Robust extraction of proton charge radius from electron-proton scattering data

Xuefei Yan,[1,2,*] Douglas W. Higinbotham,[3] Dipangkar Dutta,[4] Haiyan Gao,[1,2,5] Ashot Gasparian,[6] Mahbub A. Khandaker,[7] Nilanga Liyanage,[8] Eugene Pasyuk,[3] Chao Peng,[1,2] and Weizhi Xiong[1,2]

[1] *Duke University, Durham, North Carolina 27708, USA*
[2] *Triangle Universities Nuclear Laboratory, Durham, North Carolina 27708, USA*
[3] *Thomas Jefferson National Accelerator Facility, 12000 Jefferson Avenue, Newport News, Virginia 23606, USA*
[4] *Mississippi State University, Mississippi State 39762, USA*
[5] *Duke Kunshan University, Jiangsu 215316, China*
[6] *North Carolina A&T State University, Greensboro, North Carolina 27411, USA*
[7] *Idaho State University, Idaho 83209, USA*
[8] *University of Virginia, Charlottesville, VA 22904, USA*

(Dated: February 27, 2018)

Background: Extracting the proton charge radius from electron scattering data requires determining the slope of the charge form factor at Q^2 of zero. But as experimental data never reach that limit, numerous methods for making the extraction have been proposed, though often the functions are determined after seeing the data which can lead to confirmation bias.

Purpose: To find functional forms that will allow for a robust extraction of the input radius for a wide variety of functional forms in order to have confidence in the extraction from upcoming low Q^2 experimental data such as the Jefferson Lab PRad experiment.

Method: We create a general framework for inputting form-factor functions as well as various fitting functions. The input form factors are used to generate pseudo-data with fluctuations intended to mimic the binning and random uncertainty of a given set of real data. All combinations of input functions and fit functions can then be tested repeatedly against regenerated pseudo-data. Since the input radius is known, this allows us to find fit functions that are robust for radius extractions in an objective fashion.

Results: For the range and uncertainty of the PRad data, we find that a two-parameter rational function, a two-parameter continued fraction and the second order polynomial expansion of z can extract the input radius regardless of the input charge form factor function that is used.

Conclusions: We have created an easily expandable framework to search for functional forms that allow for a robust extraction of the radius from a given binning and uncertainty of pseudo-data generated from a wide variety of trial functions. This method has enabled a successful search for the best functional forms to extract the radius from the upcoming PRad data, and can be used for other experiments.

* xy33@phy.duke.edu

I. INTRODUCTION

A lot of efforts have been devoted to the measurement of the charge radius of the proton (R), but results from different experiments and/or analyses exhibit sizable differences. In high-precision muonic hydrogen Lamb shift experiments, R was measured to be 0.8409 ± 0.0004 fm [1, 2], while the current value from CODATA, determined from regular atomic Lamb shift and electron-proton (ep) scattering experiments, is $R = 0.8751 \pm 0.0061$ fm [3], though this result does not yet reflect the latest published electron scattering or regular atomic Lamb shift results [4, 5]. This difference has been known as the proton radius puzzle [6–8].

To extract the proton radius R from the ep-scattering data, one fits the electric form factor (G_E) as a function of the squared four-momentum transfer to the system (Q^2), and determines the slope of the G_E function at Q^2 of zero. The relation between the slope and the radius is defined to be

$$R \equiv \left(-6 \left. \frac{\mathrm{d}G_E(Q^2)}{\mathrm{d}Q^2} \right|_{Q^2=0} \right)^{1/2} . \tag{1}$$

Unfortunately, experimental electron scattering cannot reach the $Q^2 = 0$ limit; thus, many different methods have been proposed for the extraction of the radius from the data.

Some recent global analyses of ep-scattering data found $R \approx 0.84$ fm, in agreement with the muonic Lamb shift results [9–13]. Though using the same experimental data, these analyses extract systematically smaller radius than the results of other groups [14–20]. It has been pointed out that the difference between the results comes mainly due to differences in how the high-order moments $\langle r^{2n} \rangle$ ($n > 1$) are taken into account [21]. A summary table of the higher order moments from a number of these fits can be found in the recent work of Alarcón and Weiss [22].

In the study of Kraus *et al.* [23], it was shown that when using low-order polynomial expansion of Q^2 to fit pseudo-data generated with known R, the value from the fit, R(fit), is systematically smaller than the input value, R(input), though they also showed that the variance from the low order fits can be significantly smaller than the unbiased higher order fits. Sick *et al.* [24] also showed that when high-order moments $\langle r^{2n} \rangle$ ($n > 1$) were treated differently, significantly different R values were extracted from the world data of ep scattering.

Describing the form-factor G_E with a multi-parameter polynomial expansion of Q^2 up to an order Q^{2N} seems to be the natural choice for the fits, since each moment $\langle r^{2n} \rangle$ ($1 \leq n \leq N$) is related to an independent parameter. And though this description seems to be model independent, as Kraus *et al.* have shown, it does not ensure a correct R extraction from data when it is used for fitting [23].

Beyond multi-parameter polynomials, one can also use functional forms of G_E based on models of the proton charge distribution. The problem of this approach is that it can be difficult to quantify how much the R extraction is affected by the imperfectness of the assumed model. In addition, constructing a model description of the full charge distribution of the real proton is a far more complex problem than simply trying to extract only the "real" R value.

To resolve the mystery of fitting-function choice, we propose a robust method to find function(s) that can extract the "real" R from a broad set of input functions and input radii. The expected binning and uncertainty of the PRad experiment [25] will be taken as an example of applying this method.

II. METHOD

If a perfect functional form of G_E for the real proton were available, its parameters could be determined by fitting the function to the experimental data, and the "real" R could be extracted. This ideal case is simulated by examining the fitting results R(fit) using the same functional form as that used to generate the pseudo-data, and a simulated random statistical variation is added to the central G_E value. This process is repeated multiple times in order to obtain a distribution of R(fit).

However, a perfect functional form is not available, and one has to search for functional forms that can extract the "real" R from certain data sets with minimal dependence on the knowledge of the "real" functional form of G_E. For simplicity, we call this feature robustness.

In this study, a number of fittings based on different functional forms are applied to pseudo-data generated from various models and/or parameterizations. By studying the distributions of the fitting results, one can find whether a functional form is robust: it is robust if it correctly extracts the R value regardless of the parameterizations used to generate the pseudo-data.

A program library consisting of three parts (generator, fluctuation-adder and fitter) has been built to test the robust extraction of R [26]. This program library is coded in C++, using the packages of Minuit and CERN ROOT [27].

The three components of this library are described in detail in the following subsections.

A. Generator

The generator of this library has been built to generate G_E values at given Q^2 with the simple standard functions, parameterizations of experimental data, as well as full theoretical calculations. Other functions can easily be added to this library, The currently installed functions include:

a. Dipole The dipole functional form of G_E [28] is expressed as

$$G_E(Q^2) = \left(1 + \frac{Q^2}{p_1}\right)^{-2}, \tag{2}$$

where $p_1 = 12/R^2$ is a parameter related to the radius R. This functional form corresponds to an exponential charge distribution of the proton, and the relation between moments is

$$\langle r^{2n}\rangle = \frac{(n+1)(2n+1)}{6}\langle r^2\rangle\langle r^{2n-2}\rangle, \tag{3}$$

where $n > 1$.

b. Monopole The monopole functional form of G_E [28] is expressed as

$$G_E(Q^2) = \left(1 + \frac{Q^2}{p_1}\right)^{-1}, \tag{4}$$

where $p_1 = 6/R^2$. This functional form corresponds to a Yukawa charge distribution of the proton, and the relation between moments is

$$\langle r^{2n}\rangle = \frac{n(2n+1)}{3}\langle r^2\rangle\langle r^{2n-2}\rangle, \tag{5}$$

where $n > 1$.

c. Gaussian The Gaussian functional form of G_E [28] is expressed as

$$G_E(Q^2) = \exp(-Q^2/p_1), \tag{6}$$

where $p_1 = 6/R^2$. This functional form corresponds to a Gaussian charge distribution of the proton, and the relation between moments is

$$\langle r^{2n}\rangle = \frac{2n+1}{3}\langle r^2\rangle\langle r^{2n-2}\rangle, \tag{7}$$

where $n > 1$.

d. Kelly-2004 The parameterization from Ref. [29] is expressed as

$$G_E(Q^2) = \frac{1 + a_1\tau}{1 + b_1\tau + b_2\tau^2 + b_3\tau^3}, \tag{8}$$

where $\tau = Q^2/4m_p^2$, and m_p is the proton mass. The parameters a_1, b_1, b_2 and b_3 can be found in Table I of Ref. [29]. The radius in this parameterization is $R = 0.8630$ fm.

e. Arrington-2004 The parameterization from Ref. [30] is expressed as

$$G_E(Q^2) = \left(1 + \sum_{i=1}^{N} p_{2i}Q^{2i}\right)^{-1}, \tag{9}$$

where parameters p_{2i} up to $i = 6$ can be found in Table I of Ref. [30]. The radius in this parameterization is $R = 0.8682$ fm.

f. Arrington-2007 The parameterization from Ref. [31] is a fifth-order continued-fraction (CF) expansion expressed as:

$$G_E(Q^2) = \frac{1}{1 + \frac{p_1Q^2}{1 + \frac{p_2Q^2}{1+\cdots}}}, \tag{10}$$

where the parameters p_i (index i from 1 to 5) can be found in Table I in Ref. [31]. The radius in this parameterization is $R = 0.8965$ fm.

