Decoding β-decay systematics: A global statistical model for β^- half-lives*

N. J. Costiris† and E. Mavrommatis‡
Physics Department, Division of Nuclear Physics & Particle Physics, University of Athens, GR-15771 Athens, Greece

K. A. Gernoth§
Institut für Theoretische Physik, Johannes-Kepler-Universität, A-4040 Linz, Austria and
School of Physics & Astronomy, Schuster Building, The University of Manchester, Manchester, M13 9PL, United Kingdom

J. W. Clark‖
McDonnell Center for the Space Sciences & Department of Physics, Washington University, St. Louis, Missouri 63130, USA
Complexo Interdisciplinar, Centro de Mathemática e Aplicações Fundamentals, University of Lisbon, 1649-003 Lisbon, Portugal and
Departamento de Física, Instituto Superior Técnico, Technical University of Lisbon, 1096 Lisbon, Portugal
(Received 4 May 2008; revised manuscript received 19 January 2009; published 28 October 2009)

Statistical modeling of nuclear data provides a novel approach to nuclear systematics complementary to established theoretical and phenomenological approaches based on quantum theory. Continuing previous studies in which global statistical modeling is pursued within the general framework of machine learning theory, we implement advances in training algorithms designed to improve generalization, in application to the problem of reproducing and predicting the half-lives of nuclear ground states that decay 100% by the β^- mode. More specifically, fully connected, multilayer feed-forward artificial neural network models are developed using the Levenberg-Marquardt optimization algorithm together with Bayesian regularization and cross-validation. The predictive performance of models emerging from extensive computer experiments is compared with that of traditional microscopic and phenomenological models as well as with the performance of other learning systems, including earlier neural network models as well as the support vector machines recently applied to the same problem. In discussing the results, emphasis is placed on predictions for nuclei that are far from the stability line, and especially those involved in r-process nucleosynthesis. It is found that the new statistical models can match or even surpass the predictive performance of conventional models for β-decay systematics and accordingly should provide a valuable additional tool for exploring the expanding nuclear landscape.

DOI: 10.1103/PhysRevC.80.044332 PACS number(s): 23.40.−s, 21.10.Tg, 26.30.−k, 07.05.Mh

I. INTRODUCTION

This work is devoted to the development of artificial neural network models that, after being trained with a subset of the available experimental data on β decay from nuclear ground states, demonstrate significant reliability in the prediction of β^- half-lives for nuclides absent from the training set. The work represents an exploratory study of the degree to which the existing data determine the mapping from proton and neutron numbers to the corresponding β^- half-life.

There is an urgent need among nuclear physicists and astrophysicists for reliable estimates of β^--decay half-lives of nuclei far from stability [1,2]. Among nuclear physicists this need is driven both by the experimental programs of existing and future radioactive ion beam facilities and by the stresses placed on established nuclear structure theory as totally new areas of the nuclear landscape are opened for exploration. For nuclear astrophysicists, such information is intrinsic to an understanding of the onset of the nucleosynthesis of heavy elements above Fe, notably the r process [3–5]. Both the element distribution on the r-process path and the time scale of the r process are highly sensitive to the β^--decay properties of the neutron-rich nuclei involved.

In the nuclear chart there are spaces for some 6000 nuclides between the β-stability line and the neutron-drip line. Recent years have seen significant progress at radioactive ion beam facilities in laboratory production and study of nuclei in the r-process path, and this effort will surely accelerate as more advanced facilities come online [1,2]. Nevertheless, experimental information on the properties of many r-process nuclei will remain inaccessible for some time to come. Accordingly, realistic models of nucleosynthesis must at present rely heavily on theoretical predictions for β-decay lifetimes and other key quantities.

A number of useful approaches to modeling β^--decay half-lives have been proposed and applied. These include the more phenomenological treatments, such as the gross theory (GT), as well as microscopic approaches based on the shell model and the proton-neutron quasiparticle random-phase approximation (pnQRPA) in various versions. More recently, hybrid macroscopic-microscopic and relativistic models have come onto the scene. Some of these approaches emphasize only global applicability, while others seek self-consistency or comprehensive inclusion of nuclear correlations. Table 1 of Ref. [6] provides a convenient summary of a number of the competing models of β-decay systematics.

*URL: http://www.pythaim.phys.uoa.gr; pythaim@phys.uoa.gr
†ncost@phys.uoa.gr
‡emavrom@phys.uoa.gr
§klaus.a.gernoth@manchester.ac.uk
‖jwc@wuphys.wustl.edu

In gross theory, developed by Takahashi, Yamada, and Kondoh [7], gross properties of β^- decay over a wide nuclidic region are predicted by averaging over the final states of the daughter nucleus. Subsequently, various refinements and modifications of this treatment have been introduced. The most current of these is the so-called semigross theory (SGT), in which the shell effects of only the parent nucleus are taken into account [8]. However, in the calculations of β^--decay half-lives within the shell model, the detailed structure of β strength function is considered. Results exist for lighter nuclei and nuclei at $N = 50$, 82, and 126. (See Refs. [9,10] for recent calculations.) Due to the limits set by the size of the configuration space, calculations are not possible for heavy nuclei.

Several groups have carried out extensive *pn*QRPA studies including pairing. Efforts along this line by Klapdor and coworkers [11] began in the framework of the Nilsson single-particle model, including the Gamow-Teller residual interaction in Tamm-Dancoff approximation (TDA), with pairing treated at the BCS level [12]. This approach has been complemented and refined by Staudt *et al.* [13] and Hirsch *et al.* [14], using *pn*QRPA with the Gamow-Teller residual interaction. The later study by Homma *et al.* [15], denoted NBCS + *pn*QRPA, includes a schematic interaction also for the first-forbidden (ff) decay. The Klapdor group has extended the *pn*QRPA theory to calculate β-decay half-lives in stellar environments using configurations beyond 1p-1h [16].

The starting point of the β-decay calculations of Möller and coworkers is the study of nuclear-ground-state masses and deformations based on the finite-range droplet model (FRDM) and a folded-Yukawa single-particle potential [17]. The β-decay half-lives for the allowed Gamow-Teller transitions have been obtained from a *pn*QRPA calculation after the addition of pairing and Gamow-Teller residual interactions, in a procedure denoted FRDM + *pn*QRPA [18,19]. In the latest calculations the effect of the ff decay has been added by using the gross theory (*pn*QRPA + ffGT) [20]. Nonrelativistic *pn*QRPA calculations that aim at self-consistency include the Hartree-Fock-Bogoliubov + continuum QRPA (HFB + QRPA) calculations performed with a Skyrme energy-density functional for some spherical even-even semimagic nuclides with $N = 50$, 82, 126 [21]. The extended Thomas-Fermi plus Strutinski integral method (ETFSI) (an approximation to the HF method based on a Skyrme-type force plus a δ-function pairing force) has been elaborated and applied to large-scale predictions of β^- half-lives [22]. Recently, the density functional + continuum QRPA (DF + CQRPA) approximation, with the spin-isospin effective NN interaction of the finite Fermi system theory operating in the ph channel, has been developed for ground-state properties and Gamow-Teller and ff transitions of nuclei far from the stability line and applied near closed neutron shells at $N = 50$, 82, 126 and in the region "east" of ^{208}Pb [6,23]. In the relativistic framework, a *pn*QRPA calculation (*pn*RQRPA) based on a relativistic Hartree-Bogoliubov description of nuclear ground states with the density-dependent effective interaction DD-MEI* has been employed to obtain Gamow-Teller β^--decay half-lives of neutron-rich nuclei in the $N \simeq 50$ and $N \simeq 82$ regions relevant to the r process [24]. Recently, an extension of the above framework to include momentum-dependent nucleon self-energies was applied in the calculation of β-decay half-lives of neutron-rich nuclei in the $Z \simeq 28$ and $Z \simeq 50$ regions [25].

Despite continuing methodological improvements, the predictive power of these conventional, "theory-thick" models is rather limited for β^--decay half-lives of nuclei that are mainly far from stability. The predictions often deviate from experiment by one or more orders of magnitude and show considerable sensitivity to quantities that are poorly known. In this environment, statistical modeling based on advanced techniques of statistical learning theory or "machine-learning," notably artificial neural networks (ANNs) [26,27] and support vector machines (SVMs) [27–29], offers an interesting and potentially effective alternative for global modeling of β^--decay lifetimes. Such approaches have proven their value for a variety of scientific problems in astronomy, high-energy physics, and biochemistry that involve function approximation and pattern classification [30,31]. Statistical modeling implementing machine-learning algorithms is "theory-thin," because it is driven by data with minimal guidance from mechanistic concepts; thus it is very different from the "theory-thick" approaches summarized above. Any nuclear observable X can be viewed as a mapping from the atomic and neutron numbers Z and N identifying an arbitrary nuclide, to the corresponding value of the observable (the β half-life, in the present study). In machine learning, one attempts to approximate the mapping $(Z, N) \rightarrow X$ based only on an available subset of the data for X, i.e., a body of *training data* consisting of known examples of the mapping. One attempts to *infer* the mapping, in the sense of Bayesian probability theory as expounded by Jaynes [32]. Thus, one is asking the question: "To what extent do the data, and only the data, determine the mapping $(Z, N) \rightarrow X$?" The answer (or answers) to this question should surely be of fundamental interest, when confronted with databases as large, complex, and refined as those existing in nuclear physics.

A learning machine consists of (i) an input interface where, for example, input variables Z and N are fed to the device in coded form, (ii) a system of intermediate elements or units that process the input, and (iii) an output interface where an estimate of the corresponding observable of interest, say, the β half-life T_β appears for decoding. Given an adequate body of training data (consisting of input "patterns" or vectors and their appropriate outputs), a suitable learning algorithm is used to adjust the parameters of the machine, e.g., the weights of the connections between the processing elements in the case of a neural network. These parameters are adjusted in such a way that the learning machine (a) generates responses at the output interface that closely fit the half-lives of the training examples and (b) serves as a reliable predictor of the half-lives of the test nuclei absent from the training set. In the more mundane language of function approximation, the learning-machine model provides a means for *interpolation* or *extrapolation*.

Neural-network models have already been constructed for a range of nuclear properties including atomic masses, neutron separation energies, ground-state spins and parities, and branching probabilities for different decay channels, as well as β^--decay half-lives [30,31,33–36]. Very recently,

global statistical models of some of these properties have also been developed based on support vector machines [37–39]. In time, there has been steady improvement of the quality of these models, such that the documented performance of the best examples approaches or even surpasses that of the traditional "theory-thick" models in predictive reliability. By their nature, they should not be expected to compete with traditional phenomenological or microscopic models in generating new physical insights. However, their prospects for revealing new regularities are by no means sterile, because the explicit formula created by the learning algorithm for the physical observable being modeled is available for analysis.

We present here a new global model for the half-lives of nuclear ground states that decay 100% by the β^- mode, developed by implementing the most recent advances in machine-learning algorithms. Section II describes the elements of the model, the training algorithm employed, steps taken to improve generalization, the data sets adopted, and the coding schemes used at input and output interfaces. Performance measures for assessing the quality of global models of β lifetimes are reviewed in Sec. III. The results of our large-scale modeling studies are reported and evaluated in Sec. IV. Detailed comparisons are made with experiment, with a selection of the theory-driven GT and *pn*QRPA global models, and with previous ANN and SVM models. This assessment is followed by the presentation of specific predictions for nuclei that are situated far from the line of stability, focusing in particular on those involved in r-process nucleosynthesis. Finally, Sec. V summarizes the conclusions of the present study and considers the prospects for further improvements in statistical prediction of half-lives.

II. THE MODEL

A. Network architecture and dynamics

Artificial neural networks, whose structure is inspired by the anatomy of natural neural systems, consist of interconnected dynamical units (sometimes called neurons) that are typically arranged in a distinct layered topology. Also in analogy with biological neural systems, the function of the network, for example, pattern recognition, is determined by the connections between the units. In the work to be reported, we have focused exclusively on feed-forward networks, in which information flows unidirectionally from an input layer through one or more intermediate (hidden) layers to an output layer. Lateral and feedback connections are absent, but otherwise the network is fully connected. The activation of hidden units is nonlinear, whereas the output units transform their inputs linearly. The architecture of such a network is indicated by the notation

$$[I - H_1 - H_2 - \cdots - H_L - O\,|W\,], \tag{1}$$

where I is the number of inputs, H_i is the number of neurons in the i^{th} hidden layer, O is the number of units in the output layer, and W is the total number of parameters needed to complete the specification of the network, consisting of the weights of the connections and the biases of the units. Figure 1 depicts a typical fully connected network of the class used in our statistical modeling, in this case having architecture $[3-5-5-1|56]$.

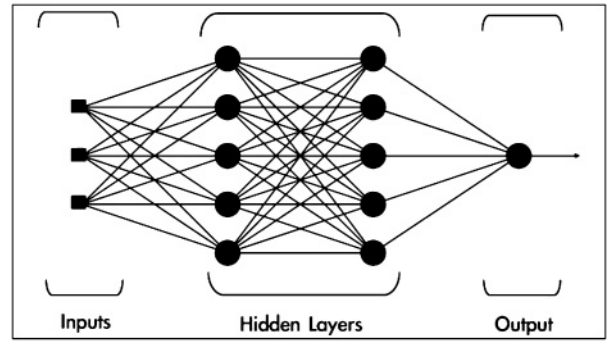

FIG. 1. Architecture of a typical fully connected feed-forward network having an input layer with three units, two hidden layers each, containing five units, and a single output unit, thus of structure $[3-5-5-1|56]$.

