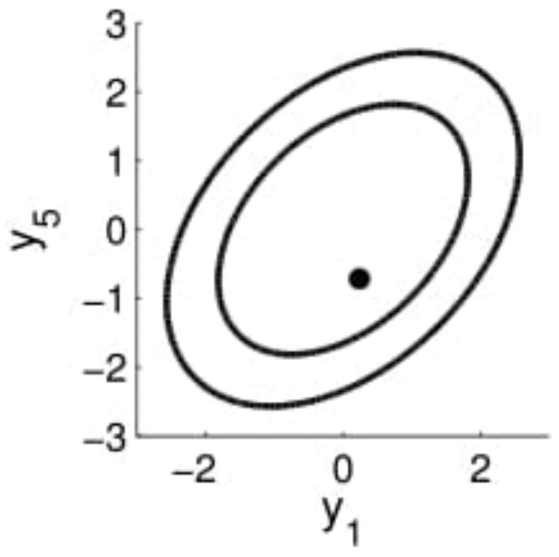
https://thegradient.pub/gaussian-process-not-quite-for-dummies/

Gaussian Process: Distribution over functions

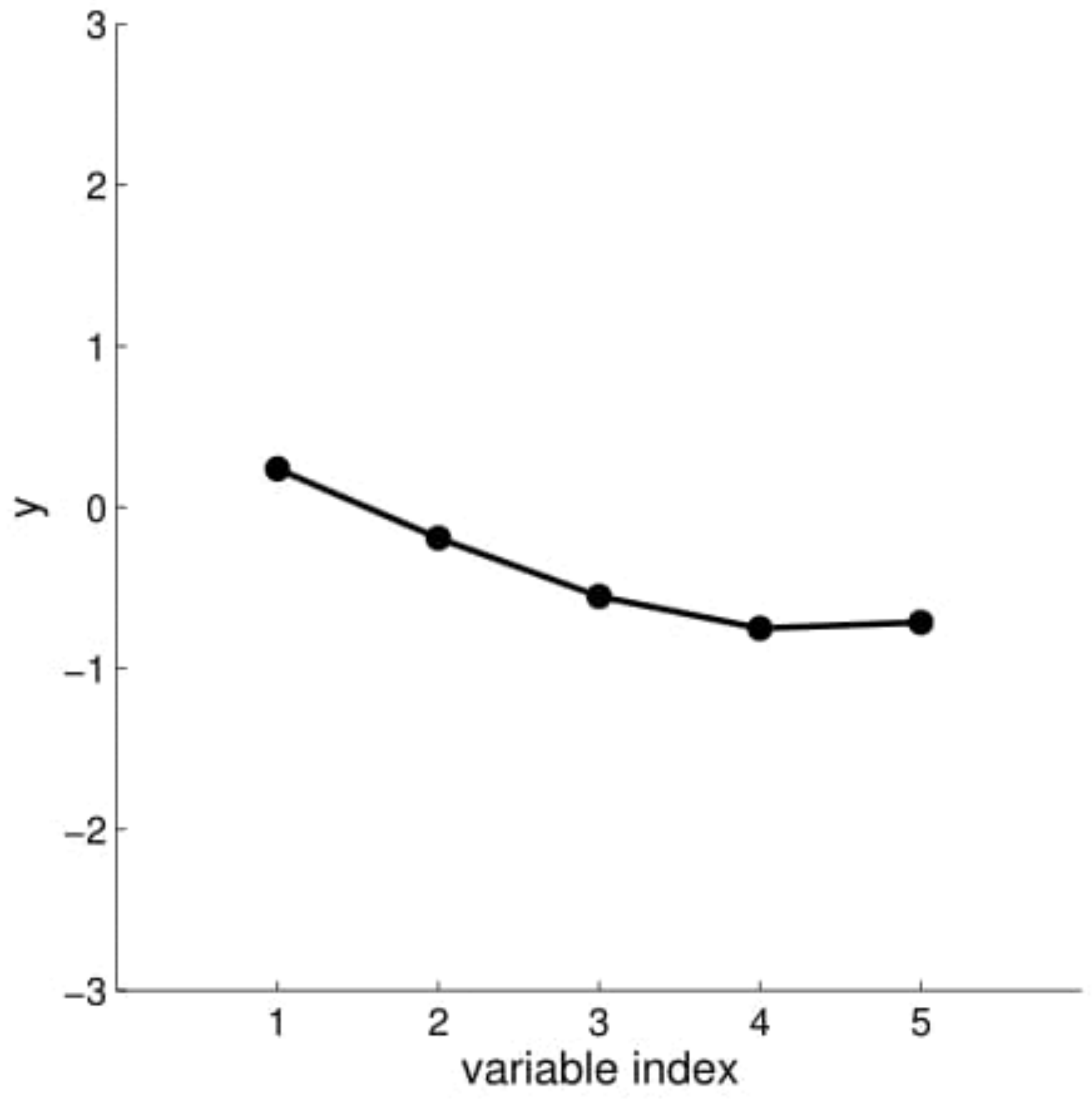
Formal definition: A Gaussian process is a collection of random variables, any finite number of which also have a joint Gaussian distribution.