g. Venkat-2011 The parameterization from Ref. [33] is expressed as

$$G_E(Q^2) = \frac{1 + a_1\tau + a_2\tau^2 + a_3\tau^3}{1 + b_1\tau + b_2\tau^2 + b_3\tau^3 + b_4\tau^4 + b_5\tau^5}, \tag{11}$$

where parameters a_i (i = 1, 2 and 3) and b_i (i = 1, 2, ..., 5) can be found in Table II of Ref. [33]. The radius in this parameterization is $R = 0.8779$ fm.

h. Bernauer-2014 This parameterization is a refit of the full final set from Ref. [20], and is very close to the result in Ref. [32]. It is expressed as a 10th-order polynomial expansion of Q^2:

$$G_E(Q^2) = 1 + \sum_{i=1}^{10} p_i Q^{2i}, \tag{12}$$

where the refitted parameters p_i are very close to the parameters in Sec. J of Ref. [32]. The radius in this parameterization is $R = 0.8868$ fm.

i. Alarcón-2017 As a fully realistic charge form factor, we used the model of Alarcón and Weiss [22, 34] referred to herein as Alarcón-2017. This model uses the recently developed method combining chiral effective field theory and dispersion analysis. Solely for the purpose of testing extraction techniques, the radius in the model was fixed to a series of values: 0.84 fm from muonic hydrogen, 0.875 fm from CODATA, and 0.85 fm as the central value from the range of radii allowed by the model. Unlike the other models, where a simple function could be programmed, here we use a fine table of charge values and then fit it with a cubic spline. The result of the fit can then be called in a similar manner to the other functions.

j. Ye-2018 The parameterization in Ref. [35] pre-fixed the radius to $R = 0.879$ fm, together with other manual constraints and fitted some global data to get the parameters. The parameterization and the values of the parameters can be found in Ref. [35] and its supplemental materials. This parameterization will be referred to as Ye-2018 in this study. The authors of Ref. [35] also provided a separate set of parameterization with a different pre-fixed radius, $R = 0.85$ fm. This re-fixed paramertization will be referred to as Ye-2018 (re-fix) in this study.

B. Fluctuation adder

The library allows adding bin-by-bin and overall fluctuations to the G_E vs. Q^2 tables, mimicking the real data from experiments. It includes fluctuations according to a user-defined random Gaussian distribution, $\mathcal{N}(\mu, \sigma^2)$.

In the bin-by-bin case, the uncertainty δG_E of each bin is defined by the user. The library sets $\mu = 0$ and $\sigma = \delta G_E$, and generates fluctuations according to $\mathcal{N}(\mu, \sigma^2)$ in each bin.

In the overall case, the user can manually set the values of μ and σ, and the library generates an overall scaling factor according to $\mathcal{N}(\mu, \sigma^2)$ for all the bins in a table.

A few other types of fluctuations (such as uniform and Breit-Wigner distributions) are also included in the library for other test purposes.

C. Fitter

This library uses the Minuit package of CERN ROOT to fit the G_E vs. Q^2 tables, with a number of functional forms as listed below.

a. Dipole The dipole fitter is expressed as

$$f_{\text{dipole}}(Q^2) = p_0 G_E(Q^2) = p_0 \left(1 + \frac{Q^2}{p_1}\right)^{-2}, \tag{13}$$

where p_0 is a floating normalization parameter, and p_1 is a fitting parameter related to the radius $R = \sqrt{12/p_1}$.

b. Monopole The monopole fitter is given by

$$f_{\text{monopole}}(Q^2) = p_0 G_E(Q^2) = p_0 \left(1 + \frac{Q^2}{p_1}\right)^{-1}, \tag{14}$$

and $R = \sqrt{6/p_1}$.

c. Gaussian The Gaussian fitter has the form

$$f_{\text{Gaussian}}(Q^2) = p_0 G_E(Q^2) = p_0 \exp(-Q^2/p_1), \tag{15}$$

and $R = \sqrt{6/p_1}$.

d. Multi-parameter polynomial-expansion of Q^2 The fitter of the multi-parameter polynomial-expansion of Q^2 is written as

$$f_{\text{polyQ}}(Q^2) = p_0 G_E(Q^2) = p_0 \left(1 + \sum_{i=1}^{N} p_i Q^{2i}\right), \tag{16}$$

where p_0 is a floating normalization parameter, p_1 is a fitting parameter related to the radius by $R = \sqrt{-6p_1}$, parameters for higher order terms (p_i with $i > 1$) are free fitting parameters, and N is defined by the user.

e. Multi-parameter rational-function of Q^2 The fitter of the multi-parameter rational-function of Q^2 is expressed as

$$f_{\text{rational}}(Q^2) = p_0 G_E(Q^2) = p_0 \frac{1 + \sum\limits_{i=1}^{N} p_i^{(a)} Q^{2i}}{1 + \sum\limits_{j=1}^{M} p_j^{(b)} Q^{2j}}, \tag{17}$$

where p_0 is a floating normalization parameter, $p_i^{(a)}$ and $p_j^{(b)}$ are free fitting parameters, and radius can be found as $R = \sqrt{6(p_1^{(b)} - p_1^{(a)})}$. The orders N and M are defined by the user.

f. CF expansion The CF expansion fitter is expressed as [24]

$$f_{\text{CF}}(Q^2) = p_0 G_E(Q^2) = p_0 \frac{1}{1 + \frac{p_1 Q^2}{1 + \frac{p_2 Q^2}{1 + \cdots}}}, \tag{18}$$

where p_0 is a floating normalization parameter, p_i ($i > 0$) are free fitting parameters, and $R = \sqrt{6p_1}$. The user can define the maximum i of the expansion.

g. Multi-parameter polynomial-expansion of z The z-transformation is expressed as [16]

$$z = \frac{\sqrt{T_c + Q^2} - \sqrt{T_c - T_0}}{\sqrt{T_c + Q^2} + \sqrt{T_c - T_0}}, \tag{19}$$

where $T_c = 4m_\pi^2$, m_π is set to be 140 MeV (close to the π^0 mass as in Ref. [16]), and T_0 is a free parameter representing the point mapping onto $z = 0$ (T_0 is set to 0 in this study). With the new variable z, G_E can be parameterized as

$$f_{\text{polyz}}(Q^2) = p_0 G_E(Q^2) = p_0 \left(1 + \sum_{i=1}^{N} p_i z^i\right), \tag{20}$$

where p_0 is a floating normalization parameter, p_1 is a fitting parameter related to the radius by $R = \sqrt{-3p_1/2T_c}$, p_i are free fitting parameters, and N is defined by the user.

III. TESTS

The tests of this method are performed using the bin centers of Q^2 and the bin-by-bin uncertainties of the PRad experiment [25]. Two sets of binning information [36] are used in the tests; only the expected statistical uncertainties from PRad are used, as the PRad analysis of the systematic uncertainties is not finished yet.

Fig. 1 presents the fits with the dipole fitter of pseudo-data generated with the dipole generator [R(input) = 0.85 fm] in the PRad binning. No fluctuation is added to the central values of G_E in the first panel. The fitting results with fluctuations (using the fluctuation-adder) are shown in the second and third panels. It is observed that when there is no fluctuation, the fit curve goes through all the pseudo-data points perfectly, and the input R value is obtained. However, when there are fluctuations, the results of the fit can differ from the input.

In the following subsections, the procedure of the tests are presented, followed by the test results. Because the test results using the two bin sets are very close, only the results from using one of them are presented in this paper.

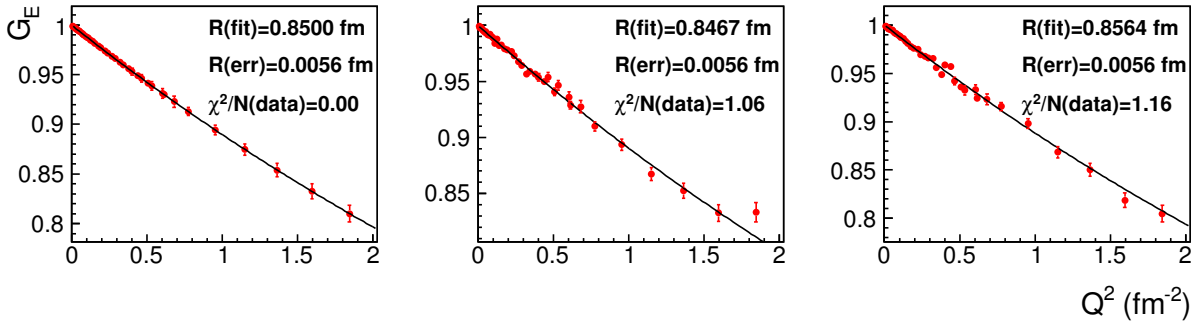

FIG. 1. (color online). Dipole fits of pseudo-data generated with dipole functional form with and without fluctuations [R(input) = 0.85 fm]. Pseudo-data in the leftmost panel has no fluctuation. The second and third panels show two different instances of fluctuations added to the central value of G_E. The fitting result [R(fit)], fitting uncertainty [R(err)] and χ^2 per data point [χ^2/N(data)] are presented in each panel.

