Statistical error analysis for phenomenological nucleon-nucleon potentials

R. Navarro Pérez,* J. E. Amaro,† and E. Ruiz Arriola‡
Departamento de Física Atómica, Molecular y Nuclear and Instituto Carlos I de Física Teórica y Computacional Universidad de Granada, E-18071 Granada, Spain
(Received 2 April 2014; published 23 June 2014)

Nucleon-nucleon potentials are common in nuclear physics and are determined from a finite number of experimental data with limited precision sampling the scattering process. We study the statistical assumptions implicit in the standard least-squares χ^2 fitting procedure and apply, along with more conventional tests, a tail-sensitive quantile-quantile test as a simple and confident tool to verify the normality of residuals. We show that the fulfillment of normality tests is linked to a judicious and consistent selection of a nucleon-nucleon database. These considerations prove crucial to a proper statistical error analysis and uncertainty propagation. We illustrate these issues by analyzing about 8000 proton-proton and neutron-proton scattering published data. This enables the construction of potentials meeting all statistical requirements necessary for statistical uncertainty estimates in nuclear structure calculations.

DOI: 10.1103/PhysRevC.89.064006 PACS number(s): 21.30.Fe, 02.50.Tt, 03.65.Nk, 25.40.−h

I. INTRODUCTION

Nucleon-nucleon potentials are the starting point for many nuclear physics applications [1]. Most of the current information is obtained from *np*- and *pp*-scattering data below the pion production threshold and deuteron properties for which abundant experimental data exist. The *NN* scattering amplitude reads

$$M = a + m(\sigma_1 \cdot \mathbf{n})(\sigma_2 \cdot \mathbf{n}) + (g - h)(\sigma_1 \cdot \mathbf{m})(\sigma_2 \cdot \mathbf{m}) + (g + h)(\sigma_1 \cdot \mathbf{l})(\sigma_2 \cdot \mathbf{l}) + c(\sigma_1 + \sigma_2) \cdot \mathbf{n}, \quad (1)$$

where a, m, g, h, and c depend on energy and angle; σ_1 and σ_2 are the single-nucleon Pauli matrices; $\mathbf{l}$, $\mathbf{m}$, and $\mathbf{n}$ are three unitary orthogonal vectors along the directions of $\mathbf{k}_f + \mathbf{k}_i$, $\mathbf{k}_f - \mathbf{k}_i$, and $\mathbf{k}_i \wedge \mathbf{k}_f$, respectively; and $(\mathbf{k}_f, \mathbf{k}_i)$ are the final and initial relative nucleon momenta. From these five complex energy- and angle-dependent quantities 24 measurable cross sections and polarization asymmetries can be deduced [2]. Conversely, a complete set of experiments can be designed to reconstruct the amplitude at a given energy [3]. The finite amount, precision, and limited energy range of the data, as well as the many different observables, call for a standard statistical χ^2-fit analysis [4,5]. This approach is subjected to assumptions and applicability conditions that can only be checked *a posteriori* in order to guarantee the self-consistency of the analysis. Indeed, scattering experiments deal with counting Poissonian statistics and for a moderately large number of counts a normal distribution is expected. Thus, one hopes that a satisfactory theoretical description $O_i^{\rm th}$ can predict a set of N-independent observed data O_i given an experimental uncertainty ΔO_i as

$$O_i = O_i^{\rm th} + \xi_i \Delta O_i \quad (2)$$

with $i = 1, \ldots, N$ and where ξ_i are independent random *normal* variables with vanishing mean value $\langle \xi_i \rangle = 0$ and unit variance $\langle \xi_i \xi_j \rangle = \delta_{ij}$, implying that $\langle O_i \rangle = O_i^{\rm th}$. Establishing the validity of Eq. (2) is of utmost importance since it provides a basis for the statistical interpretation of the error analysis. In this work we will study to what extent this normality assumption underlying the validity of the full χ^2 approach is justified. This will be done by looking at the statistical distribution of the fit residuals of about 8000 *np* and *pp* published scattering data collected since 1950. Using the normality test as a necessary requirement, we show that it is possible to fulfill Eq. (2) with a high confidence level and high statistics. Furthermore, we discuss the consequences and requirements regarding the evaluation, design, and statistical uncertainties of phenomenological nuclear forces. We illustrate our points by determining for the first time a smooth nuclear potential with *error bands* directly inferred from experiment. We hope that these estimates will be useful for *NN* potential users interested in quantifying a definite source of error in nuclear structure calculations.[1]

The history of χ^2 statistical analyzes of *NN*-scattering data around pion production threshold started in the mid-1950s [7] (an account up to 1966 can be traced from Ref. [8]). A modified χ^2 method was introduced [9] in order to include data without absolute normalization. The steady increase along the years in the number of scattering data and their precision generated mutually incompatible data and hence a rejection criterion was introduced [10–12], allowing us to discard inconsistent data. Upgrading an ever-increasing consistent database poses the question of normality, Eq. (2), of a large number of selected data. The normality of the absolute value of residuals in *pp* scattering was scrutinized and satisfactorily fulfilled [13,14] as a necessary consistency condition. The Nijmegen group made an important breakthrough 20 years ago by performing the

*rnavarrop@ugr.es
†amaro@ugr.es
‡earriola@ugr.es

[1]We note that in a Physical Review A editorial [6] the importance of including error estimates in papers involving theoretical evaluations has been stressed.

very first partial-wave analysis (PWA) fit with $\chi^2/\mathrm{DOF} \sim 1$ and applying a 3σ-rejection criterion. This was possible after including em corrections, vacuum polarization, magnetic-moment interactions, and a charge-dependent (CD) one pion exchange (OPE) potential. With this fixed database, further high-quality potentials have been steadily generated [15–18] and applied to nuclear structure calculations. However, high-quality potentials, i.e., those whose discrepancies with the data are confidently attributable to statistical fluctuations in the experimental data, have been built and used as if they were errorless. As a natural consequence, the computational accuracy to a relative small percentage level has been a goal *per se* in the solution of the few and many body problem regardless on the absolute accuracy implied by the input of the calculation. While this sets high standards on the numerical methods there is no *a priori* reason to assume the computational accuracy reflects the realistic physical accuracy and, in fact, it would be extremely useful to determine and identify the main source of uncertainty; one could thus tune the remaining uncertainties to this less demanding level.

It should be noted that the χ^2 fitting procedure, when applied to limited upper energies, fixes most efficiently the long-range piece of the potential which is known to be mainly described by OPE for distances $r \gtrsim 3\,\mathrm{fm}$. However, weaker constraints are put in the midrange $r \sim 1.5$–2.5 fm region, which is precisely the relevant interparticle distance operating in the nuclear binding. To date and to the best of our knowledge, the estimation of errors in the nuclear force stemming from the experimental scattering data uncertainties and its consequences for nuclear structure calculations has not been seriously confronted. With this goal in mind we have upgraded the *NN* database to include all published *np*- and *pp*-scattering data in the period 1950–2013, determining in passing the error in the interaction [19,20].

The present paper represents an effort towards filling this gap by providing statistical error bands in the *NN* interaction based directly on the experimental data uncertainties. In order to do so, the specific form of the potential needs to be fixed.[2] As such, this choice represents a certain bias and hence corresponds to a likely source of systematic error. Based on the previous high-quality fits which achieved $\chi^2/\nu \lesssim 1$ [15–18] we have recently raised suspicions on the dominance of such errors with intriguing consequences for the quantitative predictive power of nuclear theory [22–24]; a rough estimate suggested that *NN* uncertainties propagate into an unpleasantly large uncertainty of $\Delta B/A \sim 0.1$–0.5 MeV, a figure which has not yet been disputed by an alternative estimate. In view of this surprising finding, there is a pressing need to pin down the input uncertainties more accurately based on a variety of sources.[3] This work faces the evaluation of statistical errors after checking that the necessary normality conditions of residuals and hence Eq. (2) are confidently fulfilled. From this point of view, the present investigation represents an initial step, postponing a more complete discussion on systematic uncertainties for a future investigation.

The PWA analysis carried out previously by us [22–24] was computationally inexpensive due to the use of the simplified δ-shell potential suggested many years ago by Avilés [29]. This form of potential effectively coarse grains the interaction and drastically reduces the number of integration points in the numerical solution of the Schrödinger equation (see, e.g., Ref. [30]). However, it is not directly applicable to some of the many numerical methods available on the market to solve the few and many body problem where a smooth potential is required. Therefore, we will analyze the 3σ self-consistent database in terms of a more conventional potential form containing the same 21 operators extending the AV18 as we did in Refs. [22–24]. Testing for normality of residuals within a given confidence level for a *phenomenological potential* is an issue of direct significance to any statistical error analysis and propagation. Actually, we will show that for the fitted observables to the 3σ self-consistent experimental database O_i^{exp}, with quoted uncertainty $\Delta O_i^{\mathrm{exp}}$, $i = 1, \ldots, N = 6713$ (total number of *pp* and *np* scattering data), our theoretical fits indeed satisfy that the residuals

$$R_i = \frac{O_i^{\mathrm{exp}} - O_i^{\mathrm{th}}}{\Delta O_i^{\mathrm{exp}}} \tag{3}$$

follow a normal distribution within a large confidence level. In order to establish this we will use a variety of classical statistical tests [4,5], such as the Pearson, Kolmogorov-Smirnov (KS), the moments method (MM), and, most importantly, the recently proposed tail-sensitive (TS) quantile-quantile test with confidence bands [31]. By comparing with others, the TS test turns out to be the most demanding with regard to the confidence bands. Surprisingly, normality tests have only seldom been applied within the present context, so our presentation is intended to be at a comprehensive level. A notable exception is given in Refs. [13,14] where the moments method in a *pp* analysis up to $T_{\mathrm{LAB}} = 30$ and 350 MeV is used for $N = 389$ and 1787 data, respectively, to test that the squared residuals R_i^2 in Eq. (3) follow a χ^2 distribution with one degree of freedom. Note that this is insensitive to the sign of R_i and thus blind to asymmetries in a normal distribution. Here we test normality of R_i for a total of $N = 6713$ *np* and *pp* data up to $T_{\mathrm{LAB}} = 350$ MeV.

The paper is organized as follows. In Sec. II we review the assumptions and the rejecting and fitting processes used in our previous works to build the 3σ self-consistent database and expose the main motivation to carry out a normality

[2]This is also the case in the quantum mechanical inverse scattering problem, which has only unique solutions for specific assumptions on the form of the potential [21] and with the additional requirement that some interpolation of scattering data at nonmeasured energies is needed. One needs then the information on the bound state energies and their residues in the scattering amplitude. We will likewise impose that the only bound state is the deuteron and reject fits with spurious bound states.

[3]There is a growing concern on the theoretical determination of nuclear masses from nuclear mean-field models with uncertainty evaluation [25] (for a comprehensive discussion see, e.g., Refs. [26,27]), echoing the need for uncertainty estimates in a Physical Review A editorial [6] and the Saltelli-Funtowicz seven rules checklist [28].

test of the fit residuals. In Sec. III we review some of the classical normality tests and a recently proposed tail-sensitive test, which we apply comparatively to the complete as well as the 3σ self-consistent database, providing a *raison d'être* for the rejection procedure. After that, in Sec. IV we analyze a fit of a potential whose short-distance contribution is constructed by a sum of Gaussian functions, with particular attention to the error bar estimation, a viable task since the residuals pass satisfactorily the normality test. Finally, in Sec. V, we come to our conclusions and provide an outlook for further work.

II. STATISTICAL CONSIDERATIONS

There is a plethora of references on data and error analysis (see, e.g., Refs. [4,5]). We will review the fitting approach in such a way that our points can be more easily stated for the general reader.

A. Data uncertainties

Scattering experiments are based on counting Poissonian statistics and, for a moderately large number of counts, a normal distribution sets in. In what follows O_i will represent some scattering observable. For a set of N-independent measurements of different scattering observables $O_i^{\rm exp}$ experimentalists quote an *estimate* of the uncertainty $\Delta O_i^{\rm exp}$ so the *true* value $O_i^{\rm true}$ is contained in the interval $O_i^{\rm exp} \pm \Delta O_i^{\rm exp}$ with a 68% confidence level. In what follows we assume for simplicity that there are no sources of systematic errors. Actually, when only the pair $(O_i^{\rm exp}, \Delta O_i^{\rm exp})$ is provided without specifying the distribution, we will assume an underlying normal distribution,[4] so

$$P\left(O_i^{\rm exp}\right) = \frac{\exp\left[-\frac{1}{2}\left(\frac{O_i^{\rm true}-O_i^{\rm exp}}{\Delta O^{\rm exp}}\right)^2\right]}{\sqrt{2\pi}\,\Delta O_i^{\rm exp}} \tag{4}$$

is the probability density of finding measurement $O_i^{\rm exp}$.

B. Data modeling

The problem of data modeling is to find a theoretical description characterized by some parameters $F_i(\lambda_1,\ldots,\lambda_P)$ which contain the true value $O_i^{\rm true} = F_i(\lambda_1^{\rm true},\ldots,\lambda_P^{\rm true})$ with a given confidence level characterized by a bounded p-dimensional manifold in the space of parameters $(\lambda_1,\ldots,\lambda_P)$. For a normal distribution the probability of finding any of the (*independent*) measurements $O_i^{\rm exp}$, assuming that $(\lambda_1,\ldots,\lambda_P)$ are the true parameters, is given by

$$P\left(O_i^{\rm exp}|\lambda_1\ldots\lambda_P\right) = \frac{\exp\left[-\frac{1}{2}\left(\frac{F_i(\lambda_1,\ldots,\lambda_P)-O_i^{\rm exp}}{\Delta O^{\rm exp}}\right)^2\right]}{\sqrt{2\pi}\,\Delta O_i^{\rm exp}}. \tag{5}$$

Thus the joined probability density is

$$\begin{aligned} P\left(O_1^{\rm exp}\ldots O_N^{\rm exp}|\lambda_1\ldots\lambda_P\right) &= \prod_{i=1}^{N} P\left(O_i^{\rm exp}|\lambda_1\ldots\lambda_P\right) \\ &= C_N e^{-\chi^2(\lambda_1,\ldots,\lambda_P)/2}, \end{aligned} \tag{6}$$

where $1/C_N = \prod_{i=1}^{N}(\sqrt{2\pi}\,\Delta O_i^{\rm exp})$. In such a case the maximum likelihood method [4,5] corresponds to take the χ^2 as a figure of merit given by

$$\chi^2(\lambda_1,\ldots,\lambda_P) = \sum_{i=1}^{N}\left(\frac{O_i^{\rm exp} - F_i(\lambda_1,\ldots,\lambda_P)}{\Delta O_i^{\rm exp}}\right)^2 \tag{7}$$

and look for the minimum in the fitting parameters $(\lambda_1,\ldots,\lambda_P)$,

$$\chi^2_{\rm min} = \min_{\lambda_i}\chi^2(\lambda_1,\ldots,\lambda_P) = \chi^2(\lambda_{1,0},\ldots,\lambda_{P,0}). \tag{8}$$