Consider a multivariate Gaussian distribution: $p(\mathbf{y} \mid \Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}\right)$

5D Gaussian distribution — A different representation



$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



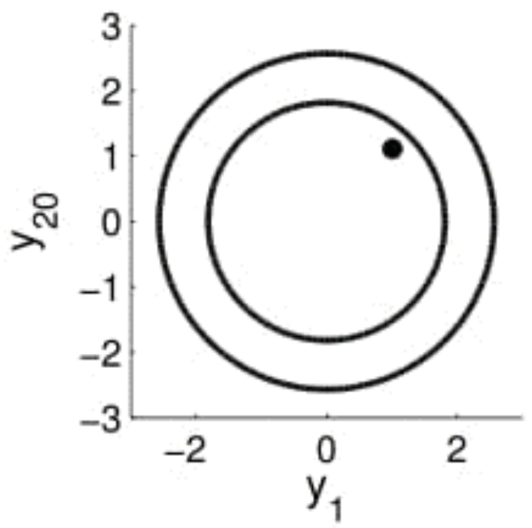
https://thegradient.pub/gaussian-process-not-quite-for-dummies/

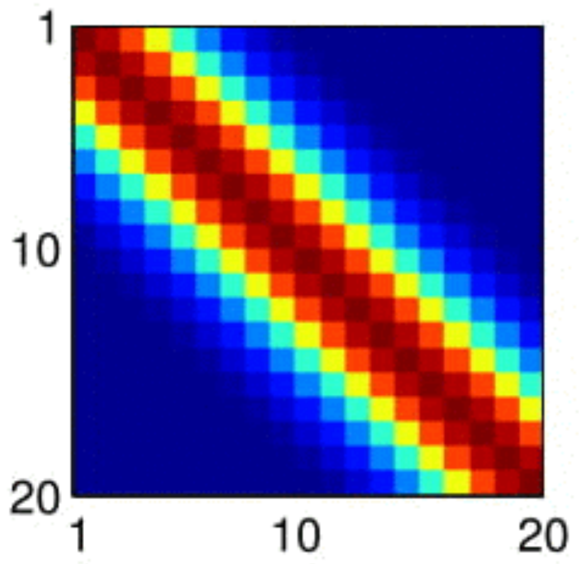
Gaussian Process: Distribution over functions

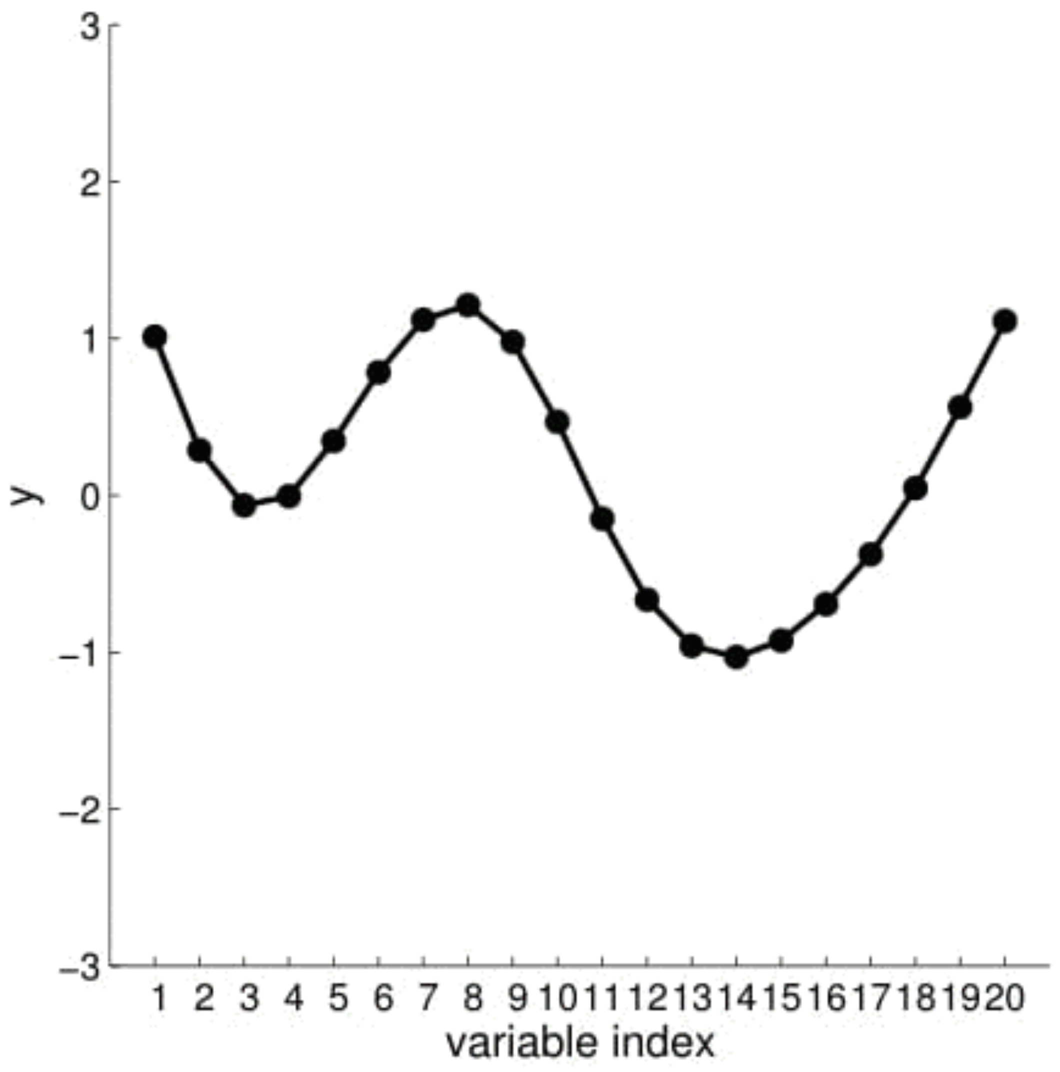
Formal definition: A Gaussian process is a collection of random variables, any finite number of which also have a joint Gaussian distribution.