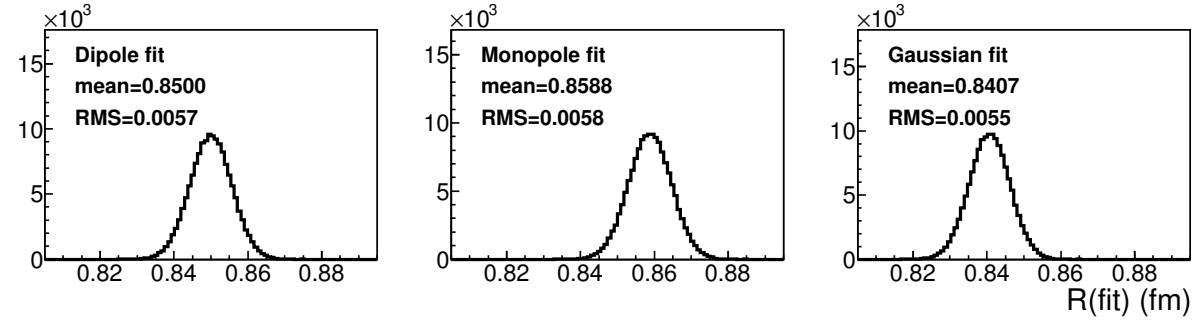

FIG. 2. (color online). Dipole, monopole and Gaussian fits of pseudo-data tables generated with the dipole functional form and added fluctuations [R(input) = 0.85 fm].

A. Procedure

The following procedure is carried out for the test.

a. Generation Firstly, one G_E model is used to generate pseudo-data (using the generator), at the bin centers of Q^2 that the user inputs into the program.

b. Fluctuation-adding Then, bin-by-bin and overall fluctuations are added to the G_E vs. Q^2 tables in a random manner (using the fluctuation-adder), to mimic the real data. The bin-by-bin uncertainties are taken from the bin-set file, and an overall scaling uncertainty of 5% (far larger than expected in the PRad result) is added in the tests to show that this method works even if there is such a big scaling uncertainty.

c. Fitting Finally, the G_E vs. Q^2 tables are fitted with a number of functional forms (using fitters) to extract R from the pseudo-data with fluctuations.

The steps of generation, fluctuation-adding and fitting are repeated for 150,000 times for each combination of generators and fitters. The 150,000 fitting results of R(fit) for each combination comprise a distribution with a central value R(mean) and a root-mean-square (RMS) width. Because the fitting uncertainty of R determined by Minuit (for each of the 150,000 fits) is very close to the RMS width of the R(fit) distribution, we will use the RMS values to represent the one-σ fitting-uncertainty.

B. Fits with strong model assumptions

The dipole, monopole and Gaussian functional forms imply strong model assumptions regarding the charge distribution of the proton.

Fig. 2 shows the R(fit) distributions of the dipole, monopole and Gaussian fits when the dipole generator is used [R(input) = 0.85 fm]. It is observed that when the dipole fitter is used, R(mean) $\approx R$(input), but when the monopole or Gaussian fitter is used, R(mean) significantly deviates from R(input).

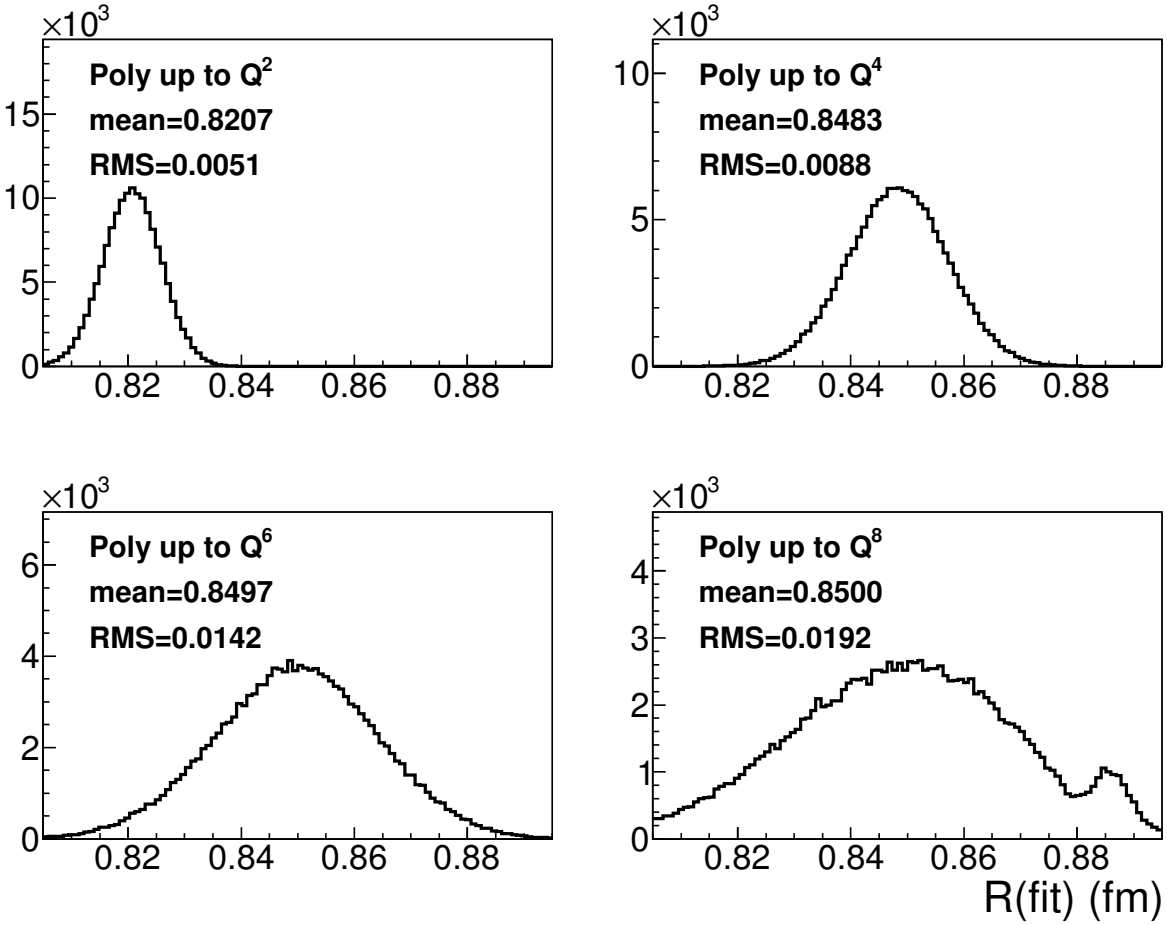

FIG. 3. (color online). Polynomial-expansion-of-Q^2 fits of pseudo-data tables generated with dipole functional form and fluctuations [R(input) = 0.85 fm].

Tables I, II and III summarize the fitting results using the dipole, monopole and Gaussian fitter, respectively, when different generators are used.

It is clear that these functional forms with strong model assumptions are not able to provide a robust extraction of R, because the result can be significantly affected by the imperfectness of these models' descriptions of the proton charge distribution.

C. Fits with polynomial expansions of Q^2

Fitting G_E vs. Q^2 tables from data with polynomial expansions has been widely carried out, but also criticized [10–12, 21, 23].

Fig. 3 shows the R(fit) spectra of the polynomial-expansion fits up to N = 1, 2, 3 and 4 [as in Eq. (16)] when the dipole generator is used [R(input) = 0.85 fm].

Tables IV, V, VI and VII summarize the fitting results using the polynomial-expansion fitter with N = 1, 2, 3 and 4 respectively, when different generators are used.

It is observed that when the order of expansion is too low (N = 1), R(mean) is systematically and significantly smaller than R(input), for all the generators used in the tests. When higher and higher orders are included (N = 2, 3 and 4), R(mean) gets closer and closer to R(input), regardless of the type of generator. At the same time, as the number of parameters increases, the constraint on R(fit) in each fit decreases, and larger fitting uncertainties are obtained.

The optimal choice of N depends on the Q^2 range, distances between bin centers and uncertainty level in the data table. Some efforts have been taken to build an algorithm that automatically and systematically determines the

