



A novel approach for nearly-coincident events rejection

M. Borghesi^{2,3,a} , M. De Gerone¹, M. Faverzani^{2,3}, M. Fedkevych¹, E. Ferri^{2,3}, G. Gallucci¹, A. Giachero^{2,3}, A. Nucciotti^{2,3}, A. Puiu^{4,5}

¹ Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Genova, 16146 Genoa, Italy

² Dipartimento di Fisica “G. Occhialini”, Università di Milano, Bicocca, 20126 Milan, Italy

³ Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Milano-Bicocca, 20126 Milan, Italy

⁴ Gran Sasso Science Institute (GSSI), 67100 L’Aquila, Italy

⁵ INFN, Laboratori Nazionali del Gran Sasso, 67010 Assergi, L’Aquila, Italy

Received: 11 January 2021 / Accepted: 16 April 2021 / Published online: 4 May 2021

© The Author(s) 2021

Abstract We present a novel technique, called DSVP (Discrimination through Singular Vectors Projections), to discriminate spurious events within a dataset. The purpose of this paper is to lay down a general procedure which can be tailored for a broad variety of applications. After describing the general concept, we apply the algorithm to the problem of identifying nearly coincident events in low temperature microcalorimeters in order to push the time resolution close to its intrinsic limit. In fact, from simulated datasets it was possible to achieve an effective time resolution even shorter than the sampling time of the system considered. The obtained results are contextualized in the framework of the HOLMES experiment, which aims at directly measuring the neutrino mass with the calorimetric approach, allowing to significantly improve its statistical sensitivity.

1 Introduction

The demand for increasing sensitivities of nowadays experiments requires the development of complex analysis tools to respond to several demands, according to the design and goals of the experiment. In many experiments, a crucial factor in achieving a high sensitivity is the ability to discriminate spurious events. This is a particularly relevant feature to keep into account for experiments where the statistics of the spurious events might be comparable with, or even overcome, the statistics of the proper events. This is the case, for instance, of the direct measurement of the neutrino mass with the calorimetric approach [1].

So far, few techniques are currently employed for the purpose of discriminating the spurious events from the proper ones, and they all require that events belonging to these two fam-

ilies must differ in some way from each other. In this paper we outline a novel technique, called DSVP (Discrimination through Singular Vectors Projections), based on a previous work by Alpert et al. [2]. Besides its effectiveness, one of the main advantage of this method, is that it does not rely on any particular assumption about the structure of the events, thus in principle it can be applied in various scenarios almost in a semi-automatic way.

In Sect. 2 we illustrate the DSVP method while in the second part of the article (Sect. 3 and 4) we show an application of this technique in the context of a direct calorimetric neutrino mass measurements. In particular, we will focus on the simulated data sets representing the one foreseen for detectors used in the HOLMES experiment [3], for which the expected main source of background will be unrecognized pile-up events.

2 The DSVP technique

The aim of the DSVP algorithm is to discriminate as many as possible undesirable events (i.e. spurious events which differ from a reference signal) present in a given dataset, using the information about the mean ‘morphology’ of the events present in the dataset.

In order to apply the DSVP technique, the following elements are required:

- The measured dataset, \mathbf{M} . This $n \times d$ matrix consists of the dataset of interest, where each row is an event described by d variables. Namely, the events can be seen as points in the \mathbb{R}^d space. From now on, we call the *good* events in the dataset A events, while B events are the ones to be

^a e-mail: matteo.borghesi@mib.infn.it (corresponding author)

rejected. We assume that the A events are more numerous respect to the B ones ($N_A > N_B$).¹

- The expected number of B events N_B that the algorithm should discard at most.
- A training dataset, \mathbf{T} , such that $N_A \gg N_B$. The events of this $n' \times d$ matrix can be distributed in a different region of \mathbb{R}^d respect to the events in \mathbf{M} . For instance, in the case of microcalorimeter signals, the events in \mathbf{T} can lie in a different energy range respect to the events in \mathbf{M} .