Consider a multivariate Gaussian distribution: $p(\mathbf{y} \mid \Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}\right)$

20D Gaussian distribution — A different representation



$\Sigma =$ 



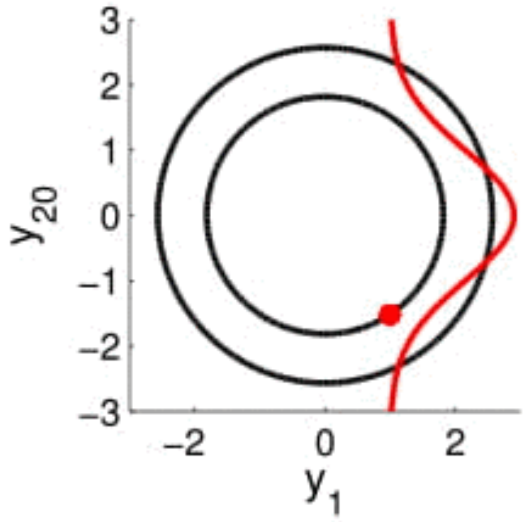
https://thegradient.pub/gaussian-process-not-quite-for-dummies/

Gaussian Process: Distribution over functions

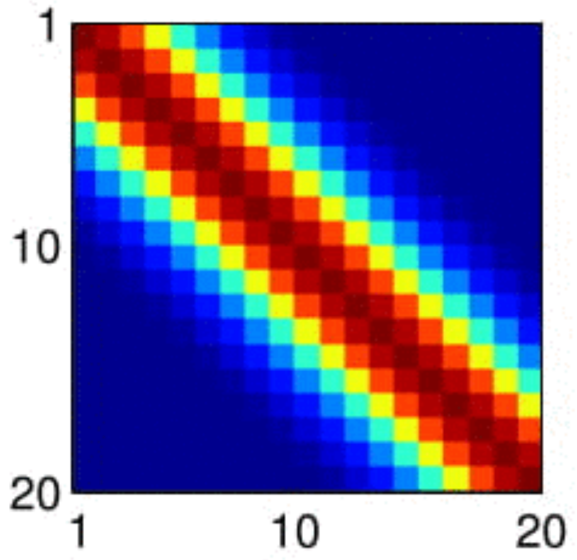
Formal definition: A Gaussian process is a collection of random variables, any finite number of which also have a joint Gaussian distribution.

Consider a multivariate Gaussian distribution: $p(\mathbf{y} \mid \Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}\right)$

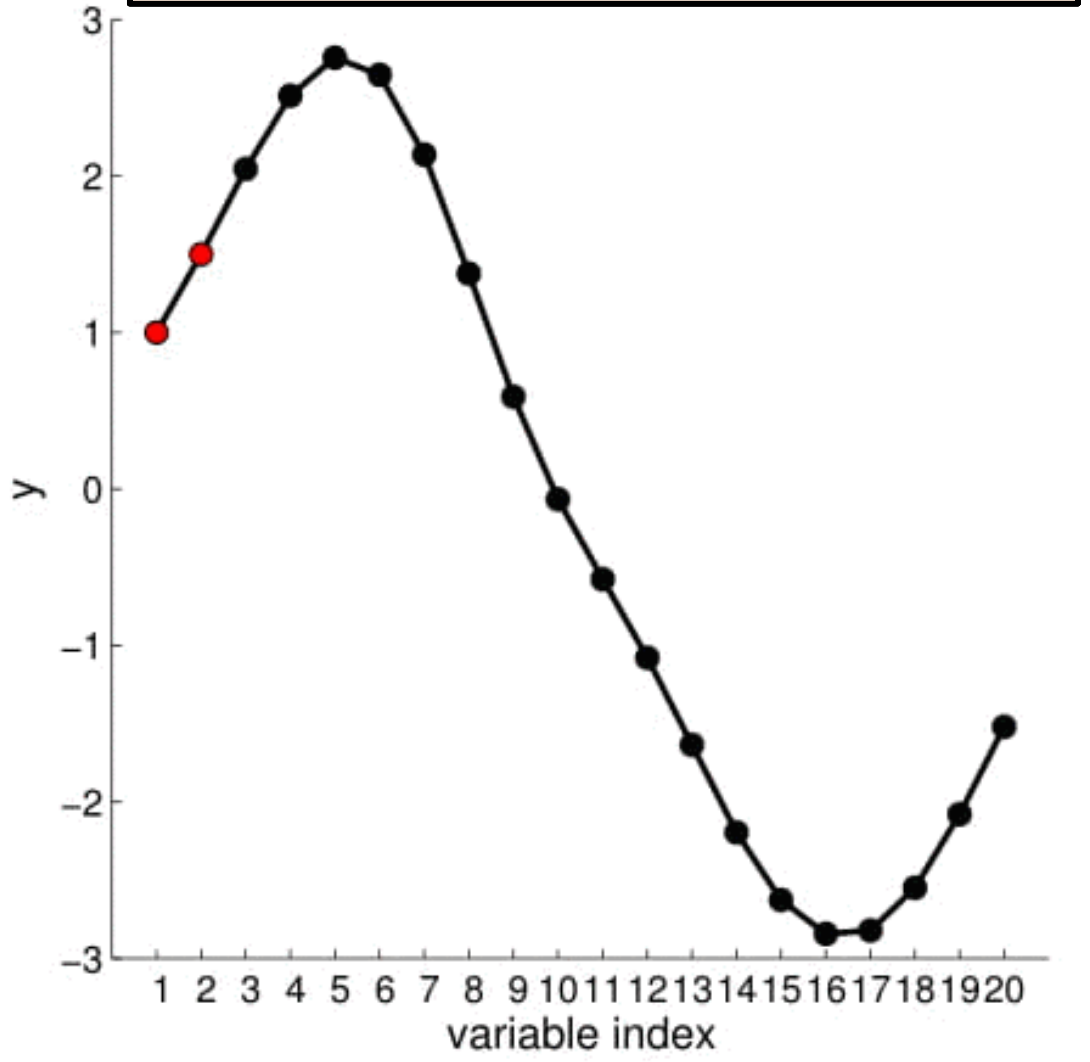
20D Gaussian distribution — A different representation