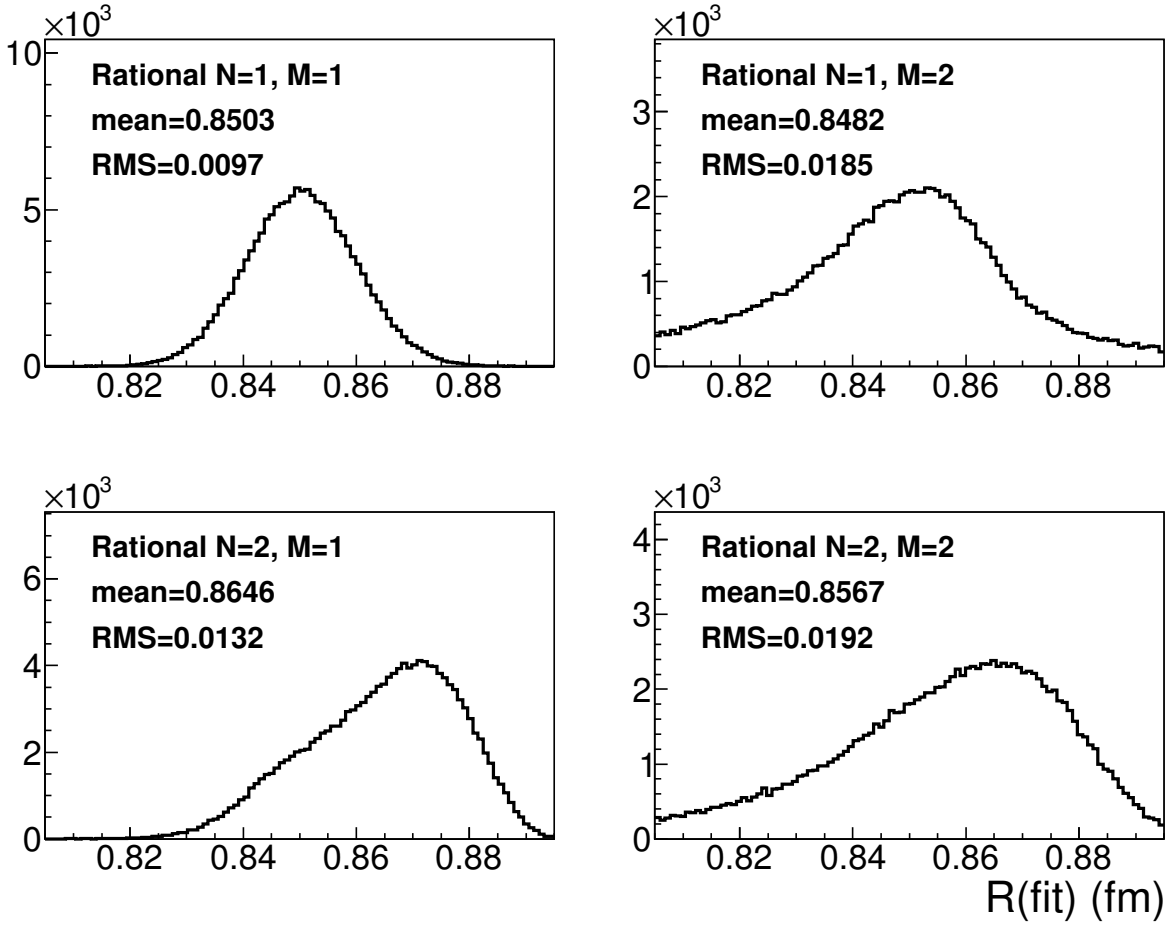

FIG. 4. (color online). Rational-function-of-Q^2 fits of pseudo-data tables generated with dipole functional form and fluctuations [R(input) = 0.85 fm].

proper order N [11], when certain data tables were fitted. Such an algorithm will be very useful, if it is tested to be successful. However, existing algorithms have not been proven successful yet. The trade-off between bias and variance is clearly observed in these tests: as N increases, δR decreases, and the uncertainty increases.

D. Fits with rational functions of Q^2

Fitting G_E vs. Q^2 tables from data with rational functions of Q^2 is also widely carried out, such as in Refs. [29, 30, 33]. The rational functions are also referred to as Padé approximations.

Fig. 4 shows the R(fit) spectra of the rational-function fits with (N, M) = (1, 1), (1, 2), (2, 1) and (2, 2) [as in Eq. (17)] when the dipole generator is used [R(input) = 0.85 fm]. It is observed that the best agreement between R(mean) and R(input) comes by setting (N, M) = (1, 1), which also provides the minimal RMS width of the R(fit) distribution.

Tables VIII, IX, X and XI summarize the fitting results using the rational-function fitter with (N, M) = (1, 1), (1, 2), (2, 1) and (2, 2) respectively, when different generators are used.

It is observed that, in these tests, the rational-function fitter ($N = 1$, $M = 1$) shows its ability to extract R robustly ($\delta R < 0.42\sigma$) regardless of the model parameterization in the generator. It also has the minimal fitting uncertainty among these four rational-function parametrizations.

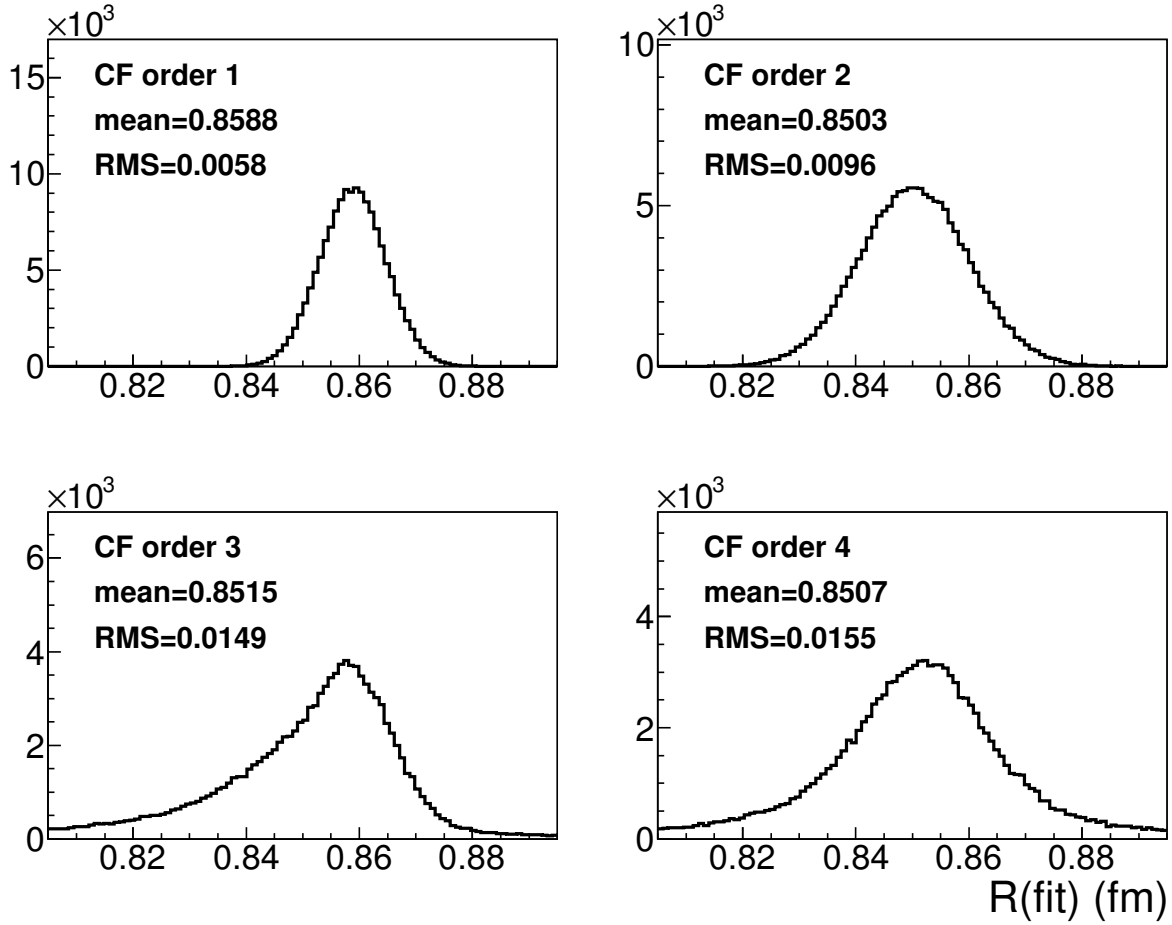

FIG. 5. (color online). CF fits of pseudo-data tables generated with dipole functional form and fluctuations [R(input) = 0.85 fm].

E. Fits with CF expansions of Q^2

Using CF expansion to fit G_E vs. Q^2 tables was proposed and applied to the world data by Sick *et al.* in 2003 [24], and it has been widely carried out since then. Ref. [24] also included broader tests and discussions regarding fitting pseudo and real data with CF expansions.

Fig. 5 shows the R(fit) spectra of the CF fits with order 1, 2, 3 and 4, when the dipole generator is used [R(input) = 0.85 fm]. It is observed that the best agreement between R(mean) and R(input) comes from the second-order CF-expansion.

Tables XII, XIII, XIV and XV summarize the fitting results using the CF fit at order 1, 2, 3 and 4, respectively, when different generators are used. In these tests (using PRad binning), CF at the second order seems to be the best: R(mean) $\approx$ R(input), regardless of the parameterizations in the generator, and the fitting uncertainties are reasonably small.

F. Fits with polynomial expansions of z

Using polynomial expansion of z instead of Q^2 is another option for R extraction [Eq. (19) needs to be used to transform Q^2 to z from the G_E vs. Q^2 tables].

Fig. 6 shows the R(fit) spectra of the fits with polynomial expansions of z [N = 1, 2, 3 and 4 as in Eq. (20)], when the dipole generator is used [R(input) = 0.85 fm]. It is observed that the agreements between R(mean) and R(input) are reasonably good except when $N = 1$.

Tables XVI, XVII, XVIII and XIX summarize the fitting results using the polynomial expansions of z with $N = 1$,

