Nucleon-Nucleon potentials are commonplace 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 $\chi^2$ 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.

PACS numbers: 03.65.Nk, 11.10.Gh, 13.75.Cs, 21.30.Fe, 21.45.+v

Keywords: NN interaction, One Pion Exchange, Statistical Analysis

\section{INTRODUCTION}

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

\begin{equation}
M = a + m(\sigma_1 \cdot n)(\sigma_2 \cdot n) + (g - h)(\sigma_1 \cdot m)(\sigma_2 \cdot m) + (g + h)(\sigma_1 \cdot l)(\sigma_2 \cdot l) + c(\sigma_1 + \sigma_2) \cdot n \tag{1}
\end{equation}

where $a, m, g, h, c$ depend on energy and angle, $\sigma_1$ and $\sigma_2$ are the single-nucleon Pauli matrices, $l, m, n$ are three unitary orthogonal vectors along the directions of $k_f + k_i$, $k_f - k_i$ and $k_i \land k_f$, respectively, and $(k_f, k_i)$ 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 \cite{2}. Inversely, a complete set of experiments can be designed to reconstruct the amplitude at a given energy \cite{3}. The finite amount, precision and limited energy range of the data as well as the many different observables calls for a standard statistical $\chi^2$-fit analysis \cite{4,5}. This approach is subjected to assumptions and applicability conditions that can only be checked \textit{a posteriori} in order to guarantee the self-consistency of the analysis. Indeed, scattering experiments deal with counting poissonian statistics and for moderately large number of counts a normal distribution is expected. Thus, one hopes that a satisfactory theoretical description $O_{i}^{\text{th}}$ can predict a set of $N$ independent observed data $O_i$ given an experimental uncertainty $\Delta O_i$ as

\begin{equation}
O_i = O_{i}^{\text{th}} + \xi_i \Delta O_i \tag{2}
\end{equation}

with $i = 1, \ldots, N$ and $\xi_i$ 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 $\langle O_i \rangle = O_{i}^{\text{th}}$. 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 $\chi^2$ 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 \cite{1}.

The history of $\chi^2$ statistical analyzes of NN scattering data around pion production threshold started in the mid fifties \cite{7} (an account up to 1966 can be traced from Ref. \cite{8}). A modified $\chi^2$ method was introduced \cite{9} in order to include data without absolute normalization. The steady increase along the years in the number of scattering data and their precision gen-

\footnotesize

\textsuperscript{a}Supported by Spanish DGI (grant FIS2011-24149) and Junta de Andalucía (grant FQM225). R.N.P. is supported by a Mexican CONACYT grant.

\textsuperscript{1}Electronic address: rnavarrop@ugr.es

\textsuperscript{2}Electronic address: amaro@ugr.es

\textsuperscript{3}Electronic address: earriola@ugr.es

\textsuperscript{1} We note that in a Physical Review A editorial \cite{6} the importance of including error estimates in papers involving theoretical evaluations has been stressed.
erated mutually incompatible data and hence a rejection criterion was introduced allowing to discard inconsistent data. The upgrading of an ever increasing consistent database poses the question of normality, \( \chi^2 / \nu \lesssim 1 \) of a large number of selected data. The normality of the absolute value of residuals in \( pp \) scattering was scrutinized and satisfactorily fulfilled as a necessary consistency condition. The Niemegen group made 20 years ago an important breakthrough with first \( \chi^2 / \text{dof} \sim 1 \) and applying a \( 3\sigma \)-rejection criterion. This was possible after including em corrections, vacuum polarization, magnetic moments interaction and a charge-dependent (CD) One Pion Exchange (OPE) potential. With this fixed database further high quality potentials have been steadily generated 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 error-less. As a natural consequence the computational accuracy to a relative small percentage level has been a goal \( \text{per se} \) 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 \( \chi^2 \) 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 One-Pion-Exchange (OPE) for distances \( r \gtrsim 3 \text{fm} \). However, weaker constraints are put in the mid-range \( r \sim 1.5 - 2.5 \text{fm} \) region, which is precisely the relevant inter-particle 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 in mind we have upgraded the NN database to include all published \( \text{np} \) and \( \text{pp} \) 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 \footnote{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 \cite{23} and with the additional requirement that some interpolation of scattering data at non-measured 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.}. 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; a rough estimate suggested that NN uncertainties propagate into an unreasonably large uncertainty of \( AB / A \sim 0.1 - 0.5 \text{MeV} \), 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 \footnote{There is a growing concern on the theoretical determination of nuclear masses from nuclear mean field models with uncertainty evaluation \cite{25} (for a comprehensive discussion see e.g. \cite{26, 27}) echoing the need for uncertainty estimates in a Physical Review A editorial \cite{28} and the Saltelli-Funtowicz seven rules check-list \cite{28}.}. 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 \( \delta \)-shell potential suggested many years ago by Avilés \cite{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. \cite{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 \( 3\sigma \)-self consistent database in terms of a more conventional potential form containing the same 21-operators extending the AV18 as we did in \[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 \( 3\sigma \) self consistent experimental data base \( O_i^{\exp} \), with quoted uncertainty \( \Delta O_i^{\exp} \), \( i = 1, \ldots \ldots N = 6713 \) (total number of \( \text{pp} \) and \( \text{np} \) scattering data), our theoretical fits indeed satisfy that the residuals

