Bachelor Degree Project

Scrutinizing the Schmidt Test and Exploring the Use of Machine Learning for Statistical Assessment of Radioactive Decay Chains Stemming from Superheavy Nuclei Research

Author
Pim Nelissen

Supervisor
Luis Sarmiento Pico

LUND UNIVERSITY

FACULTY OF SCIENCE

Division of Particle and Nuclear Physics
Faculty of Science
Lund University

Spring Semester 2024

Abstract

Experimental nuclear structure data coming from superheavy nuclei synthesis experiments often consists of correlated α-decay chains. In the absence of neutron detectors - which would fully characterize the exit channel after the fusion-evaporation reaction - the sequence of decay energies and half-lives are the 'fingerprint' of the exit channel itself. Experimental data in this region is sparse, and its interpretation can be liable to error or confirmation bias. A so-called "Schmidt test" is a method for determining the congruence of correlation times for a set of measurements of one decay step. Its outcome is not always entirely conclusive, however. This study evaluates the congruence derived from the Schmidt test using Monte Carlo simulated data with various level of contamination from incongruent data. Furthermore, the study also includes the evaluation of congruence of data stemming from single decays and multi-step decay chains. A multi-layer perceptron was trained on extracted features from simulated decay chain sets with one step. The Schmidt test performs well with larger decay sets and when the half-life of the contaminating species is longer than the original species by a factor 5 or 10. However, the test performs poorly in low counting statistics, where few recorded decay times are available. The newly proposed machine learning model outperforms the Schmidt test in certain high statistics scenarios, but also fails when few decay times are available. Its performance is also poor when the half-life of the contaminant is shorter than the original half-life. The learning behaviour of the model is analysed, showing significant contributions from higher statistical moments in training. Future work involves including chain correlations across multiple steps, α-decay energies, as well as the potential use of alternative machine learning models.

Contents

List of abbreviations

CN	Compound Nucleus
DNS	Dinuclear System
EVR	Evaporation Residue
SHN	Superheavy Nuclei
ANN	Artificial Neural Network
CDF	Cumulative Distribution Function
MC	Monte Carlo
MLP	Multi-Layer Perceptron
PDF	Probability Density Function
PRNG	Pseudorandom Number Generator
(G)ST	(Generalised) Schmidt test

1 Introduction

1.1 Superheavy nuclei research around the world

In the late 1940s, the newly proposed nuclear shell model provided an explanation to the so-called 'magic numbers', certain numbers of protons, Z, and neutrons, N, for which nuclei have closed shells, causing enhanced stability. By extrapolation, predictions were made about magic numbers beyond the ones known at the time. The numbers $Z = 126$ and $N = 184$ have been predicted to be 'magic', and, depending on the method of calculation, $Z = 114$ or $Z = 120$ have also been suggested to increase stability [1]. Developments in theory and experiment are constantly trying to shed light on this matter.

The recent decades have seen many advancements in superheavy nuclei (SHN) research, with developments continuing to be made at this far end of the nuclear landscape. Pushing the boundary of the nuclear chart requires close collaboration between nuclear theorists and experimentalists [2]. SHN are formed through the fusion of two nuclei, normally followed by the evaporation of a number of neutrons. The probability of synthesis of compound nuclei -that is, probability of fusion, evaporation, and survival- is quantified via the *cross section*, and it decreases with an increasing atomic number Z, primarily due to Coulomb repulsion between the nucleons. As such, theoretical and experimental research has focused on finding combinations of target and beam nuclei which may have an increased cross section for a SHN synthesis reaction. One recent advancement in that regard has been the use of the doubly magic (closed proton and neutron shells) ^{48}Ca isotope for the beam. This neutron-rich isotope of calcium, combined with the right actinide target, has led to the discovery of elements $Z = 113 - 118$ [2].

Unfortunately, the strategy of using ^{48}Ca seems to be reaching its practical limit. The required actinide targets for synthesis of element 113 to 118 are produced in limited quantities, primarily at the Oak Ridge National Laboratory [3]. However, manufacturing sufficient target material for synthesis of element 119 and beyond has proven exceptionally difficult. As such, current research is focused on exploring the possibility of using beam nuclei with a higher number of protons, particularly along the magic 28 neutron line. One such example is the currently used ^{51}V beam at RIKEN [4]. Increasing the proton number of the beam rather than the target is the current best effort in synthesising new elements. Recent proposals also include the use of ^{50}Ti and ^{54}Cr beams with corresponding actinide targets [5].

Few nuclear structure properties of SHNs can be determined directly. As such, there is a need for studying the characteristic α-decay energies and correlation times corresponding to different steps in the decay chain, which all end in spontaneous fission. Further complications with observing SHNs arise from the fact that there are no neutron detectors present in the setups. Due to this, the exact neutron evaporation channel can only be inferred based on reaction kinematics. The lack of certainty regarding the exit channel means that interpretation is needed in order to correlate the compound nucleus production with its individual decay chain, and especially decay chain data across different experiments. These factors leave room for error and possibly bias. In order to assist in drawing useful conclusions from limited data, the Schmidt test [6] was developed, giving an indication as to the congruence of a set of decay lifetimes, that is, that these events correspond to different samples from a unique half-life. The Schmidt test was later *generalised* to account for entire decay chains [7]. Besides analysing the behaviour of this Schmidt test, this thesis explores the development of a complimentary method for testing decay chain congruence using artificial neural

networks as potentially unbiased observers. The background information needed for this project will be covered in the next section.

2 Background

2.1 Synthesis of superheavy nuclei through fusion

In nature, the synthesis of elements heavier than hydrogen via fusion happens under extreme conditions, such as in the core of stars or the early universe. There is a strong Coulomb repulsion between nuclei due to their charge, which means that the stronger repulsion between heavier nuclei causes a rapid decline in fusion likelihood. If a certain superheavy nucleus is able to exist at all, this is due to a fine balance between the nuclear attraction and Coulomb repulsion of its constituents. Because SHNs are not found in nature (as far as we know), experimental research focuses on finding conditions where such elements can be artificially synthesised, while also providing data for testing nuclear theory. The synthesis of superheavy nuclei occurs schematically in three primary steps; (i) the bombardment of a heavy nucleus (called the *target* nucleus) with a beam of lighter nuclei; (ii) the fusion between these two nuclei, forming a new heavier nucleus known as a *compound nucleus* (CN); and (iii) the shedding of neutrons in order for the CN to reduce its excitation energy. This process is known as *evaporation*. The product left after evaporation is called the *evaporation residue* (EVR). Each of these three steps will be introduced in the following sections.

2.1.1 Bombardment and capture

Superheavy nuclei synthesis involve heavy-ion particle accelerators, where a large number of a certain nuclear species is accelerated towards a target, with the goal of generating an fusion event, otherwise known as capture of the beam nucleus. The chances of such an event happening depend on the experimental conditions such as the beam and target properties, runtime of the experiment, et cetera. When capture is successful, the entire system forms a sort of temporary state known as a dinuclear system (DNS). The excitation energy E^*_{DNS} is dependent on the capture kinematics. There are now two options (see step 1 in Figure 1): Firstly, *quasifission* can occur. This is the near-instant formation of fission products after the impact; it is the splitting of the DNS into two pieces. In this process, some neutron transfer can occur, but generally the pieces will have similar mass numbers, A, as the original beam and target [8]. However, there is also a chance of the formation of a compound nucleus (CN) through fusion. The competition between CN formation and quasifission is highly dependent on reaction parameters such as the energy of impact, the angular momentum, and the excitation energy of the DNS.

For the synthesis of superheavy elements, CN formation is a required step. Here there is an important distinction to be made between two categories of fusion, distinguished by the mass asymmetry between the target and beam nucleus together with the excitation energy of the CN.

2.1.2 Hot fusion

A fusion reaction with asymmetrical beam and target nuclei masses is known as *hot fusion*. The kinematics of such a reaction involve limited transfer of angular momentum from the beam nucleus to the target nucleus. Instead, a larger portion of the kinetic energy from the impact event goes into excitation energy of the CN. The high excitation energy of the CN is why this reaction is known as "hot fusion". Because of the large excitation energy, a CN that went through hot fusion

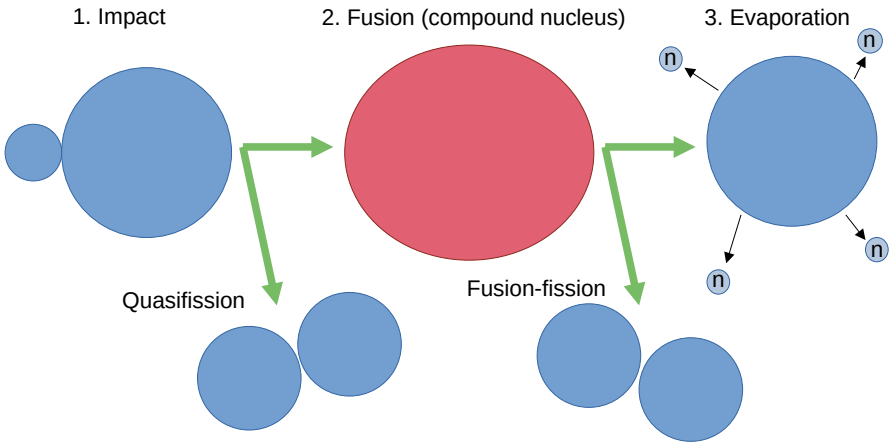

Figure 1: A qualitative schematic showing an asymmetric impact which can lead to hot fusion and the consequential evaporation of x neutrons.

and starts evaporation, will evaporate x neutrons, where the usual amount is $x \in \{2, 3, 4\}$, with 3 neutrons being the most common channel. Another characteristic is that there is a higher chance of CN formation, due to much of the kinetic energy being transferred into excitation energy of the CN. This is in contrast to fusion with more symmetric mass, where the kinematics cause an increased quasifission cross section. The increased cross section for CN formation also implies a higher absolute cross section for the evaporation channel. It is believed that these factors have been crucial in the formation of the elements $Z = 113 - 118$ [9].

2.1.3 Cold fusion

Cold fusion is the name given to a more symmetric fusion reaction. Less mass asymmetry means that the beam nucleus can transfer more of its kinetic energy to the target nucleus in the form of angular momentum. This leaves less energy available for raising the CN into an intrinsic excited state. As a result of having less excitation energy, the CN is relatively more stable, which implies a lower expected fusion-fission cross section compared to hot fusion. Moreover, once the CN is in the evaporation stage, it will evaporate fewer neutrons. There are other conditions which impact the SHN formation cross section, and these depend on the resulting energy characteristics of the compound nucleus after the fusion reaction. These characteristics determine the survivability of the compound nucleus, particularly the resistance against fission. However, this factor is still secondary to the initial fusion cross section when it comes to the SHN synthesis cross section [9]. A comparison between cold and hot fusion cross sections is made in Figure 2.