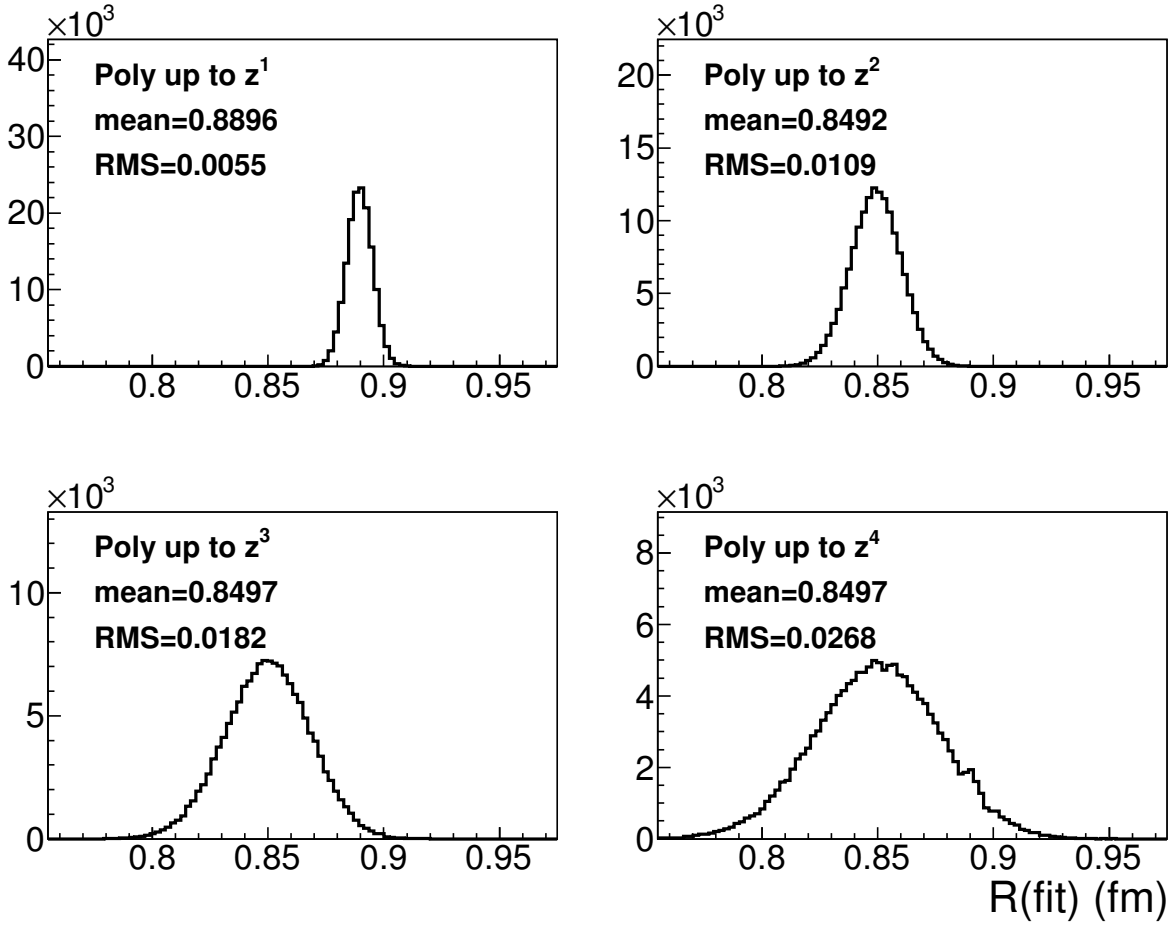

FIG. 6. (color online). Polynomial-expansions-of-z fits of pseudo-data tables generated with dipole functional form and fluctuations [R(input) = 0.85 fm].

2, 3 and 4, respectively, when different generators are used.

It is observed that when $N = 1$, R(mean) is systematically and significantly larger than R(input), for all the generators used in the tests. This is opposite to the study of polynomial expansion of Q^2 with $N = 1$.

When higher-order terms are included in the polynomial expansion of z, R(mean) is closer to R(input), but the fitting uncertainties increase. This is similar to the study of polynomial expansions of Q^2.

The polynomial expansion of z with $N = 2$ shows the good feature that it has both small bias and small variance (or fitting uncertainty) in all the tests.

IV. DISCUSSION

In this section we will discuss the robustness of the R extraction, the good fitters and the effect of fluctuations.

The method in this study helps to verify the robustness of R extraction from a certain real dataset. The binning and uncertainty information of this dataset can be used in the tests, and if some fitters are able to retrieve the input value of R regardless of the generator being used, these fitters are more likely to extract the "real" R in a robust manner from this dataset. A good description of the proton charge-distribution is not needed in this process.

In the tests of this study, the best fitters include the rational-function of Q^2 ($N = 1$, $M = 1$), and CF at the second order. These two fitters obtain R(mean) close to R(input) for all the generators, as in Tables VIII and XIII. They both have three fitting parameters (including the floating normalization parameter), and the fitting uncertainty σ [$\approx$ RMS of R(fit) distribution] around 0.0095 fm. It is noted that there is a simple mathematical transformation between these two, and more general relations between the CF expansion and the rational functions can be found in textbooks and journal articles. The polynomial expansion of z ($N = 2$) is also a good fitter. It has three fitting parameters,

and the fitting uncertainty σ is around 0.0108 fm. The difference between R(mean) and R(input) is at most 0.3σ, for each generator used in the tests. High order polynomial expansion of z and Q^2 ($N = 3$ and 4), while they show the good feature of R(mean) $\approx R$(input), their fitting uncertainties are significantly larger than the three listed above.

The root-mean-square error (RMSE) is a quantity widely used to judge the quality of a fitter considering both the bias and variance [37]. It is defined as

$$\mathrm{RMSE} = \sqrt{\mathrm{bias}^2 + \sigma^2}. \tag{21}$$

In this study, bias is represented by δR, and σ is represented by the RMS value in the tables of the fitting results. Fig. 7 summarizes the bias, σ and RMSE values of the three good fitters discussed above, one of the large-bias fitters (dipole) and one of the large-variance fitters [polynomial expansion of z ($N = 4$)]. In this figure, nine generators covering various types of G_E models are presented: different simple functions (dipole, monopole and Gaussian), rational with non-zero N and M (Kelly-2004), inverse polynomial (Arrington-2004), CF expansion (Arrington-2007), full theoretical calculations (Alarcón-2017), polynomial expansion of Q^2 (Bernauer-2014) and polynomial expansion of z (Ye-2018). The RMSE values of the three good fitters are similar. The RMSE values of the large-bias fitter, though smaller than those of the three good fitters on average, have large variations when different generators are used, which indicates that the fitter does not have the feature of robustness. The RMSE values of the large-variance fitter are significantly larger than those of the good fitters, which indicates that too many parameters are used.

The test results may change when the binning and sizes of uncertainties in the data table change. For data tables with different bin sets (covering different Q^2 ranges and/or have different distances between bin centers), and/or having different levels of uncertainties, the tests in this study can be carried out similarly, and good fitters can be found in a case-by-case manner.

Due to statistical and systematic uncertainties in a certain bin, the G_E value in one data table inevitably has some fluctuations around the true central value. These fluctuations in real data cannot be corrected back in an unbiased manner in general. Fig. 8 shows the correlation between $[\chi^2/N(\mathrm{data})]$ and $[R(\mathrm{fit}) - R(\mathrm{input})]$, where N(data) is the number of data points in the G_E vs. Q^2 table. In this figure, both the generator and the fitter use the dipole functional form. It is observed that, because of the fluctuations, very different values of R can be extracted, while similar values of $[\chi^2/N(\mathrm{data})]$ are obtained, even when the proper functional form is used (in a test with pseudo-data, the same functional form is used in the generator and the fitter). Also, a good χ^2 value does not ensure the corresponding fit extracts the radius properly. From a purely mathematical point of view, this can be understood as the difference between a good interpolating function, valid over the range of the data, and a functional form that can be used beyond the range of the data to extrapolated to the Q^2 equals zero end-point).

In the examples above, we have used repeated simulations of pseudo-data to find the best functional forms. In a real data table, where one has a single realization from a certain experiment, it is not possible to know exactly how much the fluctuation has shifted the G_E values, and/or how much effect the fluctuation brings to the R extraction.

On the other hand, one can make use of the ideas of cross validation and produce multiple data tables by different combinations of data runs, and/or using different bin sets. This should ensure that the extracted R is robust and not the result of over-training the regression to a given set of data.

V. CONCLUSION

In order to find robust methods of extracting the proton radius from data commensurate with the range and uncertainties of the upcoming PRad data, $0.0075 < Q^2 < 1.85$ fm^{-2}, we have repeatedly generated pseudo-data using numerous models and fit that data with numerous fit functions. This study has allowed us to investigate both the δR and RMS spread of the extracted proton radii in a very systematic way.

In summary, we demonstrate that interpolating functions with a good χ^2 may not extrapolate well beyond data; thus χ^2 alone cannot be used to judge which functions to use for extracting a radius. On the other hand, for the PRad experiment, we find that the (N=M=1) rational function, a two parameter continued fraction as well as the second order polynomial expansion of z can robustly extract the input radius regardless of the input function with small δR and an RMS uncertainty. In addition, the framework we created in this study can be easily expanded for more experimental extraction of R besides PRad.

ACKNOWLEDGMENTS

This work is supported in part by the U.S. Department of Energy under Contract No. DE-FG02-03ER41231, NSF MRI award PHY-1229153, Thomas Jefferson National Accelerator Facility and Duke University. We acknowledge the