We will use the training dataset \mathbf{T} to define a new vector space which will help us to highlight the features that distinguish an A event from a B one. This new vector space, called from now the projections space, has dimension k , with $k \ll d$. The events can be represented as points in the \mathbb{R}^k projection space, so that the A events are distributed differently respect to the B ones. The idea is to find a model (i.e. hypersurfaces) describing the distribution of the A points in \mathbf{M} in this new space, so that the B points can be identified as the ones with a larger distance from what predicted by the model.

In order to find the model we need to 'clean' the dataset first, obtaining a subset of \mathbf{M} , $\mathbf{M}' \subset \mathbf{M}$, which contains mostly A events at the expense of deleting also some A events.

The next step is to represent the events in \mathbf{M}' in the projection space and to find the model parameters which describes the distribution of the \mathbf{M}' ($\sim A$) events.

We then define the discrimination parameter and its threshold to recognize an A event from a B one in \mathbb{R}^k .

Finally, we take the original dataset \mathbf{M} , represent the events in the projection space, find the discrimination parameter and discard all the events that have a value of the discrimination parameter above the threshold found.

The procedure (dataset 'cleaning', model and threshold definition and B discrimination) is then repeated with the survived events. At each iteration, the \mathbf{M} dataset will contain a smaller fraction of B events.

In the following sections, each steps of the algorithm are described in detail.

2.1 Raw cleaning with PCA

In the first step, the aim is to create a suitable dataset for modeling the distribution of A events in the \mathbf{M} matrix, lowering the ratio N_B/N_A at the expense of deleting also A events. Knowing that $N_A > N_B$, the mean 'morphology' of the events is closer to the A ones. We can define a suitable parameter using the Principal Component Analysis (PCA) [4] to discard mainly B events.

The procedure used is equal to the one described in [2], which will be reported for completeness. The singular value decomposition (SVD) [5] is computed for the $n \times d$ matrix \mathbf{M} , which

is decomposed in a product of three matrices $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^T$. The columns of \mathbf{U} and \mathbf{V} are the left and right singular vectors respectively, while the entries of the diagonal matrix \mathbf{D} are the singular values. The singular values are ordered from 1 to d in order of importance. Only the first $j < d$ columns of \mathbf{D} are non-negligible. It is convenient to define a new matrix $\hat{\mathbf{U}}$ which contains only the first j columns of \mathbf{U} subtracted by their means which is equivalent to centering the data matrix, as required by the PCA.

The columns of $\hat{\mathbf{U}}$ are vectors of length n . Basically, they represent the projections of the mean-centered events contained in \mathbf{M} on the right singular vectors (i.e. the columns of \mathbf{V} , which are called principal vectors in the PCA framework) with the first column of $\hat{\mathbf{U}}$ expressing the projections on the first right singular vector and so on. The columns of \mathbf{V} are vectors of dimension d which represent the direction of greatest variance of the data in \mathbf{M} . Thanks to the properties of the PCA, an appropriate combination of the projections can be of use to define a parameter, called $norm^2$, which indicates how close an event is to the mean 'morphology' of the events in \mathbf{M} .

The precision matrix $(\sigma^2)^{-1}$ is computed from the $j \times j$ empirical covariance $\sigma^2 = \hat{\mathbf{U}}^T \hat{\mathbf{U}}$ and it is used to evaluate the parameter $norm^2$ for each event $i = 1, \dots, n$ in the matrix \mathbf{M}

$$norm_i^2 = \hat{\mathbf{U}}_{i,*} (\sigma^2)^{-1} \hat{\mathbf{U}}_{i,*}^T \quad (1)$$

Suppose that we have a guess of how many B events are expected in the dataset. We call this number N_B^{guess} . B events deviate disproportionately from the mean in this covariance-adjusted sense, so we discard those with largest $norm^2$ and repeat the procedure on the remaining data a total of l times, removing on the l -th iteration a number of events equal to $N_B^{guess}/2^l$ with the largest $norm^2$. In our tests, we use $l = 5$. The iterations guarantee that the mean morphology of the events are closer and closer to the A events each cycle, as B events are increasingly eliminated.

After the PCA, we have eliminated

$$N_{del}^{PCA} = \sum_l \frac{N_B^{guess}}{2^l} \quad (2)$$

events, where the B events are the ones predominantly discarded. The remaining events after the PCA are $m = n - N_{del}^{PCA}$. We call \mathbf{M}' the $m \times d$ the matrix of the survived events, which is mostly composed of A events.