Our theoretical estimate of $O_i^{\rm true}$ after the fit is given by

$$O_i^{\rm th} = F_i(\lambda_{1,0},\ldots,\lambda_{P,0}). \tag{9}$$

Expanding around the minimum one has

$$\chi^2 = \chi^2_{\rm min} + \sum_{ij=1}^{P}(\lambda_i-\lambda_{i,0})(\lambda_j-\lambda_{j,0})\mathcal{E}^{-1}_{ij} + \cdots, \tag{10}$$

where the $P\times P$ error matrix is defined as the inverse of the Hessian matrix evaluated at the minimum

$$\mathcal{E}^{-1}_{ij} = \frac{1}{2}\frac{\partial^2\chi^2}{\partial\lambda_i\partial\lambda_j}(\lambda_{1,0},\ldots,\lambda_{P,0}). \tag{11}$$

Finally, the correlation matrix between two fitting parameters λ_i and λ_j is given by

$$\mathcal{C}_{ij} = \frac{\mathcal{E}_{ij}}{\sqrt{\mathcal{E}_{ii}\mathcal{E}_{jj}}}. \tag{12}$$

We compute the error of the parameter λ_i as

$$\Delta\lambda_i \equiv \sqrt{\mathcal{E}_{ii}}. \tag{13}$$

Error propagation of an observable $G = G(\lambda_1,\ldots,\lambda_P)$ is computed as

$$(\Delta G)^2 = \sum_{ij}\frac{\partial G}{\partial\lambda_i}\frac{\partial G}{\partial\lambda_j}\bigg|_{\lambda_k=\lambda_{k,0}}\mathcal{E}_{ij}. \tag{14}$$

The resulting residuals of the fit are defined as

$$R_i = \frac{O_i^{\rm exp} - F_i(\lambda_{1,0},\ldots,\lambda_{P,0})}{\Delta O_i^{\rm exp}}, \quad i = 1,\ldots,N. \tag{15}$$

Assuming normality of residuals is now crucial for an statistical interpretation of the confidence level, since then $\sum_i R_i^2$ follows a χ^2 distribution. One useful application of the previous result is that we can replicate the experimental data by using Eq. (2) and in such a case $\langle\chi^2\rangle = N$. For a

[4]This may not be the most efficient unbiased estimator (see, e.g., Refs. [4,5] for a more thorough discussion). Quite generally, the theory for the noise on the specific measurement would involve many considerations on the different experimental setups. In our case the many different experiments makes such an approach unfeasible. There is a possibility that some isolated systematic errors in particular experiments are randomized when considered globally. However, the larger the set the more stringent the corresponding statistical normality test. From this point of view the verification of the normality assumption underlying Eq. (2) proves highly nontrivial.

large number of data N with P parameters one has, with a 1σ or 68% confidence level, the mean value and most likely the values nearly coincide, so one has $\langle \chi^2_{\rm min}\rangle = N - P$ and thus as a random variable we have

$$\frac{\chi^2_{\rm min}}{\nu} \equiv \frac{\sum_i \xi_i^2}{\nu} = 1 \pm \sqrt{\frac{2}{\nu}}, \tag{16}$$

where $\nu = N - P$ is the number of degrees of freedom. The goodness of fit is defined in terms of this confidence interval. However, the χ^2 test has a sign ambiguity for every single residual given that $R_i \to -R_i$ is a symmetry of the test. From this point of view, the verification of normality is a more demanding requirement.[5]

Thus a necessary condition for a least-squares fit with meaningful results is the residuals to follow a normal distribution with mean zero and variance 1, i.e., $R_i \sim N(0,1)$. It should be noted that a model for the noise need not be normal, but it must be a *known* distribution $P(z)$ such that the residuals R_i do indeed follow *such* distribution.[6]

C. Data selection

The first and most relevant problem one has to confront in the phenomenological approach to the nucleon-nucleon interaction is that the database is not consistent; there appear to be incompatible measurements. This may not necessarily mean genuinely wrong experiments but rather unrealistic error estimates or an incorrect interpretation of the quoted error as a purely statistical uncertainty.[7] Note that the main purpose of a fit is to estimate the true values of certain parameters with a given and admissible confidence level. Therefore one has to make a decision on which are the subset of data which will finally be used to determine the *NN* potential. However, once the choice has been made the requirement of having normal residuals, Eq. (3), must be checked if error estimates on the fitting parameters are truly based on a random distribution.

The situation we encounter in practice is of a large number of data, ∼8000 vs the small number of potential parameters ∼40, which are expected to successfully account for the description of the data [33]. Thus, naively there seems to be a large redundancy in the database. However, there is a crucial issue on what errors have been quoted by the experimentalists. If a conservative estimate of the error is made, there is a risk of making the experiment useless, from the point of view that any other experiment in a similar kinematical region will dominate the analysis.[8] If, on the contrary, errors are daringly too small, they may generate a large penalty as compared to the rest of the database. This viewpoint seems to favor more accurate measurements whenever they are compatible but less accurate ones when some measurements appear as incompatible with the rest. In addition, there may be an abundance bias, i.e., too many accurate measurements in some specific kinematical region and a lack of measurements in another regions. Thus, the working assumption in order to start any constructive analysis is that *most* data have *realistic* quoted errors and that those experiments with unrealistically too small or too large errors can be discerned from the bulk with appropriate statistical tools. This means that these unrealistic uncertainties can be used to reject the corresponding data.[9] If a consistent and maximal database is obtained by an iterative application of a rejection criterium, the discrepancy between theory and data has to obey a statistical distribution, see Eq. (2).

D. Data representation

For two given data with exactly the same kinematical conditions, i.e., same observable, scattering angle, and energy, the decision on whether they are compatible may be easily made by looking at nonoverlapping error bands.[10] This is frequently not the case; one has instead a set of neighboring data in the (θ, E) plane for a given observable or different observables at the same (θ, E) point. The situation is depicted in Fig. 1 (left panels) where every point represents a single pp or np measurement (for an illustrative plot on the situation by 1983 up to 1 GeV see Ref. [37]). The total number of 8124 fitting data includes 7709 experimental measurements and 415 normalizations provided by the experimentalists. Thus, the decision intertwines all available data and observables. As a consequence, the comparison requires a certain extrapolation, which is viable under a smoothness assumption of the energy dependence of the partial-wave-scattering amplitude. Fortunately, the meson exchange picture foresees a well-defined analytical branch cut structure in the complex energy plane which is determined solely from the long-distance properties of the interaction. A rather efficient way to incorporate this desirable features from the start is by using a quantum mechanical potential. More specifically, if one has $n\pi$ exchange, then at long distances $V(r) \sim e^{-nm_\pi r}$ guarantees the appearance of a left-hand branch cut at center-of-mass (c.m.) momentum

[5]One can easily see that for a set of normally distributed data R_i, while $|R_i|$ does not follow that distribution, $|R_i|^2 = R_i^2$ would pass a χ^2 test.

[6]In this case the merit figure to minimize would be

$$S(\lambda_1, \ldots, \lambda_P) = -\sum_i \log P\Big[\tfrac{O_i^{\rm exp} - F_i(\lambda_1,\ldots,\lambda_P)}{\Delta O_i^{\rm exp}}\Big].$$

For instance, in Ref. [32], dealing with πN scattering a Lorentz distribution arose as a self-consistent assumption.

[7]Indeed any measurement could become right provided a sufficiently large or conservative error is quoted.

[8]See, e.g., the recommendations of the *Guide to the Expression of Uncertainty in Measurement* by the BIPM [34] where (often generously) *conservative* error estimates are undesirable, while *realistic* error estimates are preferable. Of course, *optimal* error estimates could only arise when there is a competition between independent measurements and a bonus for accuracy.

[9]From this point of view, the small and the large errors are not symmetric; the small χ^2 (conservative errors) indicate that the fitting parameters are indifferent to these data, whereas the large χ^2 (daring errors) indicate an inconsistency with the rest of the data.

[10]For several measurements the Birge test [35] is the appropriate tool. The classical and Bayesian interpretation of this test has been discussed recently [36].

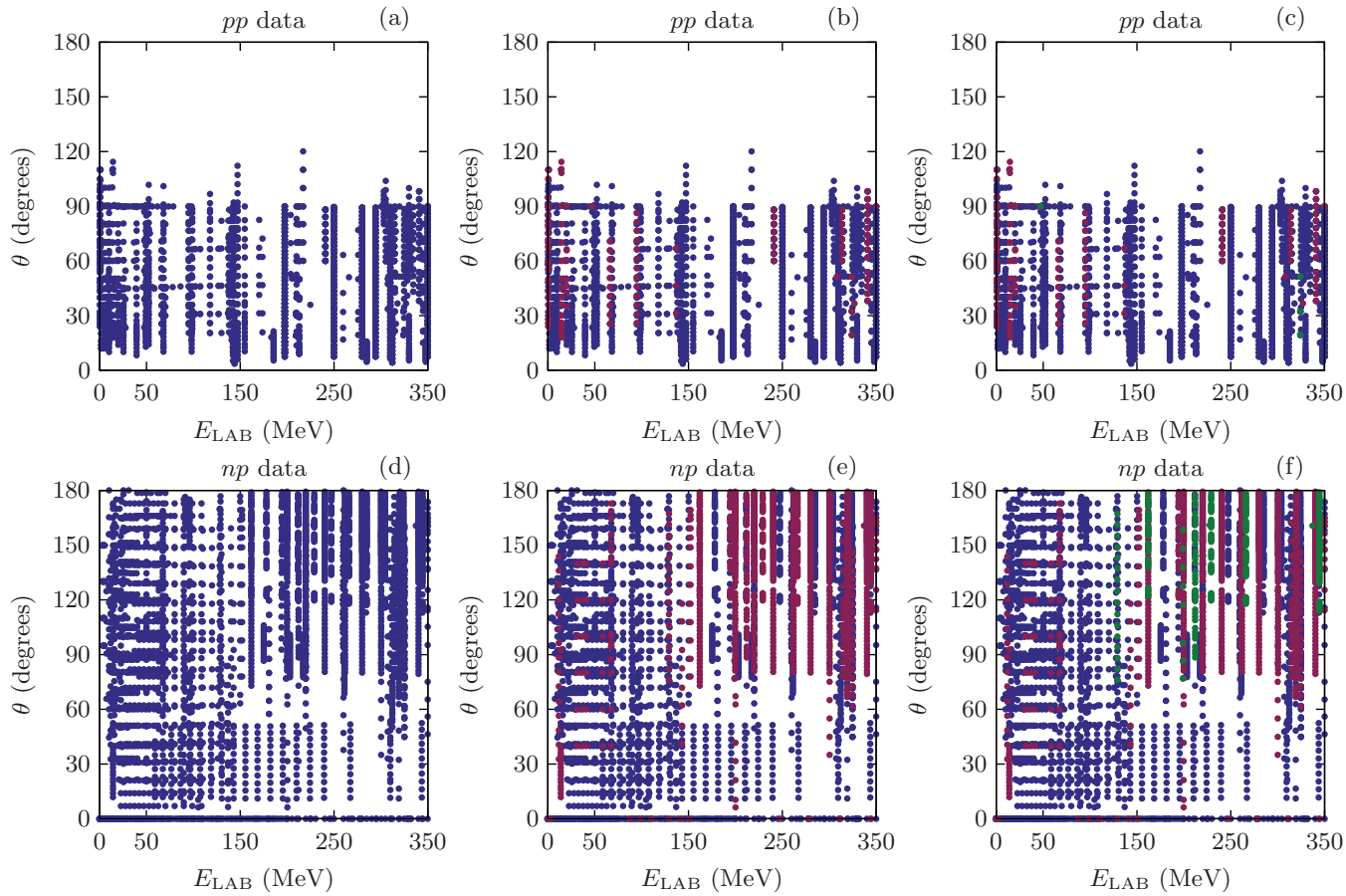

FIG. 1. (Color online) Abundance plots for pp- (top panel) and np- (bottom panel) scattering data. Full database (left panel). Standard 3σ criterion (middle panel). Self-consistent 3σ criterion (right panel). We show accepted data (blue), rejected data (red), and recovered data (green).

$p = im_\pi n/2$. Using this meson exchange picture at long distances the data world can be mapped onto a, hopefully complete, set of fitting parameters.

In order to analyze this in more detail we assume, as we did in Refs. [22–24], that the NN interaction interaction can be decomposed as

$$V(\vec{r}) = V_{\rm short}(r)\theta(r_c - r) + V_{\rm long}(r)\theta(r - r_c), \qquad (17)$$

where the short component can be written as

$$V_{\rm short}(\vec{r}) = \sum_{n=1}^{21} \hat{O}_n \left[\sum_{i=1}^{N} V_{i,n} F_{i,n}(r) \right], \qquad (18)$$

where $\hat{O}_n$ are the set of operators in the extended AV18 basis [16,22–24], $V_{i,n}$ are unknown coefficients to be determined from data, and $F_{i,n}(r)$ are some given radial functions. $V_{\rm long}(\vec{r})$ contains a CD OPE (with a common $f^2 = 0.075$ [22–24]) and electromagnetic (EM) corrections which are kept fixed throughout. This corresponds to

$$V_{\rm long}(\vec{r}) = V_{\rm OPE}(\vec{r}) + V_{\rm em}(\vec{r})\,. \qquad (19)$$

Although the form of the complete potential is expressed in the operator basis the statistical analysis is carried out more effectively in terms of some low and independent partial-waves contributions to the potential from which all other higher partial waves are consistently deduced (see Refs. [38,39]).

E. Fitting data

In our previous PWA we used the δ-shell interaction already proposed by Avilés [29] and which proved extremely convenient for fast minimization and error evaluation[11] and corresponds to the choice

$$F_{i,n}(r) = \Delta r_i \delta(r - r_i), \qquad (20)$$

where $r_i \leqslant r_c$ are a discrete set of radii and $\Delta r_i = r_{i+1} - r_i$. The minimal resolution $\Delta r_{\rm min}$ is determined by the shortest de Broglie wavelength corresponding to a pion production threshold which we estimate as $\Delta r_\pi \sim 0.6$ fm [30,33] so the needed number of parameters can be estimated *a priori*. Obviously, if $\Delta r_{\rm min} \ll \Delta r_\pi$, the number of parameters increases as well as the correlations among the different fitting coefficients, $V_{i,n}$, so some parameters become redundant or an overcomplete representation of the data, and the χ^2 value will not decrease substantially. In the opposite situation

[11] We use the Levenberg-Marquardt method where an approximation to the Hessian is computed explicitly [40] which we keep throughout.