2.2 Radioactivity

2.2.1 Radioactive decay and decay chains

Superheavy nuclei (SHNs), which are elements with $Z > 103$ protons, are weakly bound due to the short range of the strong nuclear force and the increased Coulomb repulsion. As such, these nuclei are expected to be unstable. It can then be energetically preferable for a SHN to decay in order

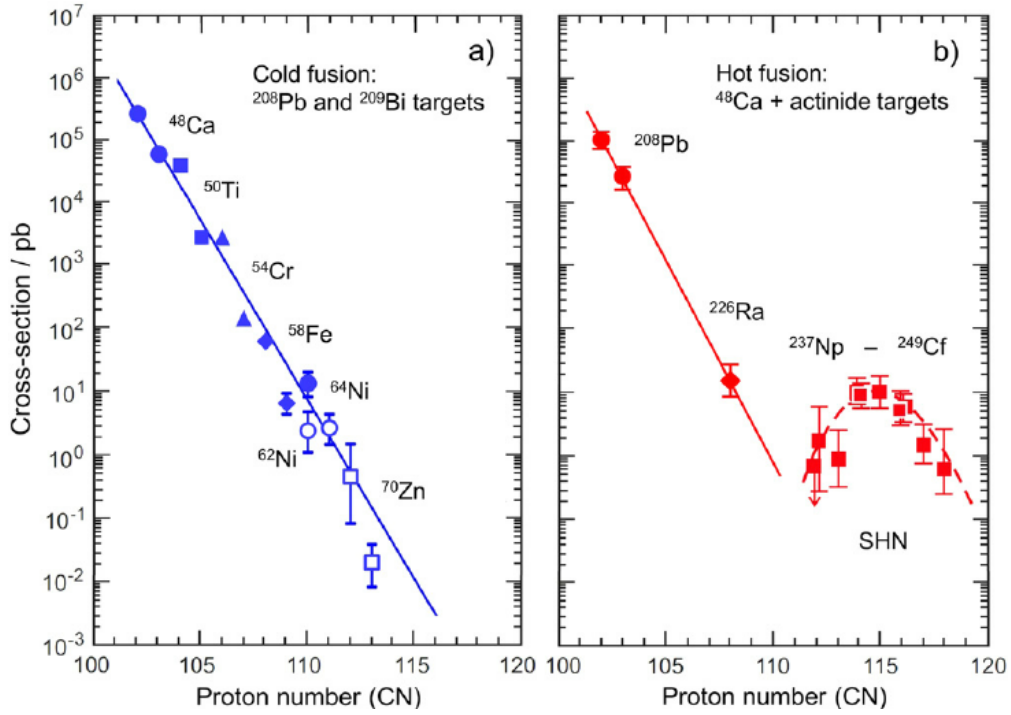

Figure 2: Comparison between cold and hot fusion cross sections as a function of atomic number Z. Taken from Ref. [1].

to reduce its mass in order to reach a more stable configuration. A common form of radioactive decay for SHNs is α-decay, which is the emission of a ${}^{4}_{2}\mathrm{He}$ nucleus:

$$ {}^{A}_{Z}X \rightarrow {}^{A-4}_{Z-2}X + {}^{4}_{2}\mathrm{He}. \tag{1}$$

α-decay is the typical mode of decay for SHNs. That is because the ${}^{4}_{2}\mathrm{He}$ nucleus is particularly high in binding energy due to shell effects, being the first doubly-magic nucleus of the nuclear chart. This makes α-decay an energetically favourable way to increase binding energy. The nucleus resulting from a SHN decaying via α emission is often itself still in the superheavy region, and may also decay via the same process. This results in a sequence of α-decays known as a *decay chain*.

2.2.2 Exponential distribution of radioactive decay

When looking at an isolated radioactive nucleus, the decay process is entirely stochastic. This means that the lifetime of the nucleus is random, following a certain probability density function P which is characterised by a decay constant λ:

$$\frac{dP}{dt} = \lambda. \tag{2}$$

When observing a large population of one particular radioactive species, λ emerges as the inverse of the mean of the observed lifetimes. Suppose that one is observing such a large population of radioactive nuclei. Let the population at some time t be denoted by $N(t)$. The population then decreases proportionally to its current population, related to the decay constant λ by

$$\frac{dN(t)}{dt} = -\lambda N(t). \tag{3}$$

This differential equation has a solution

$$N(t) = N_0 \exp(-\lambda t) \quad , \tag{4}$$

where $N_0 = N(t = 0)$ is the initial population.

A more intuitive measure of radioactivity is the half-life $T_{1/2}$, defined as the time required for a population of radioactive species to reduce $\frac{N_0}{2}$ (half of the starting value). The half-life is related to λ by the following equality:

$$T_{1/2} = \frac{\ln 2}{\lambda} \Leftrightarrow \lambda = \frac{\ln 2}{T_{1/2}} \quad , \tag{5}$$

which, when plugging back into Equation 4, gives

$$N(t) = N_0 \, \exp(-\ln 2 \cdot t/T_{1/2}). \tag{6}$$

Equation 6 now directly relates the population $N(t)$ to the initial population and the half-life $T_{1/2}$, where the rate of decay is inversely proportional to the half-life; the shorter the half-life, the faster the rate of decay is. Once a SHN has been synthesized, an α-decay step is characterized by this half-life and the energy release (Q_α). Detecting the energy of charged-particles using typical decay stations is rather precise. Hence the decay times and their implicit random nature are important for confirmation of SHN synthesis.

2.3 Experimental background

2.3.1 Evaporation residue synthesis and separation

As discussed in Section 2.1, two major variables for SHN synthesis are the beam and target selection and their reaction kinematics, but besides those, other experimental parameters play an important role; for example, the intensity of the beam directly affects the success rate of an experiment. Likewise, the thickness of the actinide target affects the production rate of SHNs. But even with successful production of a superheavy nucleus, a good experiment requires proper separation and detection mechanisms for the confirmation of the synthesis. In that regard, the primary challenge in the experimental detection of SHNs is the overwhelming presence of other ions compared to the desired EVR event. Such unwanted ions include beam, beam-like, and target-like ions, and even those stemming from non-CN reactions. It is thus necessary to separate the EVR from other reaction products.

Several methods have been developed for this separation, but we focus on recoil separation using gas-filled separators. The first application of this technology to superheavy nuclei research was in the late 1960s, at the Joint Institute for Nuclear Research (JINR) in Dubna, Russia [10]. Gas-filled separators consist of a chamber filled with a gas, configured such that EVRs after the target wheel pass through the gas in the separator. Within this chamber, the ions and gas molecules interact electromagnetically, exchanging electrons in the process such that the ions get an *average* charge. This interaction is proportional to the ion mass A and its average charge q [11]. Since the electromagnetic interaction affects the trajectory, gas-filled separators can be used to separate ions based on their mass. As such, the EVR can be separated from the unwanted products. Directing the EVR to a specific region of space where the detector setup is located allows the measurement and study of said EVR. Present-day examples of gas-filled separators for SHN experiments include

the *GAs-filled Recoil Ion Separator* (GARIS) at RIKEN National Laboratory in Japan [12], the Berkeley Gas-filled Separator (BGS) at Berkely, CA, United States [13] and the *TransActinide Separator and Chemistry Apparatus* (TASCA) at GSI in Germany [14]. The latter is displayed schematically in Figure 3.

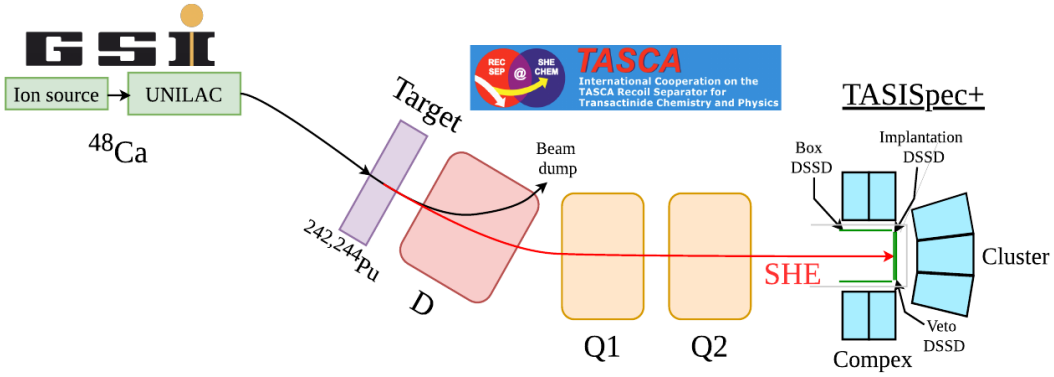

Figure 3: A schematic of a typical setup in superheavy experiments research, in particular the setup of TASISpec behind TASCA at GSI. As an example, the synthesis process for flerovium ($Z = 114$) is shown. A ^{48}Ca beam is accelerated towards a target wheel with foil composed of ^{242,244}Pu. After synthesis of a superheavy nuclei (SHN), the beam ions and other products are separated from the SHN in the TASCA gas-filled separator, using a combination of dipole and quadrupole magnets (indicated by D and Q respectively). Finally, the SHN reaches the pixelised implantation detector (here represented by TASISpec [15]). The figure is reproduced from Ref. [16].

2.3.2 Decay station measurements

Because superheavy nuclei in the range of $Z = 113 - 118$ have half-lives of $t_{1/2} < 1\,\mathrm{s}$, and elements $Z = 119$ and $Z = 120$ are predicted to follow the same trend [17], these SHNs can only be studied by observing their decay products. Measuring the α-decay correlation time and energies is one of the most commonly used ways of confirming synthesis of SHNs [18]. The most simple detection setup would be a silicon detector in the focal plane of the separator. After successful separation, the EVR is implanted into the detector. Here, the α energies and correlation times from the characteristic decay chain can be measured. In reality, the efficiency of such a setup is roughly 50 %, as the emitted α-particles may escape the implantation detector backwards, towards the origin of the beam. As such, the use of several upstream detectors is common. These are positioned around the implantation detector in such a way as to allow detection of escaping α-particles while still allowing the entrance of incoming ions. A typical detector setup is visualised in Figure 3.

Another important detail in the decay stations is the use of pixelised detectors in the focal plane, which allows for the consideration of the position of implantation and the subsequent decays. The localisation of events to a smaller geometric area reduces the risk of random correlations from background events or nuclei that did not separate successfully from the EVR.

2.3.3 Inference from experimental data

The confirmation of a chain of α-decays after the implantation of the EVRs into the detector needs analysis and interpretation. For one, due to the constant implantation of EVRs on the

implantation detector, any subsequent decays from wanted and unwanted EVRs can interfere with the measurements. This means that any radioactivity may be registered by the detector, not just the desired characteristic α-decays. This "summing effect" [19] can influence measurements significantly. This thus complicates the interpretation of experimental data. As discussed before, since neutron detectors are usually absent, it is not possible to decisively confirm whether the set of detected decay energies and correlation times originate from one unique SHN exit channel.

2.4 Decay congruence

2.4.1 Congruence of decay times

Congruence is a measure of agreement or accordance between a set of data points. For superheavy nuclei experiments, determining congruence of the recorded data is essential given the detection constraints of the experimental setups, in particular the lack of detection of the evaporated neutrons. The two most important variables that can influence congruence are the amount of data points collected (number of chains) and the quality of the data (experimental precision). Having more recorded correlation times available allows us to make stronger inferences on the data. Likewise, if little to no contamination of other species is present in the correlation data, this should in theory increase the congruence of said data. To see a quantifiable measure of congruence, one can look at the variance (or the root of the variance, the standard deviation). A high variance would imply likely incongruent data, while a very low variance could imply high congruence. Using the variance directly for SHN decay data has some drawbacks however, some of which will be discussed in the following section.

2.4.2 Schmidt test

A common way to analyse nuclear decay data is through plotting recorded events in a histogram with fixed size time bins Δt. Since the decay data is expected to follow an exponential distribution, the mean of the corresponding density distribution should give the rate of decay λ_{exp}, and the second moment (variance) can indicate whether the data originates from a single radioactive decay or not. There are some restrictions, however. Firstly, the time range must be large compared to the lifetime $\tau = 1/\lambda$. Moreover, the data must be processed to eliminate any background noise or counts originating from other radioactive species. Even with these measures in place, the variance is only a weak indication of congruence in the case of low counting statistics [6].

Earlier this century, K.H. Schmidt proposed a statistical test to determine whether a set of correlation times corresponds to a single radioactive decay [6]. This is referred to as measuring their congruence. Instead of using a linear time distribution, Schmidt proposes the use of logarithmic time bins, in order to reduce the amount of channels for large time frames. The corresponding density distribution is then given by

$$\frac{dN}{d(\ln t)} = -N_0 \exp(\ln t + \ln \lambda) \exp(\exp(-\ln t + \ln \lambda)). \tag{7}$$

This distribution is plotted in Fig. 4 along with the normal exponential distribution. With $\Theta \equiv \ln t$, the standard deviation σ_Θ of Equation 7 is given by

$$\sigma_{\Theta,n} = \sqrt{\frac{\sum_{i=1}^{n} \left(\Theta_i - \bar{\Theta}_{exp}\right)^2}{n}} \quad , \quad \bar{\Theta}_{exp} = \frac{\sum_{i=1}^{n} \Theta_i}{n}, \tag{8}$$

where Θ_i is the logarithmic lifetime of a decay i, $\bar{\Theta}_{exp}$ is the mean of all the recorded lifetimes, and n is the total number of lifetimes.