2.2 Define a model for the A-events

To discriminate the undesirable events, we now need to define a model which describes the distribution of the A points (the ones belonging to \mathbf{M}') in the projection space.

¹ See Sect. 2.1.

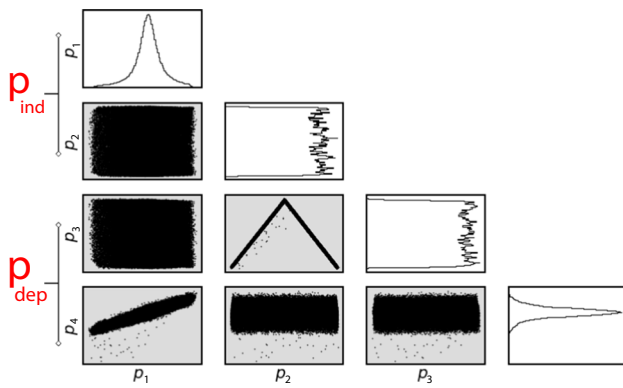


Fig. 1 An example of distribution of the points in \mathbf{T} in the projection space from Sect. 3. In this particular case, the projection space is in \mathbb{R}^4 . We decided to set $k' = 2$, thus describing the points distribution in \mathbb{R}^4 with two curves: $p_3 = f_3(p_1, p_2)$ and $p_4 = f_4(p_1, p_2)$

First, we need to define this space. We decompose the \mathbf{T} matrix using the SVD. Because the training matrix \mathbf{T} is mainly composed by A events, we assume that its first k significant right singular vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ can constitute a base of the projection space.

The events in \mathbf{M}' are projected onto these vectors. From now on, each event in \mathbf{M}' will be described by $k < d$ variables, its projections onto the right singular vector of \mathbf{T} . We indicate all the coordinates of the \mathbf{M}' points along the i -th base vector of the projection space as $p_i = \mathbf{M}' \cdot \mathbf{v}_i$.

To describe the points distribution in the new vector space, the projections p are classified into two groups: the k' independent projections, indicated as \mathbf{p}_{ind} and the dependent ones, \mathbf{p}_{dep} .

$$\{p\}_k = \underbrace{p_1, \dots, p_{k'}}_{\mathbf{p}_{ind}}, \underbrace{p_{k'+1}, \dots, p_k}_{\mathbf{p}_{dep}} \tag{3}$$

The dependent projections can be expressed as a function of the independent ones. There is no general rule to identify which projection is “independent” and which one is not, since it is related to the specific problem. The training dataset can be used to identify the dependencies among the projections, as shown in Fig. 1.

The distribution of the dependent projections can now be easily described in a $\mathbb{R}^{k'+1}$ subspace by a set of f curves

$$p_i = f_i(\mathbf{p}_{ind}) \quad ; \quad i = k' + 1, \dots, k. \tag{4}$$

Knowing precisely the set of curves $\{f\}$, we will be able to differentiate between the two distribution of events, because the projection of the B events will not follow the same curves as the one of the A events.

Usually the functional form of the different f is unknown. However, we can approximate each f curve with a Taylor expansion and let a (weighted) linear regression find the best

parameters of the expansion. In particular, we use a modified version of the random sample consensus (RANSAC) algorithm [6].²

The set of curves $\{f\}$ which describes the \mathbf{M}' events in the projection space is what we called the model.

2.3 Find a discrimination threshold

The difference between the measured dependent projections and the ones expected from the model is evaluated for each event in the \mathbf{M}' matrix. A residual norm is defined as

$$d = \sqrt{\sum_{j=k'+1}^k (p_j - f_j(\mathbf{p}_{ind}))^2}. \tag{5}$$

In order to discriminate between the A events, the one with the lowest residual norm, and the B , we need to define a threshold value, d_{thr} . Due to the fact that the \mathbf{M}' dataset is mainly made of A events the threshold is chosen as the highest value of d plus the standard deviation of the d distribution

$$d_{thr} = \max\{d\} + \text{std}\{d\}. \tag{6}$$

This threshold definition should ensure to include not only the A events in \mathbf{M}' , but also the A events in the original dataset \mathbf{M} which were eliminated by the ‘PCA cleaning’ described in Sect. 2.1. Nevertheless, this definition of threshold might need to be redefined to account for the specific problem considered.