TABLE I. Standardized moments μ_r' of the residuals obtained by fitting the complete database with the δ-shell potential and 3σ-consistent database with the OPE-δ shell, χTPE-δ shell, and OPE-Gaussian potentials. The expected values for a normal distributions are included $\pm 1\sigma$ confidence level of a Monte Carlo simulation with 5000 random samples of size N.

r	Complete database $N = 8125$		3σ OPE-δ shell $N = 6713$		3σ χTPE-δ shell $N = 6712$		3σ OPE-Gaussian $N = 6711$	
	Expected	Empirical	Expected	Empirical	Expected	Empirical	Expected	Empirical
3	0 ± 0.027	−0.176	0 ± 0.030	0.007	0 ± 0.030	−0.011	0 ± 0.030	−0.020
4	3 ± 0.053	4.305	3 ± 0.059	2.975	3 ± 0.059	3.014	3 ± 0.059	3.017
5	0 ± 0.301	−3.550	0 ± 0.330	0.059	0 ± 0.327	−0.066	0 ± 0.329	0.020
6	15 ± 0.852	42.839	15 ± 0.939	14.405	15 ± 0.948	15.110	15 ± 0.941	15.052
7	0 ± 3.923	−78.766	0 ± 4.324	0.658	0 ± 4.288	0.054	0 ± 4.300	3.077
8	105 ± 14.070	671.864	105 ± 15.591	98.687	105 ± 15.727	107.839	105 ± 15.577	106.745

$\Delta r_{\min} \gg \Delta r_\pi$ the coefficients $V_{i,n}$ do not represent the database and hence are incomplete. Our fit with an uniform $\Delta r \equiv \Delta r_\pi$ was satisfactory, as expected.

F. The 3σ self-consistent database

After the fitting process we get the desired 3σ self-consistent database using the idea proposed by Gross and Stadler [18] and worked at full length in our previous work [39]. This allows to rescue data which would otherwise have been discarded using the standard 3σ criterion contemplated in all previous analyzes [15–18,41]. The situation is illustrated in Fig. 1 (middle and right panels).

By using the rejection criterion at the 3σ level we cut off the long tails and, as a result, a fair comparison could, in principle, be made to this truncated Gaussian distribution. The Nijmegen group found that the moments method test (see below for more details) largely improved by using this truncated distribution [13]. It should be reminded, however, that the rejection criterion is applied to groups of data sets, and not to individual measurements, and in this way gets coupled with the floating of normalization. One could possibly improve on this by trying to determine individual outliers in a self-consistent way, which could make a more flexible data selection. Preliminary runs show that the number of iterations grows and the convergence may be slowed down or nonconverging by marginal decisions with some individual data flowing in and out the acceptance domain. Note also that rejection may also occur because data are themselves non normal or the disentanglement between statistical and systematic errors was not explicitly exploited. In both cases these data are useless to propagate uncertainties invoking the standard statistical interpretation, see Eq. (14).

G. Distribution of residuals

In Fig. 2 we present the resulting residuals, Eq. (3), in a normalized histogram for illustration purposes, in the cases of the original full database and the 3σ-consistent database, and compare them with a normal distribution function with the binning resolution $\Delta R = 0.2$. The complete database histogram shows an asymmetry or skewness as well as higher tails and clearly deviates from the normal distribution; meanwhile the 3σ consistent database residuals exhibit a closer agreement with the Gaussian distribution. Note that this perception from the figure somewhat depends on eyeball comparison of the three situations. We will discuss more preferable tests in the next section which are independent on this binning choice.

A handy way of checking for the normality of the residuals is looking into the standardized moments [4]. These are defined as

$$\mu_r' = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i - \mu}{\sigma}\right)^r, \tag{21}$$

where μ is the arithmetic mean and σ the standard deviation; the $r = 1$ and $r = 2$ standardized moments are zero and 1, respectively. Due to the finite size of any random sample an intrinsic uncertainty $\Delta\mu_r'(N)$ exists. This uncertainty can be estimated using Monte Carlo simulations with M random samples of size N and calculating the standard deviation of μ_r'. The result of such simulations are shown in Table I along with the moments of the residuals of the complete database with $N = 8125$ data and the 3σ self-consistent database with $N = 6713$. Clearly, the complete database shows discrepancies at 68% confidence level and hence cannot be attributed to

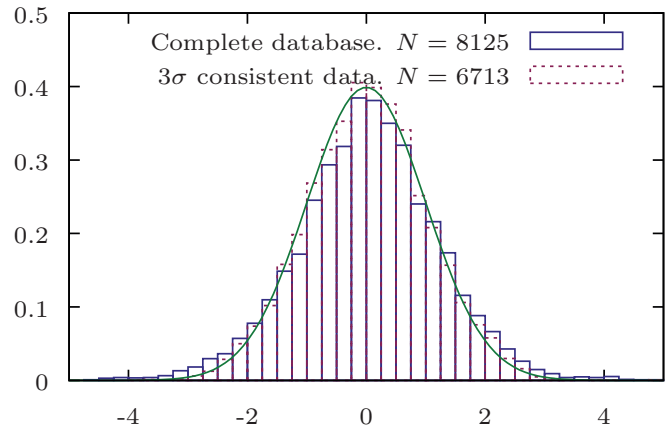

FIG. 2. (Color online) Normalized histogram of the resulting residuals after fitting the potential parameters to the complete pp and np database (blue boxes with solid borders) and to the 3σ consistent database (red boxes with dashed borders). The $N(0,1)$ standard normal probability distribution function (green solid line) is plotted for comparison.

the finite size of the sample. On the other hand,for the 3σ self-consistent database the moments fall in the expected interval. This is a first indication on the validity of Eq. (2) for our fit to this database.

The moments method was already used by the Nijmegen group [13] for the available at the time pp (about 400) data up to $T_{\rm LAB} = 30$ MeV. However, they tested the squared residuals R_i^2 in Eq. (3) with a χ^2 distribution with one degree of freedom which corresponds to testing only even moments of the normal distribution. As we have already pointed out, this is insensitive to the sign of R_i and hence may overlook relevant skewness.

H. Rescaling of errors

The usefulness of the normality test goes beyond checking the assumptions of the χ^2 fit since it allows to extend the validity of the method to naively unfavorable situations.

Indeed, if the actual value for $\chi^2_{\rm min}/\nu$ comes out outside the interval $1 \pm \sqrt{2/\nu}$, one can still rescale the errors by the so-called Birge factor [35] namely $\Delta O_i^{\rm exp} \to \sqrt{\chi^2_{\rm min}/\nu}\,\Delta O_i^{\rm exp}$ so the new figure of merit is

$$\bar{\chi}^2 = \left(\chi^2/\chi^2_{\rm min}\right)\nu, \tag{22}$$

which by definition fulfills $\bar{\chi}^2_{\rm min}/\nu = 1$. There is a common belief that this rescaling of χ^2 restores normality, when it only normalizes the resulting distribution.[12] If this was the case, there is no point in rejecting any single datum from the original database. Of course, it may turn out that one finds that residuals are nonstandardized normals. That means that they would correspond to a scaled Gaussian distribution. We will show that while this rescaling procedure works *once* the residuals obey a statistical distribution, the converse is not true; rescaling does not make residuals obey a statistical distribution.

In the case at hand we find that rescaling only works for the 3σ self-consistent database because residuals turn out to be normal. We stress that this is not the case for the full database. Of course, there remains the question on how much can errors be globally changed by a Birge factor. Note that errors quoted by experimentalists are in fact estimates and hence are subjected to their own uncertainties which ideally should be reflected in the number of figures provided in $\Delta O_i^{\rm exp}$. For $N \gg P$ one has $\nu \sim N$ and one has $\chi^2/\nu = 1 \pm \sqrt{2/N} = 1 \pm 0.016$ for $N = 8000$. Our fit to the complete database yields $\chi^2/\nu = 1.4$, which is well beyond the confidence level. Rescaling in this case would correspond to globally enlarge the errors by $\sqrt{1.4} \sim 1.2$ which is a 20% correction to the error in *all* measurements. Note that while this may seem reasonable, the rescaled residuals do not follow a Gaussian distribution. Thus, the noise on Eq. (2) remains unknown and cannot be statistically interpreted.

For instance, if we obtain $\chi^2/\nu = 1.2$ one would globally enlarge the errors by $\sqrt{1.1} \sim 1.1$ which is a mere 10% correction on the error estimate, a perfectly tolerable modification which corresponds to quoting just one significant figure on the error.[13] Thus, while $\chi^2_{\rm min}/\nu = 1 \pm \sqrt{2/\nu}$ looks as a sufficient condition for goodness of fit, it actually comes from the assumption of normality of residuals. However, one should not overlook the possibility that the need for rescaling might in fact suggest the presence of unforeseen systematic errors.

III. NORMALITY TESTS FOR RESIDUALS

There is a large body of statistical tests to quantitatively assess deviations from an specific probability distribution (see, e.g., Ref. [44]). In these procedures the distribution of empirical data X_i is compared with a theoretical distribution F_0 to test the *null hypothesis*, $H_0 : X_i \sim F_0$. If statistically significant differences are found between the empirical and theoretical distributions, the null hypothesis is rejected and its negation, the *alternative hypothesis*, $H_1 : X_i \sim F_1$, is considered valid, where F_1 is an unknown distribution that differs from F_0. The comparison is made by a *test statistic* T whose probability distribution is known when calculated for random samples of F_0; different methods use different test statistics. A decision rule to reject (or fail to reject) H_0 is made based on possible values of T; for example, if the observed value of the test statistic $T_{\rm obs}$ is greater (or smaller depending on the distribution of T) than a certain critical value T_c, the null hypothesis is rejected. T_c is determined by the probability distribution of T and the desired *significance level* α, which is the maximum probability of rejecting a true null hypothesis. Typical values of α are 0.05 and 0.01. Another relevant and meaningful quantity in hypothesis testing is the p value, which is defined as the smallest significance level at which the null hypothesis would be rejected. Therefore a small p value indicates clear discrepancies between the empirical distribution and F_0. A large p value, on the contrary, means that the test could not find significant discrepancies.

In our particular case H_0, the residuals follow a standard normal distribution, and the p value would be the probability that denying the assumption of true normality would be an erroneous decision.

A. Pearson test

A simple way of testing the goodness of fit is by using the Pearson test by computing the test statistic

$$T = \sum_{i=1}^{N_b} \frac{\left(n_i^{\rm fit} - n_i^{\rm normal}\right)^2}{n_i^{\rm th}}, \tag{23}$$

where $n_i^{\rm fit}$ are the number of residuals on each bin and $n_i^{\rm normal}$ are the number of expected residuals for the normal distribution in the same bin. T follows a χ^2 distribution with $N_b - 1$ DOF.

[12]This rescaling is a common practice when errors on the fitted quantities are not provided; uncertainties are invented with the condition that indeed $\chi^2_{\rm min}/\nu \sim 1$. The literature on phase-shift analyzes provides plenty of such examples. It is also a recommended practice in the Particle Data Group booklet when incompatible data are detected among different sets of measurements [42,43]).

[13]For instance, quoting $12.23(4) \equiv 12.23 \pm 0.04$ means that the error could be between 0.035 and 0.044, which is almost 25% uncertainty in the error. Quoting instead 12.230(12) corresponds to a 10% uncertainty in the error.

TABLE II. Results of the Pearson normality test of the residuals obtained by fitting the complete database with a δ-shell potential and the 3σ consistent database with the δ shell and the OPE-Gaussian potentials. The results of the test of the scaled residuals for every case is shown below the corresponding line. The critical value T_c corresponds to a significance level of $\alpha = 0.05$.

Database	Potential	N	T_c	$T_{\rm obs}$	p value
Complete	OPE-DS	8125	93.945	598.84	1.36×10^{-83}
				190.16	2.18×10^{-12}
3σ	OPE-DS	6713	87.108	82.67	0.09
				69.08	0.40
3σ	χTPE-DS	6712	87.108	100.70	0.004
				74.40	0.25
3σ	OPE-G	6711	87.108	84.17	0.08
				68.38	0.43

The decision on how close a given histogram is to the expected distribution depends on the specific choice of binning, which is the standard objection to this test. To perform the test we use a equiprobable binning so ΔR_i is such that $n_i^{\rm normal}$ is constant for all i, instead of the equidistant binning shown in Fig. 2 (see, e.g., Ref. [5] for more details on binning strategies). The results of the test are given in Table II and, as we see, *again* the complete database fails the test even when residuals are scaled.

B. Kolmogorov-Smirnov test

A simple and commonly used test is the Kolmogorov-Smirnov test [45,46]. The KS test uses the empirical distribution function $S(x)$, defined as the fraction of X_is that are less or equal to x and expressed by

$$S(x) = \frac{1}{N}\sum_{i=1}^{N}\theta(x - X_i), \tag{24}$$

where N is the number of empirical data. The test statistic in this procedure is defined as the greatest difference between $S(x)$ and $F_0(x)$, that is

$$T_{\rm KS} = \sup_x |F_0(x) - S(x)|. \tag{25}$$

Some of the advantages of using $T_{\rm KS}$ as a test statistic come from its distribution under the null hypothesis; since it is independent of F_0, it can be calculated analytically and a fairly good approximation exists for the case of large N. Given that large values of $T_{\rm KS}$ indicate large deviations from the theoretical distribution the decision rule will be to reject the

TABLE III. Same as Table II for the Kolmogorov-Smirnov test.

Database	Potential	N	T_c	$T_{\rm obs}$	p value
Complete	OPE-DS	8125	0.015	0.037	4.93×10^{-10}
				0.035	6.24×10^{-9}
3σ	OPE-DS	6713	0.017	0.011	0.43
				0.012	0.26
3σ	χTPE-DS	6712	0.017	0.010	0.47
				0.010	0.47
3σ	OPE-G	6711	0.017	0.013	0.22
				0.014	0.18

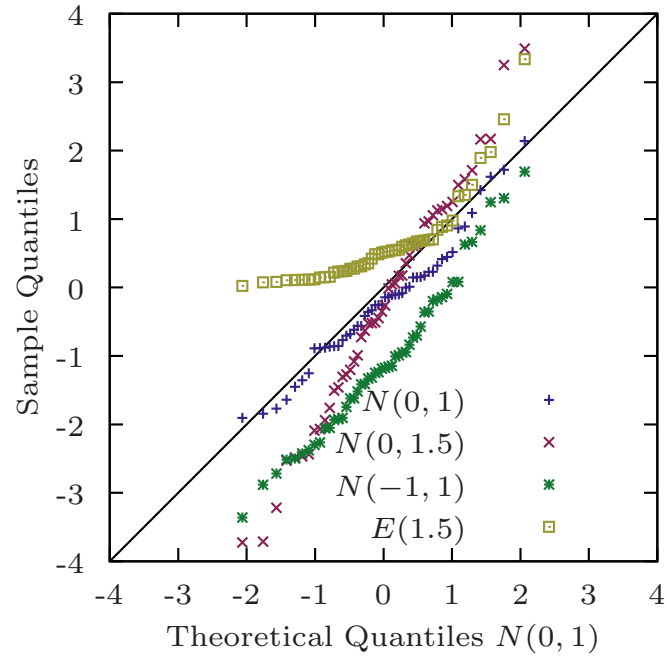

FIG. 3. (Color online) Quantile-quantile plot of different random samples against the standard normal distribution. Blue crosses are sampled from the $N(0,1)$ distribution, red diagonal crosses from $N(0,1.5)$, green asterisks from $N(-1,1)$ and yellow squares from the exponential distribution $E(1.5)$.