The shape of the logarithmic distribution only depends on t. N_0 merely scales the height of the distribution, and λ is a constant which shifts the distribution along the logarithmic time axis by a factor $-\ln(\lambda)$. This latter point is a significant detail. It means that σ_Θ should be universal for all radioactive decays, irrespective of the half life of the species in question. Thus, σ_Θ is a universal measure of congruence for individual radioactive decays. Tabulated values for the expected value and confidence interval of σ_Θ for 1-100 counts are found in Ref. [6]. For higher counting statistics, the 90% confidence interval has an analytical value of $1.28 \pm 2.15/\sqrt{n}$, where n is the number of counts.

A value of σ_Θ below the lower limit of the confidence interval may indicate that the recorded spectrum is lacking sufficient data, that is, the time range was too short to accommodate all the decay times. It could also imply that the provided data does not originate from a radioactive decay at all. On the other hand, if σ_Θ falls above the 90% confidence interval, it may indicate that two or more radioactive species are represented in the decay-time data, causing a large standard deviation.

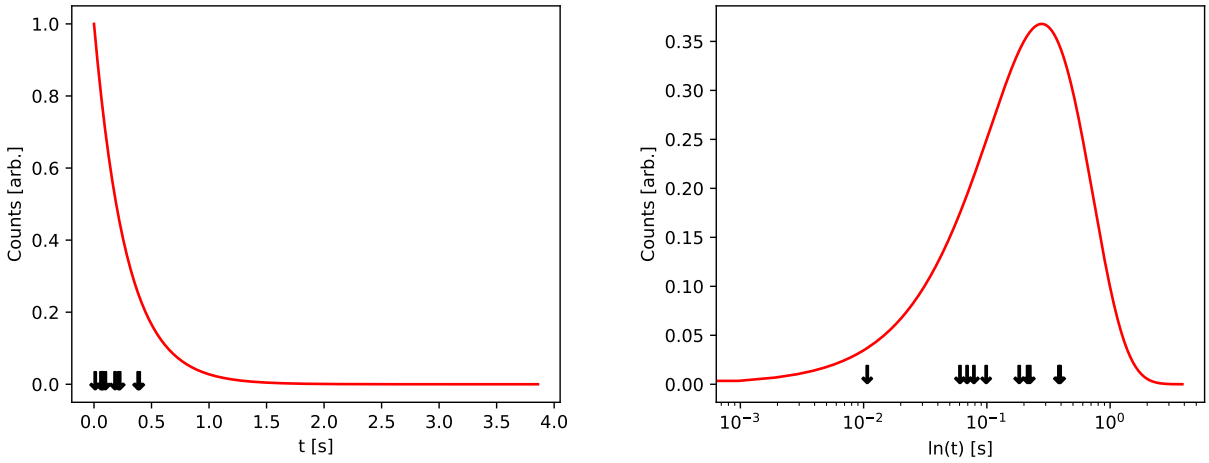

Figure 4: A visual example of the Schmidt test. On the left is a simulated exponential decay curve for ^{288}Mc with $t_{1/2} = 0.163\,\mathrm{s}$ [20]. On the right is the logarithmic time distribution proposed by Schmidt. Experimentally observed correlation times for ^{288}Mc are visualised by the arrows. These times give $\sigma_\theta = 1.02$, which is within the 90% confidence interval of $[0.65, 1.82]$, indicating the data is likely originating from a unique decay corresponding to ^{288}Mc. Correlation times were adapted from Ref. [21].

2.4.3 Generalised Schmidt test

In 2016, U. Forsberg [7] at Lund University proposed a generalised test procedure for decay chain congruence, based on the test proposed by Schmidt. Suppose our data set consists of i chains which each have n_i decay steps, where m_j is the number of available correlation times in the step j. Let θ_{i_j} be the logarithm of the j:th lifetime in chain i. The Generalised Schmidt test, just like the regular Schmidt test, correlates times t^i_j *across* across all decay times available in the step j. This

correlation is identical to the one in Equation 8. The generalisation comes from correlating steps t_j^i across the chain i as well. The measure $\xi_{m,n}$ for a set of decay chains with m_j correlation times in n_i steps is then analogous to σ_Θ in Equation 8:

$$\xi_{m,n} = \sqrt[2]{\frac{\sum_{i=1}^{m} \sqrt[n_i]{\prod_{j=1}^{n_i} \left(\theta_{i_j} - \bar{\theta_j}\right)^2}}{m}}, \bar{\theta_j} = \frac{\sum_{i=1}^{m} \theta_{i_j}}{m} \tag{9}$$

This measure correlates all steps t_j^i in a chain, and thus, questionably assigned chains can be more easily detected. Information about the entire chain is not available in the regular Schmidt test. When only one decay step is considered ($n_i = 1$), $\xi_{m,n}$ reduces to σ_Θ in Equation 8.

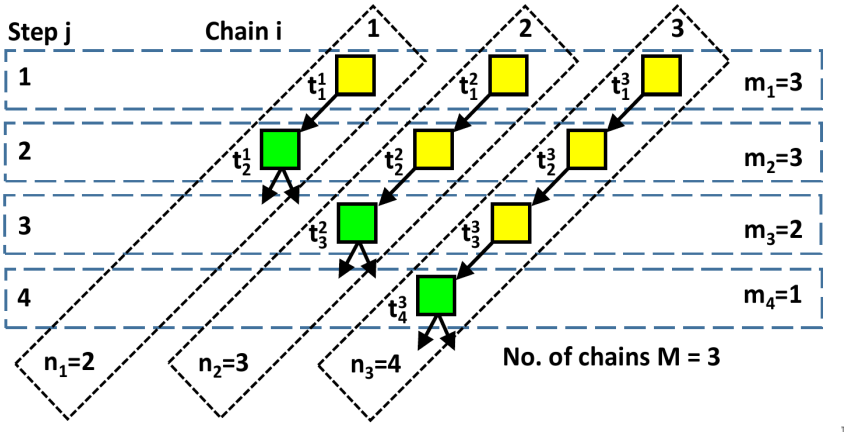

Figure 5: Example of a set of decay chains of various lengths. α-decays are indicated by yellow squares. Green squares represent spontaneous fission. Taken from Ref. [18].

2.4.4 Limitation of current methods and motivation for using machine learning

The Schmidt test can be used to draw conclusions even in very low statistics scenarios, which is commonly the case for SHN experiments. However, it has some drawbacks. In cases where there is an outlier in the correlation times of which the assignment is questionable, the measure σ_Θ may still fall within the confidence interval. The generalised Schmidt test quantifies congruence of entire chains, and as such, if there are chains which consist entirely of outlier values, it can quantify such an anomaly. This is an improvement over the regular Schmidt test. However, just like before, it is possible that statistical outlier chains do not affect the measure $\xi_{m,n}$ enough for it to fall outside the confidence interval. As such, while the (generalised) Schmidt test can be used as an assistive tool for assignment of chains, it is not a complete and definitive tool. Interpretation is still needed and it is therefore susceptible to confirmation bias.

The development of a machine learning based approach for congruence testing of lifetimes could challenge assumptions based on observation biases, in particular confirmation bias. A machine learning algorithm could also be desirable as a way of challenging the Schmidt test itself, assessing its effectiveness as a measure of congruence. An algorithm that incorporates both correlation times and decay energies could be used to directly classify chains. The following sections will focus on assessment of the Schmidt test, as well as the development and results of a machine

learning algorithm for single decays. The outlook for possible generalisations of the algorithm will be discussed too.

2.5 Artificial neural networks

Machine learning is a broad field of study. In essence, the field aims to develop algorithms which learn to identify patterns and features from a set of data, and generalise such that predictions can be inferred from new data. One type of machine learning algorithm is the *artificial neural network* (ANN). Such networks are built up of nodes, which are neuron-like mathematical operations which take the weighted sum of all the inputs and produce one output value. Suppose there is a node j with an input vector $\mathbf{x_j}$ of length P. The output y_j for node j is then

$$y_j(\mathbf{x_j}) = \varphi\left(\sum_{i=1}^{P} \omega_{ji} x_{ji} + b_j\right) \quad , \tag{10}$$

where ω_{ji} is the weight associated with input feature x_{ji}, b_j is a bias value and φ is the *activation function*. For a visualisation, see Figure 6. One can see that for $P = 1$, Equation 10 reduces to $y_j(x_j) = \varphi(\omega_j x_j + b)$, which is of the familiar linear form $y = ax + b$, with an addition of φ. This activation function φ introduces non-linearity into the node, and as such, it can be used for data sets which are not linearly separable. A common non-linear function used as an activation function is the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$.

A combination of multiple of these nodes creates a complex non-linear mathematical model. Moreover, by feeding the output of one or multiple nodes into other nodes, one can introduce more non-linearity and thus possibly solve more complicated problems [22]. Such an ANN with multiple layers is called a *multi-layer perceptron* (MLP). The MLP consists of at least three layers; *i*) the input layer, which receives the input features; *ii*) One or more hidden layers, which transform the features by Equation 10; and *iii*) an output layer, which produces the final output. Typically, each node in one layer is connected to every node in the next layer, which is referred to as a *dense* network. A visualisation of an MLP is presented in Figure 7.

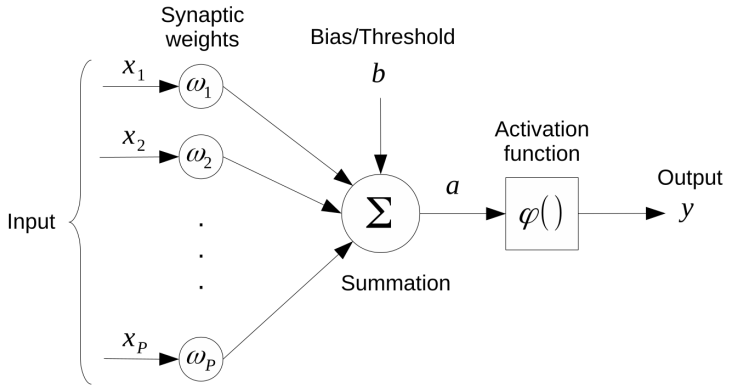

Figure 6: A visualisation of a node j as in Equation 10. Nodes are the individual building blocks of MLPs. Reproduced from Ref. [22].

2.5.1 How a multi-layer perceptron learns

In classification problems, training data is labelled with its associated class, known as the *target*. For example, if the goal is to classify images of cats and dogs, the input data consists of vectorised images of cats and dogs, and the target is either the class cat or dog (represented by a binary 0 or 1). The MLP then iteratively adjusts the weights for each node such that the provided training data produces an output as close to the labelled output as possible. Each iteration of weight updates is called an *epoch*. The closeness of the fit is quantified by a *loss function* $E(\omega)$. Weights ω are adjusted depending on the change in $E(\omega)$ after one epoch, where training will be completed when $\partial E/\partial\omega = 0$, that is, an update of the weights resulted in no decrease (or increase) in error.

One problem that may arise with training a neural network is *overfitting*. This occurs when the model optimises itself too strongly to the training data. In such a case, performance is very good on the provided data, but the model generalises poorly to test data. Likewise, *underfitting* means the model fit itself so weakly to the data that performance on both training and testing data is poor. These problems are visualised in Fig 8. Underfitting can be avoided by increasing the number of epochs (training time) or increasing the complexity of the model (number of nodes or hidden layers). Overfitting can be prevented by several methods such as decreasing complexity of the network or early stopping of the training. More advanced methods include *dropout* of nodes, which is the random omission of nodes during training, which reduces complexity, or *regularisation*, which introduces penalty terms in the loss function which penalise complexity and extreme values [22, 23, 24].