$\Sigma =$



Conditioned on (fixing) $y_1,\ y_2$



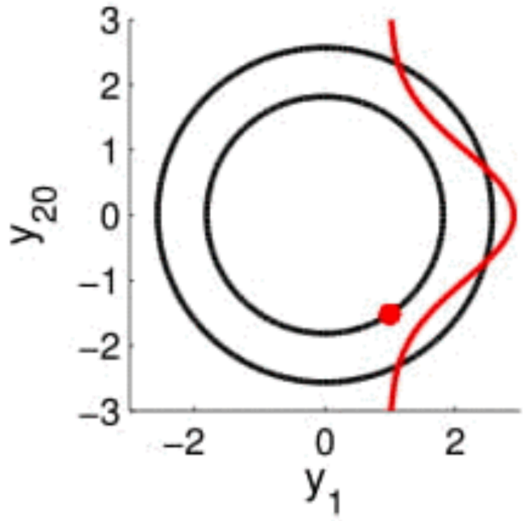
https://thegradient.pub/gaussian-process-not-quite-for-dummies/

Gaussian Process: Distribution over functions

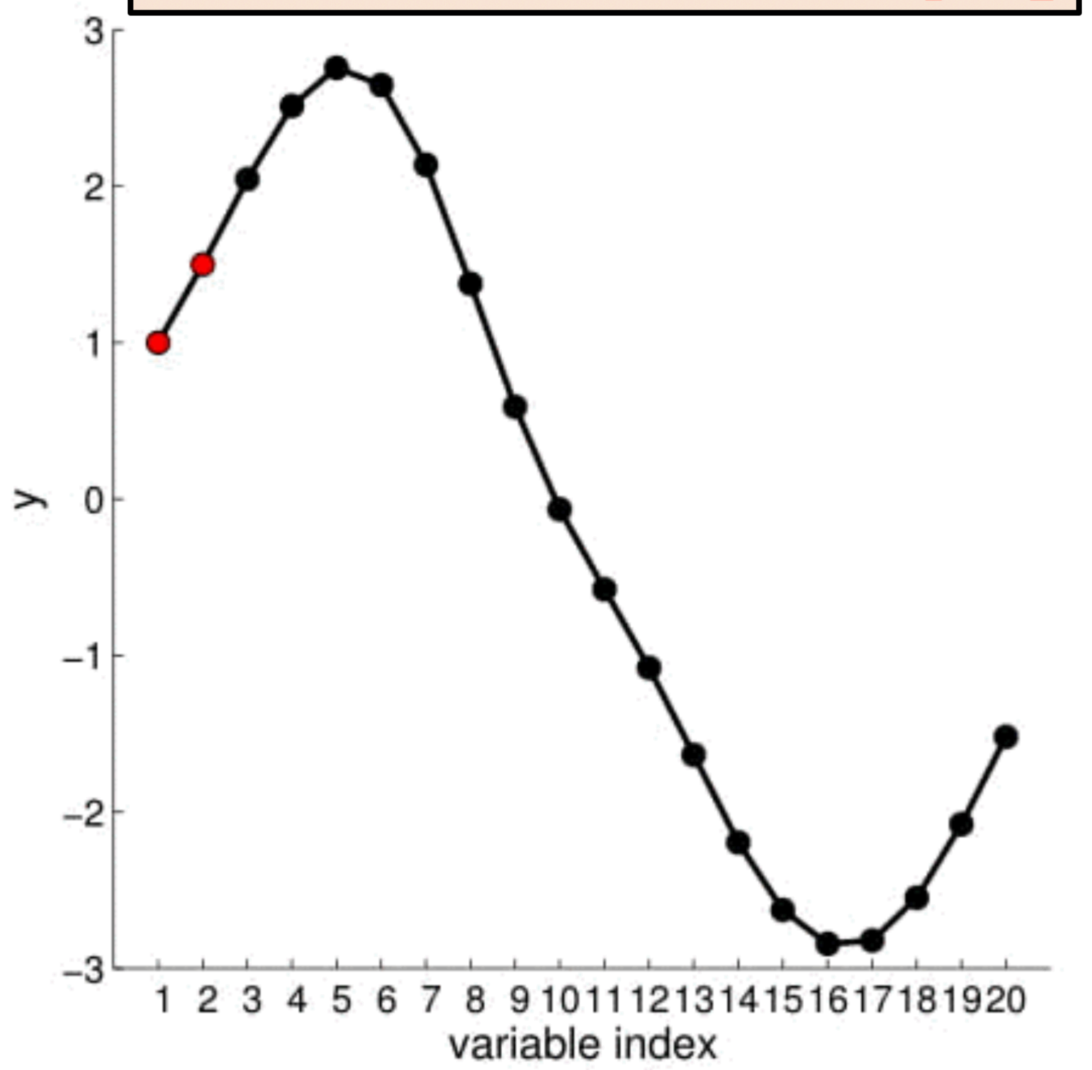
Formal definition: A Gaussian process is a collection of random variables, any finite number of which also have a joint Gaussian distribution.

