ADVANCED STATISTICAL METHODS TO FIT NUCLEAR MODELS*

M. Shelley[a], P. Becker[a], A. Gration[b], A. Pastore[a]

[a]Department of Physics, University of York
Heslington, York YO10 5DD, United Kingdom
[b]Department of Physics & Astronomy, University of Leicester
University Road, Leicester LE1 7RH, United Kingdom

We discuss advanced statistical methods to improve parameter estimation of nuclear models. In particular, using the Liquid Drop Model for nuclear binding energies, we show that the area around the global χ^2 minimum can be efficiently identified using the Gaussian Process Emulation. We also demonstrate how the Markov-chain Monte Carlo sampling is a valuable tool for visualising and analysing the associated multidimensional likelihood surface.

1. Introduction

In general, nuclear models contain parameters that must be determined to best reproduce experimental results. These parameters are typically adjusted on data by means of a least-square procedure [1–4], or more generally, using a Maximum Likelihood Estimator (MLE) [1]. MLE provides us essential information about the estimated parameters, such as errors and correlations. We refer to Ref. [3] for a more detailed discussion.

With an MLE, there are two main challenges: how to find the maximum in the parameter space, especially in the case of mutlimodal likelihood surfaces [5] and how to provide realistic estimates of errors.

In the present article, we discuss the benefit of using the statistical modelling method Gaussian Process Emulation (GPE) [6–8] and the Expected Improvement (EI) criterion [9, 10] to explore the likelihood surface [1, 2, 11] and identify the location of the maximum in parameter space. GPE has already been adopted in several scientific domains to facilitate the use of computationally expensive models [10, 12–14]. It has also been recently

---

applied to nuclear physics [15–17]. In this paper, we combine GPE with the EI criterion to estimate the parameters of a Liquid Drop Model (LDM) [18]. Through the Markov-chain Monte-Carlo (MCMC) sampling, we explain how to visualise the multidimensional likelihood surface and extract the covariance matrix without calculating explicitly the Hessian matrix [3].

The article is organised as follows: in Section 2, we introduce the basic formalism of GPE statistical method. In Section 3, we present the LDM and in Section 4, we demonstrate how GPE and EI can be used to iteratively find the maximum of the unimodal likelihood of the LDM. We finally provide our conclusions in Section 5.

2. Gaussian Process Emulation and Expected Improvement

Gaussian Process Emulation (GPE) is a regression method that performs a smooth interpolation of the points y_s of a data set, providing associated confidence intervals. In a standard fitting procedure, one assumes a fixed relation between the independent variable $\boldsymbol{x}$ and the points $y_s = f(\boldsymbol{x}|\boldsymbol{p})$, where $\boldsymbol{p}$ represents a set of adjustable parameters. It is not the case with GPE, where we treat the points y_s as a draw from a Gaussian process (GP), $\boldsymbol{y}$ with mean 0 and covariance $R(\boldsymbol{x}, \boldsymbol{x}')$. See Ref. [10] for a more general discussion.

We still need to make some assumption about the structure of the data. Such an assumption is done for GPE on the structure of the covariance matrix. In the present article, we assume that the latter is filled with the *squared exponential* covariance function, so that each matrix element corresponds to

$$R\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \sigma_k^2 \prod_{i=1}^{d} \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{x}')^2}{2\theta_i^2}\right), \qquad (1)$$

where d is the number of parameters and $\boldsymbol{x}, \boldsymbol{x}'$ are the data points.

We choose to emulate normalised training data to avoid numerical issues that can arise during emulation. We, therefore, set the maximum covariance to be $\sigma_k^2 = 1$. The *hyperparameters* θ_i, often referred to as correlation lengths or length scales, describe the level of smoothness of the emulated surface along each dimension of the parameter space. These *hyperparameters* are adjusted on training data using an MLE method. We refer to Ref. [8] for more details. Given the particular Gaussian form of the covariance matrix in Eq. (1), we observe that neighbouring points will be strongly correlated, while points that are far from each other will be uncorrelated. The length scale controls how fast such a correlation drops moving away from a given point and, therefore, moving away from the diagonal of the covariance matrix.