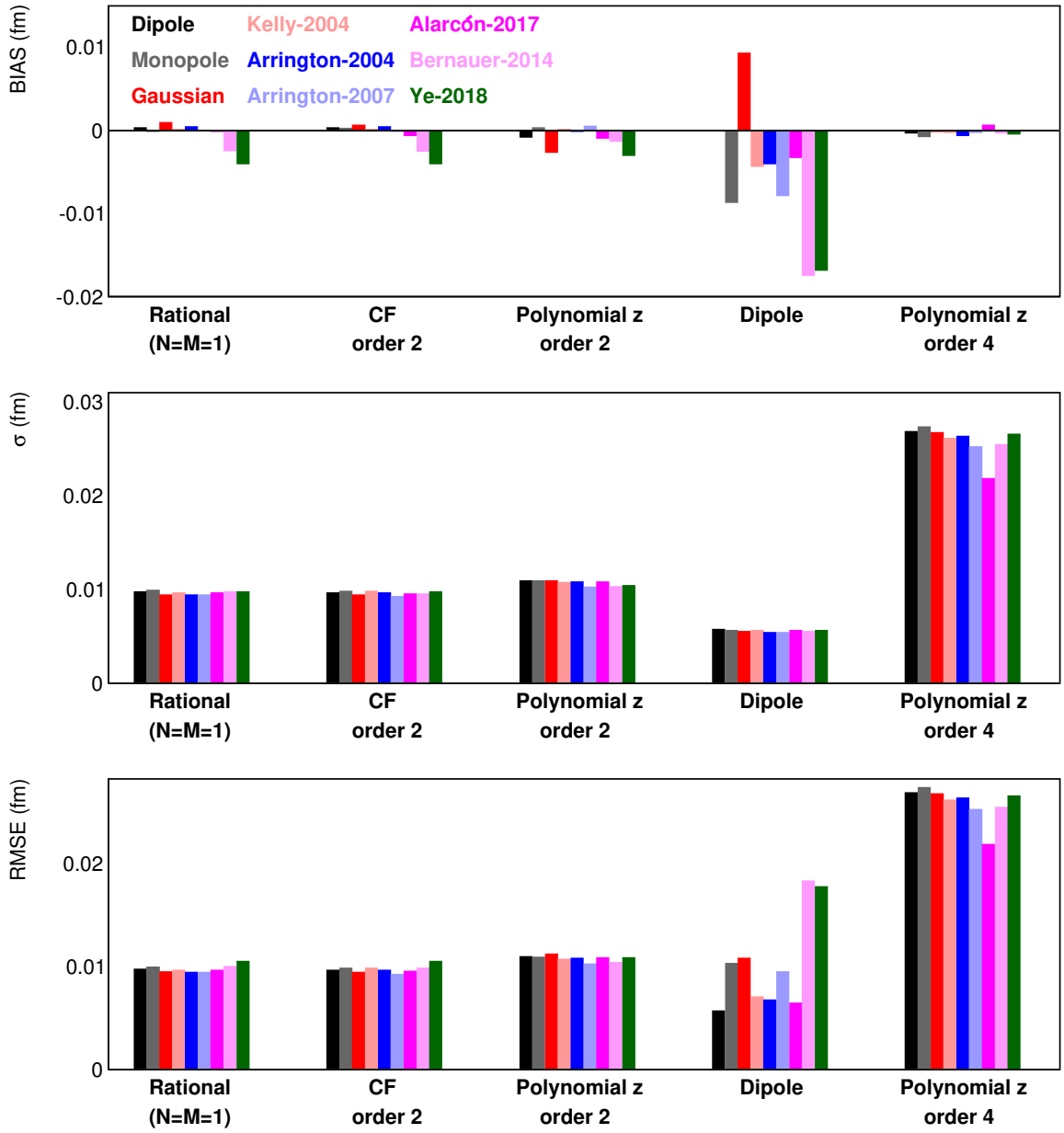

FIG. 7. (color online). Bias, σ and RMSE of the rational ($N = 1, M = 1$), CF (second order), polynomial expansion of z ($N = 2$), dipole and polynomial expansion of z ($N = 4$) fitters. The last two fitters represent typical cases of under-fit (large bias and small variance) and over-fit (small bias and large variance), respectively. The bias, σ and RMSE values of nine G_E generators with one fitter are presented by the nine colored columns in the corresponding group of the fitter. These nine G_E generators are: dipole, monopole, Gaussian, Kelly-2004, Arrington-2004, Arrington-2007, Alarcón-2017, Bernauer-2014 and Ye-2018.

helpful discussions with Jose Alarcón, John Arrington, Carl Carlson, Keith Griffioen, Ingo Sick, Christian Weiss and Zhihong Ye. We also acknowledge the support and encouragement from R. D. McKeown which motivated this work.

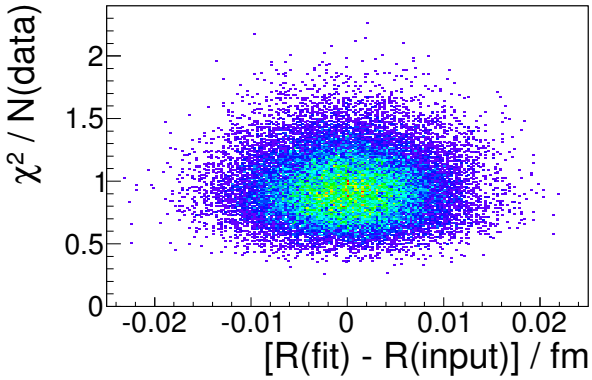

FIG. 8. (color online). Correlation between $[\chi^2/N(\text{data})]$ and $[R(\text{fit}) - R(\text{input})]$, when both the generator and the fitter use the dipole functional form.

TABLE I. The fitting results using the dipole fitter, when different generators are used. The columns represent the type of the generator (Generator), the input radius [R(input)], the mean of the R(fit) distribution [R(mean)], the difference [$\delta R = R(\text{mean}) - R(\text{input})$], and the RMS width of the R(fit) distribution.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8500	0.0000	0.0057
Monopole	0.8500	0.8414	−0.0086	0.0056
Gaussian	0.8500	0.8593	0.0093	0.0055
Kelly-2004	0.8630	0.8587	−0.0043	0.0056
Arrington-2004	0.8682	0.8642	−0.0040	0.0054
Arrington-2007	0.8965	0.8887	−0.0078	0.0054
Venkat-2011	0.8779	0.8709	−0.0070	0.0056
Bernauer-2014	0.8868	0.8694	−0.0174	0.0055
Alarcón-2017	0.8500	0.8468	−0.0032	0.0056
Alarcón-2017 (codata)	0.8750	0.8740	−0.0010	0.0054
Alarcón-2017 (μ)	0.8400	0.8366	−0.0034	0.0056
Ye-2018	0.8790	0.8622	−0.0168	0.0056
Ye-2018 (re-fix)	0.8500	0.8477	−0.0023	0.0056

TABLE II. The fitting results using the monopole fitter, when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8588	0.0088	0.0058
Monopole	0.8500	0.8501	0.0001	0.0058
Gaussian	0.8500	0.8682	0.0182	0.0057
Kelly-2004	0.8630	0.8679	0.0049	0.0058
Arrington-2004	0.8682	0.8735	0.0053	0.0056
Arrington-2007	0.8965	0.8988	0.0023	0.0056
Venkat-2011	0.8779	0.8804	0.0025	0.0057
Bernauer-2014	0.8868	0.8791	−0.0077	0.0057
Alarcón-2017	0.8500	0.8557	0.0057	0.0057
Alarcón-2017 (codata)	0.8750	0.8836	0.0086	0.0056
Alarcón-2017 (μ)	0.8400	0.8451	0.0051	0.0057
Ye-2018	0.8790	0.8716	−0.0074	0.0057
Ye-2018 (re-fix)	0.8500	0.8566	0.0066	0.0058

TABLE III. The fitting results using the Gaussian fitter, when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8407	−0.0093	0.0055
Monopole	0.8500	0.8322	−0.0178	0.0055
Gaussian	0.8500	0.8499	−0.0001	0.0054
Kelly-2004	0.8630	0.8491	−0.0139	0.0055
Arrington-2004	0.8682	0.8544	−0.0138	0.0053
Arrington-2007	0.8965	0.8780	−0.0185	0.0052
Venkat-2011	0.8779	0.8608	−0.0169	0.0054
Bernauer-2014	0.8868	0.8593	−0.0275	0.0053
Alarcón-2017	0.8500	0.8376	−0.0124	0.0057
Alarcón-2017 (codata)	0.8750	0.8640	−0.0110	0.0053
Alarcón-2017 (μ)	0.8400	0.8276	−0.0124	0.0054
Ye-2018	0.8790	0.8523	−0.0267	0.0054
Ye-2018 (re-fix)	0.8500	0.8384	−0.0116	0.0054

TABLE IV. The fitting results using the polynomial-expansion fitter ($N = 1$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8207	−0.0293	0.0051
Monopole	0.8500	0.8132	−0.0368	0.0045
Gaussian	0.8500	0.8297	−0.0203	0.0050
Kelly-2004	0.8630	0.8284	−0.0346	0.0050
Arrington-2004	0.8682	0.8338	−0.0344	0.0044
Arrington-2007	0.8965	0.8550	−0.0415	0.0049
Venkat-2011	0.8779	0.8409	−0.0370	0.0037
Bernauer-2014	0.8868	0.8374	−0.0494	0.0050
Alarcón-2017	0.8500	0.8178	−0.0322	0.0049
Alarcón-2017 (codata)	0.8750	0.8429	−0.0321	0.0043
Alarcón-2017 (μ)	0.8400	0.8105	−0.0295	0.0037
Ye-2018	0.8790	0.8310	−0.0480	0.0050
Ye-2018 (re-fix)	0.8500	0.8185	−0.0315	0.0051

TABLE V. The fitting results using the polynomial-expansion fitter ($N = 2$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8483	−0.0017	0.0088
Monopole	0.8500	0.8473	−0.0027	0.0089
Gaussian	0.8500	0.8491	−0.0009	0.0088
Kelly-2004	0.8630	0.8608	−0.0022	0.0086
Arrington-2004	0.8682	0.8654	−0.0028	0.0085
Arrington-2007	0.8965	0.8930	−0.0035	0.0083
Venkat-2011	0.8779	0.8744	−0.0035	0.0085
Bernauer-2014	0.8868	0.8796	−0.0072	0.0084
Alarcón-2017	0.8500	0.8474	−0.0026	0.0086
Alarcón-2017 (codata)	0.8750	0.8735	−0.0015	0.0082
Alarcón-2017 (μ)	0.8400	0.8383	−0.0017	0.0088
Ye-2018	0.8790	0.8707	−0.0083	0.0085
Ye-2018 (re-fix)	0.8500	0.8490	−0.0010	0.0087

TABLE VI. The fitting results using the polynomial-expansion fitter ($N = 3$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8497	−0.0003	0.0142
Monopole	0.8500	0.8498	−0.0002	0.0143
Gaussian	0.8500	0.8494	−0.0006	0.0144
Kelly-2004	0.8630	0.8630	0.0000	0.0137
Arrington-2004	0.8682	0.8680	−0.0002	0.0136
Arrington-2007	0.8965	0.8962	−0.0003	0.0135
Venkat-2011	0.8779	0.8775	−0.0004	0.0139
Bernauer-2014	0.8868	0.8858	−0.0010	0.0137
Alarcón-2017	0.8500	0.8496	−0.0004	0.0139
Alarcón-2017 (codata)	0.8750	0.8758	0.0008	0.0136
Alarcón-2017 (μ)	0.8400	0.8405	0.0005	0.0140
Ye-2018	0.8790	0.8770	−0.0020	0.0138
Ye-2018 (re-fix)	0.8500	0.8504	0.0004	0.0142