The connection from neuron j to neuron i carries a real-number weight w_{ij}. Thus, if o_j is the activity of neuron j, it provides an input $w_{ij}o_j$ to neuron i. In addition, each neuron i is assigned a bias parameter b_i, which is summed together with its input signals from other neurons j to form its total input u_i. This quantity is fed into the activation function φ_i characterizing the response of neuron i. For the neurons in hidden layers, this function is taken to have the nonlinear *hyperbolic tangent* form

$$\varphi(u) = \frac{2}{1+\exp(-2u)} - 1, \tag{2}$$

while for the neurons in the output layer the *symmetric saturating linear* form

$$\varphi(u) = \begin{cases} -1, & u < -1 \\ u, & -1 \leqslant u \leqslant 1, \\ 1, & u > 1 \end{cases} \tag{3}$$

is adopted. The *output* (or *activity*) o_i of neuron i is given by

$$o_i = \varphi\left(b_i + \sum\nolimits_j w_{ij}o_j\right). \tag{4}$$

We note that with its sign reversed, a neuron's bias can be viewed as a threshold for its activation. Also, it is sometimes convenient to regard the bias b_i as the *weight* of a connection to neuron i from a virtual unit v that is always fully "on," i.e., $o_v \equiv 1$. The weights w_{ij} and biases b_i are *adjustable* scalar parameters of the untrained network, available for optimization of the network's performance in some task, notably classification and function approximation in the case of applications to global nuclear modeling. This is usually done by minimizing some measure of the errors made by the network in response to inputs corresponding to a set of training examples or "training patterns."

The dynamics of the network is exceptionally simple. When a pattern p is presented at the input, the system computes a response according to two rules:

(a) The states of all neurons within a given layer, as specified by the outputs o_i of Eq. (4), are updated in parallel, and

(b) The layers are updated successively, proceeding from the input to the output layer.

In modeling the systematics of β lifetimes with this approach, we apply a supervised learning algorithm to optimize the weights and biases, as described in the subsections to follow. The patterns p to be learned or predicted, examples of the mapping from nuclide to lifetime, take the form

$$\left\{ \begin{pmatrix} Z_p \\ N_p \end{pmatrix}, \log_{10} T^p_{\beta,\mathrm{exp}} \right\}, \tag{5}$$

and thus consist of an association between the atomic and neutron numbers of the parent nuclide, with the base-10 log of the experimental half-life $T^p_{\beta,\mathrm{exp}}$. It is of course natural to work with the logarithm of T_β, because the observed values of T_β itself vary over many orders of magnitude.

According to the nature of statistical estimation, realized here in the application of machine learning techniques to function approximation, a neural network model is only one form in which empirical knowledge of a physical phenomenon of interest (β decay in this case) may be encoded [27]. As indicated in the introduction, the present work is at some level an investigation of the degree to which the available data determine the physical mapping from Z and N to the corresponding β-decay half-life. Actually, we do not have knowledge of the exact functional relationship involved. Thus we should write

$$\log_{10} T_\beta(Z, N) = g(Z, N) + \varepsilon(Z, N), \tag{6}$$

where $g(Z, N)$ is a function that decodes the decay systematics and ε is a random expectation error—a Gaussian noise term that represents our ignorance about the dependence of T_β on Z and N. From a heuristic perspective beyond strict mathematical definitions, this ε noise term could reflect "chaotic" influences on the phenomenon, along with missing regularities that could be more easily modeled and eventually included in the estimate of the physical quantity T_β.

The pragmatic objective of the training process in this application will be to minimize the sum of squared errors e_p committed by the network model relative to experiment, for the n patterns p from the available experimental data (D) that constitute the training set

$$E_D = \sum_{p=1}^{n} (e_p)^2 = \sum_{p=1}^{n} \left(\log_{10} T^p_{\beta,\mathrm{exp}} - \log_{10} T^p_{\beta,\mathrm{calc}} \right)^2. \tag{7}$$

Here $\log_{10} T^p_{\beta,\mathrm{calc}}$ is the neural-network output for pattern (nuclide) p, whereas $\log_{10} T^p_{\beta,\mathrm{exp}}$ is the target output. This quantity is often referred to as a *cost function* or *objective function* and can obviously be used as a measure of network performance. In practice, its form will be modified in subsection C2 to improve the network's ability to generalize or predict. A network model is said to generalize well if it performs well for inputs (nuclides) outside the training set, with the mean-square error for these "fresh" nuclei providing an appropriate measure of predictive performance.

B. The training algorithm

In supervised learning, the network is exposed, in succession, to the input patterns (nuclides) of the training set, and the errors made by the network are recorded. One pass through the training set is called an *epoch*. In *batch* training, weights and biases are incremented after each epoch according to a suitable learning algorithm, with the expectation of improving subsequent performance on the training set.

Statistical modeling inevitably involves a trade-off between closely fitting the training data and reliability in interpolation and extrapolation [27,28]. In the present application, it is not the goal of network training to achieve an exact reproduction, by the model, of the known half-lives. This would necessarily entail fitting the data precisely with a large number of parameters—which would in general require a complex ANN with many layers and/or neurons/layer. Obviously, there is no point in constructing a lookup table of the known β half-lives. Rather, the goal is to achieve an accurate representation of the regularities inherent in the training data by means of a network that is no more complicated than it need be, thereby promoting good generalization.

We employ a training algorithm within the general class of back-propagation learning procedures. There are now quite a number of well-tested procedures in this class, including steepest-descent, conjugate-gradient, Newton, and Levenberg-Marquardt training algorithms [26]. All of these approaches aim to minimize an appropriate cost function with respect to the network weights and biases. The term back-propagation refers to the process by which derivatives of network errors with respect to weights/biases can be computed starting from the output layer and proceeding backward toward the input. In general, the Levenberg-Marquardt back-propagation (LMBP) algorithm will have the fastest convergence in function approximation problems, an advantage that is especially noticeable if very accurate training is required [40].

In the Newton method, minimization of the cost function is accomplished through the update rule

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mathbf{H}_k^{-1} \mathbf{g}_k, \tag{8}$$

where $\mathbf{w}_k$ is the vector formed from the weights and biases, $\mathbf{H}_k$ is the Hessian matrix (the matrix of second derivatives of the objective function E_D with respect to the weights and biases) and $\mathbf{g}_k$ is the gradient of E_D at the current epoch k. As a Newton-based procedure attempting to approximate the Hessian matrix, the Levenberg-Marquardt algorithm [26,41] was designed to approach second-order training speed without having to compute second derivatives. When the cost function has the form of Eq. (7), the Hessian matrix for nonlinear networks can be approximated as

$$\mathbf{H} \approx \mathbf{J}^{\mathrm{T}} \mathbf{J}, \tag{9}$$

where J is the Jacobian matrix composed of the first derivatives of the network errors with respect to the weights/biases. This generates a $W \times W$ matrix, where W is the number of the free parameters (weights and biases) of the network. The gradient $\mathbf{g}$ can be computed as

$$\mathbf{g} = \mathbf{J}^{\mathrm{T}} \mathbf{e}, \tag{10}$$

where **e** is the vector whose components are the network errors e_p. [As in Eq. (7), the network error for a given input pattern is the target value of the estimated quantity, minus the value produced by the network.]

Adopting the Gauss-Newton approximation (9), the Levenberg-Marquardt algorithm then adjusts the weights according to the Newton-like updating rule

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \left[\mathbf{J}_k^{\mathrm{T}}\mathbf{J}_k + \mu_k \mathbf{I}\right]^{-1} \mathbf{J}_k^{\mathrm{T}}\mathbf{e}_k, \tag{11}$$

where **I** is the unit matrix.

The factor μ_k appearing in the Eq. (11) is an adjustable parameter that controls the step size so as to quench oscillations of the cost function near its minimum. When μ_k is very small, LMBP coincides with the Newton method executed with the approximate Hessian matrix. When μ_k is large enough, matrix **g** in Eq. (10) is nearly diagonal and the algorithm behaves like a steepest-descent method with a small step size. Steepest-descent algorithms are based on linear approximation of the cost function, while the Newton algorithm involves quadratic approximation. Newton's method is faster and more accurate near an error minimum. Therefore the preferred strategy is to shift toward Newton's method as quickly as possible. To this end, μ_k is decreased after each successful step and is increased only when a tentative step would raise the cost function. In this way, the cost function will always be reduced at each iteration of the algorithm. The algorithm begins with μ_k set to some small value (e.g., $\mu_k = 0.01$). If a step does not yield a smaller value for the cost function, the step is repeated with μ_k multiplied by some factor $\theta > 1$ (e.g., $\theta = 10$). Eventually the cost function should decrease. If a step does produce a smaller value for the cost function, then μ_k is divided by θ for the next step, so that the algorithm will approach Gauss-Newton, which should provide faster convergence. Thus, the Levenberg-Marquardt algorithm implements a favorable compromise between slow but guaranteed convergence far from the minimum and a fast convergence in the neighborhood of the minimum.

The key step in LMBP algorithm is the computation of the Jacobian matrix. To perform this computation we use a variation of the classical back-propagation algorithm. In the standard back-propagation procedure, one computes the derivatives of the squared errors with respect to the weights and biases of the network. To create the Jacobian matrix we need to compute the derivatives of the errors, instead of the derivatives of their squares, a trivial difference computationally.

C. Improving generalization

To build a viable statistical model, it is imperative to avoid the phenomenon of *overfitting*, which, for example, occurs when, under excessive training, the network simply "memorizes" the training data and makes a lookup table. Such a network fails to learn the regularities of the target mapping that are inherent in the data; the network is therefore deficient in generalization. We seek to avoid overfitting through a combination of well-established techniques, namely *cross-validation* [27] and *Bayesian regularization* [42].

1. Cross-validation

Cross-validation is a standard statistical technique based on dividing the data into three subsets [27]. The first subset is the *learning* or *training set* employed in building the model (i.e., in computing the Jacobians and updating the network weights and biases). The second subset is the *validation set*, used to evaluate the performance of the model outside the training set and guide the choice of model. The error on the validation set is monitored during the training process. When the network begins to overfit the data, the error on the validation set will typically begin to rise. If this continues to occur for a specified number of iterations, the training is stopped, and the weights and biases at the minimum of the validation error are reinstated. The third subset is the *test set*. The error on the test set is not used during the training procedure, but it is used to assess the generalization performance of the model and to compare different models. While effective in suppressing overfitting, cross-validation tends to produce networks whose response is not sufficiently smooth. This is dealt with by performing Bayesian regularization together with cross-validation.

2. Bayesian regularization

The standard Levenberg-Marquardt algorithm aims to reduce the sum of squared errors E_D, written explicitly in Eq. (7) for the β-decay problem. However, in the framework of Bayesian regularization [42], the Levenberg-Marquardt *optimization* (back-propagation) algorithm (denoted LMOBP) minimizes a linear combination of squared errors and squared network parameters,

$$F = \tilde{\beta} E_D + \tilde{\alpha} E_W, \tag{12}$$

where E_W is the sum of squares of the network weights (including biases). The multipliers $\tilde{\alpha}$ and $\tilde{\beta}$ are hyperparameters defined by

$$\tilde{\alpha}_k = \frac{\gamma_k}{2E_W} \qquad \text{and} \qquad \tilde{\beta}_k = \frac{n - \gamma_k}{2E_D}, \tag{13}$$

where

$$\gamma_k = W - 2\tilde{\alpha} \cdot \mathrm{tr}\,(\mathbf{H}_k)^{-1}, \tag{14}$$

is the number of parameters (weights and biases) that are being effectively used by the network, n is the number of errors, W is the total number of parameters characterizing the network model [see Eq. (1)], and $\mathbf{H} = \nabla^2 F$ is the Hessian matrix evaluated for the extended ("regularized") objective function (12). The full Hessian computation is again bypassed using the Gauss-Newton approximation, writing

$$\mathbf{H}_k = \tilde{\beta}_k \nabla^2 E_D + \tilde{\alpha}_k \nabla^2 E_W \approx 2\tilde{\beta}_k \mathbf{J}_k^{\mathrm{T}}\mathbf{J}_k + 2\tilde{\alpha}_k \mathbf{I}. \tag{15}$$

Thus, the Levenberg-Marquardt optimization algorithm updates the weights/biases by means of the rule

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \left[\tilde{\beta}_k \mathbf{J}_k^{\mathrm{T}}\mathbf{J}_k + (\mu_k + \tilde{\alpha}_k)\,\mathbf{I}\right]^{-1} \left(\tilde{\beta}_k \mathbf{J}_k^{\mathrm{T}}\mathbf{e}_k + \tilde{\alpha}_k \mathbf{w}_k\right). \tag{16}$$

A detailed discussion of the use of Bayesian regularization in combination with the Levenberg-Marquardt algorithm can be found in Ref. [43].

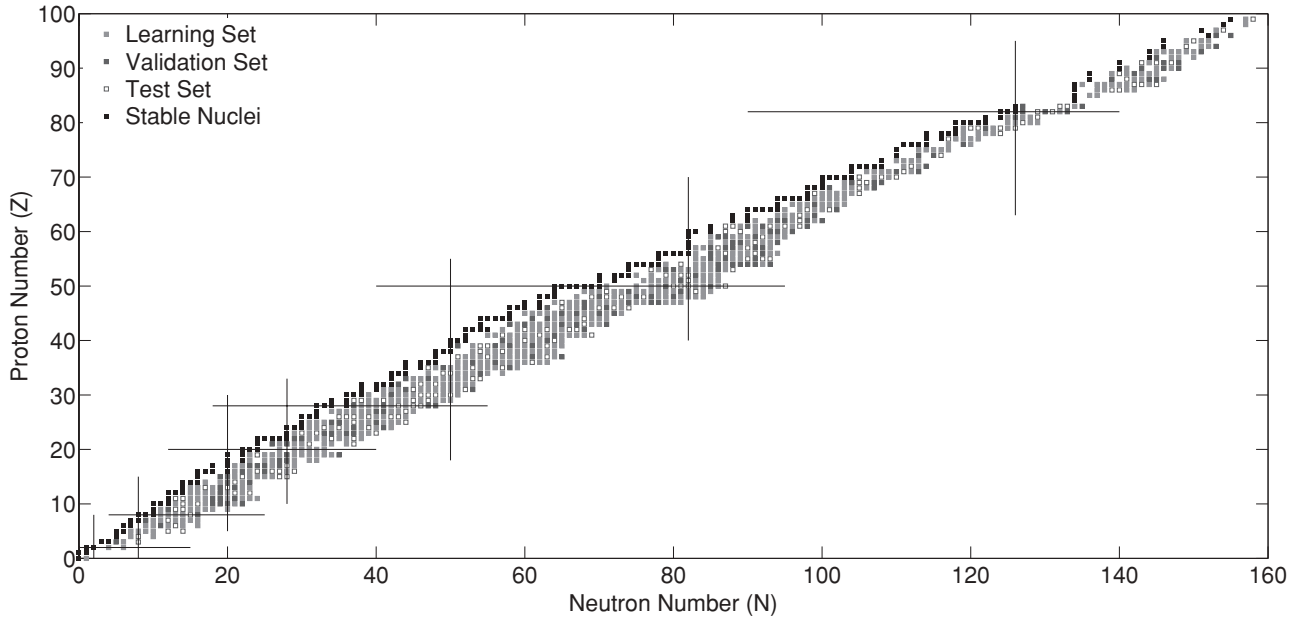

FIG. 2. Partitioning of the nuclides of the data set NuSet-B into learning, validation, and test sets, as viewed in the N-Z plane. Stable nuclides are also indicated.