The GPE prediction at an unknown location $\boldsymbol{x}^*$ is

$$\hat{\mu}(\boldsymbol{x}^*) = r^T(\boldsymbol{x}^*)R^{-1}\boldsymbol{y}\,, \tag{2}$$

where $r(\boldsymbol{x}^*) = R(\boldsymbol{x}, \boldsymbol{x}^*)$ corresponds to the vector filled with the correlation of $\boldsymbol{x}^*$ with all the data points. GPE also easily provides the 1σ confidence intervals as

$$\hat{\sigma}(\boldsymbol{x}^*) = R(\boldsymbol{x}^*, \boldsymbol{x}^*) - r^T(\boldsymbol{x})R^{-1}r(\boldsymbol{x})\,. \tag{3}$$

These formulas ignore possible uncertainties on the points y_s since we assume here that they come from a deterministic computer simulation. It is, however, possible to take them into account into the GPE procedure [19]. In the following section, we introduce a model that will be used to illustrate how the GPE practically works.

3. Liquid Drop Model

The LDM is a simple model that links the binding energy B_{th} of a given nucleus to its number of neutrons N and protons Z. It is composed of five adjustable parameters $\boldsymbol{p} = \{a_{\text{v}}, a_{\text{s}}, a_{\text{c}}, a_{\text{a}}, a_{\text{p}}\}$ that are related to bulk properties of the nucleus and defined as [18]

$$B_{\text{th}}(N, Z) = a_{\text{v}}A - a_{\text{s}}A^{2/3} - a_{\text{c}}\frac{Z(Z-1)}{A^{1/3}} - a_{\text{a}}\frac{(N-Z)^2}{A} - a_{\text{p}}\delta(N, Z)\,, \tag{4}$$

where $A = N + Z$, and

$$\delta(N, Z) = \begin{cases} +A^{-1/2} & \text{for } Z, N \text{ even}\,, \\ 0 & \text{for } A \text{ odd}\,, \\ -A^{-1/2} & \text{for } Z, N \text{ odd}\,. \end{cases} \tag{5}$$

To determine the optimal set of parameters $\boldsymbol{p}_0$, we construct the penalty function

$$\chi^2 = \sum_{N,Z \in \text{data}} \frac{(B_{\text{exp}}(N, Z) - B_{\text{th}}(N, Z))^2}{\sigma^2(N, Z)} \tag{6}$$

to be minimised. The experimental energies, $B_{\text{exp}}(N, Z)$, are extracted from Ref. [20]. Since Eq. (4) is not suitable for describing light systems, we exclude all nuclei with $A < 16$. We also choose not to take into account those with experimental errors on binding energies larger than 100 keV. This LDM model has a large variance of a few MeV on average [21, 22] making the residuals larger than any experimental errors. The weights in Eq. (6) are, therefore, unimportant and we have fixed them as $\sigma^2(N, Z) = 1$ for all nuclei. We also recall for the sake of clarity that the χ^2 function is directly linked to the likelihood function as $\mathcal{L} \propto e^{-\chi^2}$.