2.5.2 Performance measures

Apart from training data, training requires a validation data set, which is a set of inputs and targets which are not directly used for optimising the weights, but instead are used to evaluate the network. For a classification algorithm, performance can be measured as the accuracy and loss on the validation data. Accuracy refers to the ratio between the correct predictions and the total number of predictions. The loss is an average measure of how off the predictions are for the entire data set, calculated by the loss function $E(\omega)$.

3 Methodology

3.1 Simulations and Statistical testing

3.1.1 Monte Carlo simulations

As discussed in Section 2.2.2, individual radioactive decays follow some underlying exponential distribution with a half-life λ. For individual decays, the probability density function (PDF) corresponding to exponential decay is

$$\mathrm{F}_{pdf}(t) = \lambda exp(-\lambda t). \tag{11}$$

This function represents the probability of the decay occurring at a time t. By accumulating the area under the PDF and normalising the result, the cumulative distribution function (CDF) is obtained:

$$\mathrm{F}_{cdf}(t) = 1 - exp(-\lambda t). \tag{12}$$

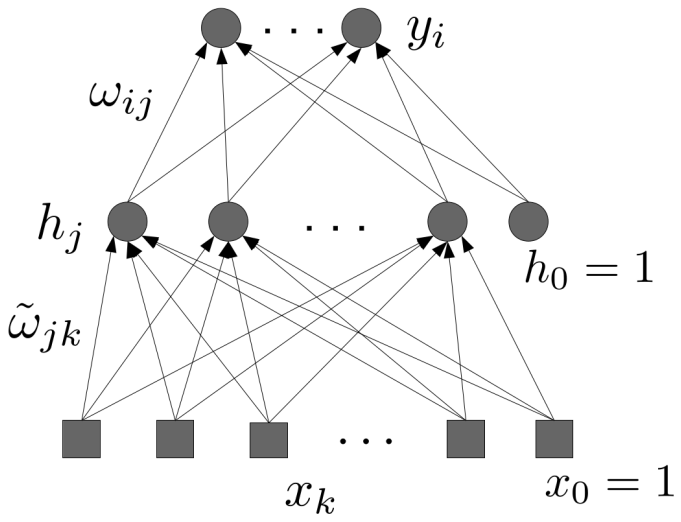

Figure 7: A dense multi-layer perceptron with k input nodes, j hidden nodes in layer h and i output nodes. x_0 and h_0 are the bias nodes for the input and hidden layer respectively. Figure reproduced from Ref. [22].

This function represents the cumulative likelihood of a decay at some time in the range $(0,\ t]$.

Individual decays are independent stochastic events and can thus be simulated using Monte Carlo (MC) techniques. The CDF is sampled using a pseudorandom number generator (PRNG), and the associated time t is taken as the simulated lifetime, representing a single decay event. As the number of samples from the CDF approaches infinity, an exponential fit to the histogram of sampled lifetimes will reproduce the PDF belonging to the underlying half-life (See Fig. 9).

An algorithm was implemented for contaminating a set of correlation times with some other contaminant species with $T_{1/2}^{contam}$, where $T_{1/2}^{contam}$ is varied by some provided factor. The number of contaminated lifetimes is also variable in this algorithm. This will allow simulation of incongruent sets of lifetimes. Besides the decay itself, the decay branch was also sampled using Monte Carlo. The function determines which decay branch is chosen based on the branching ratios, taken as a sample from a non-uniform sample space. This method is particularly relevant when concatenating more than one decay path.

3.1.2 How the Schmidt test was applied

In order to evaluate the original Schmidt test, various ratios between the original half-life $T_{1/2}$ and the contaminant half-life $T_{1/2}^{contam}$ were tested. For each combination of half-lives, 100 sets of decay times were simulated for various combinations of set size N and number of contaminated decay times C. N was varied between 2 and 100, and C was varied between 0 and 100. The goal with testing these various parameters is to see how the half-life of the contaminating species, level of contamination and the counting statistics (total available lifetimes) affect the Schmidt test measure $\sigma_{\Theta,exp}$.

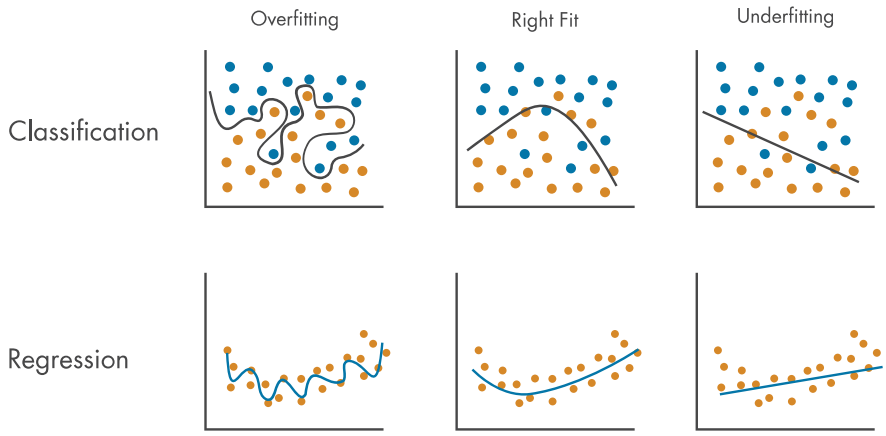

Figure 8: A visualisation of over- and underfitting in a classification and regression scenario. Reproduced from Ref. [24].

3.2 Considerations for the neural networks

3.2.1 Input vector for single decays

In order to appropriately apply neural networks to any problem, it is important to understand what input features are effective for supervised learning and what kind of output is meaningful. The primary input data relevant for decay congruence, besides the rather fixed energy released, is the correlation time (lifetime). Correlation times could be directly applied to a neural network, where each decay lifetime τ is an input feature. However, this poses a limitation, since it restricts future testing data to match the format of the training data. For example, if the algorithm is trained to determine congruence of sets of $n = 100$ correlation times, then it can only make predictions for sets with exactly the same amount of correlation times.

Hence, to overcome this limitation, an indirect feature set is proposed, which extracts features from the correlation times and its distribution. Four features describing some characteristics of a decay set i are used; maximum lifetime $\tau_{i,max}$, minimum lifetime $\tau_{i,min}$, range $(\tau_{max} - \tau_{min})$ and median $med(i)$. On top of that, standard statistical measures of the data set were computed and used as features. The measures used are; 1) mean lifetime τ_μ; 2) variance σ^2, which represents the spread of values around the mean; 3) skewness γ_τ, which is a measure of the asymmetry of the lifetimes around the mean; and 4) kurtosis κ_τ, which is known as "tailedness", quantifying extreme values which deviate a lot from the mean.

For any set of lifetimes of length n, these 8 features represent some key values in the set, as well as some characteristics of the distribution of lifetimes. This gives a universal input vector, which is appropriate for machine learning. After extracting these features, the data is standardised.

In order to compare our ML-algorithm with already existing methods, the neural network needs to output some kind of measure of confidence that the provided lifetimes are indeed congruent. As discussed in Section 2.4.2, the Schmidt test has three possibilities; *i*) the recorded spectrum lacks data (σ_Θ is too low); *ii*) the data is likely originating from one species (σ_Θ is within the confidence interval); and *iii*) the data is likely contaminated and two or more species are present in the decay (σ_Θ is too high).

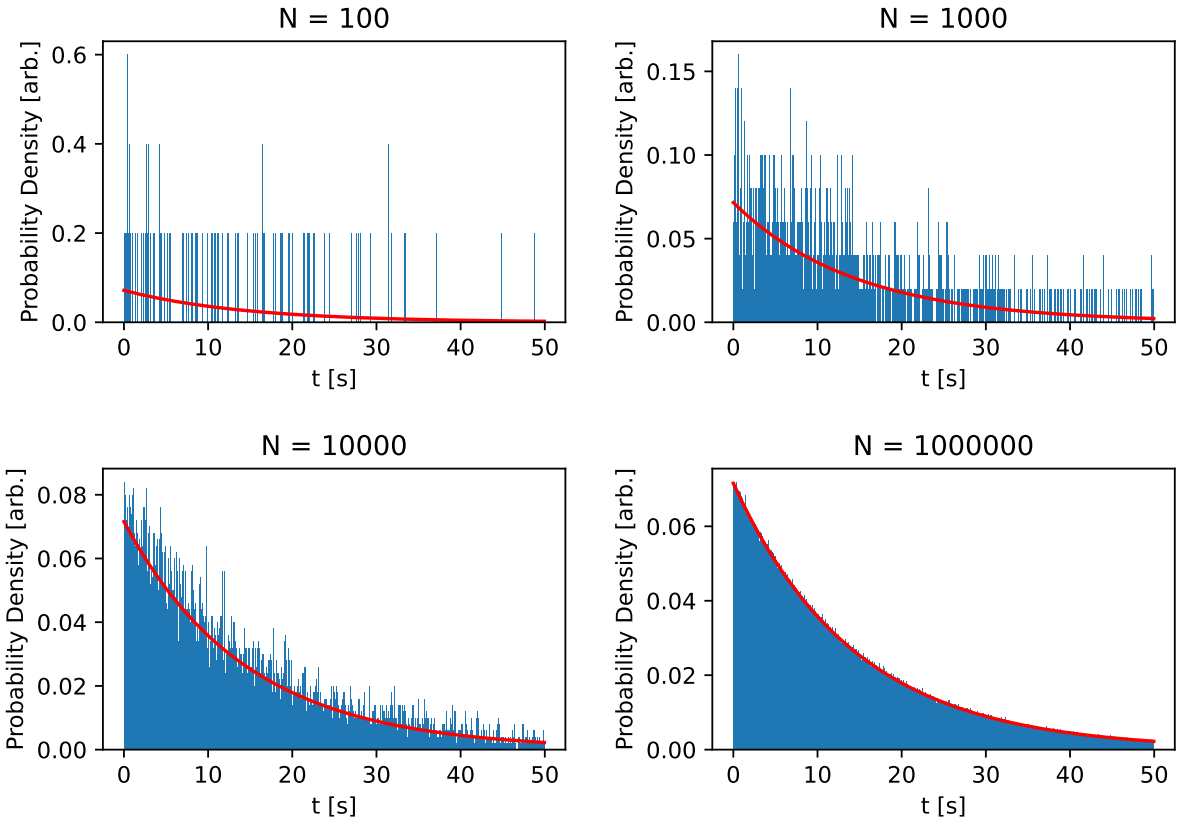

Figure 9: Monte Carlo simulations for N number of correlation events. Red represents the analytical PDF for a decay with $T_{1/2} = 10\,\text{s}$. Blue represents a density histogram of the correlation times simulated in accordance with the PDF.

MLPs are generally well-suited for classification, that is, identifying patterns in a set of data and making predictions for it [25]. The most basic use of an MLP is binary classification (1 or 0), and this is the initial approach that was tried in this project. If 1 represents likely congruent, and 0 represents likely not congruent, then the output of the MLP can be interpreted as the degree of confidence that the provided data is congruent, measured between 0 and 1. The value of 0.5 being chosen as a threshold for classification, meaning that a predicted value below the threshold indicates incongruence, and a value above the threshold is predicted congruence. Although not always the optimal threshold, 0.5 is a typical threshold value to start with in binary classification [26].

3.2.2 ML-algorithm for decay chains

A set of decay chains is a 2-dimensional array, where the columns represent decay steps j, and rows represent different recorded decay chains i such that correlation times are indexed as $\tau_{\alpha j,i}$. Extracting features as discussed in Section 3.2.1 would result in a two-dimensional feature array. This poses a problem because MLPs can only take one-dimensional arrays.