D. Training mode

Back-propagation learning, as a technique for iterative updating of network parameters, can be executed in either the *batch* or *pattern-by-pattern* (or "on-line") mode. In the on-line mode, a pattern is presented to the network and its response recorded; the Jacobian matrix is then computed and the weights/biases updated *before* the next pattern is presented. In the batch mode, however, calculation of the Jacobian and parameter updating is performed only after all training examples have been presented, i.e., at the end of each epoch. The model results reported here are based on the batch mode, the choice being made on the empirical basis of findings from a substantial number of computer experiments carried out with both strategies.

E. Data sets

The experimental data used in developing ANN models of β-decay systematics have been taken from the Nubase2003 evaluation [44] of nuclear and decay properties carried out by Audi *et al.* at the Atomic Mass Data Center. Restricting attention to those cases in which the ground state of the parent decays 100% through the β^- channel, we form a subset of the β-decay data denoted by NuSet-A, consisting of 905 nuclides sorted by half-life. The half-lives of nuclides in this set range from 0.15×10^{-2} s for ^{35}Na to 2.43×10^{23} s for ^{113}Cd. Of these NuSet-A nuclides, 543 (60%) have been chosen, at random with a uniform probability, to form the training set, and 181 (20%) of those remaining have been similarly chosen to form the validation set. The residual 181 (20%) are reserved for testing the predictive capability of the models constructed. Such partitioning of the NuSet-A database (uniform selection) was implemented to ensure that the distribution over half-lives in the whole set is faithfully reflected in the learning, validation, and test sets. Figure 2 shows an example of the results of this procedure, as viewed in the $Z - N$ diagram in the case of NuSet-B defined below.

We also formed a more restricted data set, called NuSet-B, by eliminating from NuSet-A those nuclei having half-life greater than 10^6 s. The half-lives in this subset, which consists of 838 nuclides, range from 0.15×10^{-2} s for ^{35}Na to 0.20×10^6 s for ^{247}Pu. Histograms depicting the lifetime distribution of the NuSet-B nuclides are shown in Fig. 3, having made a uniform subdivision of the data into learning, validation, and test sets, consisting respectively of 503 (∼60%), 167 (∼20%), and 168 (∼20%) examples. Having excluded the few long-lived examples from NuSet-A (situated to the left of the vertical line in Fig. 3), one is then dealing with a more homogeneous collection of nuclides, a property that facilitates the training of network models. Accordingly, we have focused our efforts on

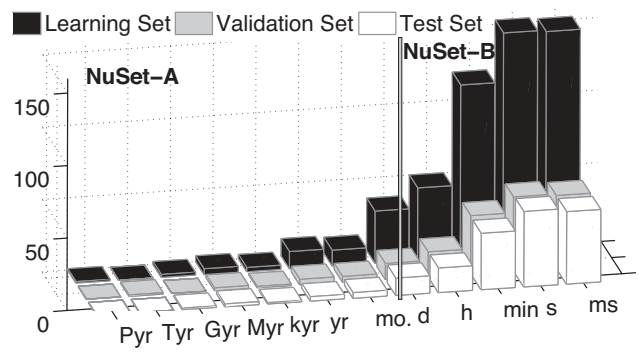

FIG. 3. Distribution of half-lives over time scale for NuSet-A nuclides. NuSet-B nuclides lie to the right of the vertical gray rectangle.

NuSet-B. Table VIII gives information on the distribution of NuSet-B nuclides with respect to the even versus odd character of Z and N.

When considering the performance of a network model for examples taken from the whole data set (whether NuSet-A or NuSet-B), we speak of operation in the *Overall Mode*. Similarly, we speak of operation in the *Learning*, *Validation*, and *Prediction Modes* when studying performance on the learning, validation, and test sets, respectively.

F. Coding schemes at input and output interfaces

In our initial experiments in the design of ANN models for β-decay half-life prediction, we employed input coding schemes that involve only the proton number Z and the neutron number N. To keep the number of weights to a minimum, we make use of analog (i.e., *floating-point*) coding of Z and N through two dedicated inputs, whose activities represent scaled values of these variables. The LMOBP algorithm works better when the network inputs and targets are scaled to the interval $[-1, 1]$ than (say) the interval $[0, 1]$ [26]. Moreover, the range of the hyperbolic tangent activation function employed by the hidden units lies in the interval $-1 \leqslant \varphi(u) \leqslant 1$. The ranges $[0, 230]$ of both Z and N are therefore scaled to fall in this range to avoid overflow conditions. The base-10 log of the β^- half-life $T_{\beta,\mathrm{calc}}$, as calculated by the network for input nuclide (Z_p, N_p), is represented by the activity of a single analog output unit. For the same reason as indicated for the input units, the range $[0.17609, 8.9771]$ of the target values $\log_{10} T^p_{\beta,\mathrm{exp}}$ is scaled again to the interval $[-1, 1]$.

Also in the primary stages of our study of β-half-life systematics, we have assumed that the half-life of a given nucleus is properly given by an expression of the form of Eq. (6). Such an expression echoes the essence of Weizsacker's semiempirical mass formula based on the liquid-drop model, with the binding energy given by a function $B(Z, N)$ representing a statistical estimate of the physical quantity, plus an additive noise term.

Taking Z and N as the only inputs to the inference machine formed by the neural network has, of course, the logistical advantage that there is no limitation to the range of prediction of nuclear properties across the nuclear landscape. If, however, such quantities as Q values and neutron separation energies (S_n) were included as inputs, one would have to calculate these quantities for choices of (Z, N) at which experimental values are not available. But this implies a departure from the "ideal" of determining the physical mapping from (Z, N) to the target nuclear property, based *only* on the existing body of experimental data for that property. The predictions of the network model would necessarily be contingent on some theoretical model to provide the additional values of the input quantities.

However, estimating a given nuclear property—the log lifetime of β decay in the present case—as a smooth function of Z and N has clear limitations. The nuclear data itself sends strong messages of the importance of pairing, shell effects ("quantal effects"), associated with the integral nature of Z and N. The problem of atomic masses provides the classic example: the liquid-drop formula must be supplemented by pairing and shell corrections to account for the existence of different mass surfaces for even-even, odd-A, and odd-odd nuclei and other effects of the integral/particulate character of Z and N.

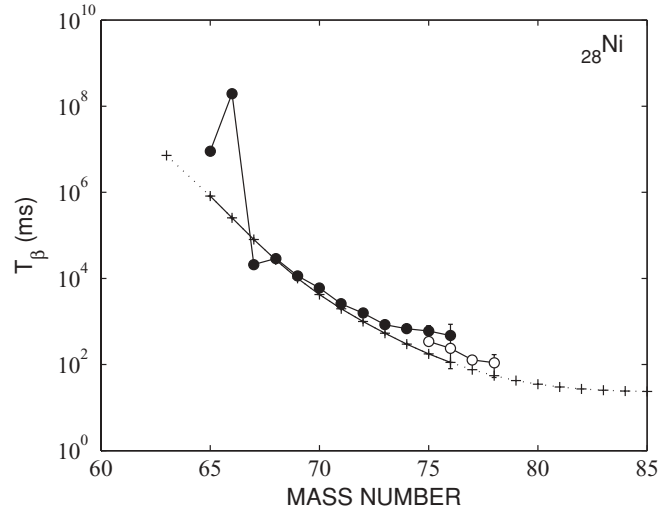

FIG. 4. Plot showing calculated and experimental β^--decay half-lives for the $_{28}$Ni isotopic chain. (Solid dots) Experimental data points. (Unfilled dots) New and more precise experimental half-lives recently deduced by Hosmer *et al.* [45]. Pluses: results generated by the $[2-5-5-5-5-1|111]$ ANN model with inputs (Z, N). Solid lines trace the calculated values of the overall mode (learning, validation, and test sets), while dotted lines trace extrapolated values produced by the model.

Examination of results from the simple coding scheme with Z and N alone serving as analog inputs is nevertheless instructive. We have applied the LMBP training algorithm to develop a network model with architecture $[2-5-5-5-5-1|111]$. As shown in Fig. 4, the model yields a smooth curve that represents a gross fit of the experimental data involved. The predictive ability of the model naturally relies on extrapolation based on this curve. These results demonstrate the need for a more refined model within which quantal effects such as pairing and shell structure are given an opportunity to exert themselves, so the natural fluctuations are followed in validation and prediction modes, as well as in the learning (or "fitting") phase.

A straightforward modification of the input interface of the network model that can at least partially fulfill this need is suggested by the extension of the liquid-drop model to include a pairing-energy term. In addition to the two input units representing Z and N as floating-point numbers, we introduce a third input unit representing a discrete parameter analogous to the pairing constant, namely

$$\delta = \begin{cases} +1, & \text{for e} - \text{e nuclei,} \\ 0, & \text{for o} - \text{mass nuclei,} \\ -1, & \text{for o} - \text{o nuclei,} \end{cases} \tag{17}$$

which distinguishes between even-Z-even-N, odd-A, and odd-Z-odd-N nuclides. This simple refinement has the conceptual benefit of remaining in the spirit of stand-alone, "theory-thin"

modeling, driven purely by data rather than data plus physical intuition and accepted theory.

The only information required is a knowledge of the Z and N values and (redundantly yet importantly) their even/odd parities. The expression replacing Eq. (6) as a representation of the inference process performed by the ANN model is evidently

$$\log_{10} T_\beta(Z, N) = \tilde{g}(Z, N, \delta) + \tilde{\varepsilon}(Z, N). \tag{18}$$

We note that it has proven advantageous in global statistical models of nuclear mass excess [35] to introduce two binary input units that encode the even/odd parity of Z and N.

We shall see that residual shell effects that may impact the behavior of half-lives for both allowed and/or forbidden transitions can be achieved to some extent through the δ input defined in Eq. (17).

G. Initialization of network parameters

Proper initialization of the free parameters of the ANN (its weights and biases) is a very important and highly nontrivial task. One needs to choose an initial point on the error surface defined by Eqs. (7) and (12) as close as possible to its global minimum with respect to these parameters and such that the output of each neuronal unit lies within the sensitive region of its activation function ϕ. We adopt a method devised by Nguyen and Widrow [46], in which the initial weights are selected so as to distribute the active region of each neuron (its "receptive field" neurobiological parlance) approximately evenly across the input space of the layer to which that neuron belongs. The Nguyen-Widrow method has clear advantages over more naive initializations in that all neurons begin operating with access to good dynamical range and all regions of the input space receive coverage from neurons. Consequently, training of the network is accelerated.

III. PERFORMANCE MEASURES

The performance of the models we have been developing is assessed in terms of several commonly used statistical measures, namely, the *root-mean-square error* ($\sigma_{\rm rms}$), the *mean absolute error* ($\sigma_{\rm ma}$), and the *normalized-mean-square error* ($\sigma_{\rm nms}$). For any given data set, these quantities provide overall measures of the deviation of the calculated values $y_i \equiv \log_{10} T_{\beta,\rm calc}$ of the log-half-life produced by the model for nuclide i, from the corresponding experimental value $\hat{y}_i \equiv \log_{10} T_{\beta,\rm exp}$. To understand the network's response in more detail, a *linear regression analysis* (LR) is also carried out in which the correlation between experimental and calculated half-life values is evaluated in terms of the correlation coefficient (R value). Definitions of these quantities follow, with n standing for the total number of nuclides in each case (the full data set or one of its subsets: the learning, validation, or test set).

Root-mean-square error

$$\sigma_{\rm rms} = \left[\frac{1}{n}\sum_{p=1}^{n}(y_p - \hat{y}_p)^2\right]^{1/2}. \tag{19}$$

Normalized-mean-square error

$$\sigma_{\rm nms} = \frac{\sum_{p=1}^{n}(y_p - \hat{y}_p)^2}{\sum_{p=1}^{n}(y_p - \bar{y}_p)^2}. \tag{20}$$

Mean absolute error

$$\sigma_{\rm ma} = \frac{1}{n}\sum_{p=1}^{n}|y_p - \hat{y}_p|. \tag{21}$$

Those models having smaller values of $\sigma_{\rm rms}$ and $\sigma_{\rm ma}$, and $\sigma_{\rm nms}$ closer to unity, are favored.

Linear regression (LR)

$$y_p = a\hat{y}_p + b. \tag{22}$$

In linear regression, the slope a and the intercept b are calculated, as well as the correlation coefficient

$$R = \frac{\sum_{p=1}^{n} Y_p \hat{Y}_p}{\left[\sum_{p=1}^{n}(Y_p - \langle Y_p\rangle)^2 \sum_{p=1}^{n}(\hat{Y}_p - \langle \hat{Y}_p\rangle)^2\right]^{1/2}}, \tag{23}$$

where $Y_p = y_p - \langle y\rangle$ and $\hat{Y}_p = \hat{y}_p - \langle \hat{y}\rangle$. Values of R greater than 0.8 indicate strong correlations.

The above indices necessarily provide only gross assessments of the quality of our models. In the literature on global modeling of β^- half-lives, several additional indices, perhaps more appropriate to the physical context, have been used to analyze performance. The collaboration led by Klapdor [11–16] has employed the quality measure

$$\bar{x}_K = \frac{1}{n}\sum_{p=1}^{n} x_p, \tag{24}$$

wherein

$$x_p = \begin{cases} T_{\beta,\rm exp}/T_{\beta,\rm calc}, & \text{if } T_{\beta,\rm exp} \geqslant T_{\beta,\rm calc} \\ T_{\beta,\rm calc}/T_{\beta,\rm exp}, & \text{if } T_{\beta,\rm exp} < T_{\beta,\rm calc}, \end{cases} \tag{25}$$

along with the corresponding standard deviation

$$\sigma_K = \left[\frac{1}{n}\sum_{p=1}^{n}(x_p - \bar{x}_K)^2\right]^{1/2}. \tag{26}$$

Again, the sums run over the appropriate set of nuclides. Perfect accuracy is attained when $\bar{x}_K = 1$ and $\sigma_K = 0$.