\[
R_i = \frac{O_i^{\exp} - O_i^{\text{th}}}{\Delta O_i^{\exp}} \tag{3}
\]

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 \cite{31}. By comparing with others the TS test turns out to be the most demanding one concerning 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 \( \text{pp} \) analysis up to \( T_{\text{LAB}} = 30 \) and \( 350 \) \text{MeV}, is used for \( N = 389 \) and \( 1787 \) data, respectively, to test that the
squared residuals $R_i^2$ in Eq. 3 follow a $\chi^2$-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 $T_{lab} = 350$ MeV.

The paper is organized as follows. In Section II we review the assumptions, the rejecting and fitting process used in our previous works to build the $3\sigma$ self-consistent database, and expose the main motivation to carry out a normality test of the fit residuals. In Section 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 $3\sigma$ self-consistent database, providing a raison d’être for the rejection procedure. After that, in Section IV we analyze a fit to our conclusions and provide some outlook for further work.

II. STATISTICAL CONSIDERATIONS

There is a plethora of references on data and error analysis (see e.g. 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 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 $O_i^{\exp}$ experimentalists quote an estimate of the uncertainty $\Delta O_i^{\exp}$ so that the true value $O_i^{\true}$ is contained in the interval $O_i^{\exp} \pm \Delta O_i^{\exp}$ 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 $(O_i^{\exp}, \Delta O_i^{\exp})$ is provided without specifying the distribution we will assume an underlying normal distribution \(^4\), so that

$$P(O_i^{\exp}) = \exp \left[ -\frac{1}{2} \left( \frac{O_i^{\exp} - O_i^{\true}}{\Delta O_i^{\exp}} \right)^2 \right]$$

\(^4\) This may not be the most efficient unbiased estimator (see e.g. 2, 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 non-trivial.

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

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 $O_i^{\true} = F_i(\lambda_1^{\true}, \ldots, \lambda_P^{\true})$ 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 $O_i^{\exp}$, assuming that $(\lambda_1, \ldots, \lambda_P)$ are the true parameters, is given by

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

Thus the joined probability density is

$$P(O_1^{\exp} \ldots O_N^{\exp}|\lambda_1 \ldots \lambda_P) = \prod_{i=1}^N P(O_i^{\exp}|\lambda_1 \ldots \lambda_P)$$

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

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

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

$$\chi^2_{\min} = \min_{\lambda_i} \chi^2(\lambda_1, \ldots, \lambda_P) = \chi^2(\lambda_1, 0, \ldots, \lambda_P, 0)$$

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

$$O_i^{\th} = F_i(\lambda_1, 0, \ldots, \lambda_P, 0).$$

Expanding around the minimum one has

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

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

$$\delta_{ij}^{-1} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial \lambda_i \partial \lambda_j}(\lambda_1, 0, \ldots, \lambda_P)$$

Finally, the correlation matrix between two fitting parameters $\lambda_i$ and $\lambda_j$ is given by

$$\rho_{ij} = \frac{\delta_{ij}}{\sqrt{\delta_{ii}\delta_{jj}}}$$
We compute the error of the parameter $\lambda_i$ as
\[
\Delta \lambda_i \equiv \sqrt{\frac{\partial^2 \chi^2}{\partial \lambda_i^2}}. \tag{13}
\]

Error propagation of an observable $G = G(\lambda_1, \ldots, \lambda_P)$ is computed as
\[
(\Delta G)^2 = \sum_{ij} \frac{\partial G}{\partial \lambda_i} \frac{\partial G}{\partial \lambda_j} \bigg|_{\lambda_k = \lambda_{k0}} \varepsilon_{ij}. \tag{14}
\]

The resulting residuals of the fit are defined as
\[
R_i = \frac{O_i^{\text{exp}} - F_i(\lambda_1, \ldots, \lambda_P)}{\Delta O_i^{\text{exp}}}, \quad i = 1, \ldots, N. \tag{15}
\]

Assuming normality of residuals is now crucial for a statistical interpretation of the confidence level, since then $\sum R_i^2$ follows a $\chi^2$-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 large number of data $N$ with $P$-parameters one has, with a 16 or 68% confidence level, the mean value and most likely value nearly coincide, so that one has $\langle \chi^2 \rangle = N - P$ and thus as a random variable we have
\[
\frac{\chi^2_{\text{min}}}{\nu} = \frac{\sum R_i^2}{N - P} = 1 \pm \sqrt{\frac{2}{\nu}} \tag{16}
\]
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 $\chi^2$-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.