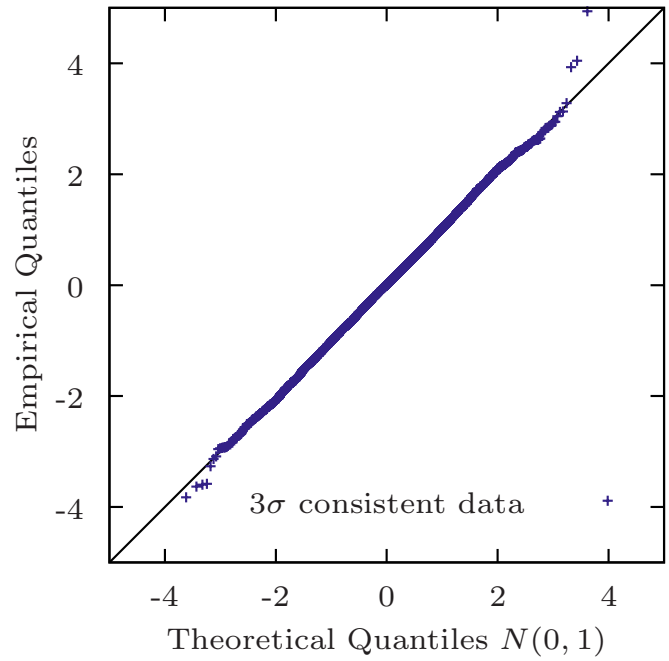

FIG. 4. (Color online) Quantile-quantile plot of the residuals obtained from fitting the 3σ consistent database against the standard normal distribution. The deviations at the tails, which are not detected using the Kolmogorov-Smirnov test, are clearly visible with this graphical tool.

TABLE IV. Same as Table II for the tail-sensitive test.

Database	Potential	N	T_c	$T_{\rm obs}$	p value
Complete	OPE-DS	8125	0.00070	0.0000	<0.0002
				3.54×10^{-25}	<0.0002
3σ	OPE-DS	6713	0.00072	0.0010	0.07
				0.0076	0.32
3σ	χTPE-DS	6712	0.00072	0.0005	0.03
				0.0156	0.50
3σ	OPE-G	6711	0.00072	0.0001	0.01
				0.0082	0.33

null hypothesis if the observed value $T_{\rm obs,KS}$ is larger than a certain critical value $T_{c,\rm KS}$. The critical value depends on α and N; for large numbers of data and a significance level of $0.05 T_{c,\rm KS} = 1.36/\sqrt{N}$. Also, a good approximation for the corresponding p value has been given [47],

$$P_{\rm KS}(T_{\rm obs}) = 2\sum_{j=1}^{\infty}(-1)^{j-1} e^{-2[(\sqrt{N}+0.12+0.11/\sqrt{N})jT_{\rm obs}]^2}. \quad (26)$$

The results of the KS normality test to the residuals obtained by fitting the potential parameters to the complete and 3σ consistent databases are shown in Table III. For the case of the complete database the observed test statistic is much larger than the critical value at the 0.05 significance level, which indicates that with a 95% confidence level the null hypothesis $H_0 : X_i \sim N(0,1)$ can be rejected; the extremely low p value gives an even greater confidence level to the rejection of H_0 very close to100%. In contrast, the observed test statistic using the 3σ-consistent data is smaller than the corresponding critical value, this indicates that there is no statistically significant evidence to reject H_0.

A shortcoming of the KS test is that the sensitivity to deviations from $F_0(x)$ is not independent from x. In fact, the KS test is most sensitive to deviations around the median value of F_0 and therefore is a good test for detecting shifts on the probability distribution, which in practice are unlikely to occur in the residuals of a least-squares fit. But, in turn, discrepancies away from the median such as spreads, compressions, or outliers on the tails, which are not that uncommon on residuals, may go unnoticed by the KS test.

C. Quantile-quantile plot

A graphical tool to easily detect the previously mentioned discrepancies is the quantile-quantile (QQ) plot, which maps two distributions quantiles against each other. The q quantiles of a probability distribution are obtained by taking $q-1$ equidistant points on the (0,1) interval and finding the values whose cumulative distribution function correspond to each

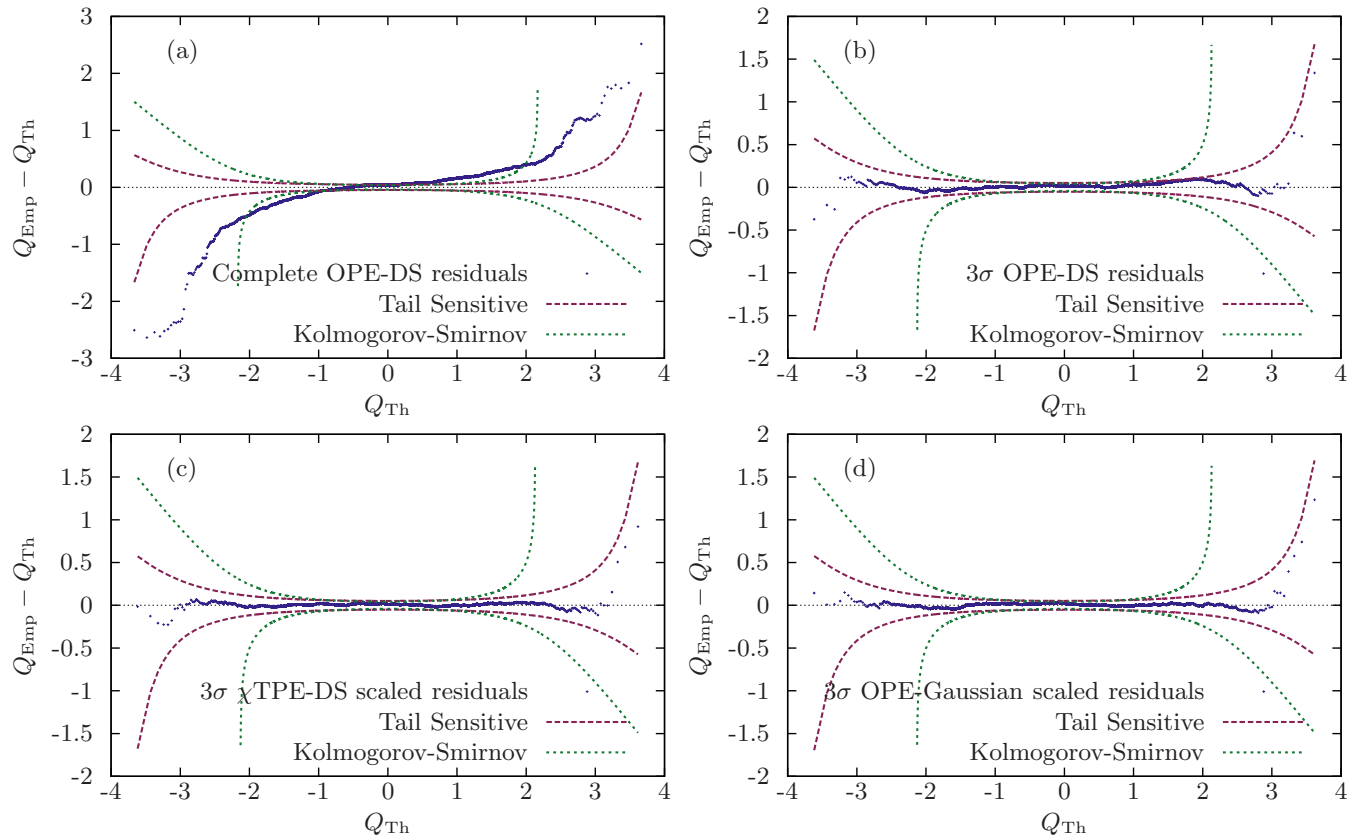

FIG. 5. (Color online) Rotated quantile-quantile plot of the residuals obtained (blue points) from fitting the complete database with the OPE-δ-shell potential (upper left panel), the 3σ self-consistent database fitted with the OPE-δ-shell potential (upper right panel), the χTPE-δ-shell potential (lower left panel), and the OPE-Gaussian potential (lower right panel). 95% confidence bands of the TS (red dashed lines) and KS (green dotted lines) tests are included.

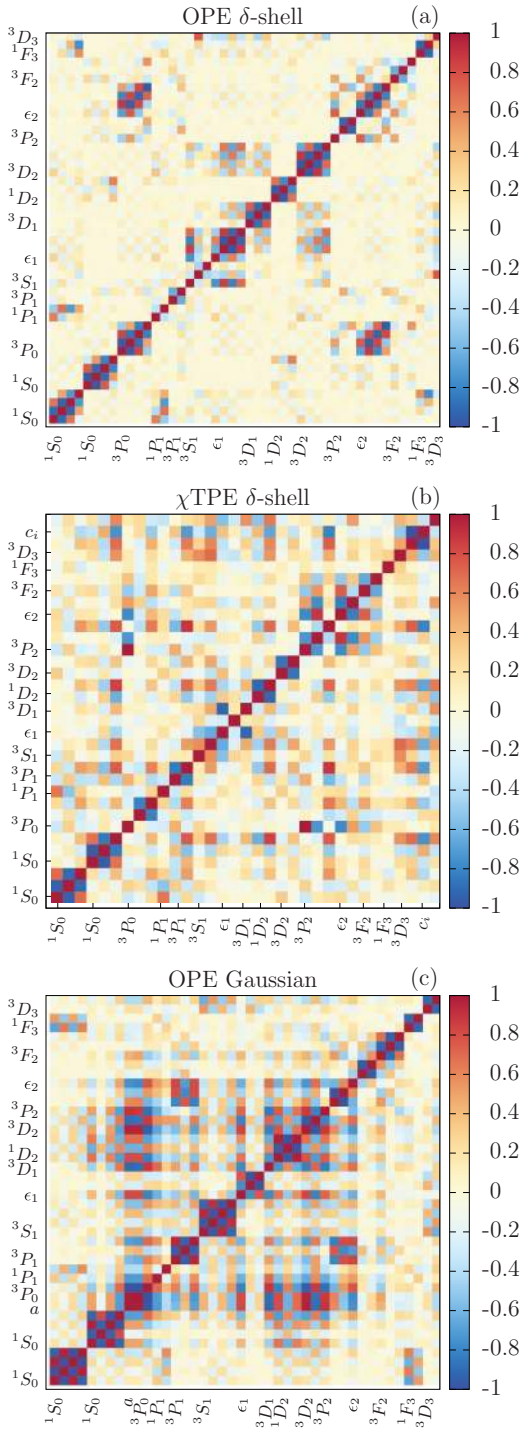

FIG. 6. (Color online) Correlation matrix $\mathcal{C}_{ij}$ for the short distance parameters in the partial wave basis $(V_i)^{LSJ}_{l,l'}$, see Eq. (18). We show the OPE-DS (upper panel) and the χTPE-DS (middle panel) potentials. The points $r_i = \Delta r_\pi (i+1)$ are grouped within every partial wave. The ordering of parameters is as in the parameter tables in Refs. [38,39] and [48] for OPE-DS 46 parameters and the χTPE 30+3 parameters (the last three are the chiral constants c_1,c_3,c_4) respectively. The OPE-Gaussian case (lower panel) also contains the parameter a. We grade gradually from 100% correlation, $\mathcal{C}_{ij} = 1$ (red), 0% correlation, $\mathcal{C}_{ij} = 0$ (yellow) and 100% anti-correlation, $\mathcal{C}_{ij} = -1$ (blue).

TABLE V. Fitting partial-wave parameters $(V_i)^{JS}_{l,l'}$ (in MeV) with their errors for all states in the JS channel. The dash indicates that the corresponding fitting $(V_i)^{JS}_{l,l'} = 0$. The parameters marked with an asterisk are set to have the tensor components vanish at the origin. The parameter a, which determines the width of each Gaussian, is also used as a fitting parameter and the value 2.3035 ± 0.0133 fm is found.

Wave	V_1	V_2	V_3	V_4
$^1S_0 np$	−67.3773	598.4930	−2844.7118	3364.9823
	±4.8885	±64.8759	±245.3275	±268.9192
$^1S_0 pp$	−52.0676	408.7926	−2263.1470	2891.2494
	±1.1057	±12.9206	±57.0254	±76.3709
$^3P_0$	−60.3589	–	520.5645	–
	±1.2182		±17.4210	
$^1P_1$	22.8758	–	256.2909	–
	±0.9182		±8.1078	
$^3P_1$	35.6383	−229.1500	928.1717	–
	±0.9194	±9.0104	±28.8275	
$^3S_1$	−42.4005	273.1651	−1487.4693	2064.7996
	±2.1344	±24.1462	±91.3195	±105.4383
ε_1	−121.8301	262.7957	−1359.3473	1218.3817*
	±3.2650	±19.0432	±50.9369	±34.8398
$^3D_1$	56.6746	–	–	–
	±1.3187			
$^1D_2$	−44.4366	220.5642	−617.6914	–
	±1.2064	±10.8326	±27.1533	
$^3D_2$	−107.3859	74.8901	–	–
	±2.9384	±7.1627		
$^3P_2$	−10.4319	–	−170.3098	132.4249
	±0.3052		±7.3280	±13.2310
ε_2	50.0324	−177.7386	748.5717	−620.8659*
	±0.8985	±8.2027	±34.7849	±27.2518
$^3F_2$	6.3917	−659.4308	3903.1138	–
	±2.6615	±41.3707	±187.9877	
$^1F_3$	28.5198	42.9715	–	–
	±3.0801	±19.5127		
$^3D_3$	−9.6022	65.9632	–	–
	±0.8870	±4.3677		

point. For example, to find the 4-quantiles of the normal distribution with zero mean and unit variance we take the points 0.25, 0.5, and 0.75 and look for values of x satisfying

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{\tilde{x}}{2}} d\tilde{x} = 0.25,\ 0.5,\ 0.75. \tag{27}$$

In this case, the 4-quantiles are −0.6745, 0, and 0.6745. For a set of ranked empirical data the easiest way to find the q-quantiles is to divide it into q essentially equal-sized subsets and take the $q-1$ boundaries as the quantiles.

To compare empirical data with a theoretical distribution function using a QQ plot the $N+1$-quantiles are used. In this way each datum can be graphed against the corresponding theoretical distribution's quantile; if the empirical and theoretical distributions are similar, the QQ plot points should lie close to the $y = x$ line. In Fig. 3 different random samples of size $N = 50$ are compared with a normal distribution. The first sample corresponds to the $N(0,1)$ distribution, and the second to the $N(0,1.5)$, and the larger spread of the data can be seen as a shift on the tails towards the bottom left and top right

TABLE VI. Operator coefficients $V_{i,n}$ (in MeV) with their errors for the OPE-Gaussian potential. The operators tT, τz, and $\sigma\tau z$ are set to zero.