4. Results

As a start, we explore the likelihood surface associated to our penalty χ^2 function using the No-U-Turn Sampler (NUTS) [23], an efficient Markov-chain Monte Carlo (MCMC) sampler [24]. In Fig. 1, we present the marginalised likelihood in vicinity of the maximum for individuals parameters on the diagonal part and the contour plots for pairs of parameters off the diagonal. The best estimation of LDM parameters with error bars, extracted as the 68% percentile of each parameter distribution, are given on top of each column.
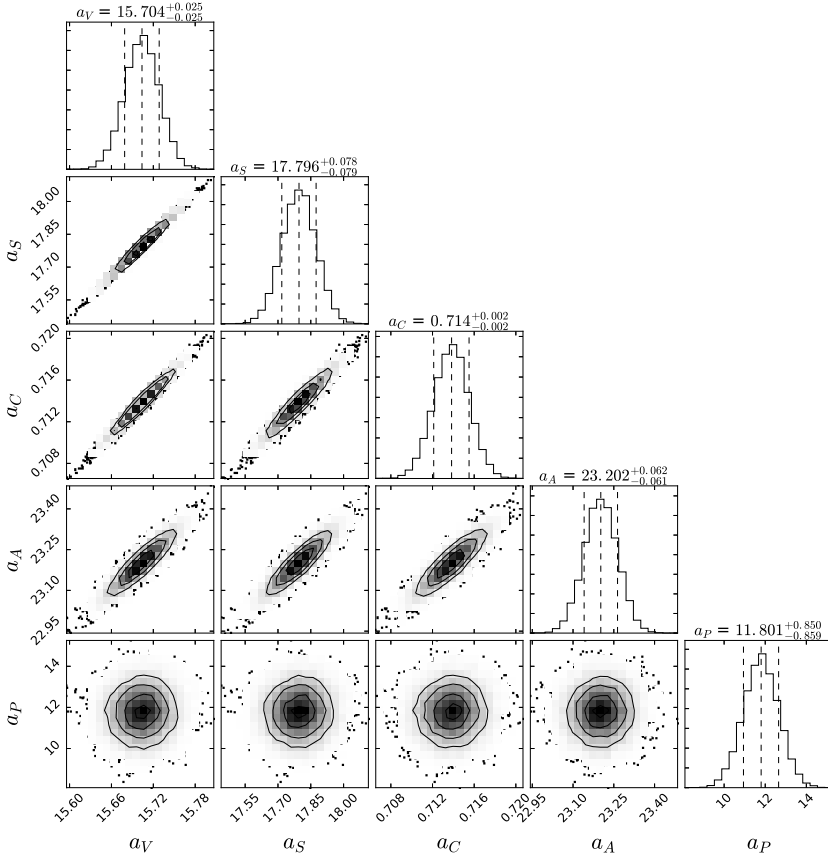

Fig. 1. Corner plot of the exact likelihood surface in vicinity of the maximum. The diagonal plots show the parameters distributions. The central dashed line corresponds to the mean of the distribution, while the side ones delimit the 1σ interval. The off-diagonal plots show the correlation between the two parameters. See the text for details.

The marginalisation of the likelihood [15] provides us with a very useful insight on the behaviour of the likelihood surface. Given its exponential nature, the MCMC algorithm is the most suitable technique to perform the required multidimensional integrals leading to the marginalisation. From the resulting distributions of the parameters $\boldsymbol{p}$, we can extract information about correlations. In particular, we have been able to directly extract the covariance matrix without performing any derivative in parameter space [1]. In Table I, we provide the correlation matrix extracted from MCMC. We observe that the results are in perfect agreement with the correlations obtained using Non-Parametric Bootstrap [22], which is also a Monte Carlo sampling based on the log-likelihood function. For this particular model, we observe a strong correlation among the parameters a_v, a_s, a_a, a_c that also reflects on the cigar-like shape of the marginalised likelihood shown in Fig. 1. The pairing term is not correlated to the others, thus the marginal likelihood has a spheroid shape.

TABLE I

Correlation matrix for LDM parameters obtained from MCMC sampling.

	a_v	a_s	a_c	a_a	a_p
a_v	1				
a_s	0.993	1			
a_c	0.984	0.962	1		
a_a	0.917	0.901	0.884	1	
a_p	0.038	0.037	0.040	0.038	1

The MCMC is a valuable tool to explore the surface of the unimodal likelihood, but in the case of a multimodal one, it may not be the most efficient method. A possible way to solve the issue is to use the GPE method to emulate the likelihood surface together with the Expected Improvement (EI) criterion [9] to focus the exploration around the maxima.

The EI criterion helps us to choose iteratively the points in parameter space one should explore in order to efficiently find a global optimum of a function. It is an optimisation technique which, after a large enough number of iterations, will converge to the global optimum, never getting stuck into a local optimum. It uses both the GPE mean and GPE confidence intervals, in a tradeoff between exploitation and exploration. The first means sampling areas of likely improvement near the current optimum predicted by the mean, the second means sampling areas of high uncertainty, where the confidence intervals are large. It employs an acquisition function, whose maximum tells us where next in parameter space to run our model. GPE is then performed again with this new point, and this process is repeated until convergence is reached according to a given convergence criteria.

For the sake of illustration, we fix three parameters, a_c, a_a, a_p, of LDM at values obtained from MCMC and emulate the log-likelihood surface, *i.e.* the χ^2, as a function of the remaining two parameters a_v, a_s.