Evaluating steps individually One rather simple solution to this problem is analysing the data along the columns, that is, evaluating each step j separately. Reducing to several sets of single

decay steps like this allows the use of the previously discussed model. With predictions for each step in the decay, statistical analysis can be done on the collective.

Evaluating across decay chains In order to have a model comparable with the Generalised Schmidt test, it is not sufficient to evaluate decay steps individually. Entire chains need to be considered. To do this with an MLP, new features would need to be determined, in order to keep the input vector to one dimension. This would require a newly designed model. Another option could be alternative models which allow for multi-dimensional input vectors.

For chain testing, only the first option is implemented; decay steps are considered individually, and can be tested with the model, after which analysis of the individual steps can infer some (albeit limited) conclusions about the whole chain. Options for evaluating across decay chains will be discussed in detail in Section 6.

3.3 Generating the training data

Since the single decay model should classify congruent and incongruent sets, a part of the training data consists of *contaminated* sets, and part of the data consists of *clean* sets. Simulations were ran for $i = 100000$ chains with $j = 1$ step, where each set consists of 100 lifetimes corresponding to some random half-life $T_{1/2} \in [1, 100]$. The contaminated sets of lifetimes were contaminated randomly with 1 to 99 correlation times of a species with half-life $T_{1/2}^{contam}$, where the half-life is some random factor different from $T_{1/2}$. Features are extracted as explained in Section 3.2.1. The clean feature vectors are labeled as class 1 (congruent), whereas the contaminated feature vectors are labeled class 0 (incongruent). The data is standardised, so that the model trains on *differences* between features, rather than the absolute values, which may be very different for vastly different half-lives. The data was standardised according to $\mathbf{x} = (x - \mu(x))/\sigma(x)$, where $\mu(x)$ is the mean of the input features and $\sigma(x)$ is the standard deviation.

3.4 Network architecture and hyperparameters

As discussed previously, the single decay model is an MLP network. The input layer is set to have 8 nodes (following the input features described in Section 3.2.1. Two hidden layers were implemented, with the first layer and second layer have 8 nodes and 4 nodes respectively. No dropout was implemented, but L2 regularisation was added in order to reduce the risk of overfitting. Additionally, an `EarlyStopping` callback was added to the model, so that overfitting can further be reduced [27]. Through trial and error, the learning rate λ was set to 0.0005, with a *binary cross-entropy* cost function and the `Adam` optimiser algorithm as an optimisation method. Since the output should represent a measure of congruence, the output activation function φ_{out} is a sigmoid function, which gives a continuous output between 0 and 1. This can be interpreted as the probability that the provided lifetimes are congruent. The training/validation split was 90/10, which means 90000 decay chain sets were used for training and 10000 for validation of the model.

4 Results

4.1 Schmidt test on simulated data

In order to explore the behaviour of the Schmidt test, the test was performed for various sets of lifetimes with different sizes N and various amounts of contaminated lifetimes C. $T_{1/2}$ was set

to 1 s, with $T_{1/2}^{contam}$ varied between 1 and 10. Each subfigure in Fig. 10 shows a heatmap for a different $T_{1/2}^{contam}$. Each pixel in one of the heatmaps is a unique combination of set size N and C contaminated decay times, where the value of the pixel represents the percentage of $i = 100$ simulated decay chain sets with $j = 1$ step for which the measure $\sigma_{\Theta,exp}$ falls within the confidence interval, indicating likely congruence. Different ranges of sizes of decay sets and the Schmidt test performance P on these sets are tabulated in Table 1.

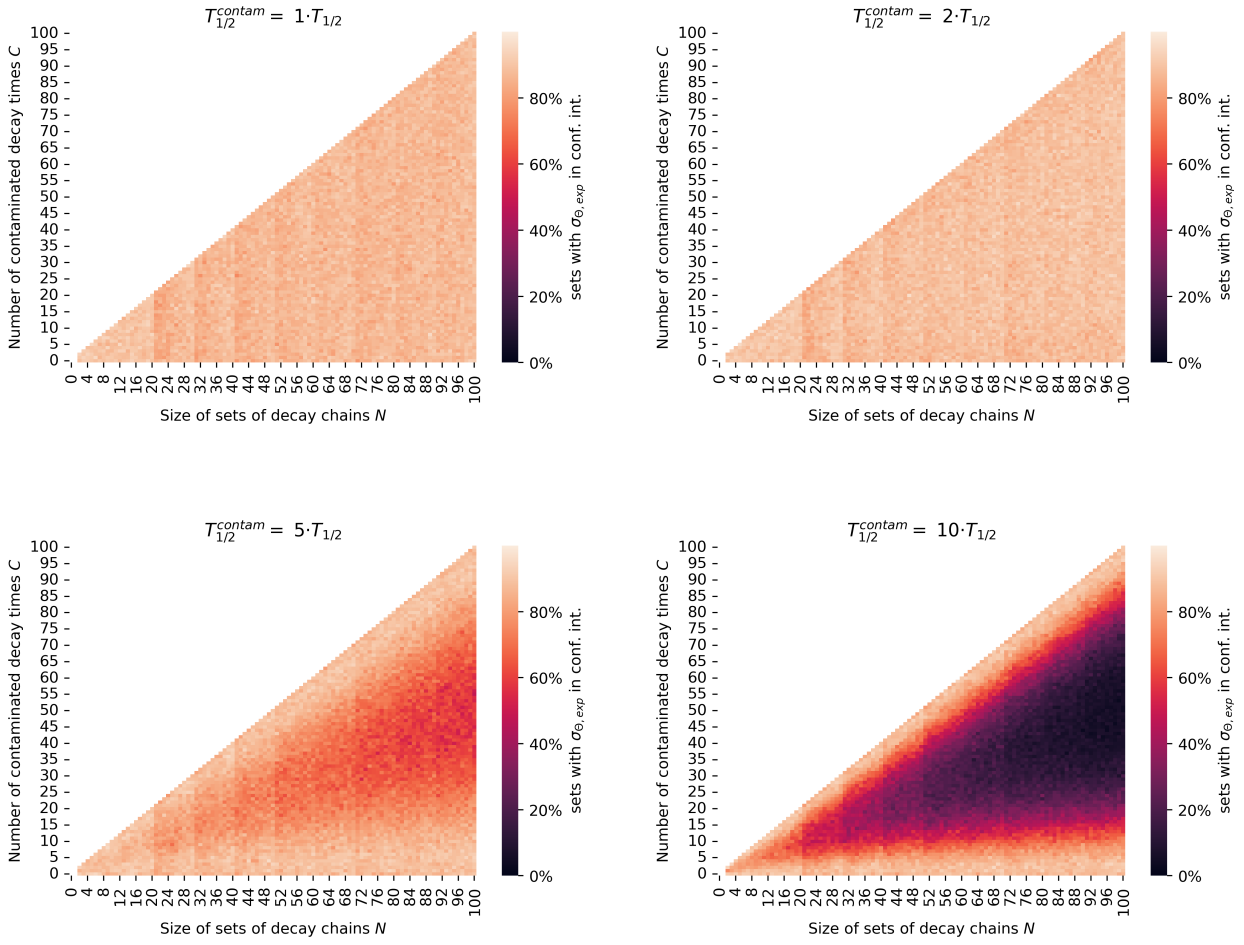

Figure 10: Heatmaps for the Schmidt test congruence for various half-lives $T_{1/2}^{contam}$. Each pixel is a unique combination of set size N and C contaminated decay times, where the value of the pixel represents the percentage of $i = 100$ simulated decay chain sets with $j = 1$ step for which the measure $\sigma_{\Theta,exp}$ falls within the confidence interval, indicating likely congruence. The longer $T_{1/2}^{contam}$ is compared to $T_{1/2}$, the more sets are deemed incongruent by the Schmidt test. Additionally, it can be seen that when nearly all of the decay times in the data set are from the contaminant species, the data becomes congruent again. The number of sets deemed congruent is somewhat symmetrical around the the line of 50% contamination.

$T_{1/2}^{contam}/T_{1/2}$	Average $P_{2\leq N\leq 10}$	Average $P_{90\leq N\leq 100}$	Average P_{total}
1	90.3%	87.2%	87.1%
2	88.7%	89.2%	88.7%
5	89.0%	73.6%	77.9%
10	84.8%	37.1%	47.9%

Table 1: The percentage P of sets deemed congruent by the Schmidt test, for various ranges of N.

4.2 Single decay model

4.2.1 Training statistics

Training of the model was performed several times, with adjustments made to the hyperparameters until a validation accuracy of 91% was reached, with a loss of 0.2386. The final parameters are found in Section 3.4. Figure 11 shows the training history, where it can be seen that the early stop was activated around 67 epochs.

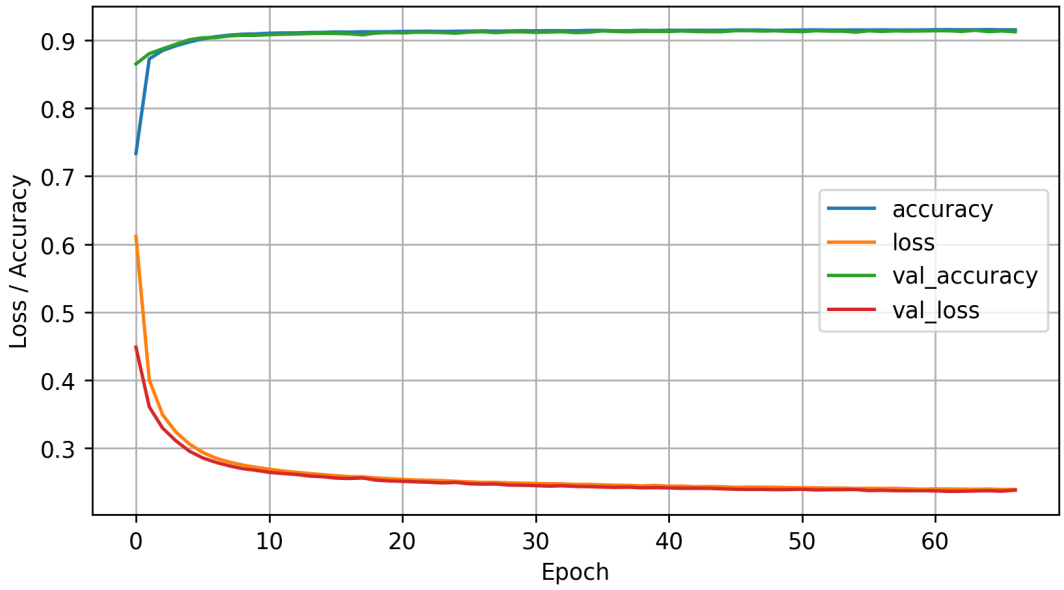

Figure 11: Training history for the model. `accuracy` and `loss` refer to the accuracy and loss $E(\omega)$ on the training set. `val_accuracy` and `val_loss` refer to the accuracy and loss $E(\omega)$ on the validation set.

4.2.2 Permutation importance

Permutation importance (PIMP) is a method for determining which input features most significantly affect the performance of a machine learning model. It can give insight into which features are most important for the classification of data [28]. It works by systematically permuting each of the input features, and re-evaluating the model with this permutation. The resulting accuracy is then subtracted from the unpermuted accuracy, resulting in a relative feature importance. A low feature importance for feature x thus means that the validation accuracy before and after permuta-

tion of feature x is nearly identical. By contrast, a large feature importance means the permutation of feature x significantly affected the validation accuracy. Figure 12 shows the feature importance of each input feature to the neural network.