TABLE VII. The fitting results using the polynomial-expansion fitter ($N = 4$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8500	0.0000	0.0192
Monopole	0.8500	0.8503	0.0003	0.0192
Gaussian	0.8500	0.8499	−0.0001	0.0194
Kelly-2004	0.8630	0.8628	−0.0002	0.0187
Arrington-2004	0.8682	0.8685	0.0003	0.0189
Arrington-2007	0.8965	0.8965	0.0000	0.0194
Venkat-2011	0.8779	0.8772	−0.0007	0.0189
Bernauer-2014	0.8868	0.8866	−0.0002	0.0197
Alarcón-2017	0.8500	0.8497	−0.0003	0.0189
Alarcón-2017 (codata)	0.8750	0.8769	0.0019	0.0185
Alarcón-2017 (μ)	0.8400	0.8426	0.0026	0.0184
Ye-2018	0.8790	0.8783	−0.0007	0.0199
Ye-2018 (re-fix)	0.8500	0.8499	−0.0001	0.0208

TABLE VIII. The fitting results using the rational-function fitter ($N = 1$, $M = 1$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8503	0.0003	0.0097
Monopole	0.8500	0.8499	−0.0001	0.0099
Gaussian	0.8500	0.8509	0.0009	0.0094
Kelly-2004	0.8630	0.8631	0.0001	0.0096
Arrington-2004	0.8682	0.8686	0.0004	0.0094
Arrington-2007	0.8965	0.8965	0.0000	0.0094
Venkat-2011	0.8779	0.8777	−0.0002	0.0096
Bernauer-2014	0.8868	0.8844	−0.0024	0.0097
Alarcón-2017	0.8500	0.8499	−0.0001	0.0096
Alarcón-2017 (codata)	0.8750	0.8758	0.0008	0.0093
Alarcón-2017 (μ)	0.8400	0.8407	0.0007	0.0096
Ye-2018	0.8790	0.8750	−0.0040	0.0097
Ye-2018 (re-fix)	0.8500	0.8514	0.0014	0.0096

TABLE IX. The fitting results using the rational-function fitter ($N = 1$, $M = 2$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8482	−0.0018	0.0185
Monopole	0.8500	0.8546	0.0046	0.0193
Gaussian	0.8500	0.8448	−0.0052	0.0164
Kelly-2004	0.8630	0.8629	−0.0001	0.0174
Arrington-2004	0.8682	0.8675	−0.0007	0.0151
Arrington-2007	0.8965	0.8963	−0.0002	0.0159
Venkat-2011	0.8779	0.8779	0.0000	0.0126
Bernauer-2014	0.8868	0.8881	0.0013	0.0177
Alarcón-2017	0.8500	0.8507	0.0007	0.0194
Alarcón-2017 (codata)	0.8750	0.8747	−0.0003	0.0131
Alarcón-2017 (μ)	0.8400	0.8457	0.0057	0.0206
Ye-2018	0.8790	0.8799	0.0009	0.0192
Ye-2018 (re-fix)	0.8500	0.8556	0.0056	0.0265

TABLE X. The fitting results using the rational-function fitter ($N = 2$, $M = 1$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8646	0.0146	0.0132
Monopole	0.8500	0.8641	0.0141	0.0162
Gaussian	0.8500	0.8620	0.0120	0.0121
Kelly-2004	0.8630	0.8736	0.0106	0.0125
Arrington-2004	0.8682	0.8767	0.0085	0.0118
Arrington-2007	0.8965	0.8991	0.0026	0.0153
Venkat-2011	0.8779	0.8829	0.0050	0.0114
Bernauer-2014	0.8868	0.8908	0.0040	0.0152
Alarcón-2017	0.8500	0.8654	0.0154	0.0194
Alarcón-2017 (codata)	0.8750	0.8813	0.0063	0.0107
Alarcón-2017 (μ)	0.8400	0.8585	0.0185	0.0169
Ye-2018	0.8790	0.8844	0.0054	0.0147
Ye-2018 (re-fix)	0.8500	0.8661	0.0161	0.0145

TABLE XI. The fitting results using the rational-function fitter ($N = 2$, $M = 2$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8567	0.0067	0.0192
Monopole	0.8500	0.8639	0.0139	0.0194
Gaussian	0.8500	0.8518	0.0018	0.0187
Kelly-2004	0.8630	0.8683	0.0053	0.0178
Arrington-2004	0.8682	0.8723	0.0041	0.0164
Arrington-2007	0.8965	0.8969	0.0004	0.0182
Venkat-2011	0.8779	0.8806	0.0027	0.0147
Bernauer-2014	0.8868	0.8895	0.0027	0.0183
Alarcón-2017	0.8500	0.8593	0.0093	0.0197
Alarcón-2017 (codata)	0.8750	0.8773	0.0023	0.0156
Alarcón-2017 (μ)	0.8400	0.8578	0.0178	0.0203
Ye-2018	0.8790	0.8828	0.0038	0.0182
Ye-2018 (re-fix)	0.8500	0.8595	0.0095	0.0209

TABLE XII. The fitting results using the CF fit (order 1), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8588	0.0088	0.0058
Monopole	0.8500	0.8501	0.0001	0.0059
Gaussian	0.8500	0.8682	0.0182	0.0059
Kelly-2004	0.8630	0.8676	0.0046	0.0058
Arrington-2004	0.8682	0.8736	0.0054	0.0057
Arrington-2007	0.8965	0.8987	0.0022	0.0056
Venkat-2011	0.8779	0.8805	0.0026	0.0056
Bernauer-2014	0.8868	0.8792	−0.0076	0.0057
Alarcón-2017	0.8500	0.8556	0.0056	0.0058
Alarcón-2017 (codata)	0.8750	0.8833	0.0083	0.0058
Alarcón-2017 (μ)	0.8400	0.8452	−0.0034	0.0059
Ye-2018	0.8790	0.8717	−0.0073	0.0057
Ye-2018 (re-fix)	0.8500	0.8565	0.0065	0.0058

TABLE XIII. The fitting results using the CF fit (order 2), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8503	0.0003	0.0096
Monopole	0.8500	0.8502	0.0002	0.0098
Gaussian	0.8500	0.8506	0.0006	0.0094
Kelly-2004	0.8630	0.8631	0.0001	0.0098
Arrington-2004	0.8682	0.8686	0.0004	0.0096
Arrington-2007	0.8965	0.8964	−0.0001	0.0092
Venkat-2011	0.8779	0.8776	−0.0003	0.0094
Bernauer-2014	0.8868	0.8843	−0.0025	0.0097
Alarcón-2017	0.8500	0.8494	−0.0006	0.0095
Alarcón-2017 (codata)	0.8750	0.8756	0.0006	0.0095
Alarcón-2017 (μ)	0.8400	0.8405	0.0005	0.0097
Ye-2018	0.8790	0.8750	−0.0040	0.0097
Ye-2018 (re-fix)	0.8500	0.8514	0.0014	0.0096

TABLE XIV. The fitting results using the CF fit (order 3), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8515	0.0015	0.0149
Monopole	0.8500	0.8525	0.0025	0.0139
Gaussian	0.8500	0.8537	0.0037	0.0167
Kelly-2004	0.8630	0.8659	0.0029	0.0250
Arrington-2004	0.8682	0.8704	0.0022	0.0229
Arrington-2007	0.8965	0.8973	0.0008	0.0108
Venkat-2011	0.8779	0.8785	0.0006	0.0213
Bernauer-2014	0.8868	0.8902	0.0034	0.0210
Alarcón-2017	0.8500	0.8509	0.0009	0.0140
Alarcón-2017 (codata)	0.8750	0.8794	0.0044	0.0141
Alarcón-2017 (μ)	0.8400	0.8448	0.0048	0.0141
Ye-2018	0.8790	0.8826	0.0036	0.0232
Ye-2018 (re-fix)	0.8500	0.8572	0.0072	0.0235

TABLE XV. The fitting results using the CF fit (order 4), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8507	0.0007	0.0155
Monopole	0.8500	0.8532	0.0032	0.0144
Gaussian	0.8500	0.8486	−0.0014	0.0148
Kelly-2004	0.8630	0.8655	0.0025	0.0255
Arrington-2004	0.8682	0.8677	−0.0005	0.0265
Arrington-2007	0.8965	0.8970	0.0005	0.0127
Venkat-2011	0.8779	0.8764	−0.0015	0.0257
Bernauer-2014	0.8868	0.8918	0.0050	0.0224
Alarcón-2017	0.8500	0.8510	0.0010	0.0144
Alarcón-2017 (codata)	0.8750	0.8768	0.0018	0.0156
Alarcón-2017 (μ)	0.8400	0.8450	0.0050	0.0151
Ye-2018	0.8790	0.8839	0.0049	0.0242
Ye-2018 (re-fix)	0.8500	0.8587	0.0087	0.0260