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 one, 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.

### 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. Note that the main purpose of the 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 $\sim 8000$ vs the small number of potential parameters $\sim 40$ which are expected to successfully account for the description of the data. 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. 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 other 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. If a consistent and maximal database is obtained by an iterative application of a rejection criterion, 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 or not may be easily

---

5 One can easily see that for a set of normally distributed data $R_i$, while $|R_i|$ does not follow that distribution, $|R_i|^2 \equiv R_i^2$ would pass a $\chi^2$-test.

6 In this case the merit figure to minimize would be
\[
S(\lambda_1, \ldots, \lambda_P) = -\sum_i \log P\left(\frac{O_i^{\text{exp}} - F_i(\lambda_1, \ldots, \lambda_P)}{\Delta O_i^{\text{exp}}}ight)
\]
For instance in Ref. [2], dealing with $\pi N$ 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 $\chi^2$ (conservative errors) indicate that the fitting parameters are indifferent to these data, whereas the large $\chi^2$ (daring errors) indicate an inconsistency with the rest of the data.
made by looking at non-overlapping error bands. This is frequently not the case; one has instead a set of neighboring data in the \((\theta, E)\) plane for a given observable or different observables at the same \((\theta, E)\) 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 1GeV see [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 \(V(r) \sim e^{-nmr}\) guarantees the appearance of a left hand branch cut at C.M. momentum \(p = im\pi n/2\). 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 can be decomposed as

\[
V(\vec{r}) = V_{\text{short}}(\vec{r})\theta(\vec{r}_c - \vec{r}) + V_{\text{long}}(\vec{r})\theta(\vec{r} - \vec{r}_c),
\]

(17)

where the short component can be written as

\[
V_{\text{short}}(\vec{r}) = \sum_{n=1}^{21} \hat{O}_n \left[ \sum_{i=1}^{N} V_{i,n} F_i(r) \right]
\]

(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 \(F_i(r)\) are some given radial functions. \(V_{\text{long}}(\vec{r})\) contains a Charge-Dependent (CD) One pion exchange (OPE) (with a common \(f^2 = 0.075\) [22–24]) and electromagnetic

---

\(^{10}\) For several measurements the Birge test [35] is the appropriate tool. The classical and Bayesian interpretation of this test has been discussed recently [36].
We use the Levenberg-Marquardt method where an approximation to the
previously, if old which we estimate as
Broglie wavelength corresponding to pion production thresh-
where
The minimal resolution
corresponds to the choice

Among the different fitting coefficients, \( V_{\text{imin}} \), so that some parameters become redundant or an over-
complete representation of the data, and the \( \chi^2 \) value will not
decrease substantially. In the opposite situation \( \Delta r_{\text{min}} \gg \Delta r_{\pi} \) the number of parameters increases but also
the correlations among the coefficients

\[ F_{i,n}(r) = \Delta r_i \delta(r - r_i) \]

where \( r_i \leq r_e \) are a discrete set of radii and \( \Delta r_i = r_{i+1} - r_i \).
The minimal resolution \( \Delta r_{\text{min}} \) is determined by the shortest de
Broglie wavelength corresponding to pion production thresh-
hold which we estimate as \( \Delta r_{\pi} \sim 0.6 \text{fm} \) so that the
needed number of parameters can be estimated \emph{a priori}. Obviously,
if \( \Delta r_{\text{min}} \ll \Delta r_{\pi} \) the number of parameters increases but also
the correlations among the different fitting coefficients,
\( V_{\text{imin}} \), so that some parameters become redundant or an over-
complete representation of the data, and the \( \chi^2 \) value will not
decrease substantially. In the opposite situation \( \Delta r_{\text{min}} \gg \Delta r_{\pi} \) 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 satis-
factory, as expected.

F. The 3\( \sigma \) self-consistent database

After the fitting process we get the desired 3\( \sigma \) self-
consistent database using the idea proposed by Gross and
Stadler \cite{18} and worked at full length in our previous
work \cite{39}. This allows to rescue data which would otherwise have been discarded using the standard 3\( \sigma \) criterion conte-
minated in all previous analyzes \cite{13, 18, 11}. The situation is
illustrated in Fig. 1 (middle and right panels).