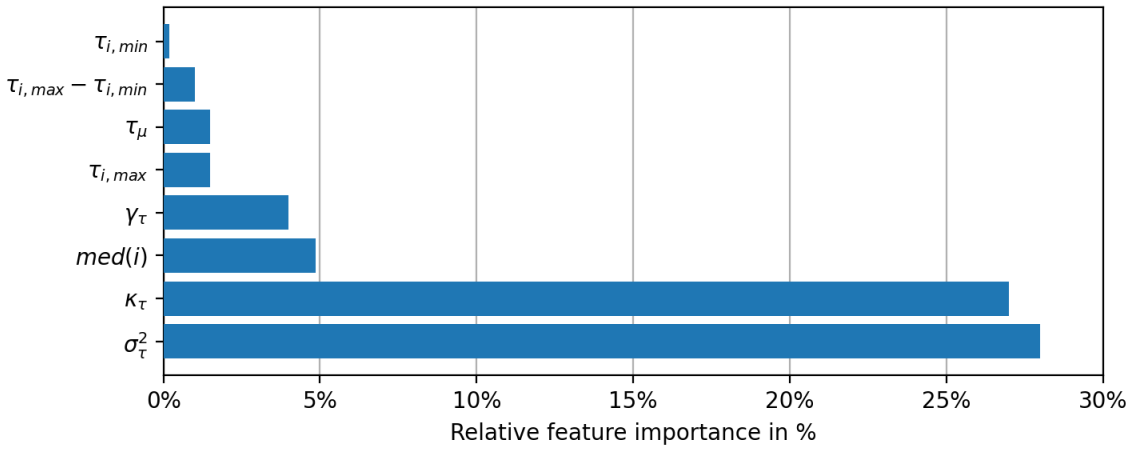

Figure 12: Feature importance for each feature, determined using permutation feature importance methods. The features are lifetime $\tau_{i,max}$, minimum lifetime $\tau_{i,min}$, range $(\tau_{max} - \tau_{min})$, median lifetime $med(i)$, mean lifetime τ_{μ}, variance σ_{τ}^2, skewness γ_{τ} and kurtosis κ_{τ}.

4.3 Statistical properties of contaminated decay sets

To analyse the features with the highest relative feature importance in Figure 12, an analysis is done of the variance σ_τ^2, skewness γ_τ and kurtosis κ_τ for the same decay sets as simulated in Section 4.1. The results are shown in Figure 13.

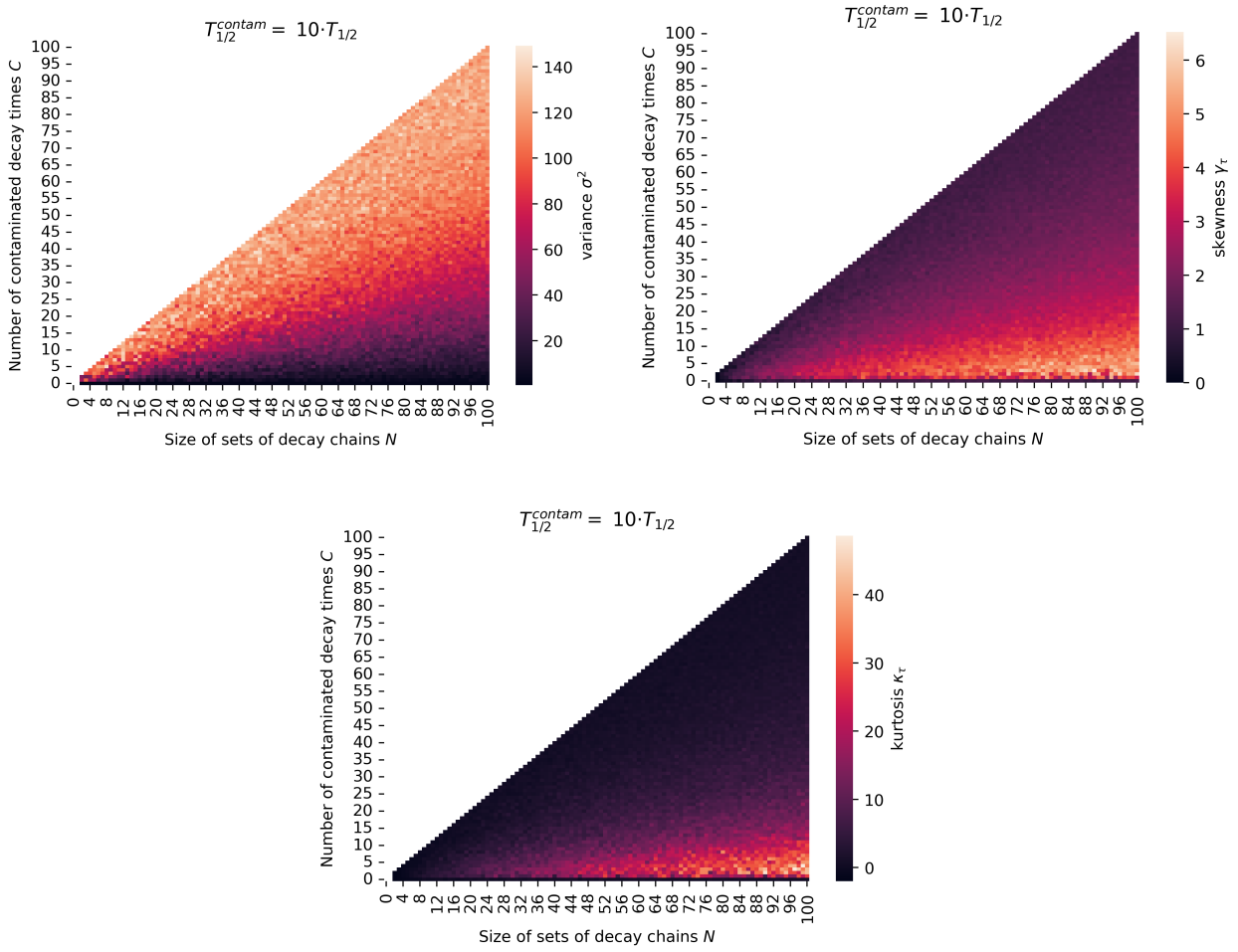

Figure 13: Heatmaps representing the value for various statistical moments on $i = 100$ simulated decay chain sets with a single decay step. Each pixel is a unique combination of set size N with C contaminated decay times. Only the heatmaps for $T_{1/2}^{contam} = 10 \cdot T_{1/2}$ are shown. For other factors, see Appendix A.

4.4 Single decay model compared to Schmidt test

To compare the model with the Schmidt test, random sets with various lifetimes was generated for a single decay step corresponding to a random $T_{1/2}$. A contaminating species with a half-life $T_{1/2}^{contam}$ is introduced, where $T_{1/2}^{contam}$ is shown for different factors (with a small random noise factor added such that the ratio is not an exact integer number). Results are shown in Figure 14 and Figure 15.

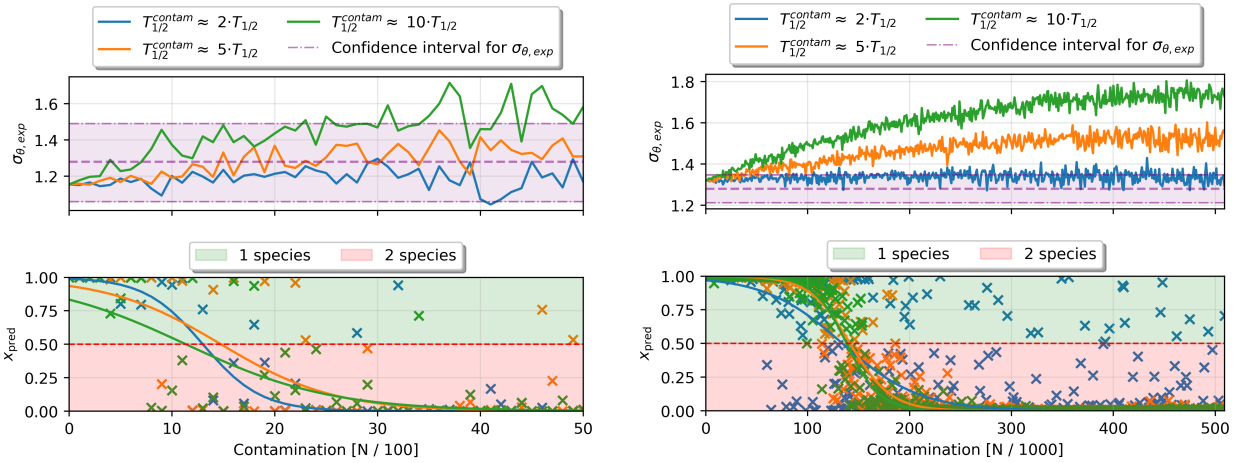

Figure 14: The Schmidt test measure $\sigma_{\Theta,exp}$ vs. the classification prediction X_{pred} of the model on a randomly generated set of 100 decay times (left); and 1000 decay times (right). Results are plotted for various levels of contamination C and various ratios between $T_{1/2}^{contam}$ and $T_{1/2}$. An inverse sigmoid function is fit to X_{pred} values in order to show the trend of predictions. It can be seen that the threshold is passed between 10% and 20% relative contamination for both $N = 100$ decay times and $N = 1000$ decay times. The ML-algorithm detects congruence earlier than the Schmidt test for the $N = 100$ case. By contrast, for $N = 1000$ the Schmidt test detects incongruence earlier than the ML-algorithm.

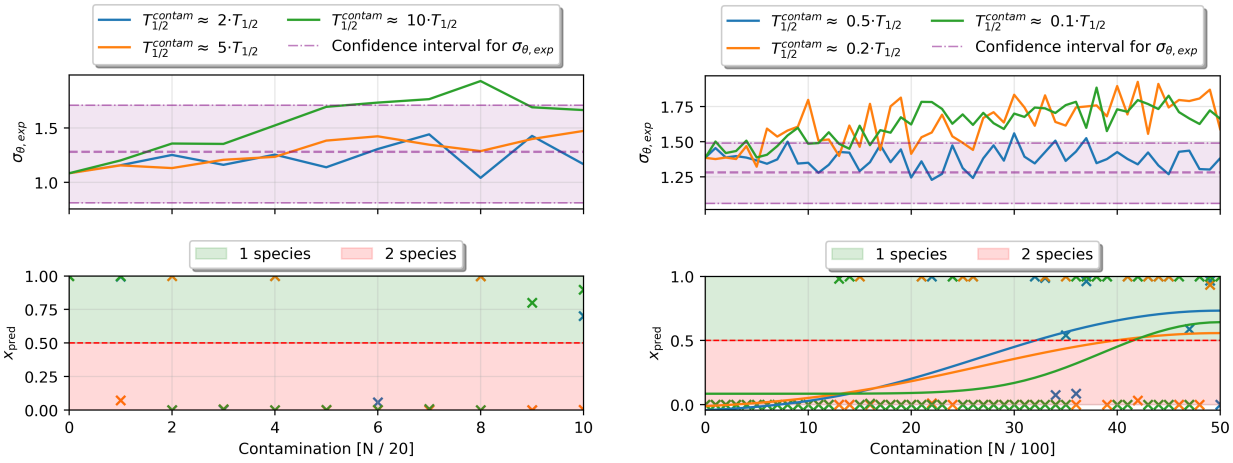

Figure 15: Same as Figure 14 but for 20 decay times (left); and 100 decay times (right), with $T_{1/2}^{contam} < T_{1/2}$. The ML-algorithm predictions for $N = 20$ are inconsistent, and no decision boundary is visible for any of the half-life ratios (the sigmoidal fit failed). The Schmidt test is also mostly unable to detect contamination on this particular set. With $T_{1/2}^{contam} < T_{1/2}$ and $N = 100$ decay times, the Schmidt test detects incongruence as effectively as when $T_{1/2}^{contam} > T_{1/2}$, whereas the machine learning algorithm is making incorrect predictions.

5 Discussion

Schmidt test Evaluating the Schmidt test for a varied collection of MC simulated data shows that the test is rather consistent when no contamination is present in a set of lifetimes. Since decay time sets are generated randomly along the CDF, there is a chance that some sets have statistical outliers for which the Schmidt test may not indicate congruence. This may explain why not 100% of sets are deemed congruent for the self-consistency test when $T_{1/2}^{contam} = T_{1/2}$. Moreover, since the confidence interval gets smaller proportional to $\sqrt{N}$ [6], it is to be expected that the fluctuations will cause σ_Θ to indicate incongruence more as the set size N increases. This explains the slightly lower congruence percentage in $90 \leq N \leq 100$ even when all correlation times originate from the same half-life.