TABLE XVI. The fitting results using the polynomial expansion of z ($N = 1$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8896	0.0396	0.0055
Monopole	0.8500	0.8809	0.0309	0.0055
Gaussian	0.8500	0.8985	0.0485	0.0055
Kelly-2004	0.8630	0.8980	0.0350	0.0055
Arrington-2004	0.8682	0.9036	0.0354	0.0056
Arrington-2007	0.8965	0.9273	0.0308	0.0053
Venkat-2011	0.8779	0.9101	0.0322	0.0054
Bernauer-2014	0.8868	0.9087	0.0219	0.0054
Alarcón-2017	0.8500	0.8858	0.0358	0.0050
Alarcón-2017 (codata)	0.8750	0.9128	0.0378	0.0051
Alarcón-2017 (μ)	0.8400	0.8764	0.0364	0.0056
Ye-2018	0.8790	0.9016	0.0226	0.0054
Ye-2018 (re-fix)	0.8500	0.8873	0.0373	0.0065

TABLE XVII. The fitting results using the polynomial expansion of z ($N = 2$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8492	−0.0008	0.0109
Monopole	0.8500	0.8503	0.0003	0.0109
Gaussian	0.8500	0.8474	−0.0026	0.0109
Kelly-2004	0.8630	0.8631	0.0001	0.0107
Arrington-2004	0.8682	0.8681	−0.0001	0.0108
Arrington-2007	0.8965	0.8970	0.0005	0.0102
Venkat-2011	0.8779	0.8781	0.0002	0.0106
Bernauer-2014	0.8868	0.8855	−0.0013	0.0103
Alarcón-2017	0.8500	0.8491	−0.0009	0.0108
Alarcón-2017 (codata)	0.8750	0.8753	0.0003	0.0106
Alarcón-2017 (μ)	0.8400	0.8401	0.0001	0.0109
Ye-2018	0.8790	0.8760	−0.0030	0.0104
Ye-2018 (re-fix)	0.8500	0.8507	0.0007	0.0107

TABLE XVIII. The fitting results using the polynomial expansion of z ($N = 3$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8497	−0.0003	0.0182
Monopole	0.8500	0.8498	−0.0002	0.0185
Gaussian	0.8500	0.8496	−0.0004	0.0185
Kelly-2004	0.8630	0.8631	0.0001	0.0179
Arrington-2004	0.8682	0.8679	−0.0003	0.0178
Arrington-2007	0.8965	0.8963	−0.0002	0.0171
Venkat-2011	0.8779	0.8776	−0.0003	0.0177
Bernauer-2014	0.8868	0.8870	0.0002	0.0173
Alarcón-2017	0.8500	0.8497	−0.0003	0.0171
Alarcón-2017 (codata)	0.8750	0.8766	0.0016	0.0169
Alarcón-2017 (μ)	0.8400	0.8417	0.0017	0.0172
Ye-2018	0.8790	0.8785	−0.0005	0.0175
Ye-2018 (re-fix)	0.8500	0.8500	0.0000	0.0178

TABLE XIX. The fitting results using the polynomial expansion of z ($N = 4$), when different generators are used. Notation as in Table I.

Generator	R(input) (fm)	R(mean) (fm)	δR (fm)	RMS (fm)
Dipole	0.8500	0.8497	−0.0003	0.0268
Monopole	0.8500	0.8493	−0.0007	0.0273
Gaussian	0.8500	0.8499	−0.0001	0.0267
Kelly-2004	0.8630	0.8628	−0.0002	0.0261
Arrington-2004	0.8682	0.8676	−0.0006	0.0263
Arrington-2007	0.8965	0.8963	−0.0002	0.0252
Venkat-2011	0.8779	0.8777	−0.0002	0.0260
Bernauer-2014	0.8868	0.8865	−0.0003	0.0254
Alarcón-2017	0.8500	0.8506	0.0006	0.0218
Alarcón-2017 (codata)	0.8750	0.8779	0.0029	0.0210
Alarcón-2017 (μ)	0.8400	0.8442	0.0042	0.0212
Ye-2018	0.8790	0.8786	−0.0004	0.0255
Ye-2018 (re-fix)	0.8500	0.8493	−0.0007	0.0265

[1] R. Pohl *et al.*, Nature **466**, 213 (2010).
[2] A. Antognini *et al.*, Science **339**, 417 (2013).
[3] P. J. Mohr, D. B. Newell, and B. N. Taylor, Rev. Mod. Phys. **88**, 035009 (2016), arXiv:1507.07956 [physics.atom-ph].
[4] M. Mihovilovic *et al.*, Phys. Lett. **B771**, 194 (2017), arXiv:1612.06707 [nucl-ex].
[5] A. Beyer, L. Maisenbacher, A. Matveev, R. Pohl, K. Khabarova, A. Grinin, T. Lamour, D. C. Yost, T. W. Hänsch, N. Kolachevsky, and T. Udem, Science **358**, 79 (2017), http://science.sciencemag.org/content/358/6359/79.full.pdf.
[6] R. Pohl, R. Gilman, G. A. Miller, and K. Pachucki, Ann. Rev. Nucl. Part. Sci. **63**, 175 (2013), arXiv:1301.0905 [physics.atom-ph].
[7] C. E. Carlson, Prog. Part. Nucl. Phys. **82**, 59 (2015), arXiv:1502.05314 [hep-ph].
[8] R. J. Hill, *Proceedings, 12th Conference on Quark Confinement and the Hadron Spectrum (Confinement XII): Thessaloniki, Greece*, EPJ Web Conf. **137**, 01023 (2017), arXiv:1702.01189 [hep-ph].
[9] I. T. Lorenz and U.-G. Meiner, Phys. Lett. **B737**, 57 (2014), arXiv:1406.2962 [hep-ph].
[10] K. Griffioen, C. Carlson, and S. Maddox, Phys. Rev. **C93**, 065207 (2016), arXiv:1509.06676 [nucl-ex].
[11] D. W. Higinbotham, A. A. Kabir, V. Lin, D. Meekins, B. Norum, and B. Sawatzky, Phys. Rev. **C93**, 055207 (2016), arXiv:1510.01293 [nucl-ex].
[12] M. Horbatsch and E. A. Hessels, Phys. Rev. **C93**, 015204 (2016), arXiv:1509.05644 [nucl-ex].
[13] M. Horbatsch, E. A. Hessels, and A. Pineda, Phys. Rev. **C95**, 035203 (2017), arXiv:1610.09760 [nucl-th].
[14] J. Arrington and I. Sick, J. Phys. Chem. Ref. Data **44**, 031204 (2015), arXiv:1505.02680 [nucl-ex].
[15] I. Sick, *From quarks and gluons to hadrons and nuclei. Proceedings, International Workshop on Nuclear Physics, 33rd course, Erice, Italy, September 16-24, 2011*, Prog. Part. Nucl. Phys. **67**, 473 (2012).
[16] G. Lee, J. R. Arrington, and R. J. Hill, Phys. Rev. **D92**, 013013 (2015), arXiv:1505.01489 [hep-ph].
[17] D. Borisyuk, Nucl. Phys. **A843**, 59 (2010), arXiv:0911.4091 [hep-ph].
[18] K. M. Graczyk and C. Juszczak, Phys. Rev. **C90**, 054334 (2014), arXiv:1408.0150 [hep-ph].
[19] J. C. Bernauer *et al.* (A1), Phys. Rev. Lett. **105**, 242001 (2010), arXiv:1007.5076 [nucl-ex].
[20] J. C. Bernauer *et al.* (A1), Phys. Rev. **C90**, 015206 (2014), arXiv:1307.6227 [nucl-ex].
[21] I. Sick and D. Trautmann, Phys. Rev. **C95**, 012501 (2017), arXiv:1701.01809 [nucl-ex].
[22] J. M. Alarcon and C. Weiss, (2017), arXiv:1710.06430 [hep-ph].
[23] E. Kraus, K. E. Mesick, A. White, R. Gilman, and S. Strauch, Phys. Rev. **C90**, 045206 (2014), arXiv:1405.4735 [nucl-ex].
[24] I. Sick, Phys. Lett. **B576**, 62 (2003), arXiv:nucl-ex/0310008 [nucl-ex].
[25] *High Precision Measurement of the Proton Charge Radius*, url: `https://userweb.jlab.org/~mezianem/PRAD/PAC39_Gasparian.pdf`.
[26] *Proton radius fitting library*, url: `https://github.com/saberbud/Proton_radius_fit_class`.
[27] R. Brun and F. Rademakers, *New computing techniques in physics research V. Proceedings, 5th International Workshop, AIHENP '96, Lausanne, Switzerland, September 2-6, 1996*, Nucl. Instrum. Meth. **A389**, 81 (1997).
[28] F. Borkowski, G. G. Simon, V. H. Walther, and R. D. Wendling, Zeitschrift für Physik A Atoms and Nuclei **275**, 29 (1975).
[29] J. J. Kelly, Phys. Rev. **C70**, 068202 (2004).
[30] J. Arrington, Phys. Rev. **C69**, 022201 (2004), arXiv:nucl-ex/0309011 [nucl-ex].
[31] J. Arrington and I. Sick, Phys. Rev. **C76**, 035201 (2007), arXiv:nucl-th/0612079 [nucl-th].
[32] J. C. Bernauer, Ph.D. thesis, Johannes Gutenberg-Universitat Mainz (2010).
[33] S. Venkat, J. Arrington, G. A. Miller, and X. Zhan, Phys. Rev. **C83**, 015203 (2011), arXiv:1010.3629 [nucl-th].
[34] J. M. Alarcon and C. Weiss, Phys. Rev. **C96**, 055206 (2017), arXiv:1707.07682 [hep-ph].
[35] Z. Ye, J. Arrington, R. J. Hill, and G. Lee, Phys. Lett. **B777**, 8 (2018), arXiv:1707.09063 [nucl-ex].
[36] *PRad bin-set files*, url: `https://github.com/saberbud/Proton_radius_fit_class/tree/master/Bin_sets`.
[37] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. (Springer, 2009).