By using the rejection criterion at the 3\( \sigma \) level we cut-off the long tails and as a result a fair comparison could in prin-
inciple be made to this truncated gaussian distribution. The Ni-
jmegen group found that the moments method test (see below for more details) largely improved by using this truncated dis-
bistribution \cite{13}. It should be reminded, however, that the rejec-
tion criterion is applied to groups of datasets, and not to in-
dividual 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. Pre-
iminary runs show that the number of iterations grows and the convergence may be slowed down or non-converging 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 disen-
tanglement between statistical and systematic errors was not explicitly exploited. In both cases these data are useless to
propagate uncertainties invoking the standard statistical inter-
pretation, see Eq. \eqref{eq:chi2}.

G. Distribution of residuals

In Fig. 2 we present the resulting residuals, Eq. \eqref{eq:residual}, in a
normalized histogram for illustration purposes, in the case of the original full database and the 3\( \sigma \) consistent database, and compare them with a normal distribution function with the binning resolution \( \Delta R = 0.2 \). The complete database his-
togram shows an asymmetry or skewness as well as higher
tails and clearly deviates from the normal distribution, mean-
while the 3\( \sigma \) consistent database residuals exhibits a closer
agreement with the Gaussian distribution. Note that this per-
cession from the figure is somewhat depending on eyeball comparison of the three situations. We will discuss more
preferable tests in the next section which are independent on this binning choice.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{(Color online) Normalized histogram of the resulting resid-
uals after fitting the potential parameters to the complete pp and np
data base (blue boxes with solid borders) and to the 3\( \sigma \) consistent
data base (red boxes with dashed borders). The N(0, 1) standard nor-
mal probability distribution function (green solid line) is plotted for
comparison.}
\end{figure}

A handy way of checking for the normality of the resid-
uals is looking into the standardized moments \cite{4}. These are
defined as

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

where \( \mu \) is the arithmetic mean and \( \sigma \) the standard deviation;
the \( r = 1 \) and \( r = 2 \) standardized moments are zero and one

\begin{footnotesize}
\begin{itemize}
\item We use the Levenberg-Marquardt method where an approximation to the
Hessian is computed explicitly \cite{44} which we keep throughout.
\end{itemize}
\end{footnotesize}
This rescaling is a common practice when errors on the fitted quantities are not provided; uncertainties are invented with the condition that indeed 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 non-standardized normals. That means that they would correspond to a scaled gauss 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 ΔΩi. For N ≫ P one has ν ∼ N and one has χ2/ν = 1 ± 2/√N = 1 ± 0.016 for N = 8000. Our fit to the complete database yields χ2/ν = 1.4 which is well beyond the confidence level. Rescaling in this case would correspond to globally enlarge the errors by √1.4 ∼ 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 χ2/ν = 1.2 one would globally enlarge the errors by √1.1 ∼ 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. Thus, while χ2/ν = 1 ± 2/√ν 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.

\[ \chi^2 = \left( \frac{\chi^2}{\chi^2_{\text{min}}} \right) \nu \]  

(22)
TABLE II: Results of the Pearson normality test of the residuals obtained by fitting the complete data base with a delta-shell potential and the $3\sigma$ consistent database with the delta-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 | $N$     | $T_c$   | $T_{\text{obs}}$ | $p$-value |
|------------|-----------|---------|---------|------------------|------------|
| Complete   | OPE-DS    | 8125    | 93.945  | 598.84           | 1.36 x 10^{-8} |
|            |           |         |         | 190.16           | 2.18 x 10^{-12} |
| $3\sigma$  | OPE-DS    | 6713    | 87.108  | 82.67            | 0.09       |
|            |           |         |         | 69.08            | 0.40       |
| $3\sigma$  | $\chi$TPE-DS | 6712   | 87.108  | 100.70           | 0.004      |
|            |           |         |         | 74.40            | 0.25       |
| $3\sigma$  | OPE-G     | 6711    | 87.108  | 84.17            | 0.08       |
|            |           |         |         | 68.38            | 0.43       |

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. [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 different 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 $T_{\text{obs}}$ 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 $\alpha$, which is the maximum probability of rejecting a true null hypothesis. Typical values of $\alpha$ 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 $H_0$ is that 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

$$T = \sum_{i=1}^{N_0} \frac{(n_i - n_i^{\text{normal}})^2}{n_i}$$

where $n_i^{\text{fit}}$ are the number of residuals on each bin and $n_i^{\text{normal}}$ are the number of expected residuals for the normal distribution in the same bin. $T$ follows a $\chi^2$-distribution with $N_0 - 1$ d.o.f. The decision on how close is a given histogram 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 that $\Delta R_i$ is such that $n_i^{\text{normal}}$ 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 data base fails the test even when residuals are scaled.

