THE RATING QUALITY FOR THEORETICAL DESCRIPTION OF EXPERIMENTAL DATA

S. O. Omelchenko, V. M. Pugatch

Institute for Nuclear Research, National Academy of Sciences of Ukraine, Kyiv, Ukraine

1. Introduction. A new multi-parameter score-rating methodology for assessing the quality of theoretical description of experimental data in heavy-ion physics is proposed. The approach overcomes the limitations of the traditional single-value criterion (χ^2/ndf), which provides only an integral measure of agreement between theory and experiment. The methodology is based on dividing the phase space into seven physically motivated kinematic regions of transverse momentum p_T distributions and particle ratios, corresponding to different underlying physical regimes. For each region, the quality of agreement is quantified by a score Qi $\in$[10;1000] defined on a logarithmic scale, ranging from very poor to excellent agreement. A comprehensive rating R is constructed through a systematic procedure that includes region definition, weighting according to physical significance, aggregation of local scores, uncertainty estimation, stability checks, and visualization. This framework enables a transparent and comparative assessment of theoretical models, revealing their region-specific performance and complementarity.

2. Main Idea of the Methodology. The phase space is divided into N=7 kinematic areas Fi (i=1…7) based on physical regimes: thermal spectrum (zone 1, p_T < 0.8 GeV/c), radial flow (zones 2-3), hard processes critical for QGP (zone 4, 2.5-4.0 GeV/c), medium-energy jets (zone 5), high-energy jets with quenching (zone 6), and perturbative QCD regime (zone 7, p_T > 10 GeV/c). For each area, local statistics Ri = χ^2i / νi is calculated, where νi = Ni - k (Ni is the number of data points, k is model parameters). Based on Ri value, a score Qi is assigned according to a logarithmic scale: Ri ≤ 1.00 → Qi = 1000 (perfect), 1.00 < Ri ≤ 1.25 → 900, up to Ri > 10.0 → Qi = 10 (very poor) (see Fig.1 and Table 1 for π+ mesons).

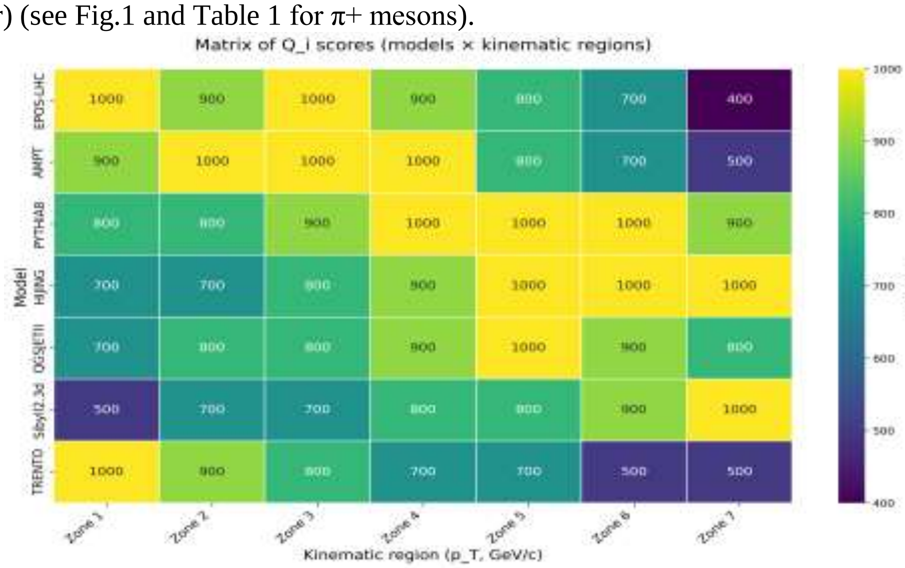

Fig. 1. Heatmap of Qi for models vs kinematic areas. Heatmap matrix illustrating Qi scores for seven theoretical models (EPOS-LHC, AMPT, PYTHIA8, HIJING, QGSJETII, Sibyll2.3d, TRENTO) across seven p_T zones. Viridis color scale: yellow indicates high consistency (Qi ≥ 900), purple indicates low (Qi ≤ 500), revealing model complementarity.