In a more incisive assessment, also pursued by Klapdor and coworkers, one calculates the percentage m of nuclides having measured ground-state half-life $T_{\beta,\rm exp}$ within a prescribed range (e.g., not greater than 10^6, 60, or 1 s), for which the half-life generated by the model is within a prescribed tolerance factor f (in particular, 2, 5, or 10) of the experimental value.

A measure M similar to $\bar{x}_K$, but defined in terms of $\log_{10} T_\beta$ rather than T_β, has been used by Möller and collaborators

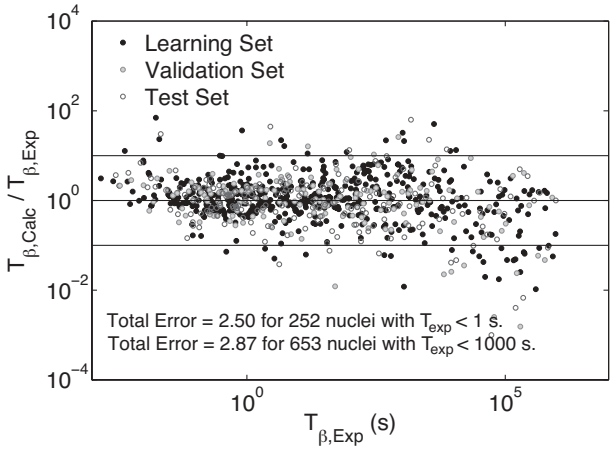

FIG. 5. Ratios of calculated to experimental half-life values for nuclides in the learning (black), validation (gray), and test (white) sets selected from NuSet-B, plotted versus half-life $T_{\beta,\mathrm{exp}}$. Calculated values are generated by the standard ANN model of this work with architecture $[3-5-5-5-5-1|116]$. Total error equals $\Sigma^{(10)}$ [see Eq. (32)].

[19,20]; specifically,

$$M = \frac{1}{n}\sum_{p=1}^{n} r_p, \tag{27}$$

where $r_p = y_p/\hat{y}_p$. This quantity gives the average position of the points in Fig. 5 for the respective data sets. Its associated standard deviation

$$\sigma_M = \left[\frac{1}{n}\sum_{p=1}^{n}(r_p - M)^2\right]^{1/2}, \tag{28}$$

is also examined, and the "total" error of the model for the data set in question is taken to be

$$\Sigma = \left[\frac{1}{n}\sum_{p=1}^{n}(y_p - \hat{y}_p)^2\right]^{1/2}, \tag{29}$$

which is the same as the σ_{rms} defined in Eq. (19). Model quality is also expressed in terms of exponentiated versions of these last three quantities, namely the mean deviation range

$$M^{(10)} = 10^{M}, \tag{30}$$

the mean fluctuation range

$$\sigma_{\mathrm{M}^{(10)}} = 10^{\sigma_M}, \tag{31}$$

and total error range $\Sigma^{(10)}$:

$$\Sigma^{(10)} = 10^{\Sigma}. \tag{32}$$

Superior models should have Σ, M, and σ_M near zero and $M^{(10)}$, $\sigma_M^{(10)}$, and $\Sigma^{(10)}$ near unity. Again, in a closer analysis of model capabilities, these indices are evaluated within prescribed half-life domains.

IV. RESULTS AND DISCUSSION

As already indicated, statistical modeling of β^--decay systematics is more effective when the range of lifetimes considered is more restricted. Accordingly, the following detailed presentation and analysis will focus on the properties and performance of the best ANN model developed using the NuSet-B database, which is restricted to nuclides with β^- half-life below 10^6 s. The quality of this model will be compared, in considerable detail, with that of traditional theoretical global models cited in the introduction, earlier ANN models, and models provided by another class of learning machines (SVMs).

After a large number of computer experiments on networks developed with different architectures, input/output coding schemes, activation functions, initialization prescriptions, and training algorithms [47], we have arrived at an ANN model well suited to approximate reproduction of the observed β^--decay half-life systematics and prediction of half-lives of nuclides unfamiliar to the network. The preferred network is of architecture $[3-5-5-5-5-1|116]$. The hyperbolic tangent sigmoid is taken as the activation function of neurons in hidden layers, and a saturated linear function is adopted in the output layer. In training, the techniques for improving generalization that were described in Sec. II, namely Bayesian regularization and cross-validation, were implemented in combination with the Levenberg-Marquardt optimization algorithm (LMOBP) and the Nguyen-Widrow initialization method. The network was taught in batch mode and the training phase was continued for 696 epochs. Of the 116 degrees of freedom corresponding to the network weights and biases, 98 survive the training process; this is the value of the number γ_k defined in Eq. (14).

A. Comparison with experiment

In this subsection, we evaluate the performance of our ANN model by direct comparison with the available experimental data. Table I collects results for the overall quality measures [Eqs. (19)–(21)] commonly used in statistical analysis as well as the values of the correlation coefficient R [see Eq. (23)]. We may quote for comparison the root-mean-square errors of 1.08 (learning mode) and 1.82 (prediction mode) obtained in an earlier ANN model of β-decay systematics [33].

These overall measures are silent with respect to specific physical merits or shortcomings of the model. However, such information can be revealed by suitable plots of the results from applications of the model, as exemplified in Figs. 5–9.

TABLE I. Performance measures for the learning, validation, test, and whole sets, achieved by the favored ANN model of β^--decay half-lives, a network with architecture $[3-5-5-5-5-1|116]$ trained on nuclides from NuSet-B.

Performance measure	Learning set	Validation set	Test set	Whole set
σ_{rms}	0.53	0.60	0.65	0.57
σ_{nms}	1.004	0.995	1.012	0.999
σ_{ma}	0.38	0.41	0.46	0.40
R value	0.964	0.953	0.947	0.958

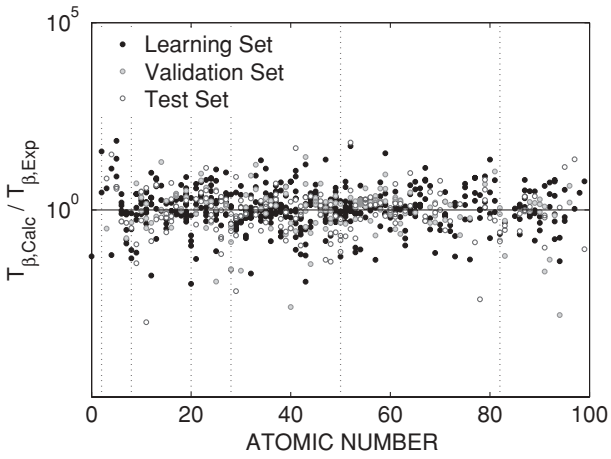

FIG. 6. Same as Fig. 5, but ratios of calculated to experimental half-lives are plotted against the atomic number Z. The dashed lines indicate the magic numbers.

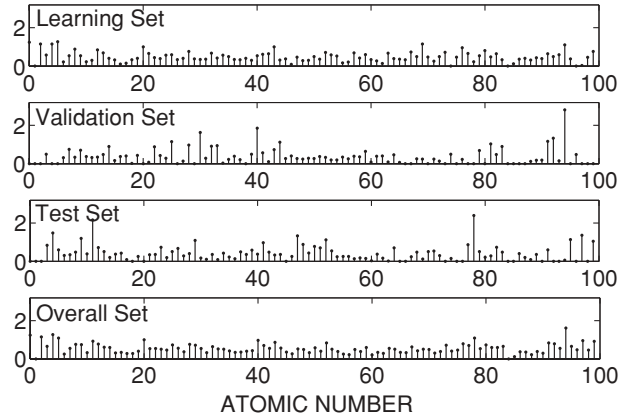

FIG. 8. Values of $\sigma_{\rm rms}$ obtained in each isotopic chain with the favored $[3-5-5-5-5-1|116]$ network model, respectively, for the learning, validation, and test sets, and the full NuSet-B database, plotted against atomic number Z.

Figures 5 and 6 present the ratios of calculated to experimental half-life values. The deviations from the measured values are clearly visible as departures from the solid line $T_{\beta,\rm calc}/T_{\beta,\rm exp}=1$. Both figures show that the model response follows the general trend of the experimental half-lives. The increased scatter of the points at longer lifetimes is presumably due to the sparse representation of long-lived β emitters in the training data. However, shell effects are included in the right direction as shown in Figs. 6–8. The accuracy of model output versus distance from stability can be inferred from Fig. 7. The local isotopic $\sigma_{\rm rms}$ (Fig. 8) and the absolute deviation of calculated from experimental $\log_{10} T_\beta$ values (Fig. 7) indicate a balanced behavior of network response in all β^--decay regions. However, Fig. 7 shows that some less accurate results are obtained very near the β-stability line, a feature also present in the traditional models of Refs. [15,20]. For nuclei with very small or very large mass values there are no significant deviations. Last, the regression analysis we have performed, in which linear fits are made for the learning, validation, and test sets as well as the full NuSet-B database, serves to demonstrate in a different way the relatively small discrepancies between calculated and observed β^--decay half-lives (see Fig. 9). Moreover, the resultant R values (see Table I) imply that the observed

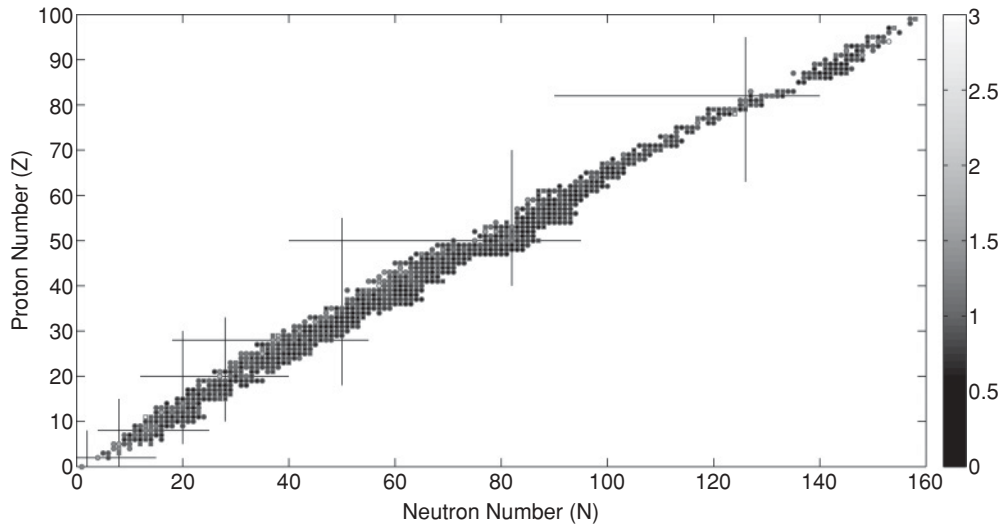

FIG. 7. Absolute errors, relative to experiment, of the calculated β-decay half-lives of all nuclides (p) in the full NuSet-B database plotted versus proton and neutron numbers Z and N for the $[3-5-5-5-5-1|116]$ network model. The bar on the right indicates the mapping from the absolute error values $|e_p| = |\log_{10} T^p_{\beta,\rm exp} - \log_{10} T^p_{\beta,\rm calc}|$ to the gray scale. Test nuclides are indicated as squares.

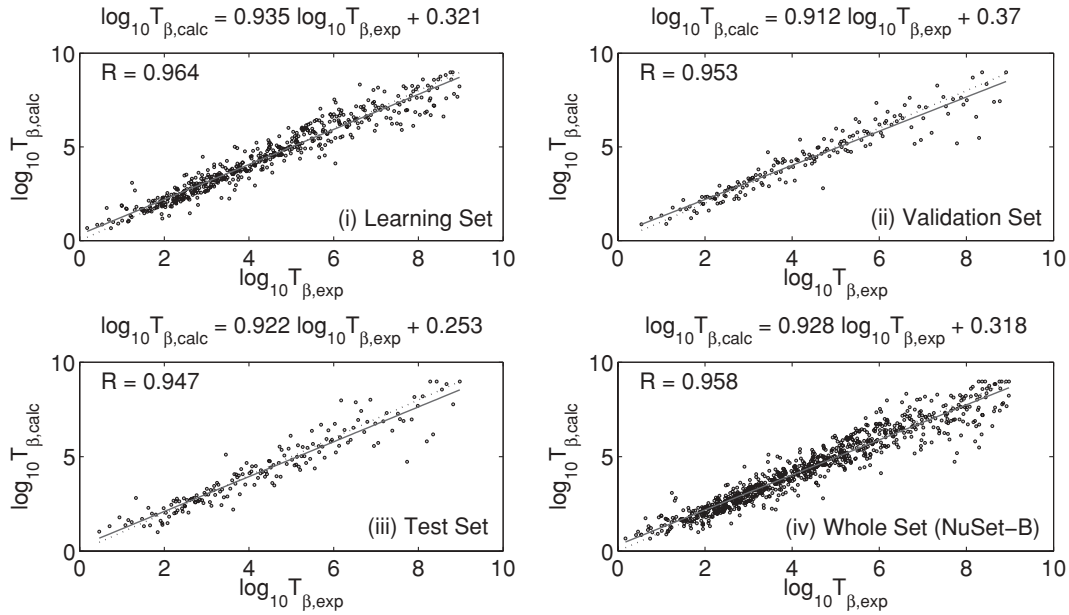

FIG. 9. Regression analysis for the (i) learning, (ii) validation, (iii) test sets (prediction mode) and for the (iv) full database (overall mode). Solid lines represent exact agreement with the data (i.e., $\log_{10} T_{\beta,\mathrm{calc}} = \log_{10} T_{\beta,\mathrm{exp}}$), while dashed lines indicate the corresponding best linear fits. The dataset used is NuSet-B. The units used for T_β are ms. The corresponding values of the parameters a and b of Eq. (22) and the correlation coefficient R of [Eq. (23)] are given in each panel.

systematics is smoothly and uniformly mirrored in the model's responses.