We design the initial data set by taking 60 points for each dimension to ensure a good quality emulation from the outset of the optimisation. We impose a uniform prior distribution on parameter space, by restricting to a fairly large interval $a_{v(s)} \in [0, 30]$ MeV. The data set is selected using a standard space-filling method [25]. After running GPE, we feed the resulting likelihood surface to the MCMC sampler. The result is illustrated in Fig. 2. In the top left panel, we show the result obtained using the initial data set,
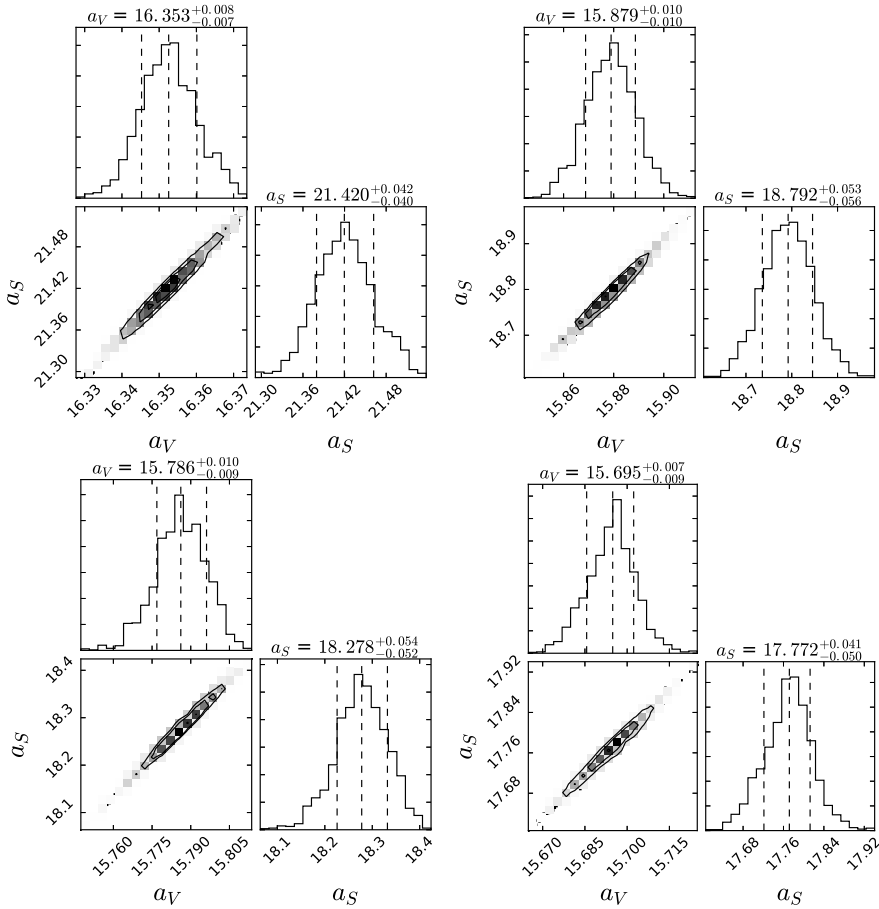

Fig. 2. Corner plots of the likelihood obtained from the emulated χ^2 surface for the a_v and a_s parameters, after the first (top left), 11 (top right), 21 (bottom left) and 51 (bottom right) iterations of EI. See the text for details.

i.e. without any EI. We notice that the means of the a_v and a_s distribution are quite different from the expected optimal values (see Fig. 1), that we attribute to an incorrect emulation of the likelihood surface.

We now apply an iterative procedure guided by EI: in the top right panel, we add extra 11 points and then we perform a new MCMC run. In this case, the volume parameter gets extremely close to its optimal value and errors.

The iterative EI procedure was then carried out in the bottom left panel (21 iterations) and bottom right (51 iterations) until the mean value of the distributions of the parameters a_v, a_s falls within the error bars provided by the complete MCMC sampling. As expected, in the limit of large number of points, GPE will provide a very good approximation of the exact surface.