2.4 Apply the model

Now all the components to make the algorithm work are present: a base for the projection space, a set of curves to model the points distribution in that space and a discrimination threshold.

We will now use these tools on the original dataset \mathbf{M} , namely:

1. Take the inner product of the events in \mathbf{M} with the base of the projection space, determining p_k .
2. Evaluate the residual norm d using the curves describing the A projections distributions.
3. The events with a residual norm above the threshold are discarded.

After the third step, we will have discarded N_{del} events. The events deleted by the third step will be almost, if not all,

² Due to the fact that the number of outliers (B events) from step Sect.2.1 is expected to be negligible, any type of weighted linear regression can in principle be used.

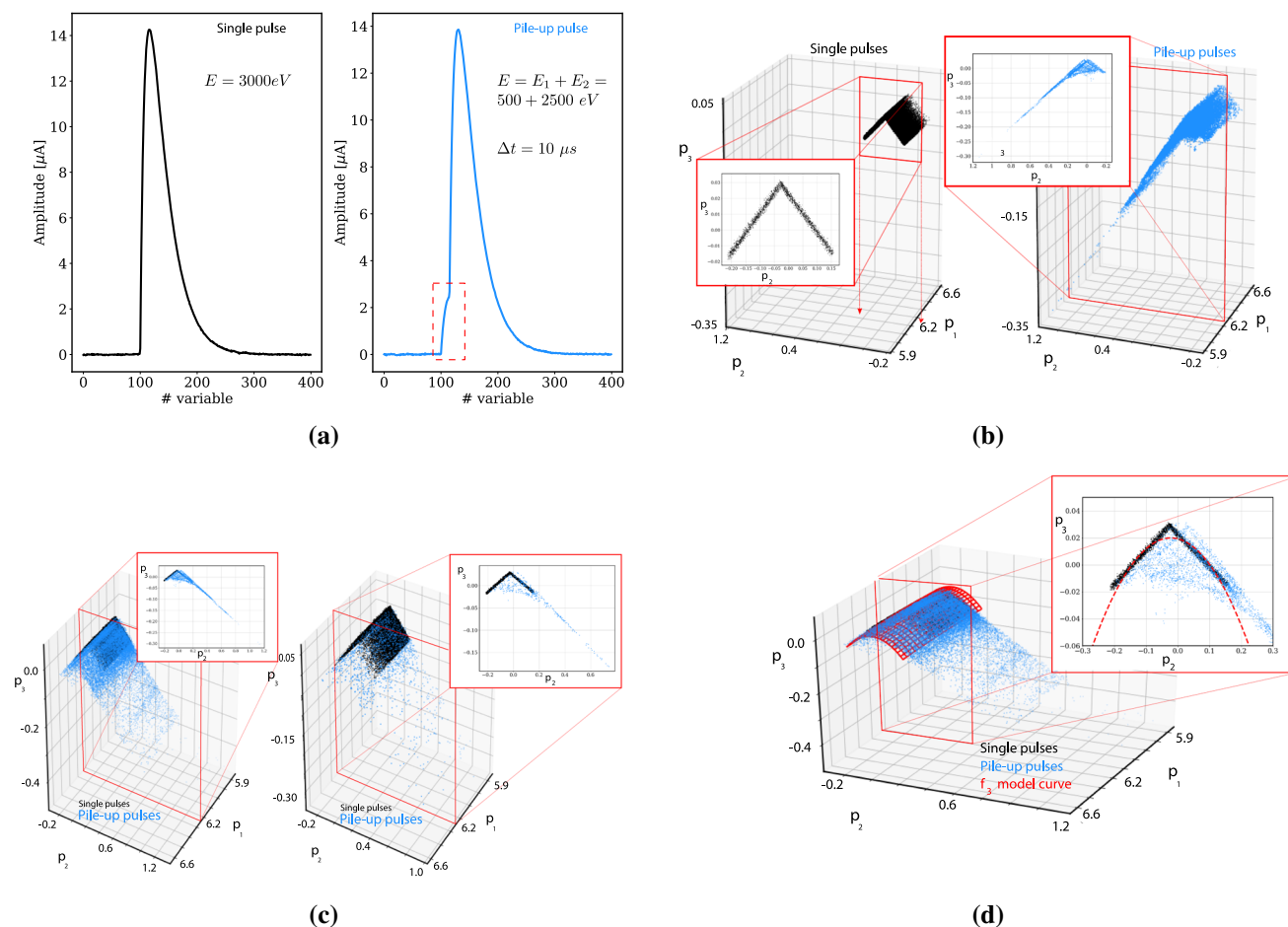


Fig. 2 **a** Each event (row) of the matrix \mathbf{M} is initially described by 400 variables, i.e., samples; we can represent each event as shown in figure or as points in \mathbb{R}^{400} . \mathbf{M} contains two different types of events: single pulses (A) of energy E and pile-up pulses with different arrival time (B) with energies E_1 and E_2 such that $E_1 + E_2 = E$. **b** The events of \mathbf{M} are represented in the projections space. In this space, the two types

of events follow two different distributions. **c** In the left (right) panel the matrix \mathbf{M} (\mathbf{M}') is represented in the projection space. It is possible to appreciate how the PCA has drastically reduced the fraction of pile-up. **d** The curve $f_3 = f_3(p_1, p_2)$, which describes the distribution of the events in \mathbf{M}' , is used to discriminate between the single pulse and pile-up pulses

spurious B events. All the previous steps (PCA, model and threshold definition) are now repeated with a reduced number of expected B events, $N^{B'} = N^B - N_{del}$. The iterations successively improve the representation of A events, as B events are increasingly eliminated. The algorithm stops when $N_{del} = 0$ or when $N^{B'} = 0$.