Consider a multivariate Gaussian distribution: $p(\mathbf{y} \mid \Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}\right)$

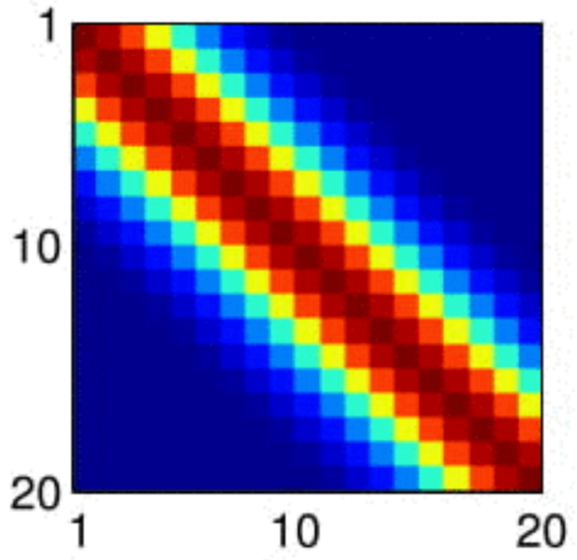
20D Gaussian distribution — A different representation



Conditioned on (fixing) $y_1,\ y_2$



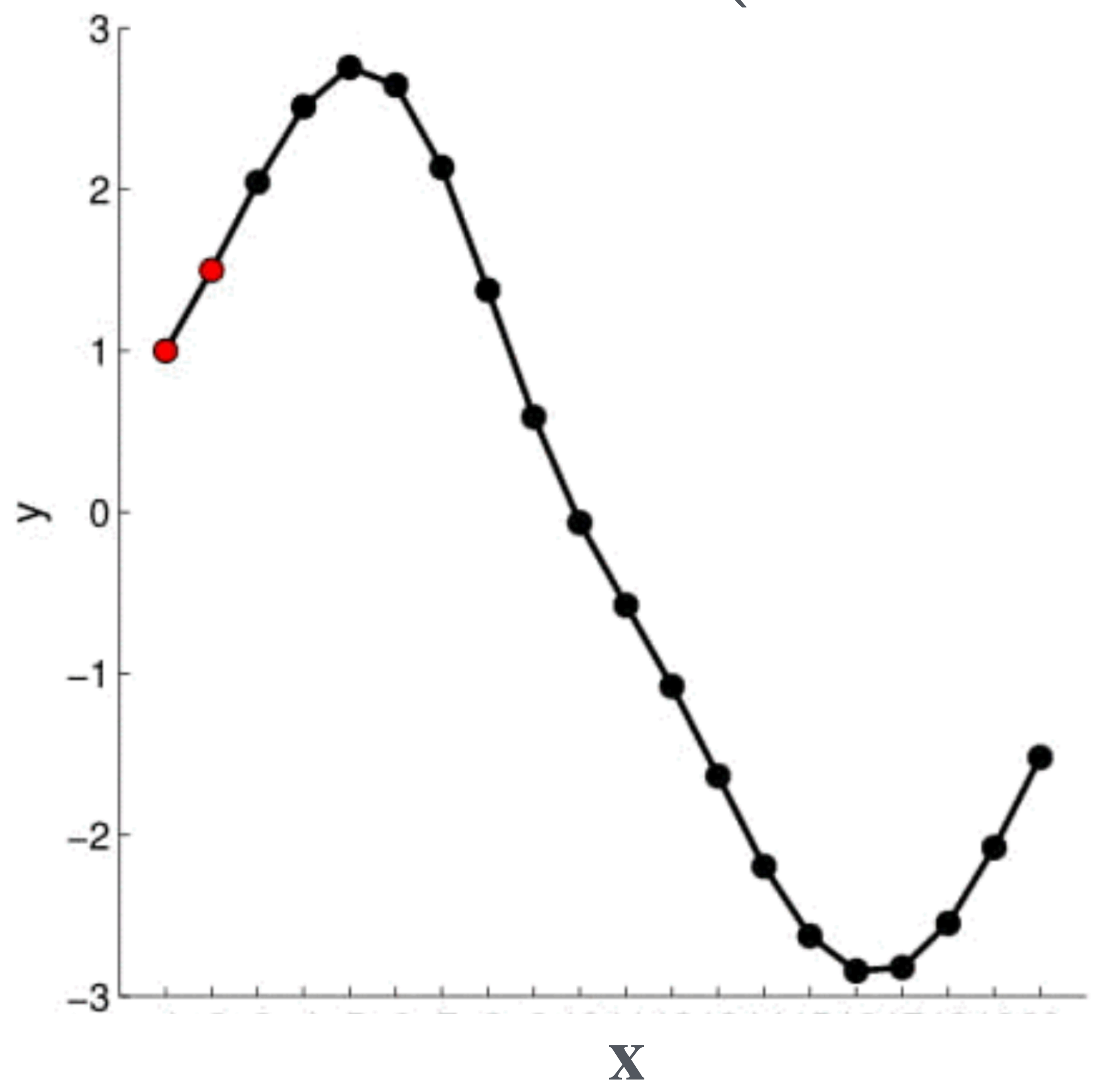
$\Sigma =$



Computing Σ from the **covariance kernel**

$$\Sigma \equiv K(x_i, x_j) = \bar{c}^2 \exp\left(-\frac{\|x_i - x_j\|^2}{2\ell^2}\right)$$