B. Comparison with RPA and GT global models: a detailed analysis

In this subsection, the performance of the favored network model of β^- lifetime systematics is compared with that of prominent theory-thick global models.

Adopting the quality measures [Eqs. (27)–(32)] introduced by Möller and collaborators, we first compare the performance of our global ANN model $[3-5-5-5-5-1\,|116\,]$ with the global microscopic models based on the proton-neutron quasiparticle random-phase approximation (pnQRPA), in particular, the NBCS + pnQRPA model of Homma *et al.* [15] and the FRDM + pnQRPA model of Möller *et al.* [19]. The efficacy of the ANN model is also compared with that of the microstatistical semigross theory (SGT) as implemented by Nakata *et al.* [8]. Table II lists the ANN values for $M^{(10)}$ and $\sigma_M^{(10)}$ specific to odd-odd, odd-A, and even-even nuclides and calculated using the whole and test sets. Table III collects the $M^{(10)}$ and $\sigma_M^{(10)}$ values for the three theory-thick models in the same format. As seen in these tables, both pnQRPA and SGT models tend to overestimate the β^- half-lives of odd-odd nuclei, while the FRDM calculation tends to underestimate the shorter half-lives for even-even and odd mass nuclei. The ANN model, however, tends to overestimate the half-lives of even-even nuclides, although to a smaller degree; this shortcoming is due, at least in part, to the relative scarcity of even-even parents.

Table IV contains values of the performance measures defined in Eqs. (27)–(32) for three global models of β^--decay half-life. Here the entries are not separated according to even-even, odd-A, or odd-odd class membership of the nuclides involved. Included are results for calculations within the FRDM + pnQRPA model, updated to a more recent mass evaluation [20], together with corresponding values for a hybrid "micromacroscopic" pnQRPA + ffGT treatment, which combines the QRPA model of allowed Gamow-Teller β decay with the gross theory of first-forbidden (ff) decay [20]. To permit a direct comparison with the ANN model, we also report in this table the results for ANN performance figures determined independently of the even-even, odd-A, odd-odd nuclidic class distinction, focusing attention only on the subdivision into half-life ranges for both the overall and the prediction mode. The improved FRDM + pnQRPA model underestimates long half-lives, whereas the pnQRPA + ffGT approach slightly underestimates half-lives over the full range considered. The tabulated quality indices indicate that the ANN responses are in closer agreement with experiment more frequently than the FRDM + pnQRPA calculations, while the ANN model and the pnQRPA + ffGT approaches perform about equally well.

The performance of our ANN model may also be evaluated in terms of the quality measures $\bar{x}_K$ and σ_K employed by Klapdor and coworkers and defined in Eqs. (24)–(26). Table V includes values of these quantities for the network model, along with values for the pnQRPA calculation of Staudt *et al.* [13] and for the NBCS + pnQRPA approach of Homma *et al.* [15]. Detailed comparison shows that, judging from these indices, there is only a modest decline in the quality of ANN responses in going from the overall mode to the prediction mode and that the performance of the pnQRPA model is distinctly better than that of the neural network for shorter half-lives but worse for longer half-life values. We note, however, that the pnQRPA model could be regarded

TABLE II. Analysis of the deviation between calculated and experimental β^--decay half-lives of the [3 − 5 − 5 − 5 − 5 − 1|116] standard ANN model in the overall and prediction modes, based on the quality measures $M^{(10)}$ and $\sigma_{M^{(10)}}$ of Eqs. (30) and (31) used by Möller and coworkers. The second column denotes the even/odd character of the parent nucleus in Z and N, while n is the number of nuclides with experimental half-lives lying in the prescribed range (first column).

$T_{\beta,\mathrm{exp}}$ (s)	Class	n	$M^{(10)}$	$\sigma_{M^{(10)}}$
	(a) ANN model, overall mode			
<1	o-o	76	1.04	2.53
	odd	125	1.16	2.25
	e-e	51	1.87	2.45
<10	o-o	121	1.11	2.96
	odd	187	1.10	2.31
	e-e	87	1.65	2.56
<100	o-o	158	1.08	3.06
	odd	261	1.08	2.45
	e-e	110	1.58	2.31
<1000	o-o	191	1.12	3.06
	odd	329	1.07	2.73
	e-e	133	1.63	2.60
$<10^6$	o-o	238	0.93	3.87
	odd	437	0.97	3.67
	e-e	163	1.25	3.44
	(b) ANN model, prediction mode			
<1	o-o	11	0.86	1.98
	odd	32	1.05	2.40
	e-e	7	2.36	3.26
<10	o-o	20	0.86	3.76
	odd	42	0.92	2.61
	e-e	17	1.80	2.58
<100	o-o	28	0.76	3.20
	odd	57	0.97	2.91
	e-e	21	1.58	2.98
<1000	o-o	35	0.78	3.13
	odd	68	0.84	3.07
	e-e	28	1.49	3.04
$<10^6$	o-o	46	0.58	4.71
	odd	87	0.86	4.07
	e-e	35	1.14	4.33

as overparameterized compared to more up-to-date models, because the strengths of the NN interactions are derived from a local fitting of the experimental data in each chain. Turning to the NBCS + *pn*QRPA calculation, it is evident from Table V that the ANN model generally exhibits smaller discrepancies between calculated and observed β^--decay half-lives. For example, the network model has the ability to reproduce approximately 50% of experimentally known half-lives shorter than 10^6 s within a factor of 2. It should be noted, however, that the NBCS + *pn*QRPA model has fewer adjustable parameters [15].

Viewed as a whole, the analyses presented in Tables II–V demonstrate that in a clear majority of cases in which the statistical model of β^- half-lives is presented with test nuclides *absent* from the training and validation sets, it makes *predictions* that are closer to experiment than the corresponding results from traditional models based on quantum many-body theory and phenomenology. This is ascribed to some extent to the larger number of adjustable parameters of the current model.

TABLE III. Same analysis as presented in Table II but instead assessing the quality of traditional theoretical models, corresponding specifically to (a) the NBCS + *pn*QRPA calculation of Homma *et al.* [15], (b) the FRDM + *pn*QRPA calculation of Möller and coworkers [19], and (c) the SGT calculation by Nakata *et al.* [8]. These assessments are limited to nuclides with experimental half-lives below 1000 s.

$T_{\beta,\mathrm{exp}}$ (s)	Class	n	$M^{(10)}$	$\sigma_{M^{(10)}}$
	(a) NBCS + *pn*QRPA calculation [15]			
<1	o-o	28	1.75	4.96
	odd	31	0.60	2.24
	e-e	10	1.15	2.36
<10	o-o	66	1.89	4.60
	odd	81	0.92	3.84
	e-e	34	1.01	2.93
<100	o-o	85	3.15	10.51
	odd	127	1.07	4.29
	e-e	52	1.13	3.58
<1000	o-o	93	3.02	10.25
	odd	157	1.10	5.55
	e-e	63	1.39	6.10
	(b) FRDM + *pn*QRPA calculation [19]			
<1	o-o	29	0.59	2.91
	odd	35	0.59	2.64
	e-e	10	3.84	3.08
<10	o-o	59	0.76	8.83
	odd	85	0.78	4.81
	e-e	34	2.50	4.13
<100	o-o	88	2.33	49.19
	odd	133	1.11	9.45
	e-e	54	2.61	4.75
<1000	o-o	115	3.50	72.02
	odd	194	2.77	71.50
	e-e	71	6.86	58.48
	(c) SGT calculation [8]			
<1	o-o	38	1.45	2.57
	odd	56	1.75	2.32
	e-e	19	2.03	2.30
<10	o-o	83	1.94	4.10
	odd	110	1.71	2.36
	e-e	45	1.58	2.23
<100	o-o	115	2.54	8.86
	odd	174	1.95	3.15
	e-e	64	1.45	2.40
<1000	o-o	144	3.42	15.21
	odd	232	2.36	5.42
	e-e	85	1.38	2.81

TABLE IV. Comparison of values of quality indices characterizing the "theory-thin" standard neural-network model of the present work and two "theory-thick" models developed by Möller and coworkers: ANN model in overall (a) and prediction (b) modes, and (c) FRDM + *pn*QRPA and (d) *pn*QRPA + *ff*GT models of Ref. [20]. The number n of nuclides with experimental half-lives below the prescribed limit is given in the second column. The quality indices labeling columns 3–8 are defined in Eqs. (27)–(32).

$T_{\beta,\mathrm{exp}}$ (s)	n	M	$M^{(10)}$	σ_M	$\sigma_{M^{(10)}}$	Σ	$\Sigma^{(10)}$
			(a) ANN model, overall mode				
<1	252	0.09	1.24	0.39	2.44	0.40	2.50
<10	395	0.08	1.21	0.42	2.60	0.42	2.65
<100	529	0.07	1.17	0.43	2.68	0.43	2.71
<1000	653	0.07	1.18	0.45	2.84	0.46	2.88
$<10^6$	838	0.00	1.01	0.57	3.70	0.57	3.70
			(b) ANN model, prediction mode				
<1	50	0.05	1.12	0.41	2.56	0.41	2.58
<10	79	0.02	1.05	0.48	3.00	0.48	3.01
<100	106	0.00	1.00	0.49	3.08	0.49	3.08
<1000	131	−0.03	0.93	0.50	3.16	0.50	3.17
$<10^6$	168	−0.09	0.82	0.64	4.38	0.65	4.44
			(c) FRDM + *pn*QRPA calculation [20]				
<1	184	0.03	1.06	0.57	3.72	0.57	3.73
<10	306	0.14	1.38	0.77	5.87	0.78	6.04
<100	431	0.19	1.55	0.94	8.81	0.96	9.21
<1000	546	0.34	2.20	1.28	19.09	1.33	21.17
$<10^6$	–	–	–	–	–	–	–
			(d) *pn*QRPA + *ff*GT calculation [20]				
<1	184	−0.08	0.84	0.48	3.04	0.49	3.08
<10	306	−0.03	0.93	0.55	3.52	0.55	3.53
<100	431	−0.04	0.91	0.61	4.10	0.61	4.12
<1000	546	−0.04	0.92	0.68	4.81	0.68	4.82
$<10^6$	–	–	–	–	–	–	–

C. Comparison with prior ANN and SVM models

Some exploratory applications of artificial neural networks to β-decay systematics were carried out earlier by the Athens-Manchester-St. Louis collaboration and reported in Refs. [33,34]. The first of these studies arrived at a fully connected multilayer feedforward ANN model having the simple architecture $[16-10-1|181]$, and the second produced a similar model with architecture $[17-10-1|191]$. Both of these efforts employed binary encoding of Z and N at the input. They used the same experimental data sets, which are somewhat smaller in size than those used in the present work. In both cases, a quite orthodox backpropagation algorithm was applied ("vanilla backprop"), incorporating a momentum term to enhance convergence of the learning process [27]. Online (pattern-by-pattern) rather than batch update of connection weights was implemented. The principal difference between the two earlier ANN models is the addition, in the second, of an analog input unit representing the Q value of the decay. Tables VI and VII present values for performance measures of these ANN models operating in the prediction mode. (We concentrate on this aspect of performance, because it relates directly to the *extrapability* of the models.) For the $[16-10-1|181]$ network model, Table VI displays results for the quality measures used by Klapdor and coworkers, evaluated on the test set. For the $[17-10-1|191]$ model, Table VII gives results for the performance measures of Möller and coworkers, based on the responses of the model to the same test set. On comparison with the entries for $M^{(10)}$ in Table II, one sees that the performance of the 17-input network model is rather similar to that of the present six-layer ANN model, except for odd-odd nuclides—whose lifetimes are overestimated by the older network. In the case of the 16-input model, comparison of the entries for $m\%$ in Tables VI and V provides substantial evidence for the superiority of the new ANN model developed here, although this is not so clearly reflected in the respective $\bar{x}_K$ values.

The improved performance shown by the new ANN model relative to the earlier ANN models can be attributed mainly to the different architecture adopted and the more advanced training procedure that was implemented. Differences in the ways input information is provided to the networks also plays a role, but this aspect of the models is not independent of architecture. The somewhat larger experimental data sets involved in the current study are helpful, but this is not a major factor.

With respect to architecture and input encoding, the new ANN model has noteworthy strategic advantages over its predecessors. The number of degrees of freedom (weight and bias parameters) is reduced considerably in the new model by the use of analog rather than binary encoding of Z and N.

TABLE V. Comparison of performance measures characterizing the standard ANN model of the present work, when operating in the overall (a) and prediction (b) modes, with corresponding values for (c) the *pn*QRPA model of Staudt *et al.* [13] and (d) the NBCS + *pn*QRPA model of Homma *et al.* [15]. The quality indices $m\%$, $\bar{x}_K$, and σ_K are defined by Eqs. (24)–(26). The third column reports the percentage, $m\%$, of nuclides having experimental half-lives within the prescribed range (second column), for which the calculated half-life lies within a certain tolerance factor (first column) of the experimental value.

Factor	$T_{\beta,\mathrm{exp}}$ (s)	$m\%$	$\bar{x}_K$	σ_K
	(a) ANN model, overall mode			
<10	$<10^6$	92.0	2.46	1.72
	<60	96.5	2.21	1.52
	<1	97.6	2.10	1.39
<5	$<10^6$	82.8	1.99	0.95
	<60	90.2	1.88	0.84
	<1	93.7	1.88	0.80
<2	$<10^6$	53.5	1.41	0.27
	<60	60.6	1.41	0.27
	<1	61.9	1.41	0.26
	(b) ANN model, prediction mode			
<10	$<10^6$	90.5	2.69	1.85
	<60	96.1	2.48	1.64
	<1	98.0	2.24	1.30
<5	$<10^6$	79.2	2.10	0.97
	<60	87.3	2.05	0.91
	<1	94.0	2.04	0.89
<2	$<10^6$	49.4	1.48	0.28
	<60	53.9	1.48	0.27
	<1	60.0	1.50	0.27
	(c) *pn*QRPA calculation [13]			
<10	$<10^6$	72.2	1.85	1.21
	<60	96.3	1.67	1.02
	<1	99.1	1.44	0.40
<5	$<10^6$	69.7	1.68	0.76
	<60	94.5	1.56	0.66
	<1	99.1	1.44	0.40
<2	$<10^6$	56.4	1.37	0.29
	<60	82.2	1.36	0.29
	<1	90.6	1.35	0.27
	(d) NBCS + *pn*QRPA calculation [15][a]			
<10	$<10^6$	76.7	3.00	–
	<60	87.2	2.81	–
	<1	95.7	2.64	–
<5	$<10^6$	–	–	–
	<60	–	–	–
	<1	–	–	–
<2	$<10^6$	33.8	1.43	–
	<60	42.0	1.41	–
	<1	50.7	1.43	–

[a] σ_K results are not available in Ref. [15].