B. Kolmogorov-Smirnov-Test

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

$$S(x) = \frac{1}{N} \sum_{j=1}^{N} \theta(x - X_j),$$

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

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

Some of the advantages of using $T_{\text{KS}}$ 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 $T_{\text{KS}}$ indicate large deviations from the theoretical distribution the decision rule will be to reject the null hypothesis if the observed value $T_{\text{obs,KS}}$ is larger than a certain critical value $T_c$. The critical value depends on $\alpha$ and $N$; for large number of data and a significance level of 0.05 $T_c$ is $\chi^2/\sqrt{N}$. Also, a good approximation for the corresponding $p$-value has been given [47]

$$P_{\text{KS}}(T_{\text{obs}}) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2\left((\sqrt{N}+0.12+0.11/\sqrt{N})/T_{\text{obs}}\right)^2}.$$ 

The results of the KS normality test to the residuals obtained by fitting the potential parameters to the complete and $3\sigma$ 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 to the 100%. In contrast the observed test statistic using the $3\sigma$ 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 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

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt = 0.25, 0.5, 0.75.$$  \hspace{1cm} (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 data 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, 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 parts of the graph. A third samples comes from the $N(-1, 1)$ distribution and this can be seen as a downward shift of the points. A last sample is taken from the exponential distribution $E(1.5)$ which is asymmetric and positive.

Fig. 4 shows the QQ plot of the residuals from the fit to the $3\sigma$ 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.
FIG. 5: (Color online) Rotated Quantile-Quantile plot of the residuals obtained (blue points) from fitting the complete database with the OPE-Delta-Shell potential (upper left panel), the 3σ self-consistent database fitted with the OPE-Delta-Shell potential (upper right panel), the χTPE-Delta-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.

TABLE IV: Same as table II for the tail sensitive test

| Database       | Potential | N    | $T_S$ | $T_{obs}$ | p-value |
|----------------|-----------|------|-------|-----------|---------|
| Complete       | OPE-DS    | 8125 | 0.00070 | 0.00000 | < 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    |

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 − α) confidence bands to the QQ plot to quantitatively test deviations from the normal distribution. This new test, called tail sensitive (TS), 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 [31]. We will restrict ourselves to point out that, a small value of $T_{TS}$ indicates discrepancies between the empirical and normal distribution and therefore the rejection criterion for the null hypothesis is $T_{obs,TS} < T_{c,TS}$.

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 $T_{MC,obs}$ for each sample to obtain the distribution of $T_{TS}$ under the null hypothesis. The critical value for a significance level $α = 0.05$ corresponds to the $T_{obs,TS}$ that is greater than 5% of all the val-

---

14 It should also be noted that a typo in Ref. [31] is made in their steps 1c and 1e where $Φ^{-1}$ and $B_{(i,n+1-1)}$ are printed instead of $Φ$ and $B_{(i,n+1-1)}$; the latter are consistent with the rest of the text and the results presented there.
values calculated. Finally the test statistic for the empirical data \( \tau_{\text{emp}} \) can be calculated and compared to the simulated distribution to obtain the \( p \)-value. In this case the \( p \)-value is the proportion of \( \tau_{\text{MC,TS}} \) that are smaller than \( \tau_{\text{emp}} \). Since the observed \( T_{\text{TS}} \) for the complete database residuals is numerically equal to zero and smaller than all of the simulated 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 from the normal distribution which is reflected on the extremely low \( p \)-values. The 3\( \sigma \) consistent data residuals (upper right panel) show deviations from the normal distribution that are always within the TS confidence bands and therefore a to a confidence level \( \alpha = 0.05 \) there are no statistically significant differences to reject the null hypothesis.

E. Discussion

We have shown in the previous discussion evidence supporting the validity of Eq. (2) for the 3\( \sigma \)-self consistent database recently built from all published np and pp scattering data since 1950 till 2013 \cite{30,33}. The numerics can be a costly procedure since multiple optimizations must be carried out 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 delta-shells interaction already proposed by Avilés \cite{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 \( \chi \)TPE potential \cite{48} which also passes the normality test as can be seen from Fig. 5 and Tables I and II. 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 \( \delta \)-shell potentials cannot be directly implemented in some of the many powerful computational approaches to nuclear structure calculations \cite{15}.