Operator	V_1	V_2	V_3	V_4
c	−19.2829	126.2986	−648.6244	694.4340
	±0.6723	±7.7913	±33.1067	±36.8638
τ	2.3602	−25.4755	130.0301	−284.7219
	±0.4287	±5.4291	±20.0608	±19.8417
σ	6.0528	−75.1908	372.4133	−530.8121
	±0.4311	±5.2742	±19.5580	±22.4309
$\tau\sigma$	7.3632	−48.5435	273.7226	−349.0040
	±0.1794	±1.9523	±8.5410	±10.1673
t	1.9977	−22.1227	70.8515	−50.7264
	±0.2293	±2.6777	±10.1475	±7.8130
$t\tau$	15.0237	−38.3450	183.8178	−160.4965
	±0.3419	±1.8260	±5.2644	±3.7129
ls	−2.6164	39.4240	−217.0569	−109.6725
	±0.1947	±3.3849	±17.5511	±10.2746
$ls\tau$	0.0069	2.5897	−26.5807	−77.5825
	±0.0944	±1.1685	±5.5782	±3.3168
$l2$	1.4358	−23.5937	67.8942	144.1521
	±0.1809	±3.5108	±18.4785	±16.7585
$l2\tau$	−0.4106	8.3379	−82.9823	175.1091
	±0.0950	±1.4331	±6.2147	±5.7715
$l2\sigma$	−0.0990	2.2549	−51.8708	175.0991
	±0.1040	±1.5679	±6.6876	±6.2497
$l2\sigma\tau$	−0.2667	6.6299	−55.3425	100.7191
	±0.0343	±0.5087	±2.1657	±2.3042
$ls2$	0.4583	−11.6586	150.5353	−302.1105
	±0.2816	±4.9506	±22.8210	±17.1765
$ls2\tau$	0.7156	−18.8891	141.7216	−182.7536
	±0.1273	±1.8340	±7.5529	±5.7410
T	0.6379	−7.9042	24.2319	−19.7389
	±0.1996	±2.6738	±9.9460	±10.6364
σT	−0.6379	7.9042	−24.2319	19.7389
	±0.1996	±2.6738	±9.9460	±10.6364
$l2T$	−0.1063	1.3174	−4.0386	3.2898
	±0.0333	±0.4456	±1.6577	±1.7727
$l2\sigma T$	0.1063	−1.3174	4.0386	−3.2898
	±0.0333	±0.4456	±1.6577	±1.7727

parts of the graph. A third samples comes from the $N(-1,1)$ distribution and this can be seen as an downward shift of the points. A last sample is taken from the exponential distribution $E(1.5)$ which is asymmetric and positive.

Figure 4 shows the QQ plot of the residuals from the fit to the 3σ consistent database against the $N(0,1)$ distribution; deviations around the tails, which cannot be seen with the histogram in Fig. 2 and are not detected by the Pearson and KS tests, are clearly visible at the bottom left and top right corners of the plot.

D. Tail-sensitive test

Even though the QQ plot is a convenient and easy-to-use tool to detect deviations from a theoretical distribution, graphical methods often depend on subjective impressions and no quantitative description of the deviations visible in Fig. 4 can given by the QQ plot alone. A recent method by Aldor-Noiman *et al.* [31] provides $(1-\alpha)$ confidence bands to the QQ plot to quantitatively test deviations from the normal distribution. This new test, called tail sensitive, has a higher sensitivity on the tails than the KS test. In fact, the TS test rejection rate is uniformly distributed over the x variable. Although no analytic expression is given for the TS test statistic distribution, it can be easily simulated via Monte Carlo techniques. The details of such simulation are explained in Ref. [31]. We will restrict ourselves to point out that a small value of $T_{\rm TS}$ indicates discrepancies between the empirical and normal distribution and therefore the rejection criterion for the null hypothesis is $T_{\rm obs,TS} < T_{c,\rm TS}$.[14]

We applied the TS normality test to both sets of residuals, the complete database and the 3σ consistent one, and show the results on Table IV. For each test the Monte Carlo simulation consisted on taking 5000 random samples of size N with a standard normal distribution and calculating $T^{\rm MC}_{\rm obs,TS}$ for each sample to obtain the distribution of $T_{\rm TS}$ under the null hypothesis. The critical value for a significance level $\alpha = 0.05$ corresponds to the $T^{\rm MC}_{\rm obs,TS}$ that is greater than 5% of all the values calculated. Finally, the test statistic for the empirical data $T^{\rm emp}_{\rm obs,TS}$ can be calculated and compared to the simulated distribution to obtain the p value. In this case the p value is the proportion of $T^{\rm MC}_{\rm obs,TS}$ that are smaller than $T^{\rm emp}_{\rm obs,TS}$. Since the observed $T_{\rm TS}$ for the complete database residuals is numerically equal to zero and smaller than all of the simulated

[14]It should also be noted that a typo in Ref. [31] is made in their steps 1c and 1e where Φ^{-1} and $B^{-1}_{(i,n+1-i)}$ are printed instead of Φ and $B_{(i,n+1-i)}$; the latter are consistent with the rest of the text and the results presented there.

TABLE VII. Deuteron static properties compared with empirical/recommended values and high-quality potentials calculations. We list binding energy E_d, asymptotic D/S ratio η, asymptotic S-wave amplitude A_S, mean-squared matter radius r_m, quadrupole moment Q_D, and D-wave probability P_D.

	This work	Emp./Rec. [55–60]	δ-shell [38]	Nijm I [15]	Nijm II [15]	Reid93 [15]	AV18 [16]	CD-Bonn [17]
E_d (MeV)	Input	2.224575(9)	Input	Input	Input	Input	Input	Input
η	0.02448(5)	0.0256(5)	0.02493(8)	0.02534	0.02521	0.02514	0.0250	0.0256
A_S (fm$^{1/2}$)	0.8885(3)	0.8845(8)	0.8829(4)	0.8841	0.8845	0.8853	0.8850	0.8846
r_m (fm)	1.9744(6)	1.971(6)	1.9645(9)	1.9666	1.9675	1.9686	1.967	1.966
Q_D (fm^2)	0.2645(7)	0.2859(3)	0.2679(9)	0.2719	0.2707	0.2703	0.270	0.270
P_D	5.30(4)	5.67(4)	5.62(5)	5.664	5.635	5.699	5.76	4.85

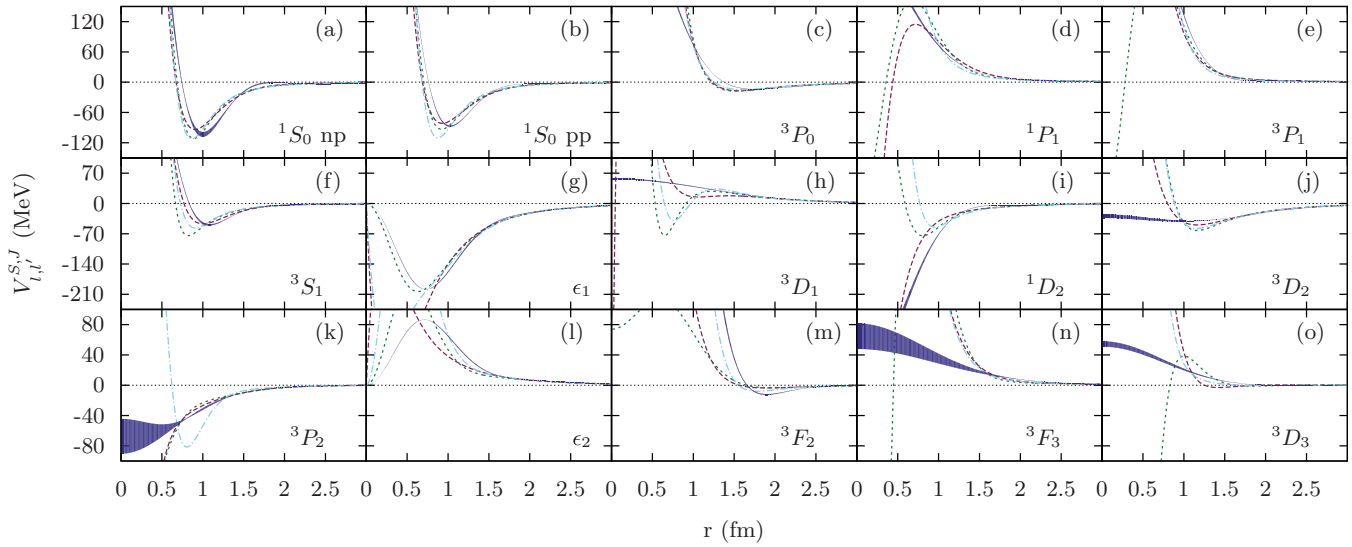

FIG. 7. (Color online) Lowest *np* and *pp* partial-wave potentials (in MeV) and their errors (solid band) as a function of the internucleon separation (in fm) for the present OPE+Gaussian analysis (blue band), Reid93 [15] (red dashed), NijmII [15] (green dotted), and AV18 [16] (light-blue dashed-dotted) as a function of the internucleon distance r (in fm).

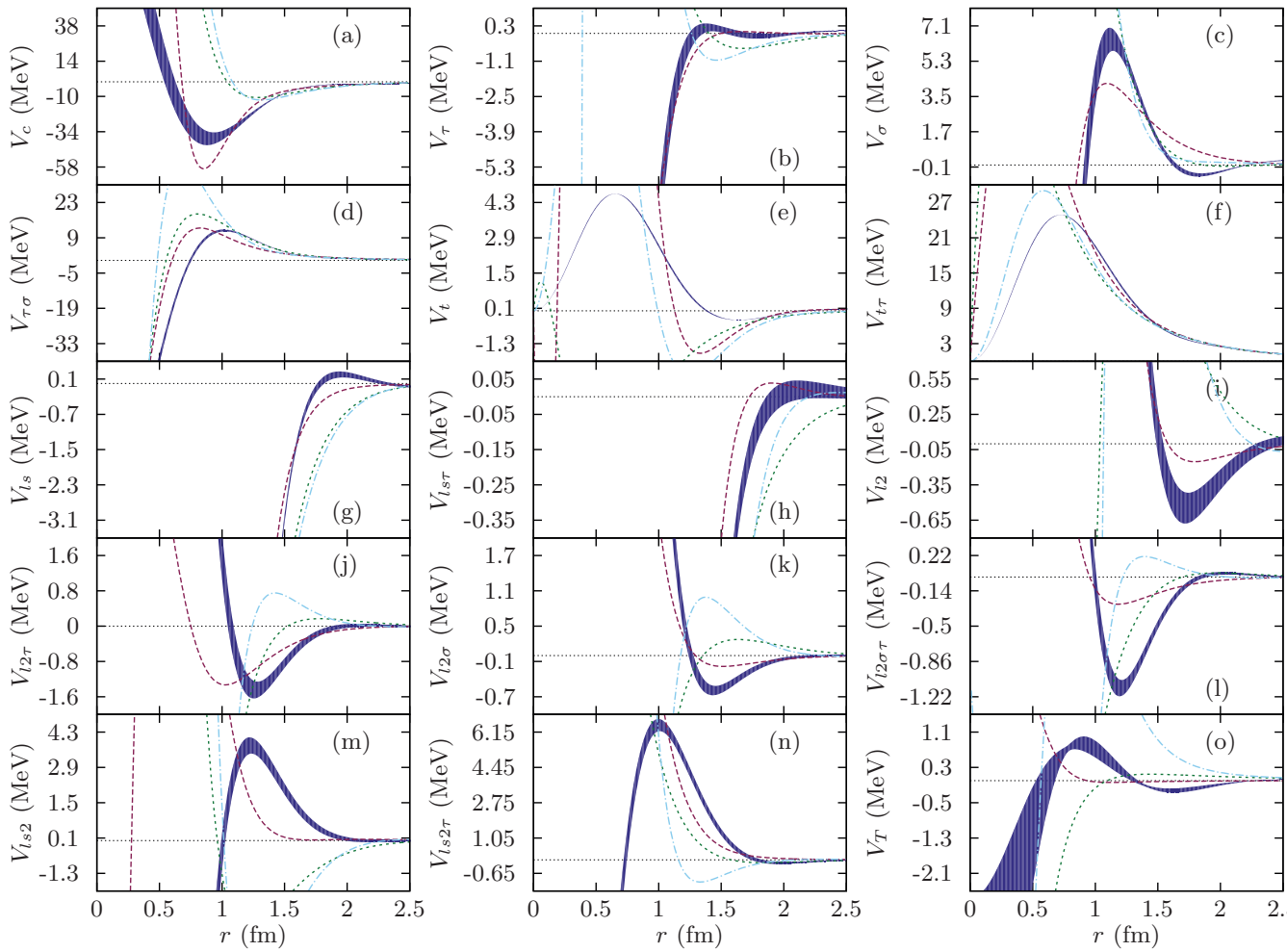

FIG. 8. (Color online) *NN* potentials (in MeV) in the operator basis with errors (solid band) as a function of the internucleon separation (in fm) for the present OPE+Gaussian analysis (blue band), Reid93 [15] (red dashed), NijmII [15] (green dotted), and AV18 [16] (light-blue dashed-dotted) as a function of the internucleon distance r (in fm).

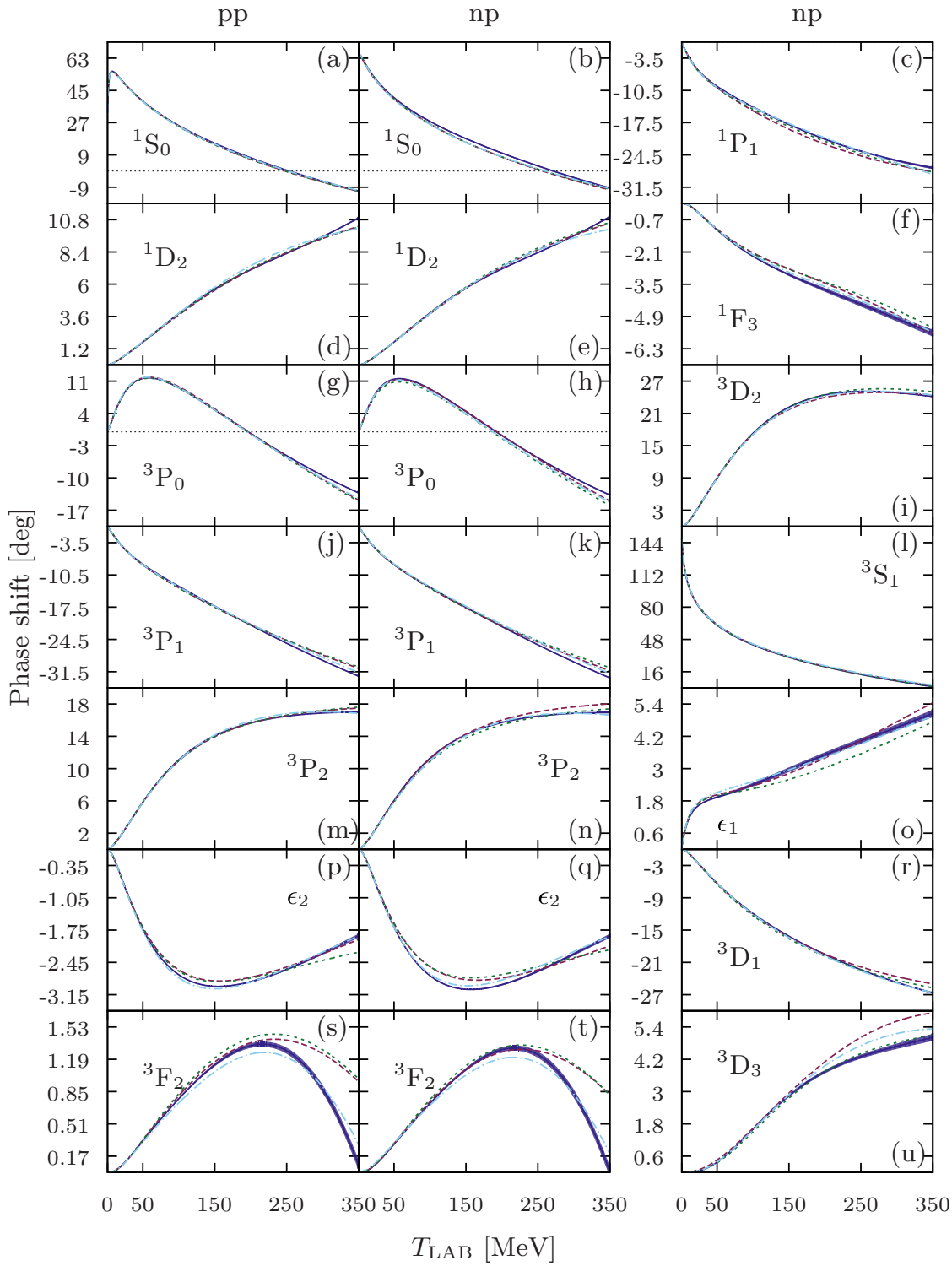

FIG. 9. (Color online) Lowest *np* and *pp* phase shifts (in degrees) and their errors for the present OPE+Gaussian analysis (blue band), Reid93 [15] (red dashed), NijmII [15] (green dotted), and AV18 [16] (light-blue dashed-dotted) as a function of the LAB energy (in MeV).

values, we can only give an upper bound to the p value. The graphical results of the TS test are presented in Fig. 5 with the 95% confidence level bands; the same bands for the KS test are drawn for comparison reasons. Since for such a large value of N the confidence bands are very narrow, a 45°-clockwise rotated QQ plot is used to visually enhance the possible deviations from a normal distribution. The complete database residuals (upper left panel) show obvious deviations

TABLE VIII. *pp* isovector phase shifts.