The Schmidt test picks up rather poorly on contamination when the ratio between the half-lives is not very large, as seen in Table 1. An explanation lies within Equation 8. When the difference between half-lives of the species is small, the correlation times recorded are statistically unlikely to vary much from the mean $\bar{\Theta}$. For low counting statistics, $N \ll 100$, the Schmidt test seems rather ineffective at detecting incongruence, regardless of the factor between the half-lives. For higher counting statistics ($N > 10$), the Schmidt test seems to detect incongruence best when the contaminated species represents about half of the lifetimes, with performance of the Schmidt test decreasing when the relative contamination level is very high or very low. Evidently, when the contamination is dominant the dataset becomes congruent again, the original population acts as the contaminant instead.

Overall, it can be concluded that the Schmidt test detects incongruence better with *i*) higher counting statistics (larger set size N); *ii*) a larger ratio between $T_{1/2}^{contam}$ and $T_{1/2}$; and *iii*) a sizeable amount of contamination.

During analysis it was noted that the heatmaps in Figure 10 seem to have a 'structure' of abrupt change at every N divisible by 10. The lines are an artifact caused by the use of discrete tabulated values. For the confidence intervals of σ_Θ, tabulated values from Ref. [6] were used. For $10 < N <= 100$, the confidence intervals for $N = 20, 30, 40..$ are tabulated, but not for values of N in between. No analytical formula for $N < 100$ is provided in the paper. As such, the most conservative tabulated confidence interval for any N was used. For example, for $N = 39$, the confidence interval for $N = 40$ was used, but for $N = 41$, the interval for $N = 50$ is used. This approach and its resulting artifacts do not affect the conclusions of this work.

Statistical properties of contaminated decay sets Figure 13 indicates that both skewness γ_τ and kurtosis κ_τ get affected significantly by as little as 1 contaminated correlation time in the region $40 \leq N \leq 100$. The Schmidt test primarily relies on the standard deviation (σ_τ), but the heatmap for kurtosis seems to indicate that higher order moments may also be helpful in separating congruent sets of decay times from incongruent ones.

ML-algorithm compared to the Schmidt test Figure 14 shows the performance of the machine learning model compared to the Schmidt test for relatively high counting statistics in SHN research ($N = 100$ and $N = 1000$). The machine learning model predicts incongruence slightly earlier than the Schmidt test in the case of a set of lifetimes of size $N = 100$. The model picks up from around 11-16 contaminated correlation times (depending on the ratio), whereas the Schmidt

test indicates congruence for all ratios of lifetimes at this level of contamination.

However, a look at Figure 15 reveals that the algorithm performs rather poorly in a low statistics test ($N = 20$). Furthermore, the case where $T_{1/2}^{contam} < T_{1/2}$ seems to be handled very poorly by the model. Figure 15 shows that the ML-algorithm fails to predict well when the contaminating species is a fraction of the original species, with virtually no correct predictions done. This is despite a 91% validation accuracy during training. This implies there could be several methodological causes for such behaviour; overfitting of the algorithm, improper balance of the training data or problems with feature scaling. Despite these issues being considered, as discussed in Section 3.3 and Section 3.4, it could be that the model is still overfitting to the training data. It is also possible that the input features considered are not enough to properly classify congruent and incongruent sets; more information may be needed than the current features used.

6 Conclusion and outlook

In situations where little statistics are available, determining congruence is a challenging task. The Schmidt test and its generalised form have proven to be beneficial tools, but the evaluation done in this report shows that the Schmidt test can be inconclusive in some scenarios. By attempting to develop a machine learning algorithm to compliment the test, we have seen that a few specific statistical features were picked up by the algorithm which are lacking in the Schmidt test. This indicates that currently ignored higher statistical moments such as skewness and kurtosis could possibly play a role in determining decay chain congruence. However, the algorithm has poor robustness, particularly in low statistics scenarios. More research would need to be done into the role that statistical moments beyond the variance could play in determining congruence of sparse decay chain data.

Inclusion of decay chain correlations The algorithm designed for single decays has limited use for decay chains with several steps, as information from entire chains is lost. As such, a new algorithm would have to be designed in order to draw comparisons to the Generalised Schmidt test. This algorithm would have to take in additional features that quantify information about the entire chain.

Inclusion of decay energies This thesis has only focused on decay times, but for the identification and classification of entire chains, the characteristic α-decay energies are important. This is because decay times can occasionally be very similar at different decay steps, in particular when *missing* steps are involved. Once again, more features would have to added to capture this information and to account for such missing steps.

Alternative models Neural networks other than MLPs could be explored. In particular, the convolutional neural network (CNN), which is commonly used for pattern recognition of images, could have a use when both correlation times and energies are considered.

Despite some limitations in very low counting statistics, the Schmidt test remains a reliable tool for verifying experimental data from superheavy nuclei research. More work is needed into what features of a set of decay chains could be used for determining congruence using a machine learning algorithm. The next iteration of this work should be able to capitalize on the findings from this work.

7 Bibliography

References

[1] Sigurd Hofmann. "Super-Heavy Nuclei". In: *Journal of Physics G: Nuclear and Particle Physics* 42.11 (Nov. 2015), p. 114001. ISSN: 0954-3899, 1361-6471. DOI: 10.1088/0954-3899/42/11/114001.

[2] Odile R Smits. "The Quest for Superheavy Elements and the Limit of the Periodic Table". In: *Nature* 6 (2024).

[3] J.B. Roberto et al. "Actinide Targets for the Synthesis of Super-Heavy Elements". In: *Nuclear Physics A* 944 (Dec. 2015), pp. 99–116. ISSN: 03759474. DOI: 10.1016/j.nuclphysa.2015.06.009.

[4] J B Roberto. "Actinide Targets for the Synthesis of Superheavy Nuclei". In: *Eur. Phys. J. A* (2023).

[5] Zu-Hua Liu and Jing-Dong Bao. "Possibility to Produce Element 120 in the 54 Cr+ 248 Cm Hot Fusion Reaction". In: *Physical Review C* 87.3 (Mar. 2013), p. 034616. ISSN: 0556-2813, 1089-490X. DOI: 10.1103/PhysRevC.87.034616.

[6] K H Schmidt. "A New Test for Random Events of an Exponential Distribution". In: *The European Physical Journal A* (2000).

[7] U. Forsberg et al. "Recoil-α-Fission and Recoil-α–α-Fission Events Observed in the Reaction 48Ca + 243Am". In: *Nuclear Physics A* 953 (Sept. 2016), pp. 117–138. ISSN: 03759474. DOI: 10.1016/j.nuclphysa.2016.04.025.

[8] V.I. Zagrebaev and W. Greiner. "Cross Sections for the Production of Superheavy Nuclei". In: *Nuclear Physics A* 944 (Dec. 2015), pp. 257–307. ISSN: 03759474. DOI: 10.1016/j.nuclphysa.2015.02.010.

[9] A. K. Nasirov et al. "Basic Distinctions between Cold- and Hot-Fusion Reactions in the Synthesis of Superheavy Elements". In: *Physics of Atomic Nuclei* 77.7 (July 2014), pp. 881–889. ISSN: 1063-7788, 1562-692X. DOI: 10.1134/S1063778814070126.

[10] I Bacho et al. "Gas-Filled Mass-Separator for Investigation of the Heavy Ion Reaction Products." In: (Jan. 1969).

[11] M. Leino. "Gas-Filled Separators – An Overview". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 204 (May 2003), pp. 129–137. ISSN: 0168583X. DOI: 10.1016/S0168-583X(02)01901-8.

[12] K. Morita et al. "Production and Decay of the Isotope 271 Ds (Z = 110)". In: *The European Physical Journal A* 21.2 (Aug. 2004), pp. 257–263. ISSN: 1434-6001, 1434-601X. DOI: 10.1140/epja/i2003-10205-1.

[13] V. Ninov, K. E. Gregorich, and C. A. McGrath. "The Berkeley Gas-filled Separator". In: *Exotic Nuclei and Atomic Masses (ENAM 98)*. Bellaire, Michigan (USA): ASCE, 1998, pp. 704–707. ISBN: 978-1-56396-804-4. DOI: 10.1063/1.57362.

[14] Dirk Rudolph. "Superheavy Element Studies with TASCA at GSI: Spectroscopy of Element 115 Decay Chains". In: *Proceedings of the Conference on Advances in Radioactive Isotope Science (ARIS2014)*. Tokyo, Japan: Journal of the Physical Society of Japan, June 2015. ISBN: 978-4-89027-110-8. DOI: 10.7566/JPSCP.6.010026.

[15] L.-L. Andersson et al. "TASISpec—A Highly Efficient Multi-Coincidence Spectrometer for Nuclear Structure Investigations of the Heaviest Nuclei". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 622.1 (Oct. 2010), pp. 164–170. ISSN: 01689002. DOI: 10.1016/j.nima.2010.06.243.

[16] Anton Såmark-Roth. "Spectroscopy along Decay Chains of Element 114, Flerovium". In: (2021).

[17] J. Khuyagbaatar et al. "Search for Elements 119 and 120". In: *Physical Review C* 102.6 (Dec. 2020), p. 064602. ISSN: 2469-9985, 2469-9993. DOI: 10.1103/PhysRevC.102.064602.

[18] Ulrika Forsberg. "Element 115". PhD thesis. Lund University, 2016.

[19] M. Asai, F.P. Heßberger, and A. Lopez-Martens. "Nuclear Structure of Elements with 100 $\leq$Z$\leq$ 109 from Alpha Spectroscopy". In: *Nuclear Physics A* 944 (Dec. 2015), pp. 308–332. ISSN: 03759474. DOI: 10.1016/j.nuclphysa.2015.06.011.

[20] Yu. Ts. Oganessian et al. "New Isotope Mc 286 Produced in the Am 243 + Ca 48 Reaction". In: *Physical Review C* 106.6 (Dec. 2022), p. 064306. ISSN: 2469-9985, 2469-9993. DOI: 10.1103/PhysRevC.106.064306.

[21] J. M. Gates et al. "Decay Spectroscopy of Element 115 Daughters: Rg 280 $\rightarrow$ Mt 276 and Mt 276 $\rightarrow$ Bh 272". In: *Physical Review C* 92.2 (Aug. 2015), p. 021301. ISSN: 0556-2813, 1089-490X. DOI: 10.1103/PhysRevC.92.021301.

[22] Mattias Ohlsson and Patrik Eden. "Introduction to Artificial Neural Networks and Deep Learning". In: (2022), pp. 6–10.

[23] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: ().

[24] The MathWorks Inc. *What Is Overfitting?*

[25] C. A. Mitrea, C. K. M. Lee, and Z. Wu. "A Comparison between Neural Networks and Traditional Forecasting Methods: A Case Study". In: *International Journal of Engineering Business Management* 1 (Jan. 2009), p. 11. ISSN: 1847-9790, 1847-9790. DOI: 10.5772/6777.

[26] Quan Zou et al. "Finding the Best Classification Threshold in Imbalanced Classification". In: *Big Data Research* 5 (Sept. 2016), pp. 2–8. ISSN: 22145796. DOI: 10.1016/j.bdr.2015.12.001.

[27] Chollet. *Keras 3 API Documentation / Callbacks API / EarlyStopping.*

[28] André Altmann et al. "Permutation Importance: A Corrected Feature Importance Measure". In: *Bioinformatics* 26.10 (May 2010), pp. 1340–1347. ISSN: 1367-4811, 1367-4803. DOI: 10.1093/bioinformatics/btq134.