The necessary conditions for a sensible interpretation of the \( \chi^2 \) 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 \( 3\sigma \) 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  

\footnote{The \( \delta \)-shell potential cannot even be plotted, which may naively seem a disadvantage. However, its Fourier transformation is smooth \cite{39} in the relevant CM momentum region of \( p_{CM} \leq 2\text{fm}^{-1} \), complying to the idea that coarse graining down to \( \Delta r \sim 0.6\text{fm} \) resolutions lacks information on shorter length scales.}

\( \text{FIG. 6: (Color online)} \) Correlation matrix \( C_{ij} \) for the short distance parameters in the partial wave basis \( (|l|,|j|) \), see Eq. (18). We show the OPE-DS (upper panel) and the \( \chi \)TPE-DS (middle panel) potentials. The points \( r_{ij} = \Delta r_p(i + 1) \) are grouped within every partial wave. The ordering of parameters is as in the parameter tables in Refs. \cite{38,39} and \cite{48} for OPE-DS 46 parameters and the \( \chi \)TPE 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, \( C_{ij} = 1 \) (red), 0% correlation, \( C_{ij} = 0 \) (yellow) and 100% anti-correlation, \( C_{ij} = -1 \) (blue).
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 $\delta$-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 Sect. 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 Ref. [22–24] for the previously developed high-quality 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) vs the large body of data, $O_{\chi}^{\text{th}} = O_{\chi}^{\text{QCD}}$ with just two parameters in the $(u,d)$ sector, $\Lambda_{\text {QCD}}$ and the quark masses $(m_u, m_d)$, or equivalently with the pion weak decay constant $f_\pi$ 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 $m_{\pi} = 701\text{MeV}$ with a shape similar to our OPE-Gaussian potential (see Sect. IV) but a depth of $-30\text{MeV}$ in the central component $V_c$ and $\Delta V_c \sim 5\text{MeV}$ for $r \gtrsim 1\text{fm}$, 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. [54]). Of course, since the lattice NN potential depends ultimately in just two parameters, $\Lambda_{\text {QCD}}$ and $m_\chi$ the different r-values in the potential functions $V_\chi(r)$ 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 be 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 $\chi$TPE-DS [48] short distance parameters in the partial wave basis $(l,f,s,l',f')$, see Eq. (13). The listing ordering is the same as the one in the parameter tables in Refs. [38, 39] and [48] for OPE-DS and $\chi$TPE-DS respectively. Note, the isolated pattern of correlations for the OPE-DS case, however as we see there are substantial correlations among different $(l,f,s,l',f')$ 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 $\chi$TPE-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.

FIG. 7: (Color online) Lowest np and pp partial waves potentials (in MeV) and their errors (solid band) as a function of the inter-nucleon 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 inter-nucleon 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 inter-nucleon 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 inter-nucleon distance $r$ (in fm).

IV. THE OPE-GAUSSIAN POTENTIAL

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

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

(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 we get $\chi^2/\nu = 1.06$. The resulting 42 fitting parameters (41 independent partial wave coefficients $V_{ij,l',l}^{JS}$ 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 $V_{i,j,l}^{JS}$ 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 $3\sigma$ 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

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.
TABLE V: Fitting partial wave parameters \( V_j^{JS} \) (in MeV) with their errors for all states in the \( Js \) channel. \( \tau \) indicates that the corresponding fitting \( V_j^{JS} = 0 \). The parameters marked with * 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 \( 2.3035 \pm 0.0133 fm \) is found.

| Wave | \( V_1 \) | \( V_2 \) | \( V_3 \) | \( V_4 \) |
|------|----------|----------|----------|----------|
| \( ^1S_{0np} \) | -67.3773 | 59.4930 | -2844.7118 | 3364.9823 |
| | \(+4.8885\) | \(+64.8759\) | \(+245.3275\) | \(+268.9192\) |
| \( ^1S_{0pp} \) | -52.0676 | 408.7926 | -2263.1470 | 2891.2494 |
| | \(+1.1057\) | \(+12.9206\) | \(+57.0254\) | \(+76.3709\) |
| \( ^2P_0 \) | -60.3589 | -520.5645 | - | - |
| | \(+1.2182\) | | | |
| \( ^1P_1 \) | 22.8758 | -256.2909 | - | - |
| | \(+0.9182\) | | | |
| \( ^3P_1 \) | 35.6383 | -229.1500 | 928.1717 | - |
| | \(+0.9194\) | \(+9.0104\) | \(+28.8275\) | - |
| \( ^3S_1 \) | -42.4005 | 273.1651 | -1487.4693 | 2064.7996 |
| | \(+2.1344\) | \(+24.1462\) | \(+91.3195\) | \(+105.4383\) |
| \( ^1\tau \) | -121.8301 | 262.7957 | -1359.3473 | 1218.3817* |
| | \(+3.2630\) | \(+19.0432\) | \(+50.9369\) | \(+34.8398\) |
| \( ^3D_1 \) | 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 | - |
| | \(+0.3052\) | \(+7.3280\) | \(+13.2310\) | - |
| \( \epsilon_2 \) | 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 | - | - |
| | \(+0.8870\) | \(+4.3677\) | | |

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 Ref. [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 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 \textit{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 \((\theta, E)\) 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 $3\sigma$ self-consistent database obtained in our PWA of np and pp scattering data below pion production threshold. 