Figure 2 shows a visual representation of some of the steps of the DSVP technique. As an example the figure reports signals from a TES microcalorimeter, as explained in Sect. 3.2. This particular case was chosen because there are just three non-negligible singular values, therefore the points in the projection space can be easily shown on a 3D plot.

The method was implemented in python, taking advantage of many of the fast modules of NumPy and SciPy. The majority of the computational time is taken by the *Raw cleaning with PCA* part, due to the fact that the SVD on the matrix \mathbf{M} is performed five times for each iteration. Nevertheless, the

algorithm is quite fast, taking ~ 7 min to compute nine iterations on a matrix \mathbf{M} composed of 1,20,000 rows and 1024 columns of float32 numbers, using only one (six years old) CPU with a base clock of 2.6 GHz.

3 HOLMES and pile-up discrimination

The algorithm described in Sect. 2 is now applied in the framework of HOLMES. The HOLMES experiment will perform a direct measurement of the neutrino mass with a sensitivity of the order of 1 eV, measuring the energy released in the electron capture (EC) of ^{163}Ho , as proposed by De Rujula and Lusignoli in [7]. It will also demonstrate the scalability of the calorimetric technique for a next generation experiments that could go beyond the current best expected sensitivity of 0.1 eV [8]. In order to reach this sensitivity, HOLMES

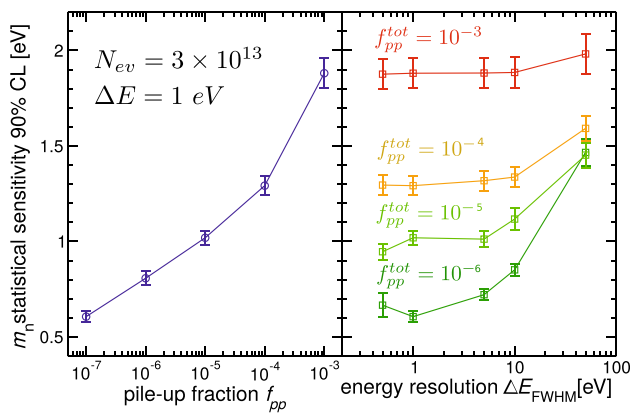


Fig. 3 Expected neutrino mass sensitivity for the HOLMES experiment. The right panel shows the sensitivity compared to the detectors energy resolution at different pile-up fraction. We called the pile-up fraction on the whole energy spectrum f_{pp}^{tot}

will use low temperature TES microcalorimeters with ^{163}Ho implanted in their absorbers, with an activity of 300 Hz per detector.