$E_{\rm LAB}$	$^1S_0$	$^1D_2$	$^1G_4$	$^3P_0$	$^3P_1$	$^3F_3$	$^3P_2$	ϵ_2	$^3F_2$	$^3F_4$	ϵ_4	$^3H_4$
1	32.666	0.001	0.000	0.133	−0.080	−0.000	0.013	−0.001	0.000	0.000	−0.000	0.000
	±0.003	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000
5	54.834	0.042	0.000	1.578	−0.899	-0.004	0.205	−0.052	0.002	0.000	−0.000	0.000
	±0.006	±0.000	±0.000	±0.002	±0.001	±0.000	±0.001	±0.000	±0.000	±0.000	±0.000	±0.000
10	55.223	0.163	0.003	3.729	−2.053	−0.031	0.628	−0.201	0.013	0.001	−0.004	0.000
	±0.010	±0.000	±0.000	±0.005	±0.002	±0.000	±0.002	±0.000	±0.000	±0.000	±0.000	±0.000
25	48.694	0.688	0.040	8.616	−4.892	−0.233	2.440	−0.815	0.103	0.018	−0.049	0.004
	±0.014	±0.001	±0.000	±0.016	±0.007	±0.000	±0.005	±0.001	±0.000	±0.000	±0.000	±0.000
50	39.040	1.701	0.152	11.601	−8.186	−0.704	5.823	−1.735	0.328	0.099	−0.197	0.026
	±0.018	±0.003	±0.000	±0.030	±0.013	±0.001	±0.009	±0.003	±0.001	±0.001	±0.000	±0.000
100	25.452	3.820	0.414	9.567	−13.010	−1.546	11.074	−2.727	0.774	0.444	−0.553	0.107
	±0.034	±0.008	±0.001	±0.052	±0.017	±0.008	±0.013	±0.007	±0.007	±0.004	±0.001	±0.000
150	15.567	5.642	0.702	4.732	−17.296	−2.070	14.058	−2.980	1.132	0.991	−0.881	0.201
	±0.050	±0.014	±0.005	±0.064	±0.026	±0.019	±0.020	±0.010	±0.015	±0.009	±0.002	±0.002
200	7.490	7.058	1.032	−0.388	−21.412	−2.308	15.663	−2.875	1.337	1.642	−1.158	0.292
	±0.064	±0.022	±0.011	±0.064	±0.037	±0.031	±0.025	±0.017	±0.024	±0.014	±0.004	±0.005
250	0.500	8.276	1.385	−5.174	−25.335	−2.371	16.506	−2.603	1.289	2.272	−1.381	0.380
	±0.080	±0.026	±0.017	±0.066	±0.052	±0.044	±0.032	±0.023	±0.032	±0.019	±0.005	±0.011
300	−5.699	9.537	1.713	−9.460	−29.016	−2.385	16.892	−2.253	0.891	2.768	−1.556	0.478
	±0.102	±0.032	±0.022	±0.087	±0.073	±0.061	±0.044	±0.031	±0.041	±0.026	±0.006	±0.018
350	−11.239	10.974	1.959	−13.221	−32.431	−2.461	16.977	−1.875	0.091	3.056	−1.691	0.608
	±0.130	±0.059	±0.027	±0.124	±0.101	±0.084	±0.060	±0.042	±0.054	±0.045	±0.006	±0.025

from the normal distribution which is reflected on the extremely low p values. The 3σ consistent data residuals (upper right panel) show deviations from the normal distribution that are always within the TS confidence bands and therefore to a confidence level $\alpha = 0.05$ there are no statistically significant differences to reject the null hypothesis.

E. Discussion

We haveshown in the previous discussion evidence supporting the validity of Eq. (2) for the 3σ self-consistent database recently built from all published np- and pp-scattering data from 1950 to 2013 [30,33]. The numerics can be a costly procedure since multiple optimizations must be carried out

TABLE IX. *np* isovector phase shifts.

$E_{\rm LAB}$	$^1S_0$	$^1D_2$	$^1G_4$	$^3P_0$	$^3P_1$	$^3F_3$	$^3P_2$	ϵ_2	$^3F_2$	$^3F_4$	ϵ_4	$^3H_4$
1	62.074	0.001	0.000	0.180	−0.108	−0.000	0.021	−0.001	0.000	0.000	−0.000	0.000
	±0.018	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000	±0.000
5	63.652	0.040	0.000	1.653	−0.940	−0.004	0.248	−0.048	0.002	0.000	−0.000	0.000
	±0.045	±0.000	±0.000	±0.002	±0.001	±0.000	±0.001	±0.000	±0.000	±0.000	±0.000	±0.000
10	60.004	0.154	0.002	3.747	−2.073	−0.026	0.705	−0.185	0.011	0.001	−0.003	0.000
	±0.065	±0.000	±0.000	±0.006	±0.003	±0.000	±0.002	±0.000	±0.000	±0.000	±0.000	±0.000
25	51.043	0.669	0.032	8.506	−4.896	−0.201	2.586	−0.768	0.089	0.015	−0.039	0.003
	±0.107	±0.001	±0.000	±0.017	±0.007	±0.000	±0.005	±0.001	±0.000	±0.000	±0.000	±0.000
50	40.920	1.701	0.131	11.433	−8.251	−0.634	6.025	−1.688	0.295	0.089	−0.169	0.020
	±0.167	±0.003	±0.001	±0.031	±0.013	±0.001	±0.009	±0.003	±0.001	±0.001	±0.000	±0.000
100	27.691	3.863	0.365	9.314	−13.211	−1.447	11.261	−2.747	0.724	0.428	−0.505	0.090
	±0.268	±0.008	±0.007	±0.053	±0.018	±0.008	±0.014	±0.007	±0.007	±0.004	±0.001	±0.000
150	18.146	5.697	0.594	4.380	−17.569	−1.977	14.170	−3.042	1.083	0.981	−0.834	0.176
	±0.313	±0.014	±0.027	±0.064	±0.027	±0.020	±0.020	±0.010	±0.016	±0.009	±0.002	±0.002
200	10.161	7.111	0.838	−0.809	−21.717	−2.236	15.705	−2.938	1.295	1.643	−1.124	0.261
	±0.309	±0.022	±0.056	±0.064	±0.038	±0.032	±0.025	±0.017	±0.024	±0.014	±0.004	±0.005
250	3.068	8.331	1.118	−5.626	−25.658	−2.322	16.495	−2.644	1.248	2.280	−1.369	0.347
	±0.304	±0.026	±0.085	±0.067	±0.053	±0.045	±0.032	±0.024	±0.032	±0.019	±0.005	±0.011
300	−3.345	9.601	1.434	−9.916	−29.352	−2.356	16.840	−2.271	0.841	2.775	−1.566	0.448
	±0.345	±0.033	±0.102	±0.089	±0.074	±0.062	±0.045	±0.031	±0.042	±0.026	±0.006	±0.018
350	−9.144	11.052	1.763	−13.666	−32.782	−2.447	16.891	−1.879	0.022	3.053	−1.720	0.583
	±0.441	±0.062	±0.105	±0.127	±0.103	±0.085	±0.061	±0.043	±0.055	±0.047	±0.006	±0.025

TABLE X. *np* isoscalar phase shifts.

$E_{\rm LAB}$	$^1P_1$	$^1F_3$	$^3D_2$	$^3G_4$	$^3S_1$	ϵ_1	$^3D_1$	$^3D_3$	ϵ_3	$^3G_3$
1	−0.186	−0.000	0.006	0.000	147.624	0.102	−0.005	0.000	0.000	−0.000
	±0.000	±0.000	±0.000	±0.000	±0.009	±0.000	±0.000	±0.000	±0.000	±0.000
5	−1.493	−0.010	0.218	0.001	117.905	0.638	−0.177	0.002	0.012	−0.000
	±0.004	±0.000	±0.000	±0.000	±0.020	±0.003	±0.000	±0.000	±0.000	±0.000
10	−3.058	−0.064	0.843	0.012	102.230	1.086	−0.661	0.007	0.080	−0.003
	±0.010	±0.000	±0.001	±0.000	±0.028	±0.007	±0.001	±0.000	±0.000	±0.000
25	−6.337	−0.421	3.698	0.170	80.068	1.653	−2.735	0.058	0.552	−0.053
	±0.034	±0.000	±0.005	±0.000	±0.041	±0.018	±0.005	±0.003	±0.000	±0.000
50	−9.603	−1.143	8.951	0.722	62.105	1.955	−6.276	0.376	1.609	−0.264
	±0.071	±0.003	±0.020	±0.000	±0.053	±0.035	±0.013	±0.013	±0.002	±0.000
100	−14.089	−2.291	17.299	2.181	42.633	2.428	−11.922	1.599	3.451	−0.989
	±0.113	±0.022	±0.049	±0.005	±0.065	±0.066	±0.030	±0.038	±0.011	±0.004
150	−17.844	−3.102	22.164	3.665	30.269	2.980	−16.143	2.830	4.700	−1.898
	±0.129	±0.052	±0.060	±0.019	±0.066	±0.085	±0.045	±0.054	±0.024	±0.013
200	−21.036	−3.775	24.449	5.065	20.890	3.517	−19.526	3.690	5.536	−2.851
	±0.148	±0.080	±0.073	±0.041	±0.067	±0.093	±0.059	±0.061	±0.034	±0.029
250	−23.623	−4.421	25.137	6.379	13.208	4.007	−22.339	4.222	6.150	−3.787
	±0.181	±0.100	±0.096	±0.066	±0.088	±0.099	±0.072	±0.074	±0.039	±0.048
300	−25.653	−5.078	24.920	7.604	6.681	4.476	−24.681	4.578	6.648	−4.692
	±0.222	±0.116	±0.121	±0.086	±0.131	±0.114	±0.088	±0.099	±0.047	±0.067
350	−27.236	−5.734	24.242	8.712	1.036	4.956	−26.586	4.876	7.067	−5.568
	±0.266	±0.145	±0.147	±0.097	±0.183	±0.137	±0.107	±0.130	±0.065	±0.082

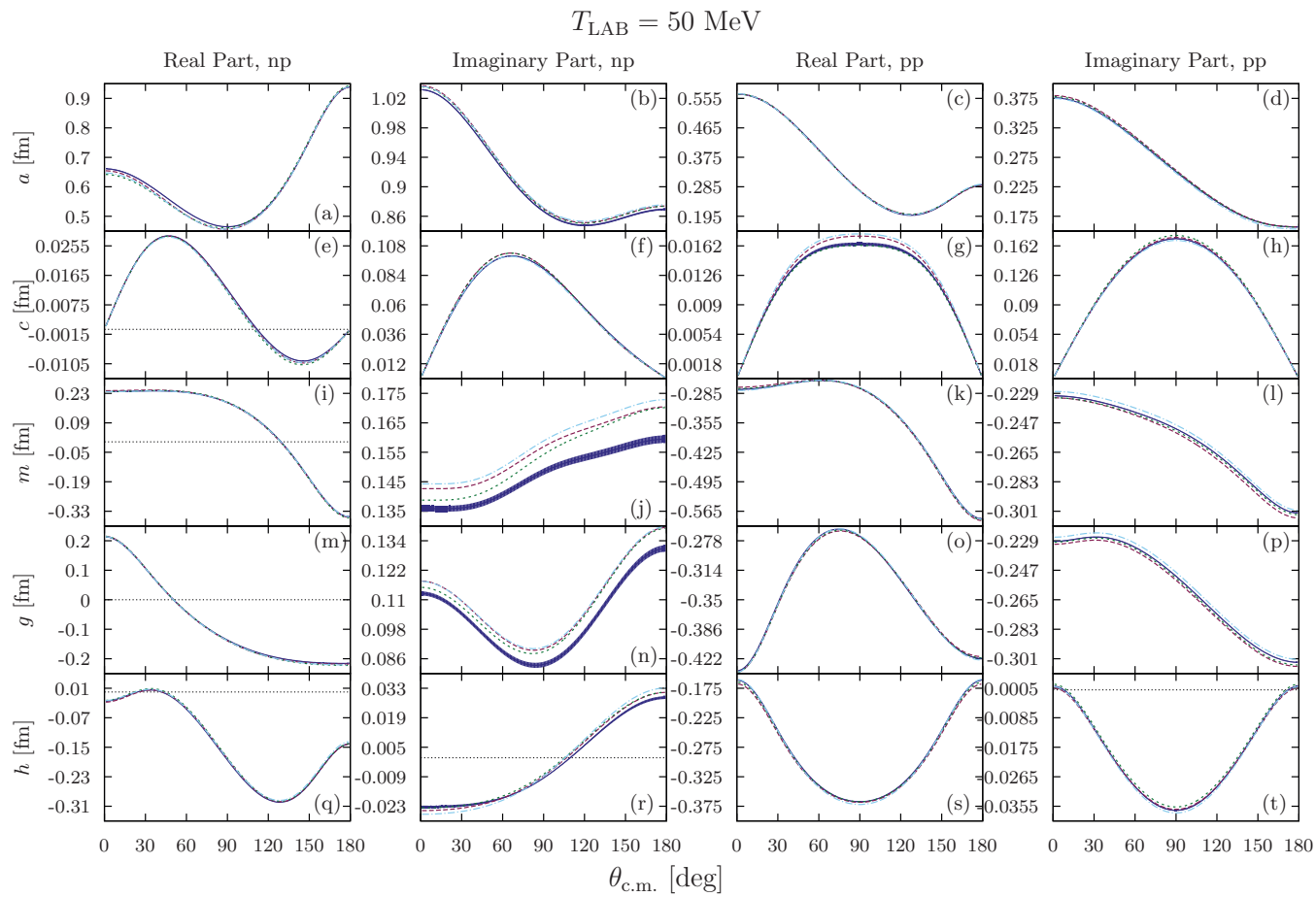

FIG. 10. (Color online) *np* (left) and *pp* (right) Wolfenstein parameters (in fm) as a function of the center-of-mass angle (in degrees) and for $E_{\rm LAB} = 50$ MeV. We compare our fit (blue band) with Reid93 [15] (red dashed), NijmII [15] (green dotted), and AV18 [16] (light-blue dashed-dotted).