$\mathbf{x}$ is a vector of length n ($= 20$ here)



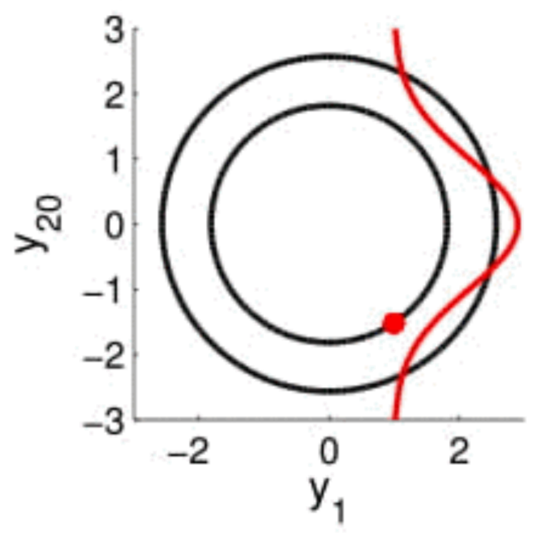
https://thegradient.pub/gaussian-process-not-quite-for-dummies/

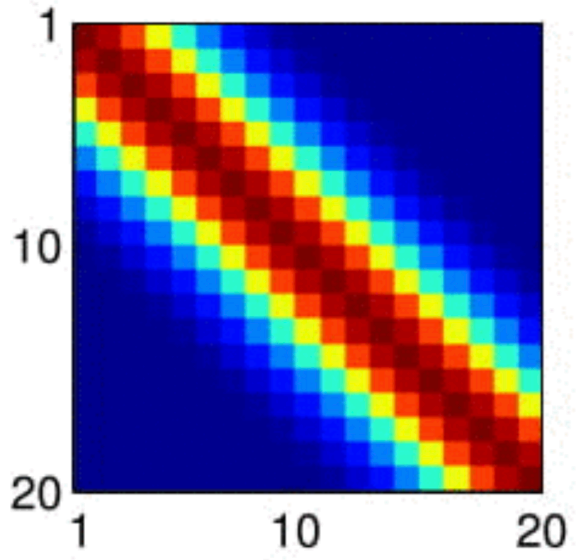
Gaussian Process: Distribution over functions

Formal definition: A Gaussian process is a collection of random variables, any finite number of which also have a joint Gaussian distribution.

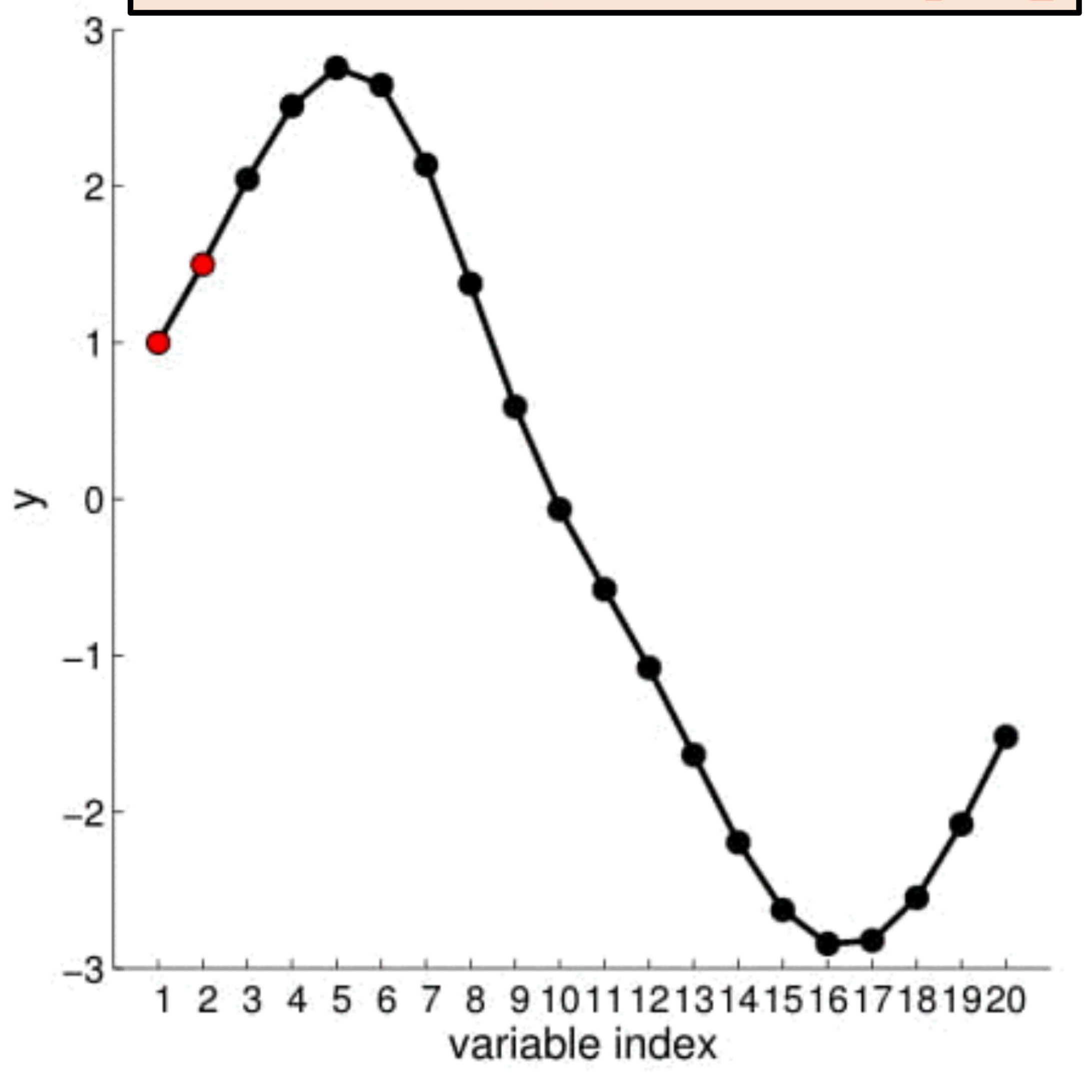
Consider a multivariate Gaussian distribution: $p(\mathbf{y} \mid \Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^T\Sigma^{-1}\mathbf{y}\right)$