The effect of a non-zero neutrino mass on the ^{163}Ho EC decay spectrum can be appreciated only in a energy region very close to the end point, where the count rate is low and the fraction of nearly-coincident events, called pile-up events, to single events is greater than one. If a pile-up event is composed of two events of energies E_1 and E_2 which occur within a time interval shorter than the time resolution of the detector, it is recorded as a single event with energy $E \simeq E_1 + E_2$. Thus, if not correctly identified, pile-up events will distort the decay spectrum of ^{163}Ho , lowering the sensitivity to m_ν . The neutrino mass sensitivity of HOLMES has been evaluated through Monte Carlo simulations [9], see Fig. 3. Once fixed the number of recorded events to 3×10^{13} , the simulations have shown that the sensitivity on neutrino mass is not strongly dependent on the energy resolution of the detector (as long as $\Delta E < 10$ eV), but rather on the pile-up fraction f_{pp} , i.e. the ratio between the number of pile-up events to single events. Its reduction is crucial for the success of the experiment.

Using the terminology of Sect. 2, in the HOLMES experiment an *A* event is a signal caused by a single energy deposition in the microcalorimeter detector, while a *B* event is a signal caused by nearly coincident events. Each signal is a collection of records I_i of the current flowing through the detector sampled at an instant $t_i = i \times t_{samp}$, where t_{samp} is the sampling time of the readout system. An example of a microcalorimeter signal is shown in Fig. 2a. With the current setup, the sampling time is fixed at 2 μs .

We tested the algorithm robustness and efficiency through many simulations which aim at emulating the results expected by the HOLMES experiment. The HOLMES TES microcalorime-

ters do not have the ^{163}Ho implanted yet, therefore a real data test will be done at later times.

3.1 Energy spectrum and ROI definitions

^{163}Ho decays via electron capture to an atomic excited state of ^{163}Dy which relaxes mostly emitting atomic electrons (i.e. the fluorescence yield is less than 10^{-3} [7]). The de-excitation energy E_c spectrum probability density is proportional to

$$\frac{d\lambda_{EC}}{dE_c} \propto \sqrt{(Q - E_c)^2 - m_\nu^2} \tag{7}$$

where m_ν is the effective neutrino mass. The pulses are generated according to the spectrum in [7] with $Q = 2.833$ keV [10], $m_\nu = 0$ and with energy between 2.650 keV and 2.900 keV. Second order effects like shake-up and shake-off [11] have not been considered in this work. Despite the optimal region of interest (ROI) aimed at determining the neutrino mass will be determined only when actual data will be collected, this energy range can be considered as a reasonable ROI.

Each detector must be treated separately with the DSVP technique in order to account for their slightly different characteristics. Thus, to create a statistic expected for a single detector with a target activity of 300 Hz over two years of data taking, we generated 40000 (~ 80000) single (double) pulse events. The arrival time of the pile-up pulses is uniformly distributed between 0 and 10 μs .

3.2 Detector models

For this study we modeled three different TES microcalorimeters with the one-body model [12] or with the two-body dangling model [13]. In both cases the current pulse profile is obtained by solving the system of the electro-thermal differential equations applying the fourth-order Runge-Kutta method (RK4) and considering the transition resistance as proposed by [14] for taking into account the TES non-linear behavior. To these pulses a noise waveform, generated as an autoregressive moving average ARMA (p,q) process with a power spectrum given by the Irwin-Hilton model, is added. To test the DSVP effectiveness with slightly different signal shapes, the physical parameters in the differential equations are chosen to describe three types of detectors (Fig. 4):

- (a) the detectors in [2] which are characterized by a non linear response and one thermal body.
- (b) the target detectors of HOLMES [15] have nearly-linear response and behave according to a two thermal body model.