A Heatmaps for statistical moments of contaminated decay sets

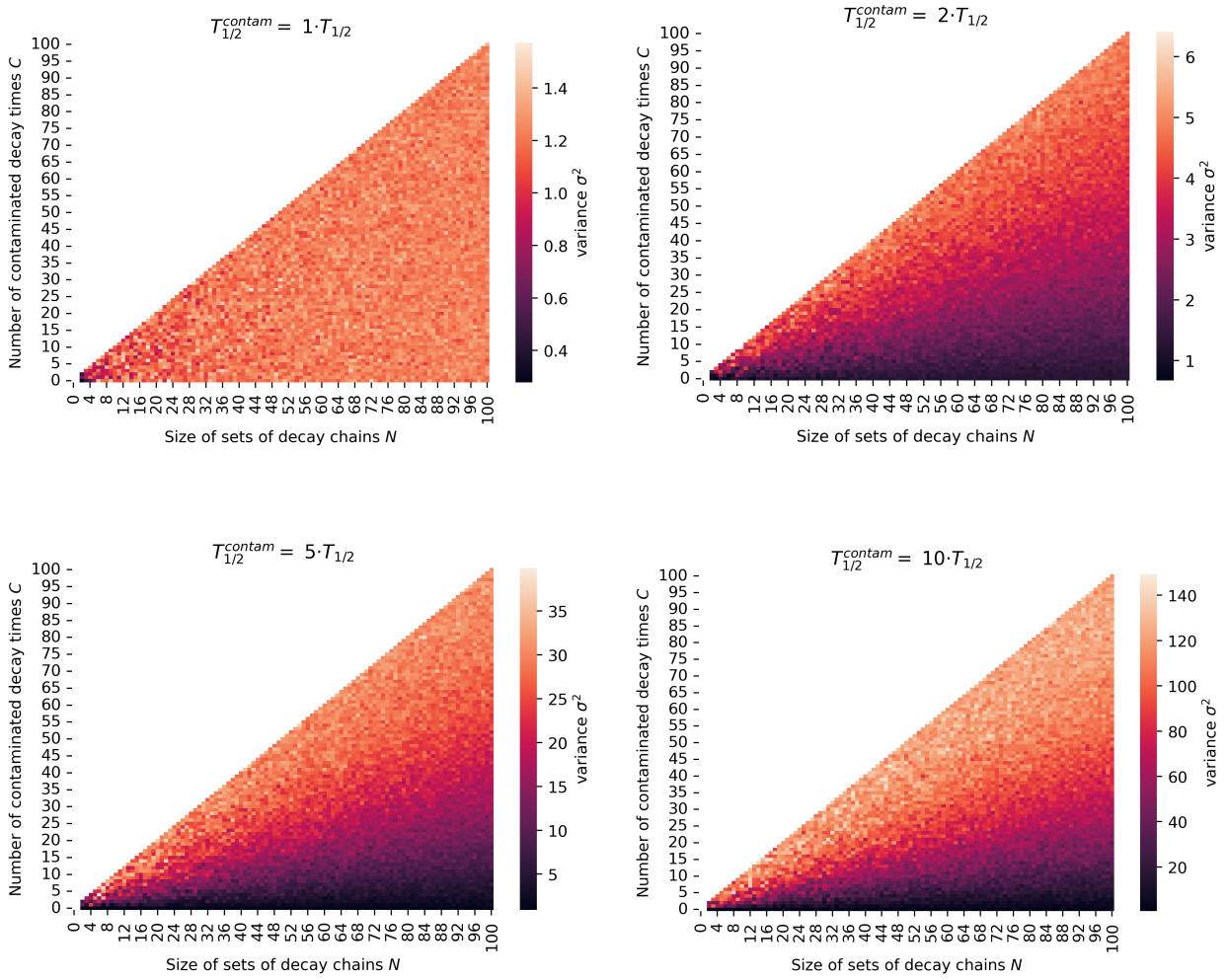

Figure 16: Heatmap of the variance σ^2 for varying ratios between $T_{1/2}$ and $T_{1/2}^{contam}$. The value of each pixel represents the mean variance for 100 simulated chain sets of size N with C contaminated decay times.

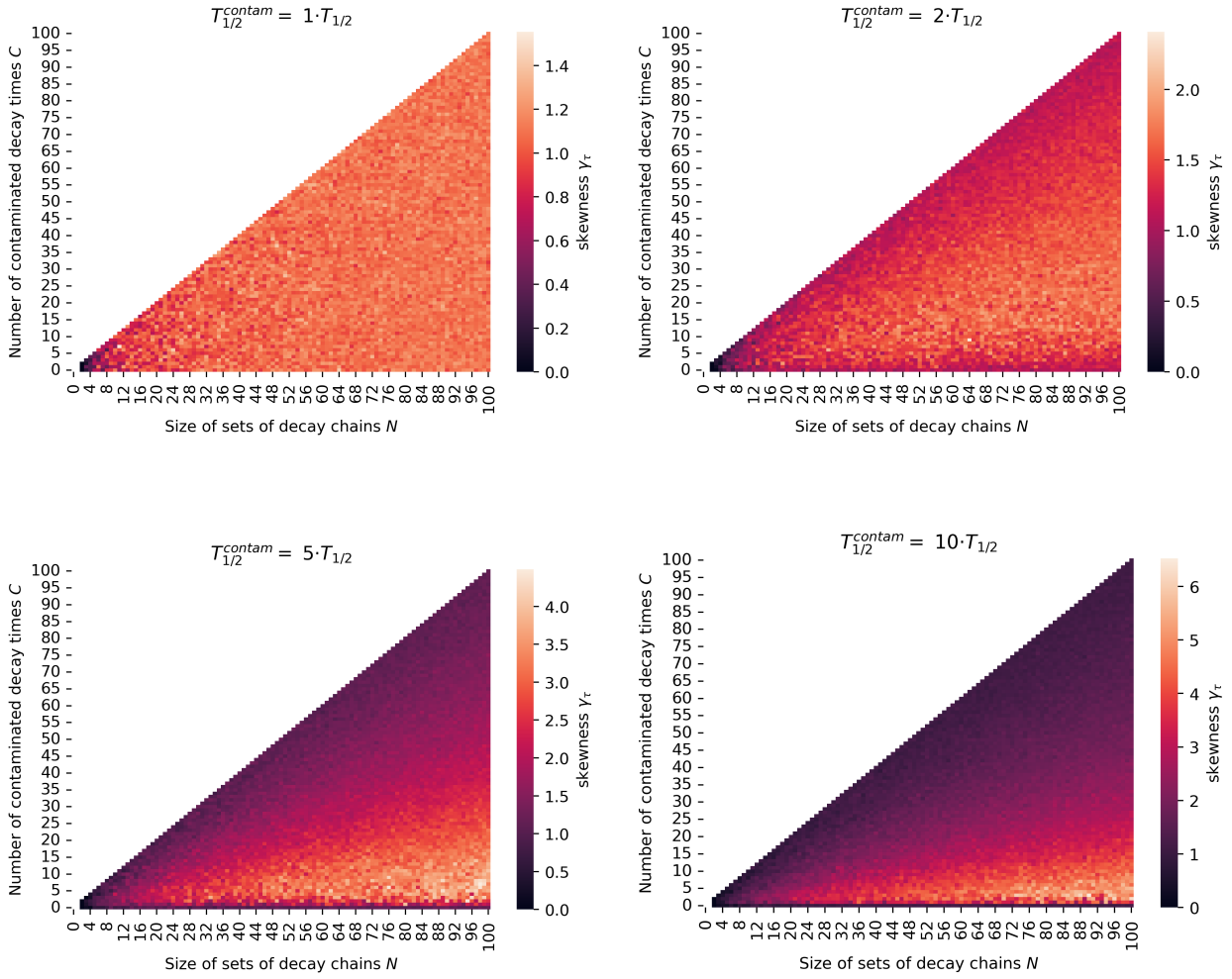

Figure 17: Heatmap of the skewness γ_τ for varying ratios between $T_{1/2}$ and $T_{1/2}^{contam}$. The value for each pixel represents the mean variance for 100 simulated chain sets of size N with C contaminated decay times.

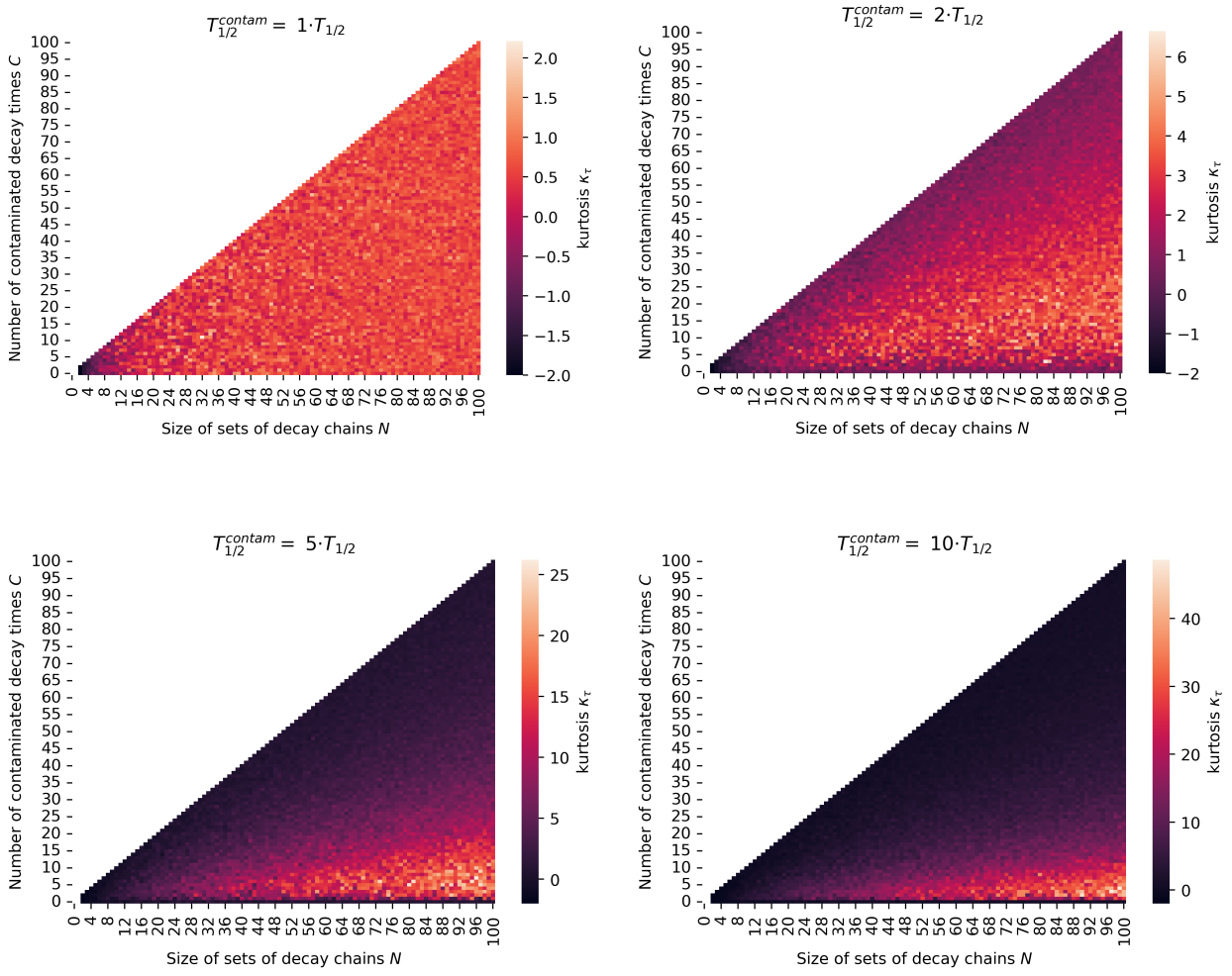

Figure 18: Heatmap of the kurtosis κ_τ for varying ratios between $T_{1/2}$ and $T_{1/2}^{contam}$. The value for each pixel represents the mean variance for 100 simulated chain sets of size N with C contaminated decay times.