20D Gaussian distribution — A different representation



$\Sigma =$ 

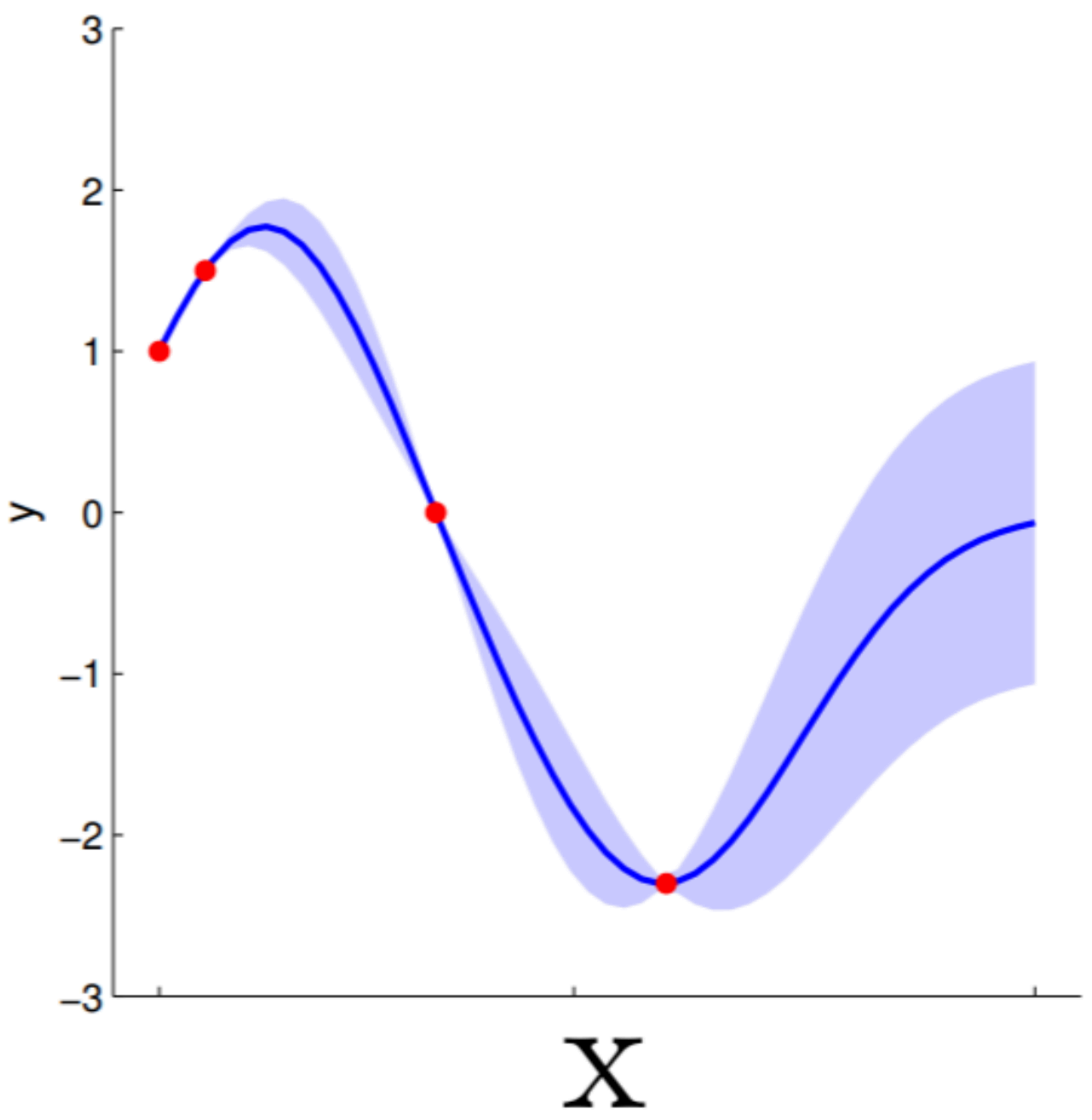
Conditioned on (fixing) $y_1,\ y_2$



Computing Σ from the **covariance kernel**

$$\Sigma \equiv K(x_i, x_j) = \bar{c}^2 \exp\left(-\frac{\|x_i - x_j\|^2}{2\ell^2}\right)$$