We remind that this normality test actually checks for the assumption, underlying any least squares $\chi^2$ fit, that the data themselves follow a normal distribution. With this fixed database one can then 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 multi-nucleon 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. 

We thank Elías Moreno for a statistician’s point of view, Antonio Bueno for disclosing the experimentalist’s feelings and Lorenzo Luis Salcedo for an introduction to the Bayesian approach. We also thank Eduardo Garrido for numerical checks.
TABLE IX: np isovector phaseshifts.

| $E_{LAB}$ | $\delta_{N_1}$ | $\delta_{D_2}$ | $\delta_{G_4}$ | $\delta_{P_6}$ | $\delta_{P_7}$ | $\delta_{F_7}$ | $\delta_{F_8}$ | $\delta_{J_2}$ | $\delta_{J_3}$ | $\delta_{J_4}$ |
|-----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 1         | ±0.018         | ±0.000         | ±0.000         | ±0.180         | ±0.108         | ±0.000         | ±0.000         | ±0.000         | ±0.000         | ±0.000         |
| 5         | ±0.045         | ±0.000         | ±0.000         | ±0.275         | ±0.253         | ±0.107         | ±0.000         | ±0.000         | ±0.000         | ±0.000         |
| 10        | ±0.065         | ±0.000         | ±0.000         | ±0.850         | ±0.706         | ±0.170         | ±0.000         | ±0.000         | ±0.000         | ±0.000         |
| 25        | ±0.143         | ±0.000         | ±0.000         | ±0.787         | ±0.658         | ±0.107         | ±0.000         | ±0.000         | ±0.000         | ±0.000         |
| 50        | ±0.167         | ±0.000         | ±0.000         | ±0.935         | ±0.823         | ±0.209         | ±0.000         | ±0.000         | ±0.000         | ±0.000         |
| 100       | ±0.329         | ±0.000         | ±0.000         | ±0.488         | ±0.435         | ±0.120         | ±0.000         | ±0.000         | ±0.000         | ±0.000         |
| 150       | ±0.379         | ±0.000         | ±0.000         | ±0.540         | ±0.487         | ±0.170         | ±0.000         | ±0.000         | ±0.000         | ±0.000         |
| 250       | ±0.560         | ±0.000         | ±0.000         | ±0.640         | ±0.595         | ±0.200         | ±0.000         | ±0.000         | ±0.000         | ±0.000         |
| 300       | ±0.553         | ±0.000         | ±0.000         | ±0.640         | ±0.595         | ±0.200         | ±0.000         | ±0.000         | ±0.000         | ±0.000         |
| 350       | ±0.547         | ±0.000         | ±0.000         | ±0.640         | ±0.595         | ±0.200         | ±0.000         | ±0.000         | ±0.000         | ±0.000         |

TABLE X: np isoscalar phaseshifts.

| $E_{LAB}$ | $\delta_{P_1}$ | $\delta_{F_1}$ | $\delta_{D_2}$ | $\delta_{G_4}$ | $\delta_{S_1}$ | $\delta_{D_1}$ | $\delta_{D_3}$ | $\delta_{J_3}$ | $\delta_{J_3}$ |
|-----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 1         | −0.186         | −0.000         | 0.000          | 147.624        | 1.102          | −0.005         | 0.000          | 0.000          | 0.000          |
| 5         | −1.493         | −0.010         | 0.218          | 117.905        | 0.638          | −0.177          | 0.007          | 0.080          | 0.000          |
| 10        | −3.058         | −0.064         | 0.843          | 102.230        | 1.086          | −0.661          | 0.007          | 0.080          | 0.000          |
| 25        | −6.337         | −0.421         | 3.698          | 80.068         | 1.653          | −2.735          | 0.058          | 0.552          | 0.000          |
| 50        | −9.603         | −1.143         | 8.951          | 62.105         | 1.955          | −6.276          | 0.376          | 1.609          | 0.000          |
| 100       | −14.089        | −2.291         | 17.299         | 31.792         | 2.181          | −11.922         | 1.599          | 3.451          | 0.000          |
| 150       | −17.844        | −3.102         | 22.164         | 30.269         | 2.980          | −16.143         | 2.830          | 4.700          | 0.000          |
| 200       | −21.036        | −3.775         | 24.449         | 30.605         | 2.089          | −19.526         | 3.690          | 5.536          | 0.000          |
| 250       | −23.623        | −4.421         | 25.137         | 30.269         | 2.089          | −23.329         | 4.222          | 6.150          | 0.000          |
| 300       | −25.653        | −5.078         | 24.920         | 6.764          | 6.681          | −24.681         | 4.578          | 6.648          | 0.000          |
| 350       | −27.256        | −5.734         | 24.242         | 8.712          | 1.030          | −26.586         | 4.876          | 7.067          | 0.000          |
| 400       | −30.656        | −6.390         | 23.462         | 8.712          | 1.030          | −26.586         | 4.876          | 7.067          | 0.000          |