3. Weight Coefficients and Aggregation. Weight coefficients wi reflect physical significance and data quality: Category A (wi=2.5) for critical zones 2.5-4.0 GeV/c (QGP), Category B (wi=1.5) for main zones, Category C (wi=1.0) for auxiliary, Category D (wi=0.5) for peripheral zones. Sum of weights equals 10 for normalization. Aggregated metrics include: $Q_{weighted} = \Sigma(wi \times Qi) / \Sigma wi$ (weighted average), $Q_{geometric} = (\prod Qi)$^(1/7) (penalizes non-uniformity), $Q_{minimum}$ = min(Q1,...,Q7) (identifies weaknesses). Complex rating: $R = 0.45 \cdot Q_{weighted} + 0.30 \cdot Q_{geometric} + 0.20 \cdot Q_{minimum} - 0.05 \cdot \sigma$, where σ is dispersion penalty.

Table 1. Rating scores Qi and complex rating R for models.

Model	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Qw	R
EPOS-LHC	1000	900	1000	900	800	700	400	855	708
AMPT	900	1000	1000	1000	800	700	500	900	725
PYTHIA8	800	800	900	1000	1000	1000	900	925	791
HIJING	700	700	800	900	1000	1000	1000	885	755
QGSJETII	700	800	800	900	1000	900	800	855	730
Sibyll2.3d	500	700	700	800	800	900	1000	765	650
TRENTO	1000	900	800	700	700	500	500	755	690

4. Results and Applications. The methodology was applied to LHCb data for K_s^0 and Λ hyperons in *p-Pb* collisions at 5.02 TeV. For mesons (K_s^0), PYTHIA8 achieves highest rating (R = 830) due to realistic nuclear PDFs describing forward/backward asymmetry. For baryons (Λ), PHSD model (Additionally, PHSD (Parton-Hadron-String Dynamics) [11]) ranks best (R = 875) through explicit coalescence dynamics capturing baryon anomaly at intermediate p_T (2-5 GeV/c). Analysis reveals: (1) no universal model exists – complementarity is essential; (2) forward/backward asymmetry provides critical test of nuclear shadowing; (3) Λ/ K_s^0 ratio discriminates hadronization mechanisms (fragmentation vs coalescence).

5. Conclusions. The proposed multi-parameter score-rating methodology overcomes single χ^2/ndf limitations by: (1) preserving statistical rigor while revealing zone-specific performance; (2) enabling transparent ranking of dozens of models; (3) identifying complementarity to guide hybrid model development; (4) providing stability checks and uncertainty estimates. Methodology is ready for practical application to LHC Run 3 and beyond data and can accelerate progress in theoretical description of quark-gluon plasma.

References

[1] ALICE Collaboration, Phys.Lett.B 846 (2024) 137875.
[2] U.Heinz, R.Snellings, Ann.Rev.Nucl.Part.Sci. 63 (2013) 123.
[3] J.E. Bernhard et al., Phys.Rev.C 103 (2021) 054904.
[4] K.Werner et al., Phys.Rev.C 85 (2013) 064907.
[5] Z.W.Lin et al., Phys.Rev.C 72 (2005) 064901.
[6] JET Collaboration, J.Phys.G 47 (2020) 065101.
[7] CMS Collaboration, JHEP 04 (2017) 039.
[8] T.Sjöstrand et al., Comput.Phys.Commun. 191 (2015) 159.
[9] X.N.Wang, M.Gyulassy, Phys.Rev.D 44 (1991) 3501.
[10] J.S.Moreland et al., Phys.Rev.C 92 (2015) 011901.
[11] W.Cassing, E.L.Bratkovskaya, Nucl.Phys.A 831 (2009) 215.