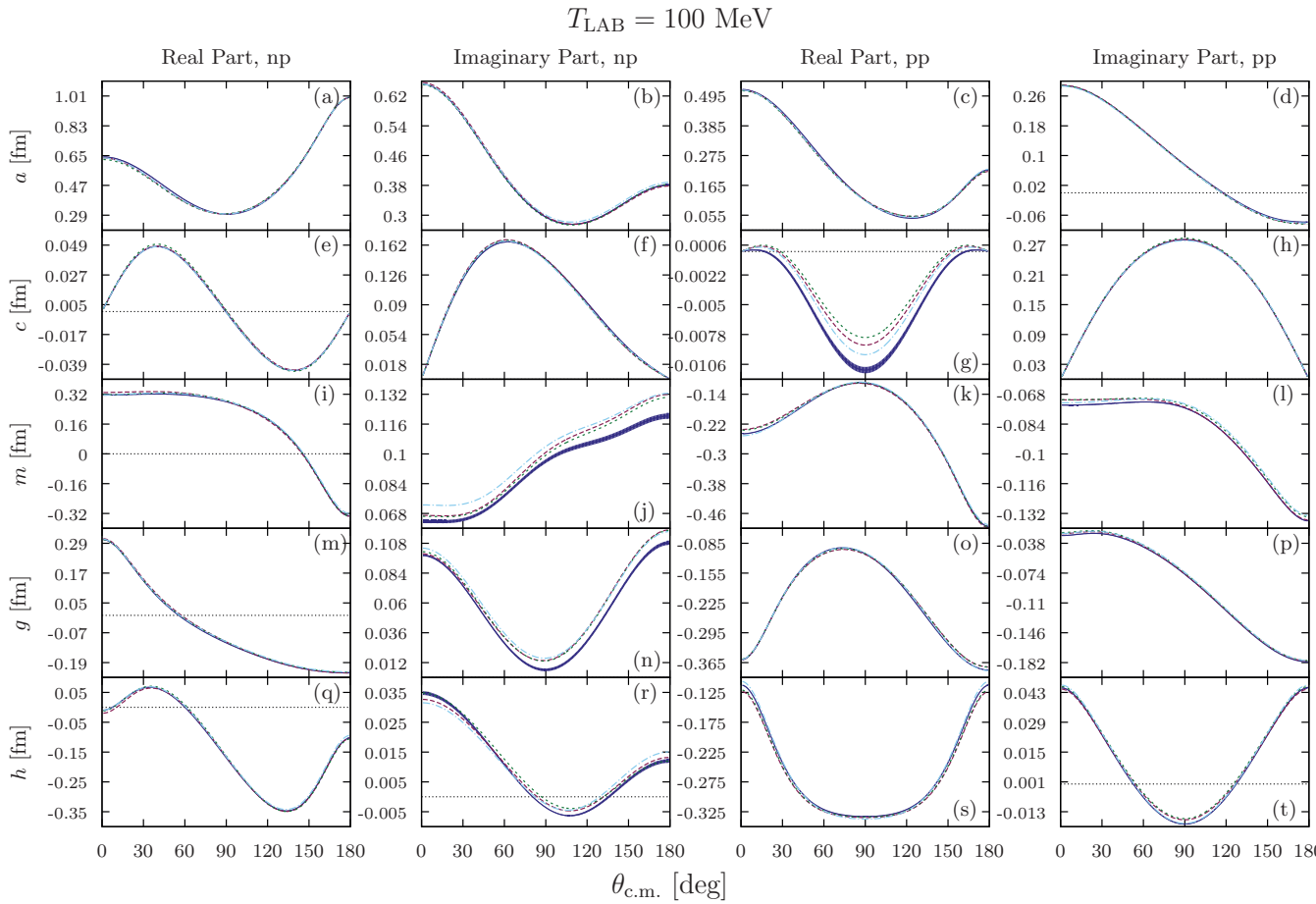

FIG. 11. (Color online) Same as in Fig. 10 but for $E_{\rm LAB} = 100$ MeV.

and different subsets of data of the complete database must be tested and confronted. As outlined above, our analysis was carried out using a physically motivated coarse-grained potential and, more specifically, a δ-shells interaction already proposed by Avilés [29]. This scheme proved extremely convenient for fast minimization and error evaluation.

As a first application, with the currently fixed database, we have also addressed the calculation of the chiral constants which appear in the χTPE potential [48] (which also passes the normality test, as can be seen from Fig. 5 and Tables I–IV). We note that the small rescaling by the Birge factor $\sqrt{1.07}$ is requested to pass the Pearson and TS tests. As we have mentioned, this form of δ-shell potentials cannot be directly implemented in some of the many powerful computational approaches to nuclear structure calculations.[15]

The necessary conditions for a sensible interpretation of the χ^2 fit according to Eq. (2) requires testing for normality of residuals of a fit to a consistent database. In all, the present situation regarding both the selection of data with the self-consistent 3σ criterion and the normality of residuals turns out to be highly satisfactory. In our view, this combined consistency of the statistical assumptions and the theory used to analyze it provides a good starting point to proceed further in the design of theory-friendly smooth *NN* interactions as well as a sound estimate of their statistical uncertainties.

Of course, the normality of residuals applies to *any* fit aiming at representing the data. Thus, any potential which pretends to represent the data ought to pass the test. In the next section we propose a potential whose short-distance part is made of a superposition of Gaussian functions and, unlike the δ-shell potential, can be plotted. We will check that our proposed potential does in fact pass the normality test.

There is an issue concerning the statistical approach on *what* would be the "true" potential since the concept of true parameters of a given model is invoked (see the discussion in Sec. II B). On the one hand, the very definition of potential is subject to ambiguities because the scattering information only determines an interaction once its specific form has been chosen [21]. This reflects the well-known off-shell ambiguities which by definition are inaccessible to experiment [49]. On the other hand, nuclear structure calculations are carried out with potentials statistically representing the scattering data. This is a source for a systematic uncertainty which was unveiled in Refs. [22–24] for the previously developed high-quality

[15]The δ-shell potential cannot even be plotted, which may naively seem a disadvantage. However, its Fourier transformation is smooth [30] in the relevant center-of-mass momentum region of $p_{\rm c.m.} \lesssim 2$ fm^{-1}, complying to the idea that coarse graining down to $\Delta r_\pi \sim 0.6$ fm resolutions lacks information on shorter length scales.

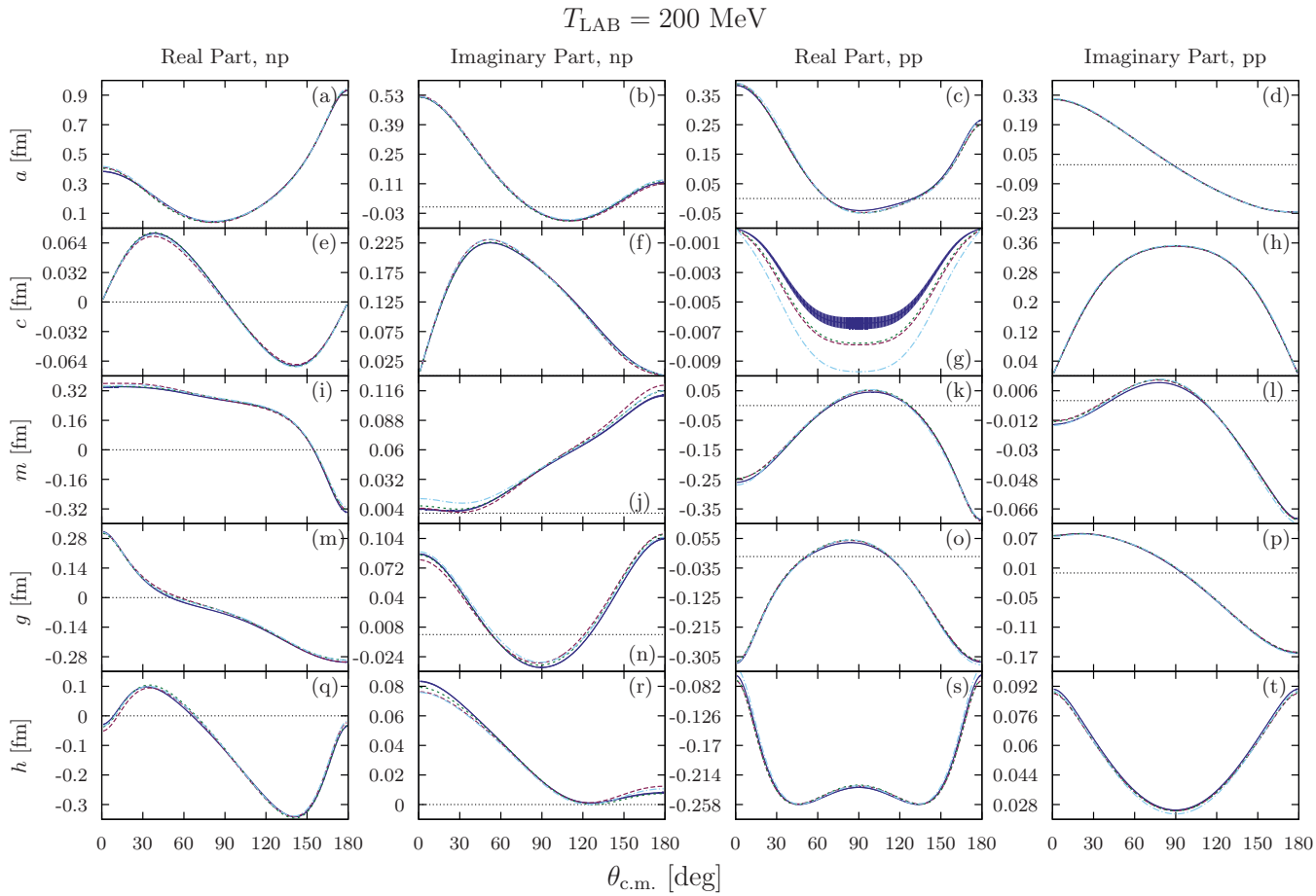

FIG. 12. (Color online) Same as in Fig. 10 but for $E_{\rm LAB} = 200$ MeV.

interactions. The upgrade of this systematic uncertainty study using the present statistical analysis is left for future research.

Ultimately, QCD is the theory to validate Eq. (2) versus the large body of data, $O_i^{\rm th} = O_i^{\rm QCD}$, with just two parameters in the (u,d) sector, $\Lambda_{\rm QCD}$, and the quark masses (m_u,m_d), or, equivalently, with the pion weak decay constant f_π and the pion masses $(m_{\pi^0},m_{\pi^\pm})$. Remarkably, nuclear potentials have been evaluated on the lattice recently [50–52]. The HAL QCD Collaboration [53] finds a local potential for the unphysical pion mass $m_\pi = 701$ MeV with a shape similar to our OPE-Gaussian potential (see Sec. IV) but a depth of -30 MeV in the central component V_c and $\Delta V_c \sim 5$ MeV for $r \gtrsim 1$ fm, and, consequently, the $^1S_0$ phase-shift obtained by directly solving the Schrödinger equation is smaller as compared to ours with much larger errors. This potential approach uses the Nambu-Bethe-Salpeter wave function which ultimately depends on the choice of the interpolating composite nucleon fields (for a recent overview of the pros and cons of the potential approach to lattice QCD see, e.g., Ref. [54]). Of course, since the lattice *NN* potential depends ultimately in just two parameters, $\Lambda_{\rm QCD}$ and m_q the different r values in the potential functions $V_n(r)$ *must be* correlated. In the phenomenological approach correlations among the fitting parameters are indeed found or built in. Some of them are the trivial ones due to the OPE potential which just depends on the pion masses $(m_{\pi^0},m_{\pi^\pm})$, but others correspond to the inner short-distance parameters, suggesting that the number of parameters can de reduced *solely* from the phenomenological potential analysis of the data. In Fig. 6 we represent pictorially the resulting correlation matrix both for the OPE-DS fit [38,39] as well as for χTPE-DS [48] short-distance parameters in the partial-wave basis $(V_i)_{l,l'}^{LSJ}$, see Eq. (18). The listing ordering is the same as the one in the parameter tables in Refs. [38,39] and [48] for OPE-DS and χTPE-DS, respectively. Note the isolated pattern of correlations for the OPE-DS case, however, as we see there are substantial correlations among different $(V_i)_{l,l'}^{LSJ}$ within a given partial wave, suggesting the possibility of reducing the number of parameters. Indeed, we observe that this parameter reduction takes place from 46 to 33 when going from the OPE-DS case to the χTPE-DS potential [48], which incorporates specific QCD features such as chiral symmetry. The resulting correlation pattern becomes now more spread over the full short-distance parameter space.