Despite the numerous hidden layers in the architecture $[3-5-5-5-5-1|116]$ of the current model, it has 65 parameters less than the 16-input model and 75 less than the 17-input model. Relative to the 16-input model, the presence of the additional input δ is advantageous in providing information on the even/odd character of Z and N. Relative to the 17-input model there is the advantage, conceptual as well as strategic, that the new model does not rely on Q-value input. Experimental Q values are not known for all the nuclides of interest, so the need to call on external theoretical predictions for input variables is eliminated.

Based on a general principle of machine learning, it is to be expected that the reduction in complexity achieved in the current ANN model compared to previous versions will result in improved predictive performance. For an ANN, reduced complexity means reduction in the number of weight and bias parameters. Other things being equal, i.e., if satisfactory accuracy of fit to the training sample is preserved, a model with less computational complexity is less susceptible to overfitting and superior in generalization [27].

TABLE VI. Performance measures for the $[16-10-1|181]$ ANN model constructed by Mavrommatis *et al.* [33] and operating in the prediction mode. The quality indices $\bar{x}_K$ and σ_K, introduced by Klapdor and coworkers, are defined in Eqs. (24) and (26), respectively, while $m\%$ is the percentage of nuclides having experimental half-lives within the prescribed range (second column), for which the calculated half-life lies within the tolerance factor (first column) of the experimental value.

Factor	$T_{\beta,\mathrm{exp}}$ (s)	$m\%$	$\bar{x}_K$	σ_K
<10	$<10^6$	82.8	2.78	1.83
	<60	88.1	2.80	1.83
	<1	90.0	2.88	1.88
<5	$<10^6$	72.4	2.22	1.07
	<60	76.2	2.20	1.01
	<1	76.7	2.23	1.02
<2	$<10^6$	39.7	1.39	0.29
	<60	42.9	1.44	0.32
	<1	43.3	1.46	0.32

TABLE VII. Performance measures for the $[17-10-1|191]$ ANN model constructed by Clark *et al.* [34] and operating in the prediction mode. The quality indices $M^{(10)}$ and $\sigma_{M^{(10)}}$, introduced by Möller and coworkers, are defined in Eqs. (30) and (31).

$T_{\beta,\mathrm{exp}}$ (*s*)	Class	$M^{(10)}$	$\sigma_{M^{(10)}}$
<1	o-o	2.05	2.31
	odd	1.08	2.38
	e-e	1.79	2.71
<10	o-o	2.26	5.42
	odd	1.19	2.44
	e-e	1.31	2.30
<100	o-o	1.76	5.19
	odd	1.12	3.15
	e-e	0.98	2.67
<1000	o-o	2.22	6.25
	odd	1.22	5.50
	e-e	0.93	4.78

Modifications and improvements of the training procedure relative to previous work include implementation of the LMOBP algorithm (with batch updating), together with the Nguyen-Widrow method for parameter initialization. To assess the effect of using the LMOBP instead of simpler learning algorithms, we performed additional modeling studies in which we kept all other aspects of the current model development in place, but trained with simple back-propagation or vanilla backprop (incorporating a momentum term) instead of LMOBP. The following results for the respective error measures $\sigma_{\rm rms}$ in learning, validation, and prediction are typical: 1.08, 1.06, and 1.03 (simple back-propagation) and 0.93, 0.85, and 0.85 (vanilla backprop) compared with 0.53, 0.60, and 0.65 (LMOBP).

As mentioned in the Introduction, initial studies of the classification and regression problems presented by nuclear systematics have recently been carried out [37–39] using the relatively new methodology of SVMs. SVMs, which belong to the class of kernel methods [27], are learning systems that have a rigorous basis in the statistical learning theory developed by Vapnick and Chervonenkis [28] (VC theory). There are similarities to multilayer feed-forward neural networks, notably in architecture, but there are also important differences having to do with the prospect of better control over the trade-off between complexity and generalization ability within the SVM framework. Importantly, within this framework there is an automated process for determining the explicit weights of the network in terms of a set of support vectors optimally distilled from among the training patterns [48]. The few remaining parameters are embodied in the inner-product kernel that allows one to deal efficiently with the high-dimensional feature space appropriate to the problem to be solved. The SVM methodology was originally developed for classification problems but has been extended to function approximation (regression) [27].

The recent applications of SVMs to global modeling of nuclear properties, including atomic masses, α-decay chains of superheavy nuclei, ground-state spins and parities, and β^- lifetimes, demonstrate considerable promise for this approach. As in the present work, cross-validation is performed, separating the full database into learning, validation, and test sets. In the existing studies, the data for a given property is divided into four non-overlapping subsets containing input-output pairs for even-even, even-odd, odd-even, and odd-odd classes of nuclides distinguished by the parity of Z and N.

Table VIII provides values of the conventional $\sigma_{\rm rms}$ performance measure (19), both for the SVM model of β^--decay systematics constructed by Clark *et al.* [38] and for the present ANN model. The SVM model demonstrates better performance based on this comparison, with a few exceptions involving the even-even nuclides. However, this comparison is somewhat misleading, because a larger fraction of the data was used for training, leaving numerically smaller validation and test sets in the SVM construction. Moreover, the SVM model of Ref. [38] is actually composed of *four* SVMs developed separately to model the data in each of the four combinations of Z and N parities. This strategy can lead to spurious fluctuations in the predictions of lifetimes for nuclides of isotopic and isotonic chains, as seen in detailed inspection of the outputs of the SVM model. The results appearing in part (a) of Table VIII were obtained for the single ANN model we have developed by training on data from all four (Z, N) parity classes; the results are broken down into these classes only for convenience of comparison. Recently, however, ANN models have been constructed separately for the four parity classes [49], using the same data sets, architecture, input scheme, and training procedures as employed for the ANN model highlighted in the present work. The quality of the combined ANN model as reflected in the corresponding $\sigma_{\rm rms}$ values is very similar to that of the composite SVM model of Clark *et al.*

TABLE VIII. Root-mean-square errors ($\sigma_{\rm rms}$) for (a) the $[3-5-5-5-5-1|116]$ standard ANN model of the present work and (b) the SVM model constructed by Clark *et al.* [38]. Here n is the number of nuclides in each of the data (sub-)sets.

Class	Learning set		Validation set		Test set	
	n	$\sigma_{\rm rms}$	n	$\sigma_{\rm rms}$	n	$\sigma_{\rm rms}$
			(a) ANN model			
EE	95	0.52	33	0.52	35	0.64
EO	121	0.55	46	0.77	47	0.57
OE	141	0.46	42	0.53	40	0.66
OO	146	0.56	46	0.52	46	0.71
Total	503	0.53	167	0.58	168	0.65
			(b) SVMs calculation, Clark *et al.* [38]			
EE	131	0.55	16	0.57	16	0.62
EO	179	0.41	22	0.42	22	0.51
OE	172	0.41	21	0.47	21	0.47
OO	190	0.52	24	0.4	24	0.52
Total	672	0.47	83	0.46	83	0.53

D. Extrapability of the ANN model

It is of course desirable to have a model that reproduces experimentally known β^- half-lives of nuclei across the known nuclear landscape. One can certainly achieve that goal with a sufficiently complex model that involves a sufficient number of adjustable parameters. However, as we have already stressed, excess complexity generally implies poor predictive ability and especially poor extrapability—lack of the ability to extrapolate away from existing data. Accordingly, a much more important and challenging goal is to develop a global model, statistical or otherwise, with minimal complexity consistent with good generalization properties. The extent to which this goal can be achieved with machine-learning techniques for different nuclear properties is yet to be decided. Of course, one can test the performance of a favored network model on outlying nuclei (outlying with respect to the valley of stability), nuclei that are unknown to the network, but have known values for the property of interest. Adequate performance in such tests can provide some measure of confidence in predictions made by the model for nearby nuclei that have not yet been reached by experiment.

Very recent β^- lifetime measurements carried out for "outlying" nuclides not contained in NuSet-B present us with an opportunity to make such a test here. These nuclides are included in the new database NUDAT (version 2.4) [50].

TABLE IX. Predictions for β-decay half-lives T_β recently measured for neutron-rich nuclei, as given by the [3 − 5 − 5 − 5 − 5 − 1|116] standard ANN model and by the pnQRPA + ffGT model of Möller *et al.* [20]. The second column lists the experimental half-life values. The overall error measures $\sigma_{\rm rms}$ of the two models on the test data are given at the bottom of the table.

Nucleus	T_β (ms)		
	Experiment Ref. [50]	ANN model this work	pnQRPA + ffGT Ref. [20]
^{36}Mg	3.9 ± 1.3	14.5	15.9
^{37}Al	10.7 ± 1.3	15	9.9
^{38}Al	7.6 ± 0.6	7.7	4.7
^{39}Al	7.6 ± 1.6	7.2	4.3
^{39}Si	47.5 ± 2.0	27.7	101.5
^{40}Si	33.0 ± 1.0	48.5	30.6
^{42}Si	12.5 ± 3.5	15.5	43.4
^{44}P	18.5 ± 2.5	12.7	17.2
^{46}S	50 ± 8	39.5	30.8
^{47}Cl	101 ± 6	77.9	51.5
^{48}Ar	475 ± 40	447.6	181.9
^{49}Ar	170 ± 50	130.2	54.9
^{64}V	19 ± 8	23.4	7.6
^{73}Co	41 ± 4	103.6	30.7
^{115}Tc	73^{+32}_{-22}	84.2	70.7
^{118}Ru	123^{+48}_{-35}	69.1	211.8
^{120}Rh	136^{+14}_{-13}	196.2	82.7
^{121}Rh	151^{+67}_{-58}	90.7	62.3
^{122}Pd	175 ± 16	227.2	951.2
^{124}Pd	38^{+38}_{-19}	124.2	288.7
^{163}Eu	7.8 ± 5 (s)	7.8 (s)	17.2 (s)
^{164}Eu	4.2 ± 2 (s)	3.3 (s)	8.6 (s)
^{165}Eu	2.3 ± 2 (s)	3.1 (s)	5.7 (s)
^{199}Ir	6^{+5}_{-4} (s)	73 (s)	370.6 (s)
	$\sigma_{\rm rms}$	0.31	0.53

Table IX compares the predictions of the current ANN model and the values given by the pnQRPA + ffGT model of Ref. [20] with the reported experimental lifetimes for the 24 nuclides involved. The respective rms error measures on this data set for these models are 0.31 ("theory-thin") and 0.53 ("theory-thick"). These results increase the degree of confidence in the extrapability of the ANN model.

Useful information on the extrapability and other features of the [3 − 5 − 5 − 5 − 5 − 1|116] ANN model developed in the present work is provided by Figs. 10–15, which track the half-lives estimated by the model for the nuclides in the Fe, Ag, Sn, Ni, Cd, and Bi isotopic chains. Similar plots for relevant isotonic chains at $N = 50$, 82, and 126 are provided in Figs. 16–18. Corresponding pnQRPA + ffGT estimates are included for comparison. Also included are some results (labeled GT*) from calculations by Pfeiffer, Kratz, and Möller [51] based on the early GT of Takahashi *et al.* [7], with updated mass values [17,52] (GT*). There are no unambiguous criteria that can be used to gauge the performance of these models. Judging from the observed behavior of the known nuclei, it should generally be the case that the more neutron-rich an exotic isotope, the shorter its

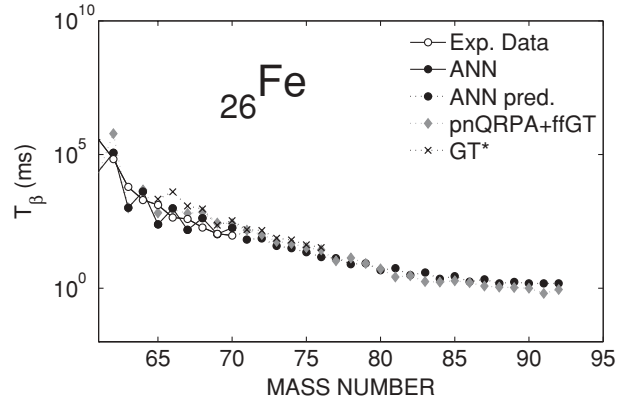

FIG. 10. Half-lives given by the indicated models for the isotopic chain of $_{26}$Fe, compared with experimental data. ANN refers to the standard network model developed based on the NuSet-B data and pnQRPA + ffGT and GT* to models of Ref. [20]. ANN points connected by solid lines are results for nuclides in the learning, validation, and test sets (NuSet-B); those connected by dotted lines are predictions for unknown nuclides *absent* from NuSet-B.

half-life. This expected downward tendency of the isotopic chains is predicted by all the models. One also expects to see some even-odd staggering in the points for neighboring isotopes (isotones), at least for nuclides not far from stability. The ANN model produces such behavior, but it tends to be somewhat exaggerated. (Presumably the δ input is exerting too strong an effect.) Similar behavior, though less pronounced, appears in the results from continuum-quasiparticle-RPA (CQRPA) approaches [22] and in the results of other theoretical calculations [7,20].

Two additional modeling studies have been carried out to probe the extrapability of statistical models of the type developed in the present work. In the first of these studies, another ANN (namely ANN-II) model was created using exactly the same architecture, coding schemes, and training procedure as adopted for the "standard" model explored in detail in previous sections. However, instead of using experimental data for the training, validation, and test sets, we made use of the *theoretical* half-lives generated by the pnQRPA + ffGT model of Ref. [20]. To deviate minimally from the construction of the standard model, the same learning

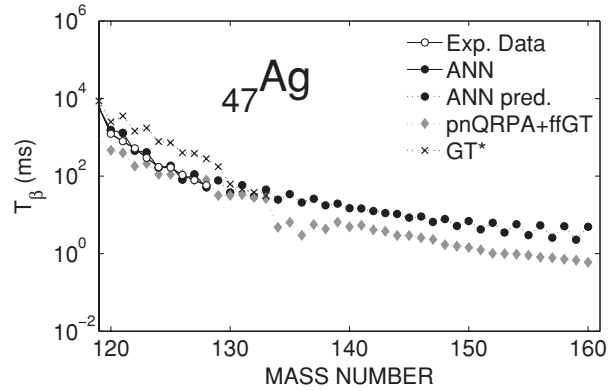

FIG. 11. Same as in Fig. 10 but for the isotopic chain of $_{47}$Ag.