[1] R. Machleidt, Adv.Nucl.Phys. 19, 189 (1989).
[2] W. Glöckle, The quantum mechanical few-body problem (Springer Berlin, 1983).
[3] L. Puznik, R. Ryndin, and J. Smorodinsky, Nuclear Physics 3, 436 (1957).
[4] M. J. Evans and J. S. Rosenthal, Probability and statistics: The science of uncertainty (Macmillan, 2004).
[5] W. T. Eadie and F. James, Statistical methods in experimental physics (World Scientific, 2006).
[6] T. Editors, Phys. Rev. A 83, 040001 (2011), URL http://link.aps.org/doi/10.1103/PhysRevA.83.040001.
[7] H. Stapp, T. Ypsilantis, and N. Metropolis, Phys.Rev. 105, 302 (1957).
[8] R. Arndt and M. Macgregor, Methods in Computational Physics 6, 253 (1966).
[9] J. Perring, Nuclear Physics 42, 306 (1963).
[10] R. A. Arndt and M. H. MacGregor, Physical Review 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. **B50**, 285 (1972).
[13] J. Bergervoet, P. van Campen, W. van der Sanden, and J. J. de Swart, Phys.Rev. **C38**, 15 (1988).
[14] V. G. Stoks, R. Timmermans, and J. de Swart, Phys.Rev. **C47**, 512 (1993), nucl-th/9211007.
[15] V. G. J. Stoks, R. A. M. Klomp, C. P. F. Terheggen, and J. J. de Swart, Phys. Rev. **C49**, 2950 (1994), nucl-th/9406039.
[16] R. B. Wiringa, V. Stoks, and R. Schiavilla, Phys.Rev. **C51**, 38 (1995), nucl-th/9408016.
[17] R. Machleidt, Phys.Rev. **C63**, 024001 (2001), nucl-th/0006014.
[18] F. Gross and A. Stadler, Phys.Rev. **C78**, 014005 (2008), 0802.1552.
[19] R. Navarro Perez, J. Amaro, and E. Ruiz Arriola (2012), 1202.6624.
[20] R. Navarro Perez, J. Amaro, and E. Ruiz Arriola, PoS **CD12**, 104 (2013), 1301.6949.
[21] K. Chadan and P. C. Sabatier, Inverse problems in quantum scattering theory (Springer Publishing Company, Incorporated, 2011).
[22] R. Navarro Perez, J. Amaro, and E. Ruiz Arriola (2012), 1202.6624.
[23] R. Navarro Perez, J. Amaro, and E. Ruiz Arriola, PoS **QNP2012**, 145 (2012), 1206.3508.
[24] J. Amaro, R. N. Perez, and E. R. Arriola (2013), 1310.4756.
[25] J. Toivanen, J. Dobaczewski, M. Kortelainen, and K. Mizuyama, Phys.Rev. **C78**, 034306 (2008), 0806.1914.
[26] J. Dudek, B. Szpak, B. Fornal, and A. Dromard, Physica Scripta **2013**, 014002 (2013).
[27] J. Dobaczewski, W. Nazarewicz, and P. G. Reinhard (2014), 1402.4657.
[28] A. Saltelli and S. Funtowicz, Issues in Science and Technology, Fall **2013**, 79 (2013).
[29] J. B. Aviles, Phys. Rev. **C6**, 1467 (1972).
[30] R. Navarro Perez, J. Amaro, and E. Ruiz Arriola, Prog.Part.Nucl.Phys. **67**, 359 (2012), 1111.4328.
[31] S. Aldor-Noiman, L. D. Brown, A. Buja, W. Rolke, and R. A. Stine, Amer. Statist. **67**, 249 (2013).
[32] E. Matsinos, Phys.Rev. **C56**, 3014 (1997).
[33] J. Birge, Physical Review **40**, 207 (1932).
[34] R. N. Kacker, A. Forbes, R. Kessel, and K.-D. Sommer, Metrologia **45**, 257 (2008).
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).
\( T_{\text{LAB}} = 50 \text{ MeV} \)

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

\( T_{\text{LAB}} = 100 \text{ MeV} \)

FIG. 11: (Color on-line) Same as in Fig. 10 but for \( E_{\text{LAB}} = 100 \text{MeV} \).
$T_{\text{LAB}} = 200 \text{ MeV}$

FIG. 12: (Color on-line) Same as in Fig. 10 but for $E_{\text{LAB}} = 200 \text{MeV}$.

$T_{\text{LAB}} = 350 \text{ MeV}$

FIG. 13: (Color on-line) Same as in Fig. 10 but for $E_{\text{LAB}} = 350 \text{MeV}$.