IV. THE OPE-GAUSSIAN POTENTIAL

In the present section we provide a rather simple local form of the potential Eqs. (17) and (18) based on Gaussian functions

$$F_{i,n}(r) = e^{-r^2/(2a_i^2)}, \tag{28}$$

where we have taken the parameters as $a_i = a/(1+i)$. The parameter a is used as a fitting variable. With this potential

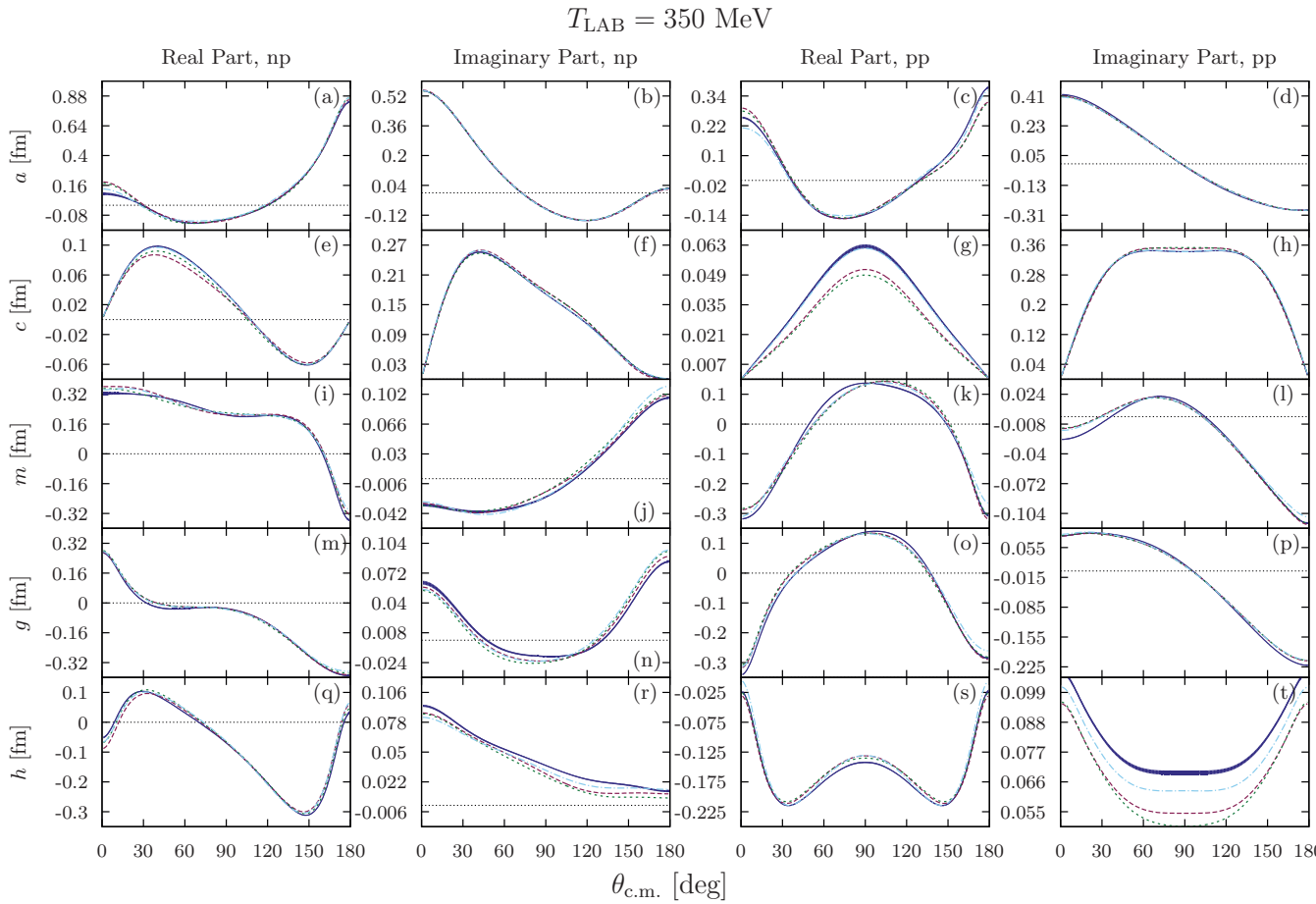

FIG. 13. (Color online) Same as in Fig. 10 but for $E_{\rm LAB} = 350$ MeV.

we get $\chi^2/\nu = 1.06$. The resulting 42 fitting parameters (41 independent partial-wave coefficients $(V_i)^{JS}_{l,l'}$ and the Gaussian width a are listed with their uncertainties in Table V. The $V_{i,n}$ operator coefficients are given in Table VI.[16] The linear transformation from partial-wave coefficients $(V_i)^{JS}_{l,l'}$ to the $V_{i,n}$ operator coefficients has been given explicitly in Ref. [39]. In Fig. 6 we depict the correlation matrix, Eq. (12), for the partial-wave parameters listed in Table V, where a similar correlation pattern to the OPE-DS one is observed. Deuteron properties for this potential compared with calculations using other potentials and empirical or recommended values can be looked up in Table VII.

The rotated QQ plot of the scaled residuals for the OPE-Gaussian fit to the 3σ self-consistent database can be seen in Fig. 5. As we can see the TS test is passed satisfactorily. On a more quantitative level we show on Table I the moments test. The resulting p value of the different normality tests are given in Tables II, III, and IV for the Pearson, KS, and TS tests, respectively. As we see all tests are satisfactorily passed except for the TS where a tiny scaling of the residuals by a Birge factor of $\sqrt{\chi^2/\nu} = 1.03$, corresponding to a global enlargement of the provided experimental errors by 3%, allows to restore normality. Thus, we are entitled to propagate the uncertainties of the data to derived quantities through the determined parameters $V_{i,n}$ with errors and their corresponding correlations, see Eq. (14).

In Figs. 7 and 8 we show the OPE-Gaussian potential in partial wave and operator basis, respectively, with the error bands propagated with the corresponding correlation matrix from the fit to the experimental data. As we see, these error bands are smaller than the discrepancy with Reid93 [15], NijmII [15], and AV18 [16]. This may be a hint that systematic errors induced by the bias involved in the choice of the several potentials, as first noted in Refs. [22–24], may indeed play a relevant role in the total evaluation of nuclear uncertainties.

In Fig. 9 we present the lowest np and pp phase shifts and their errors based on the OPE-Gaussian potential and compared with the Reid93 [15], NijmII [15], and AV18 [16] potential phases. In Tables VIII, IX, and X the low-angular-momentum phases as a function of the LAB energy with their errors propagated from the fit are listed.

The resulting Wolfenstein parameters, Eq. (1), for the OPE-Gaussian potential are depicted in Figs. 10, 11, 12, and 13 for LAB energies 50,100,200, and 350 MeV, respectively, with their corresponding errors. For comparison we also show the same quantities calculated with the

[16]The many digits are provided to guarantee numerical reproducibility of results, since we find strong correlations among the parameters. We thank Eduardo Garrido numerical checks.

1993 high-quality Reid93 [15], NijmII [15], and AV18 [16] potentials.

V. CONCLUSIONS AND OUTLOOK

We summarize our main points. The determination of uncertainties in theoretical nuclear physics is one of the most urgent issues to be solved in order to establish the predictive power of *ab initio* nuclear structure calculations. One certain source for these uncertainties is the errors of the phenomenological *NN* interaction stemming from the finite accuracy of experimental scattering data as well as local scarcity in certain regions of the (θ, E) plane and an abundance bias in some other regions. Any statistical analysis of this sort *assumes* a model both for the signal and the noise which can only be checked *a posteriori*. In order to carry out such an analysis the lack of bias in the data and the model has to be established with a given confidence level. If normal errors on the data are assumed, the check can be made by applying normality tests to the residuals between the fitted model and the experimental data. We have used some classical tests and the highly demanding recently proposed tail-sensitive quantile-quantile test with a confidence level of 95%. Based on the outcome there is no serious reason to doubt on the normality of residuals of the 3σ self-consistent database obtained in our PWA of np- and pp-scattering data below pion production threshold.

We note that this normality test actually checks for the assumption, underlying any least-squares χ^2 fit, that the data themselves follow a normal distribution. With this fixed database one then can look for different representations of the potential which facilitate a straightforward implementation in any of the many available powerful methods which are currently available for solving the multinucleon problem.

We provide a user-friendly potential which consists of a short-range local part with 21-operators multiplying a linear superposition of Gaussian functions. The resulting fitted potential passes the normality tests satisfactorily and, hence, can be used to estimate statistical uncertainties stemming from *NN*-scattering data.

Our findings here seem to confirm a previous study of us when we compare the current OPE-Gauss potential including *statistical error bands* with previous potentials such as NijmII, Red93, or AV18 (without statistical bands); errors in the potential are dominated by the form of the potential rather than by the experimental data. Nonetheless, a thorough study of these kind of errors requires repeating the present analysis with an identical database with the *most general* potentials and functional forms and looking for discrepancies in the nuclear structure calculations outcome.

ACKNOWLEDGMENTS

We thank Elías Moreno for a statistician's point of view, Antonio Bueno for an experimentalist's point of view, and Lorenzo Luis Salcedo for an introduction to the Bayesian approach. We also thank Eduardo Garrido for numerical checks.

[1] R. Machleidt, Adv. Nucl. Phys. **19**, 189 (1989).
[2] W. Glöckle, *The Quantum Mechanical Few-Body Problem* (Springer, Berlin, 1983).
[3] L. Puzikov, R. Ryndin, and J. Smorodinsky, Nucl. Phys. **3**, 436 (1957).
[4] M. J. Evans and J. S. Rosenthal, *Probability and Statistics: The Science of Uncertainty* (Macmillan, New York, 2004).
[5] W. T. Eadie and F. James, *Statistical Methods in Experimental Physics* (World Scientific, Singapore, 2006).
[6] The Editors, Phys. Rev. A **83**, 040001 (2011).
[7] H. Stapp, T. Ypsilantis, and N. Metropolis, Phys. Rev. **105**, 302 (1957).
[8] R. Arndt and M. Macgregor, Methods Comput. Phys. **6**, 253 (1966).
[9] J. Perring, Nucl. Phys. **42**, 306 (1963).
[10] R. A. Arndt and M. H. MacGregor, Phys. Rev. **141**, 873 (1966).
[11] M. H. MacGregor, R. A. Arndt, and R. M. Wright, Phys. Rev. **169**, 1128 (1968).
[12] R. Arndt and L. Roper, Nucl. Phys. B **50**, 285 (1972).
[13] J. R. Bergervoet, P. C. van Campen, W. A. van der Sanden, and J. J. de Swart, Phys. Rev. C **38**, 15 (1988).
[14] V. Stoks, R. Timmermans, and J. J. de Swart, Phys. Rev. C **47**, 512 (1993).
[15] V. G. J. Stoks, R. A. M. Klomp, C. P. F. Terheggen, and J. J. de Swart, Phys. Rev. C **49**, 2950 (1994).
[16] R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla, Phys. Rev. C **51**, 38 (1995).
[17] R. Machleidt, Phys. Rev. C **63**, 024001 (2001).
[18] F. Gross and A. Stadler, Phys. Rev. C **78**, 014005 (2008).
[19] R. Navarro Perez, J. E. Amaro, and E. Ruiz Arriola, Phys. Lett. B **724**, 138 (2013).
[20] R. Navarro Perez, J. E. Amaro, and E. Ruiz Arriola, PoS **CD12**, 104 (2013).
[21] K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer, Berlin, 2011).
[22] R. Navarro Perez, J. E. Amaro, and E. Ruiz Arriola, arXiv:1202.6624.
[23] R. Navarro Perez, J. E. Amaro, and E. Ruiz Arriola, PoS **QNP2012**, 145 (2012).
[24] J. E. Amaro, R. Navarro Perez, and E. Ruiz Arriola, Few-Body Syst., doi:10.1007/s00601-013-0756-4.
[25] J. Toivanen, J. Dobaczewski, M. Kortelainen, and K. Mizuyama, Phys. Rev. C **78**, 034306 (2008).
[26] J. Dudek, B. Szpak, B. Fornal, and A. Dromard, Phys. Scr. **2013**, 014002 (2013).
[27] J. Dobaczewski, W. Nazarewicz, and P.-G. Reinhard, J. Phys. G **41**, 074001 (2014).
[28] A. Saltelli, S. Funtowicz, and P. Signell, Issues in Science and Technology, Fall **2013**, 79 (2013).
[29] J. B. Aviles, Phys. Rev. C **6**, 1467 (1972).
[30] R. Navarro Perez, J. E. Amaro, and E. Ruiz Arriola, Prog. Part. Nucl. Phys. **67**, 359 (2012).
[31] S. Aldor-Noiman, L. D. Brown, A. Buja, W. Rolke, and R. A. Stine, Am. Statist. **67**, 249 (2013).

[32] E. Matsinos, Phys. Rev. C **56**, 3014 (1997).
[33] R. Navarro Perez, J. E. Amaro, and E. Ruiz Arriola, Few-Body Syst., doi:10.1007/s00601-014-0817-3.
[34] Joint Committee for Guides in Metrology. Guide to the expression of uncertainty in measurement. JCGM 100-2008.
[35] R. T. Birge, Phys. Rev. **40**, 207 (1932).
[36] R. N. Kacker, A. Forbes, R. Kessel, and K.-D. Sommer, Metrologia **45**, 257 (2008).
[37] R. A. Arndt, L. D. Roper, R. A. Bryan, R. B. Clark, B. J. Ver West, and P. Signell, Phys. Rev. D **28**, 97 (1983).
[38] R. Navarro Perez, J. E. Amaro, and E. Ruiz Arriola, Phys. Rev. C **88**, 024002 (2013).
[39] R. Navarro Perez, J. E. Amaro, and E. Ruiz Arriola, Phys. Rev. C **88**, 064002 (2013).
[40] W. H. Press, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, Cambridge, 2007).
[41] V. G. J. Stoks, R. A. M. Kompl, M. C. M. Rentmeester, and J. J. de Swart, Phys. Rev. C **48**, 792 (1993).
[42] J. Beringer *et al.* (Particle Data Group), Phys. Rev. D **86**, 010001 (2012).
[43] O. Behnke, K. Kröninger, G. Schott, and T. Schörner-Sadenius, *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods* (John Wiley & Sons, New York, 2013).
[44] W. Conover, *Practical Nonparametric Statistics* (John Wiley & Sons, New York, 1980).
[45] A. N. Kolmogorov, Giornale dell Istituto Italiano degli Attuari **4**, 83 (1933).
[46] N. Smirnov, Ann. Math. Statist. **19**, 279 (1948).
[47] R. Von Mises, *Mathematical Theory of Probability and Statistics* (Academic Press, New York, 1964).
[48] R. Navarro Perez, J. E. Amaro, and E. Ruiz Arriola, Phys. Rev. C **89**, 024004 (2014).
[49] M. Srivastava and D. W. Sprung, in *Advances in Nuclear Physics* (Springer, Berlin, 1975), pp. 121–218.
[50] S. Aoki, T. Hatsuda, and N. Ishii, Prog. Theor. Phys. **123**, 89 (2010).
[51] S. Aoki (Sinya AOKI for HAL QCD Collaboration), Prog. Part. Nucl. Phys. **66**, 687 (2011).
[52] S. Aoki, Eur. Phys. J. A **49**, 81 (2013).
[53] N. Ishii *et al.* (HAL QCD Collaboration), Phys. Lett. B **712**, 437 (2012).
[54] A. Walker-Loud, PoS **LATTICE2013**, 013 (2014).
[55] C. V. D. Leun and C. Alderliesten, Nucl. Phys. A **380**, 261 (1982).
[56] I. Borbly, W. Grebler, V. Knig, P. A. Schmelzbach, and A. M. Mukhamedzhanov, Phys. Lett. B **160**, 17 (1985).
[57] N. L. Rodning and L. D. Knutson, Phys. Rev. C **41**, 898 (1990).
[58] S. Klarsfeld, J. Martorell, J. A. Oteo, M. Nishimura, and D. W. L. Sprung, Nucl. Phys. A **456**, 373 (1986).
[59] D. M. Bishop and L. M. Cheung, Phys. Rev. A **20**, 381 (1979).
[60] J. J. de Swart, C. P. F. Terheggen, and V. G. J. Stoks (1995), arXiv:nucl-th/9509032.