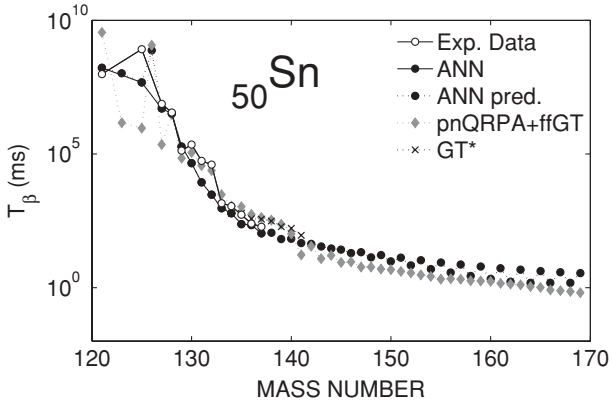

FIG. 12. Same as in Fig. 10 but for the isotopic chain of $_{50}$Sn.

and validation sets were employed for the new model, with the necessary exclusion of those nuclides having $Z < 8$ or $N < 11$. (We note that the "theory-thick" model of Ref. [20] was designed for nuclides with $Z \geqslant 8$ and $N \geqslant 11$, and its predictions were limited accordingly.) Consequently, the learning set now contains 482 rather than 503 nuclides; and the validation set, 164 rather than 167. Similarly, the original test set was reduced from 168 nuclides to 162 by the removal of light nuclei. A much larger test set of 4538 nuclides is provided by all the nuclides tabulated by Möller *et al.* [20] that have not been included in the modified learning, validation, and test sets. The new ANN, developed to model *theoretical* data, yields $\sigma_{\rm rms}$ values of 0.67 and 0.65, respectively, for the 162-member and 4538-member test sets. This is the same level of predictive performance as is achieved by the standard ANN model trained and validated on experimental data. As seen in Figs. 19–22, performance of the new ANN model along different isotopic and isotonic chains is also of satisfactory and/or comparable quality.

In the second of the new modeling studies, we have developed another ANN model, once again implementing the same design and procedures as for the standard ANN. However, for this variant we have selected learning and validation sets from an experimental database, NuSet-C, created by removing all nuclides with half-lives less than 0.1 s from the original database NuSet-B. With the resulting network, we are able to

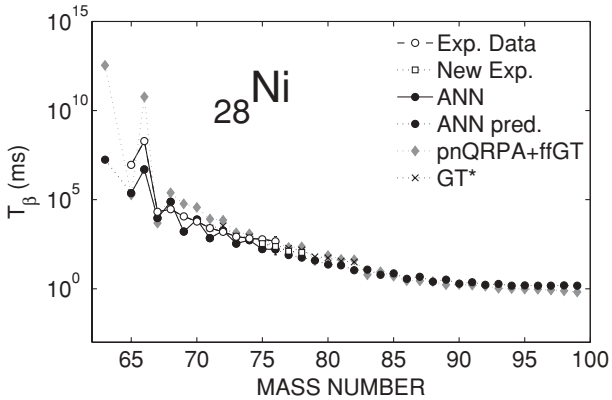

FIG. 13. Same as in Fig. 10 but for the isotopic chain of $_{28}$Ni.

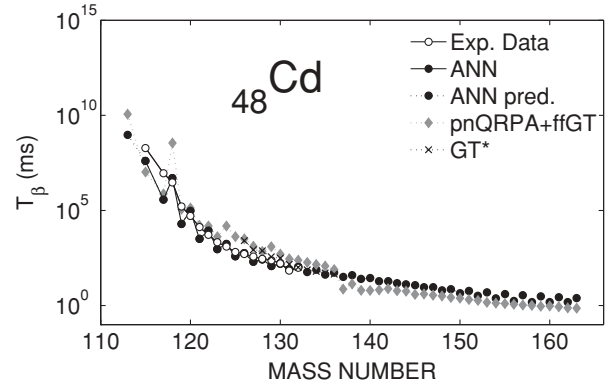

FIG. 14. Same as in Fig. 10 but for the isotopic chain of $_{48}$Cd.

test extrapability to a new class of nuclides not represented in the training data—those with very short half-lives. NuSet-C consists of 749 nuclei, of which 599 (80%) were chosen at random to form the learning set, with the remaining 150 (20%) used for validation. The primary test set is composed of the 89 nuclei eliminated from NuSet-B, all with half-lives below 0.1 s. The rms errors found for the variant ANN model on the learning, validation, and tests sets are 0.52, 0.65, and 0.60, respectively. These results are very much in line with the performance achieved by the original model (see Table I) and, significantly, show no deterioration of predictive quality for the "new" nuclides belonging to a different lifetime class. The corresponding $\sigma_{\rm rms}$ values for the pnQRPA $+ ff$GT model of Ref. [20] are 0.90, 1.02, and 0.42, respectively, allowing for the necessary deletion of nuclides with $Z < 8$ or $N < 11$ from the three sets when evaluating $\sigma_{\rm rms}$. A complementary test set is provided by the 24 neutron-rich nuclides listed in Table IX, for which half-life measurements have been made very recently. The rms measure over this set is 0.40, which is to be compared with 0.31 for the original ANN model and 0.53 for the model of Ref. [20]. The extrapability of the ANN model was also checked by tracking the behavior of calculated half-lives along isotopic and isotonic chains, with results of quality similar to that for the original model. Based on these and the other tests that have been carried out, it would seem that the ANN model extrapolates satisfactorily, at least for nearby nuclei in the nuclear landscape.

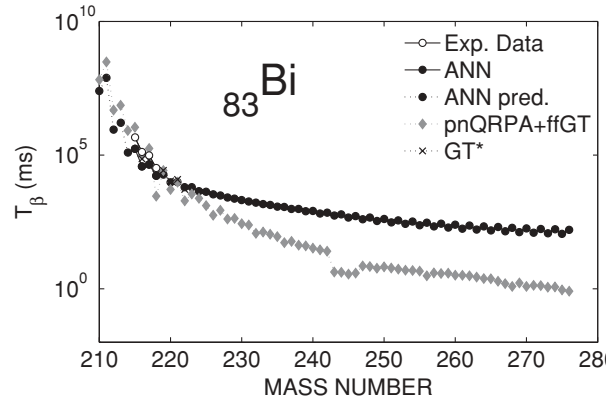

FIG. 15. Same as in Fig. 10 but for the isotopic chain of $_{83}$Bi.

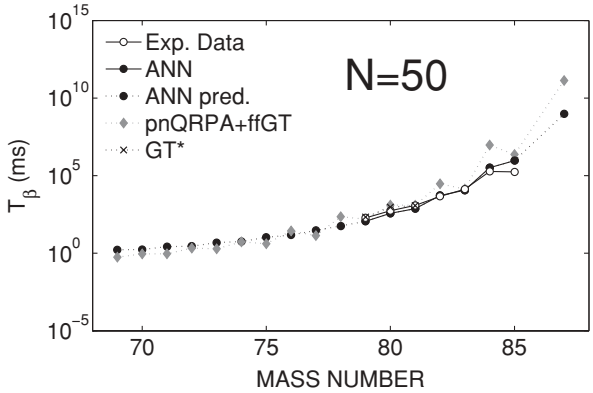

FIG. 16. Same as in Fig. 10 but for the r-ladder isotonic chain of $N = 50$. Here ANN refers to the standard network model developed on the basis of the NuSet-B data and pnQRPA $+ ff$GT and GT* to models of Ref. [20]. ANN points connected by solid lines are results for nuclides in the learning, validation, and test sets (NuSet-B); those connected by dotted lines are predictions for unknown nuclides *absent* from NuSet-B.

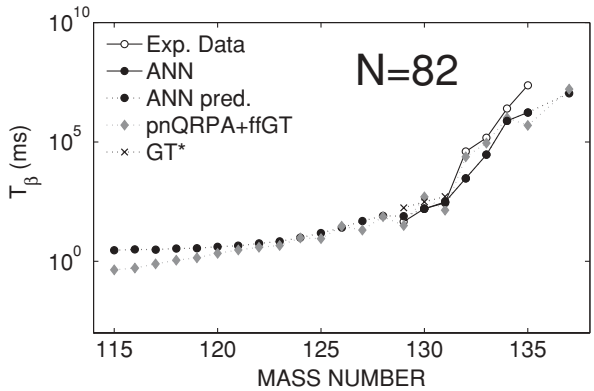

FIG. 17. Same as in Fig. 16 but for the r-ladder istonic chain of $N = 82$.

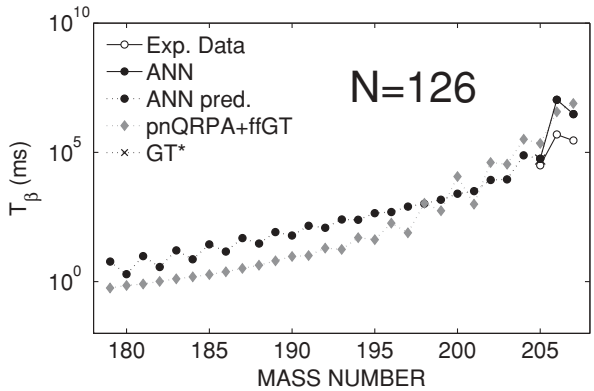

FIG. 18. Same as in Fig. 16 but for the r-ladder isotonic chain of $N = 126$.

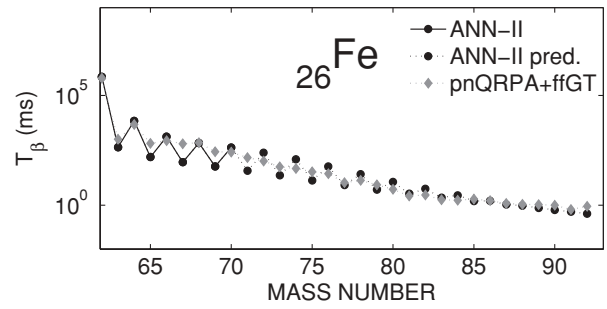

FIG. 19. Half-lives for the isotopic chain of $_{26}$Fe, as given by the pnQRPA $+ ff$GT model of Ref. [20] and by the new ANN model (namely ANN-II) designed to fit and predict data generated by this "theory-thick" model rather than the experimental data. ANN-II points connected by solid lines are results for nuclides in the corresponding learning, validation, and test sets; those connected by dotted lines are predictions for nuclides outside these sets.

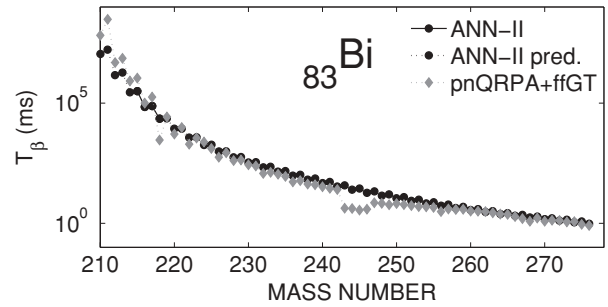

FIG. 20. Same as in Fig. 19 but for the isotopic chain of $_{50}$Bi.

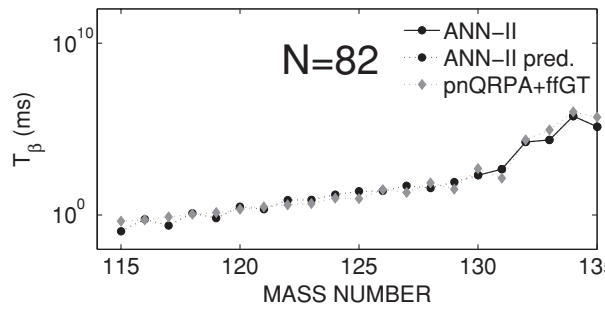

FIG. 21. Same as in Fig. 19 but for the isotonic chain of $N = 82$.

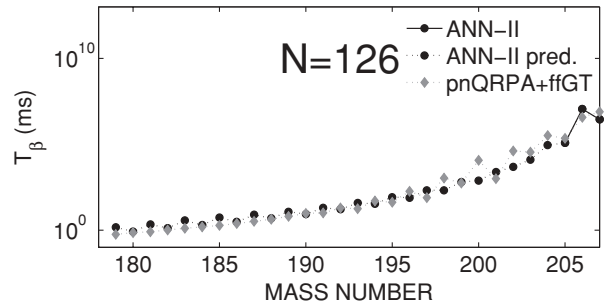

FIG. 22. Same as in Fig. 19 but for the isotonic chain of $N = 126$.

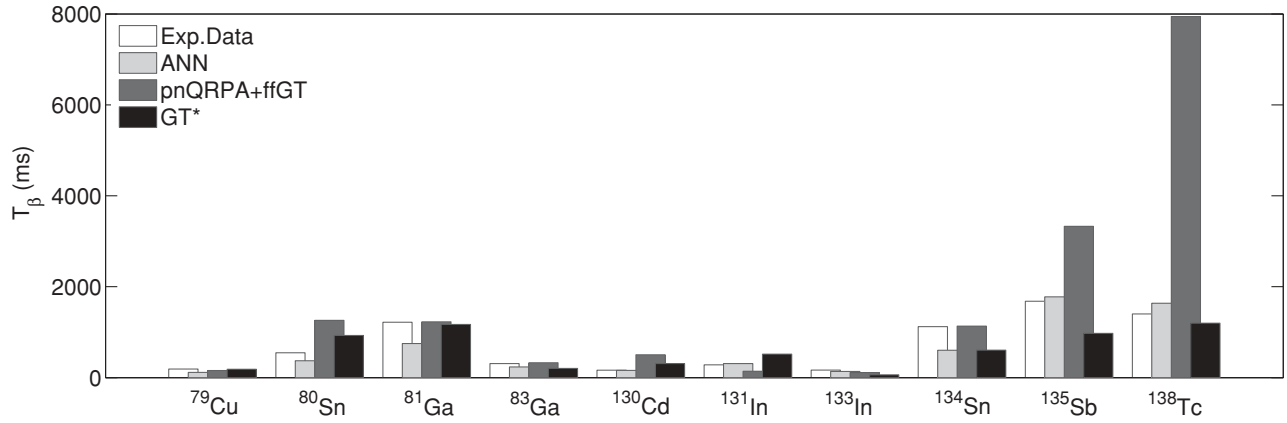

FIG. 23. Half-lives for β^--decaying nuclides that are found on or near a typical r-process path, with neutron separation energies up to 3 MeV. ANN refers to the standard model of this work and pnQRPA $+ ff$GT and GT* to models of Ref. [20].