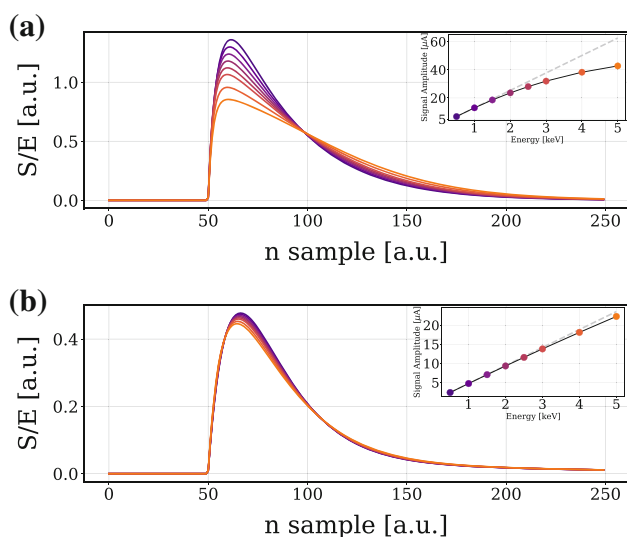


Fig. 4 Pulse profiles corresponding to different energies from 0.5 to 5 keV for two different detectors with non-linear (det. a), and nearly-linear response (det. b). In order to compare the signals, all the pulses are normalized by dividing their amplitude by the energy

- (c) same nominal design as (b) except for the production process, causing a significantly weaker link toward the thermal bath. Thus, the signals have a slower decay time and a lower signal amplitude respect to (b). Despite this difference, the detectors show a linear response to energy deposition with a two-body feature.

The detectors of type (b) are the most promising for the HOLMES goals. Therefore, for this detector three different configurations were tested, in which the rise time was changed, adjusting the inductance of the circuit and keeping the other detector parameters constant.

3.3 DSVP and HOLMES

We indicate the ratio between the number of pile-up pulses and the number of single pulses in the ROI as f_{pp}^{ROI} . From simulations, setting a time resolution of 10 μ s a value of $f_{pp}^{ROI} \simeq 2$ is expected. In order to apply the algorithm, the **M** matrix, which contains the ROI events, must have $N_A > N_B$, thus f_{pp}^{ROI} needs to be lowered below one. To reduce this ratio many different strategies can be adopted. In the following a non exhaustive list is reported.

- *Adding an additional calibration source* By adding a source characterized by a monochromatic X-ray emission in the ROI, the number of single pulses in the ROI can be increased while keeping the number of pile-up pulses unchanged. This approach can be very useful because it reshapes the energy spectrum, potentially reducing the probability of discarding single events with energy very

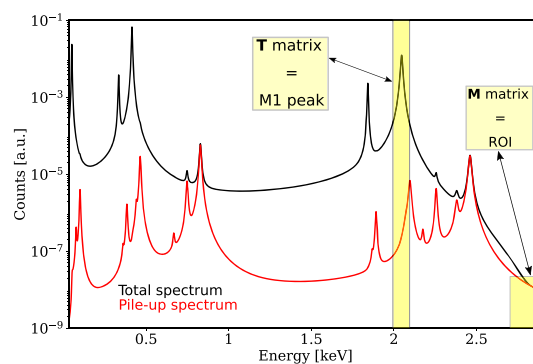


Fig. 5 De-excitation simulated spectra of ^{163}Ho with $f_{pp}^{tot} = 3 \times 10^{-3}$. Near the end-point the single pulse counts are outnumbered by pile-up counts

close to the end-point. A similar approach was investigated by Alpert [2].

- *Volumetric cuts* The events of the training dataset **T** are distributed in a finite volume in the k -dimensional projection space. The single pulses in the ROI reasonably lie within the same portion of space, while the pile-up are expected to be distributed in a different region. Thus if we select only the points in the projection space lying inside the volume which includes the training dataset, we could easily eliminate a large fraction of pile-up events. Before evaluating their projection on the **T** right singular vectors, the **T** and ROI events are normalized to set their amplitude equal to one. Then, we define the region in the k -space in which the **T** events are distributed. We increase it by a little amount in order to account small non-linearity effects. Finally, we select only the events in the ROI included inside this region. This method can achieve good time resolution, but it works only if the detector response does not depart from linearity too much, so in our simulation in detectors (b), (c) but not (a).
- *Filtering* Few filtering techniques allow to achieve effective time resolution close to the sampling time. Among these, a particular Wiener filter, as described in [16], is probably the best technique to achieve this goal.