5. Conclusions

In this article, we have explained how advanced statistical methods may be a very useful tool to estimate parameter in theoretical nuclear models. By using a simple Liquid Drop Model, we have discussed the use of MCMC method to visualise multidimensional likelihood surfaces and how to extract information concerning the covariance matrix without performing numerical derivatives in parameter space. This is a very interesting method, since contrary to the standard Hessian method [26], we do not need to assume a parabolic shape of the surface around the optimum.

Since MCMC is likely to not adequate deal with multimodal likelihood surfaces, we have also investigated the use of GPE combined with the EI to identify in an efficient way the position of the optimum in parameter space and then use the MCMC to obtain informations on the parameter distributions. By combining these two methods, we now plan to perform a more complete study using the likelihood surface of a Skyrme functional [27].

REFERENCES

[1] R.J. Barlow, *A Guide to the Use of Statistical Methods in the Physical Sciences*, John Wiley, 1989.

[2] M. Kortelainen *et al.*, *Phys. Rev. C* **82**, 024313 (2010).

[3] J. Dobaczewski, W. Nazarewicz, P. Reinhard, *J. Phys. G: Nucl. Part. Phys.* **41**, 074001 (2014).

[4] P. Becker *et al.*, *Phys. Rev. C* **96**, 044330 (2017).

[5] B.W. Silverman, *J. R. Statist. Soc. B* **43**, 97 (1981).

[6] A. O'Hagan, *J. R. Statist. Soc. B* **40**, 1 (1978).

[7] J. Sacks, W.J. Welch, T.J. Mitchell, H.P. Wynn, *Stat. Sci.* **4**, 409 (1989).

[8] C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes for Machine Learning*, *Adaptive Computation and Machine Learning*, MIT Press, Cambridge, Mass, 2006.

[9] D.R. Jones, M. Schonlau, W.J. Welch, *J. Global Optim.* **13**, 455 (1998).

[10] A. Gration, M.I. Wilkinson, *Mon. Not. R. Astron. Soc.* **485**, 4878 (2019).

[11] Y. Gao *et al.*, *Phys. Rev. C* **87**, 034324 (2013).

[12] M.C. Kennedy, C.W. Anderson, S. Conti, A. O'Hagan, *Reliab. Eng. Syst. Safe.* **91**, 1301 (2006).

[13] N. Gibson *et al.*, *Mon. Not. R. Astron. Soc.* **419**, 2683 (2012).

[14] S. Sale, J. Magorrian, *Mon. Not. R. Astron. Soc.* **445**, 256 (2014).

[15] J. McDonnell *et al.*, *Phys. Rev. Lett.* **114**, 122501 (2015).

[16] A. Pastore, M. Shelley, S. Baroni, C. Diget, *J. Phys. G: Nucl. Part. Phys.* **44**, 094003 (2017).

[17] L. Neufcourt, Y. Cao, W. Nazarewicz, F. Viens, *Phys. Rev. C* **98**, (2018).

[18] K.S. Krane, D. Halliday, *Introductory Nuclear Physics*, Vol. 465, Wiley, New York 1988.

[19] I. Andrianakis, P.G. Challenor, *Comput. Stat. Data Anal.* **56**, 4215 (2012).

[20] M. Wang *et al.*, *Chin. Phys. C* **36**, 1603 (2012).

[21] G. Bertsch, D. Bingham, *Phys. Rev. Lett.* **119**, 252501 (2017).

[22] A. Pastore, *J. Phys. G: Nucl. Part. Phys.* **46**, 052001 (2019) [arXiv:1810.05585 [nucl-th]].

[23] M.D. Hoffman, A. Gelman, *J. Mach. Learn. Res.* **15**, 1593 (2014).

[24] R.M. Neal, Probabilistic Inference Using Markov Chain Monte Carlo Methods, Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto, 1993.

[25] M.D. McKay, R.J. Beckman, W.J. Conover, *Technometrics* **21**, 239 (1979).

[26] X. Roca-Maza, N. Paar, G. Colò, *J. Phys. G: Nucl. Part. Phys.* **42**, 034033 (2015).

[27] E. Perlińska, S. Rohoziński, J. Dobaczewski, W. Nazarewicz, *Phys. Rev. C* **69**, 014316 (2004).