E. The r-process path

Predictions from the ANN model developed here, and improvements on it, may prove to be useful for quantitative studies involving r-process nucleosynthesis. The β half-lives (T_β) and β-delayed neutron emission probabilities (P_n) of those isotopes lying in the r-process path are the two key β-decay parameters that bear on the β-strength function (S_β) [5]. Accordingly, an approach having global applicability for accurate prediction of β half-lives is needed for detailed dynamical r-process calculations. Moreover, reliable β-half-life calculations are of special interest for the r-ladder isotones $N = 50$, 82, and 126 where solar abundances peak, because they determine the r-process time scale. As mentioned earlier, in Figs. 16–18 we plot the half-lives of closed-neutron-shell nuclei in these significant r-process regions as predicted by our ANN model, in comparison with corresponding results from pnQRPA $+ ff$GT and GT* calculations [20]. In particular, it is interesting to compare the various estimates of the half-life of the doubly magic r-process nucleus ^{78}Ni ($Z = 28$, $N = 50$). The result given by the ANN model is consistent with the recent measurement by Hosmer *et al.* [45]. In Fig. 23, half-lives of β^--decaying nuclides that are found near or on a typical r-process path with neutron separation energy below 3 MeV and derived by means of the standard ANN model are compared with those from pnQRPA $+ ff$GT and GT* calculations [20] and with the experimental data. The results given by the ANN model are close to the experimental values.

Recent measurements of the half-lives of the heavier r-process nuclides ^{202,199,198}Ir, ^{200,199}Os, and ^{196,195,194}Re [53] tend toward lower values than those expected from the corresponding results from pnQRPA $+ ff$GT calculations [20]. The half-lives for the five Ir and Os isotopes generated by the principal ANN model developed here are also lower than those of Ref. [20]. As seen in Fig. 18, this trend holds for most $N = 126$ isotones with $A \geqslant 200$. (However, it may be noted that for nuclides with $A < 200$, the ANN model gives somewhat higher values than reported in Ref. [20].) By contrast, DF3-QRPA calculations [23,53] produce lower half-life values than the pnQRPA $+ ff$GT model [20] for almost the whole r-ladder $N = 126$ isotonic chain. Indeed, DF3-QRPA yields predictions closer to experiment than either the latter model or the ANN model for the newly measured nuclides indicated above. Further experiments, including refined spectroscopy, are needed to clarify the situation in this region.

A more detailed examination of the results of ANN models in relation to r-process nuclei will be published elsewhere.

V. CONCLUSION AND PROSPECTS

A statistical approach to the global modeling of nuclear properties has been proposed and implemented for treatment of the systematics of β^- lifetimes of the ground states of nuclei that decay exclusively in this mode. Specifically, artificial neural networks (ANNs) of multilayer feedforward architecture are taught to reproduce the experimentally measured lifetimes of nuclides from a chosen large data set. Training of the networks is carried out in such a way that their intrinsic generalization capabilities can be exploited to predict lifetimes of nuclides outside the data set used for learning.

We have been able to develop an ANN model of this kind that demonstrates very good properties in terms of both the standard performance measures used in statistical analysis and more problem-specific quality measures that have been introduced to assess traditional theoretical models for calculating β^- lifetimes on a global scale. In a purely results-oriented sense (accurate fitting of given data and good prediction for nuclei not involved in the fitting process), the performance of this model matches or surpasses that of traditional models based on nuclear theory and phenomenology. This success opens the prospect that statistical modeling based on machine learning can provide a valuable tool in the exploration of β^- half-lives of newly created nuclei beyond the valley of stability.

Experience gained previously with neural-network modeling of nuclear systematics (especially the modeling of masses [30,35,36]) strongly suggests that significant further improvements on the current ANN model of β^- systematics

are possible, as more sophisticated training algorithms and machine-learning strategies are continuously being developed. Thus we plan further studies along the same lines with multilayer feedforward perceptrons, while also exploring the potential of SVMs.

It is to be stressed that this program can be no substitute for aggressive pursuit of traditional, "theory-thick" global modeling, which inevitably provides greater insight into the underlying physics responsible for values taken by the targeted nuclear properties. The statistical approach can best serve in complementary and supportive roles. We point out that hybrid statistical-theoretical models show special promise, as demonstrated in Ref. [36]. In that recent work, a $[4-6-6-6-1\,|169\,]$ ANN is used to model the *differences* between measured mass-excess values and the theoretical values given by the finite-range droplet model (FRDM) of Ref. [17], thereby enabling improved prediction of masses away from stability.

Finally, as this last remark exemplifies, the prospects for fruitful application of statistical, machine-learning methods extend to a wide range of nuclear properties beyond the systematics of β-decay lifetimes.

ACKNOWLEDGMENTS

This research has been supported in part by the US National Science Foundation under Grant No. PHY-0140316 and by the University of Athens under Grant No. 70/4/3309. We wish to thank G. Audi and his team for very helpful communications. J.W.C. is grateful to Complexo Interdisciplinar of the University of Lisbon and to the Department of Physics of the Technical University of Lisbon for gracious hospitality during a sabbatical leave and to Fundação para a Ciência e a Tecnologia of the Portuguese Ministério da Ciência, Tecnologia e Ensino Superior as well as Fundação Luso-Americana for research support during the same period.

[1] *The Frontiers of Nuclear Science, A Long-Range Plan for the Next Decade*, DOE/NSF Nuclear Science Advisory Committee, December 2007; *NuPECC Long Range Plan 2004, Perspectives for Nuclear Physics Research in Europe in the Coming Decade and Beyond*, edited by M. Harakeh *et al.*, NuPECC Report, NuPECC, April 2004.
[2] B. Jonson and K. Riisager, Nucl. Phys. **A693**, 77 (2001).
[3] F. Kappeler, F. K. Thielemann, and M. Wiescher, Annu. Rev. Nucl. Part. Sci. **48**, 175 (1998).
[4] M. Arnould, S. Goriely, and K. Takahashi, Phys. Rep. **450**, 97 (2007).
[5] K.-L. Kratz, K. Farouqi, and B. Pfeiffer, Prog. Part. Nucl. Phys. **59**, 147 (2007).
[6] I. N. Borzov, Nucl. Phys. **A777**, 645 (2006).
[7] K. Takahashi, M. Yamada, and T. Kondoh, At. Data Nucl. Data Tables **12**, 101 (1973).
[8] H. Nakata, T. Tachibana, and M. Yamada, Nucl. Phys. **A625**, 521 (1997).
[9] E. Caurier, K. Langanke, G. M. Pinedo, and F. Nowaski, Nucl. Phys. **A653**, 439 (1999).
[10] H. Grawe, K. Langanke, and G. M. Pinedo, Rep. Progr. Phys. **70**, 1525 (2007).
[11] H. V. Klapdor, Prog. Part. Nucl. Phys. **10**, 131 (1983); **17**, 419 (1986); **32**, 261 (1994).
[12] H. V. Klapdor, J. Metzinger, and T. Oda, At. Data Nucl. Data Tables **31**, 81 (1984).
[13] A. Staudt, E. Bender, K. Muto, and H. V. Klapdor, At. Data Nucl. Data Tables **44**, 79 (1990).
[14] M. Hirsch, A. Staudt, K. Muto, and H. V. Klapdor, At. Data Nucl. Data Tables **53**, 165 (1993).
[15] H. Homma, E. Bender, M. Hirsch, K. Muto, H. V. Klapdor-Kleingrothaus, and T. Oda, Phys. Rev. C **54**, 2972 (1996).
[16] J. U. Nabi and H. V. Klapdor, At. Data Nucl. Data Tables **71**, 149 (1999); **88**, 237 (2004).
[17] P. Moller, J. R. Nix, W. D. Myers, and W. J. Swiatecki, At. Data Nucl. Data Tables. **59**, 185 (1995).
[18] P. Moller and J. Randrup, Nucl. Phys. **A514**, 1 (1990).
[19] P. Moller, J. R. Nix, and K.-L. Kratz, At. Data Nucl. Data Tables **66**, 131 (1997).
[20] P. Moller, B. Pfeiffer, and K.-L. Kratz, Phys. Rev. C **67**, 055802 (2003), http://t16web.lanl.gov/Moller/publications/rspeed2002.html.
[21] J. Engel, M. Bender, J. Dobaczewski, W. Nazarewicz, and R. Surman, Phys. Rev. C **60**, 014302 (1999).
[22] I. N. Borzov and S. Goriely, Phys. Rev. C **62**, 035501 (2000).
[23] I. N. Borzov, Phys. Rev. C **67**, 025802 (2003).
[24] T. Niksic, T. Marketin, D. Vretenar, N. Paar, and P. Ring, Phys. Rev. C **71**, 014308 (2005).
[25] T. Marketin, D. Vretenar, and P. Ring, Phys. Rev. C **75**, 024304 (2007).
[26] C. M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon, Oxford, 1995).
[27] S. Haykin, *Neural Networks: A Comprehensive Foundation* (McMillan, New York, 1994).
[28] V. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1995).
[29] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods* (Cambridge University Press, Cambridge, UK, 2002).
[30] J. W. Clark, in *Scientific Applications of Neural Nets*, edited by J. W. Clark, T. Lindenau, and M. L. Ristig (Springer-Verlag, Berlin, 1999), p. 1.
[31] K. A. Gernoth, in *Scientific Applications of Neural Nets*, edited by J. W. Clark, T. Lindenau, and M. L. Ristig (Springer-Verlag, Berlin, 1999), p. 139.
[32] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, UK, 2003).
[33] E. Mavrommatis, A. Dakos, K. A. Gernoth, and J. W. Clark, in *Condensed Matter Theories*, edited by J. da Providencia and F. B. Malik (Nova Science, New York, 1998), Vol. 13, pp. 423–438.
[34] J. W. Clark, E. Mavrommatis, S. Athanassopoulos, A. Dakos, and K. A. Gernoth, in *Proc. of the Conf. on Fission Dynamics of Atomic Clusters and Nuclei*, edited by D. M. Brink, F. F. Karpechire, F. B. Malik, and J. da Providencia (World Scientific, Singapore, 2001), pp. 76–85, arXiv:nucl-th/0109081.
[35] S. Athanassopoulos, E. Mavrommatis, K. A. Gernoth, and J. W. Clark, Nucl. Phys. **A743**, 222 (2004).

[36] S. Athanassopoulos, E. Mavrommatis, K. A. Gernoth, and J. W. Clark, in *Advances in Nuclear Physics, Nuclear Astrophysics, Heavy Ions and Related Areas*, edited by G. A. Lalazissis and C. C. Moustakidis (HNPS, Thessaloniki, 2005), p. 65 (to be published), arXiv:nucl-th/0511088.

[37] H. Li, J. W. Clark, E. Mavrommatis, S. Athanassopoulos, and K. A. Gernoth, in *Condensed Matter Theories*, edited by J. W. Clark, R. M. Panoff, and H. Li (Nova Science, New York, 2006), Vol. 20, pp. 505–519, arXiv:nucl-th/0506080.

[38] J. W. Clark and H. Li, in *Recent Progress in Many-Body Theories*, edited by S. Hernandez and H. Cataldo (World Scientific, Singapore, 2006), Vol. 8, pp. 47–61.

[39] H. Li, Ph.D. thesis, Washington University, 2006.

[40] M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural Network Design* (PWS, Boston, 1996).

[41] M. T. Hagan and M. B. Menhaj, IEEE T. Neural Networ. **5**, 989 (1994).

[42] D. J. C. MacKay, Neural Comput. **4**, 415 (1992).

[43] F. D. Foresee and M. T. Hagan, in *Proceedings of the 1997 IEEE International Joint Conference on Neural Networks* (IEEE Press, Piscataway, New Jersey, 1997), Vol. 2, pp. 1930–1935.

[44] G. Audi, O. Bersillon, J. Blachot, and A. H. Wapstra, Nucl. Phys. **A729**, 3 (2003).

[45] P. T. Hosmer, H. Schatz, A. Aprahamian, O. Arndt, R. R. C. Clement, A. Estrade, K.-L. Kratz, S. N. Liddick, P. F. Mantica, W. F. Mueller *et al.*, Phys. Rev. Lett. **94**, 112501 (2005).

[46] D. Nguyen and B. Widrow, in *Proceedings of the 1990 IEEE International Joint Conference on Neural Networks* (Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1990), Vol. 3, pp. 21–26.

[47] N. Costiris, Diploma thesis, University of Athens, Greece, 2006.

[48] A. J. Smola and B. Schoelkopf, NeuroCOLT2 Tech. Rep. Series NC2-TR-1998-030, Royal Holloway College, London, 1998.

[49] N. Costiris, E. Mavrommatis, K. A. Gernoth, and J. W. Clark, in *Advanced in Nuclear Physics*, edited by A. Pakou *et al.* (HNPS, Ioannina, 2008), p. 243, arXiv:0809.0383.

[50] NUDAT (2.4) (all errata have been corrected from the original publications), http://www.nndc.bnl.gov.

[51] B. Pfeiffer, K.-L. Kratz, and P. Moller, Internal Report, Institut Fur Kernchemie, Mainz, 2003.

[52] G. Audi and A. H. Wapstra, Nucl. Phys. **A595**, 409 (1995).

[53] T. Kurtukian-Nieto, J. Benlliure, L. Audouin, F. Becker, B. Blank, I. N. Borzov, E. Casarejos, M. Fernandez-Ordonez, J. Giovinazzo, D. Henzlova *et al.*, Nucl. Phys. **A827**, 587c (2009).