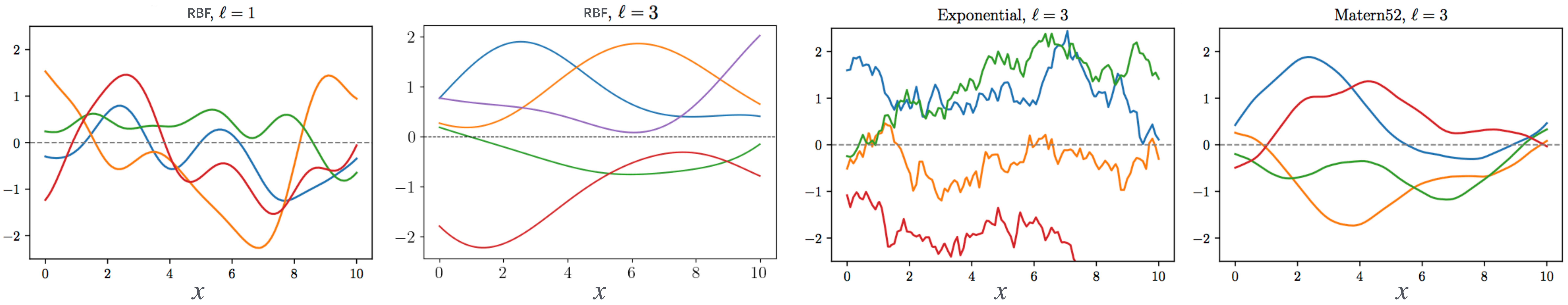
$\mathbf{x}$ is a vector of length n ($= 20$ here)



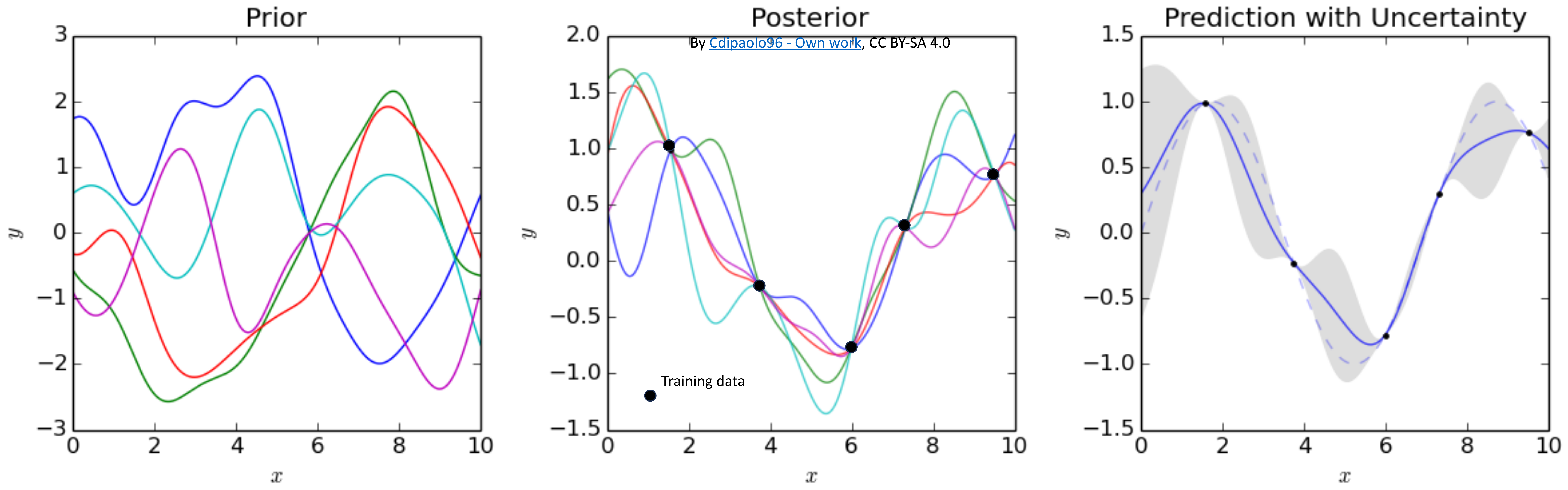
https://thegradient.pub/gaussian-process-not-quite-for-dummies/

Takeaway: A Gaussian process is a probability distribution over functions that fit a set of points. The functions have special properties determined by the covariance kernel.

GP conditioned on data



Working of GP:
Before training (prior) —> Training on data (posterior): choose only functions passing through the data and learn hyper-parameters of covariance function —> Predict with uncertainty.



Choice of covariance kernel to model theory error

- We want to model the error of the theory: $\delta(x_i, \phi) \sim y(x_i) - \eta(x_i, \theta)$

- Prior knowledge of the theory domain: "*the theory is more reliable at small x than at large x*".
  That means, for theory with true parameters θ^*, the error should grow as x increases.
  Goal is to determine this true θ^* by comparing model predictions with different θ to data.

- We consider the covariance kernel which generates functions whose magnitude increase with x.
  This emphasizes θ^* in fitting. Conversely, θ for which the discrepancy between data and model predictions does not increase with x are assigned less weight. SJ, C. Shen, R. J. Furnstahl, U. Heinz, and M. T. Pratola, PLB 870, 139946 (2025), arXiv: 2504.13144

➡ Kernel I: $K(x_i, x_j \mid \phi) \equiv s^2 + \bar{c}^2 (x_i x_j)^r \exp\left(-\frac{\|x_i - x_j\|^2}{2\ell^2}\right).$

Encodes the belief:
model is more reliable at small x than at large x.

➡ Kernel II: $K(x_i, x_j \mid \phi) \equiv \bar{c}^2 (x_i x_j)^r \exp\left(-\frac{\|x_i - x_j\|^2}{2\ell^2}\right).$

Encodes the belief:
model is correct at $x = 0$, but our confidence in its validity decreases as x increases