For the HOLMES purposes, in order to fulfill the $N_A > N_B$ condition in the ROI the most suitable and practical method are the 'wiener filter' and the 'volumetric cuts'. As indicated in Table 1, applying these algorithms to the ROI events, the f_{pp}^{ROI} can be reduced around 0.6. Most of our simulations are therefore aimed to test their applications. Nevertheless, in Sect. 4.2 the performance of the DSVP technique with an external calibration source is also shown.

As shown in Fig. 5, we use the events at M1 peak ($E \sim 2$ keV) as training region **T**. This is also the energy range in which the average signal for the Wiener filter is defined. The M1 peak is the most suitable region for two reasons: it is the peak

Table 1 Effective time resolution of the DSVP with Wiener Filter and Volumetric cuts. We indicated with (*) the algorithm used in that simulation to lower the ratio of f_{pp}^{ROI} below one. For simplicity, we always

set N_B equal to the exact number of pile-up pulses in the ROI. The errors associated with the DSVP τ_{eff} are $\leq 5\%$ and are due to the random nature of the modified minimization RANSAC algorithm

Detector type	Rise Time (μs)	t_{sample} (μs)	τ_{eff} Wiener Filter	τ_{eff} Volumetric cuts	τ_{eff} with DSVP
b.	11	2	2.26	2.12 (*)	1.55
b.	17	2	2.37 (*)	2.60	1.55
b.	22	2	2.94	2.90 (*)	2.01
b.	17	1	1.66 (*)	2.00	0.94
a.	10	2	1.82 (*)	–	1.24
c.	19	2	2.70 (*)	3.54	1.82

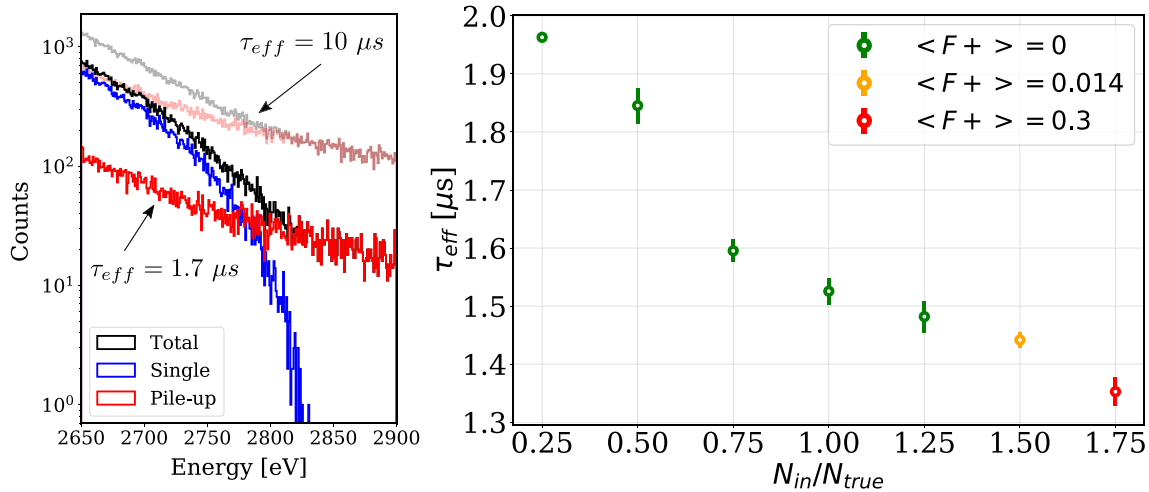


Fig. 6 Left panel: ROI energy spectrum before and after the application of the WF/Volumetric cuts and the DSVP technique. Light line represents the energy spectrum with τ_{eff} of 10 μs , while the solid line with a τ_{eff} of 1.7 μs . Right panel: The dependence of τ_{eff} and the

average percentage of false positive $F+$ from the input parameter N_B (N_{in}), which is normalized by the number of pile-up pulses present in the ROI (N_{true})

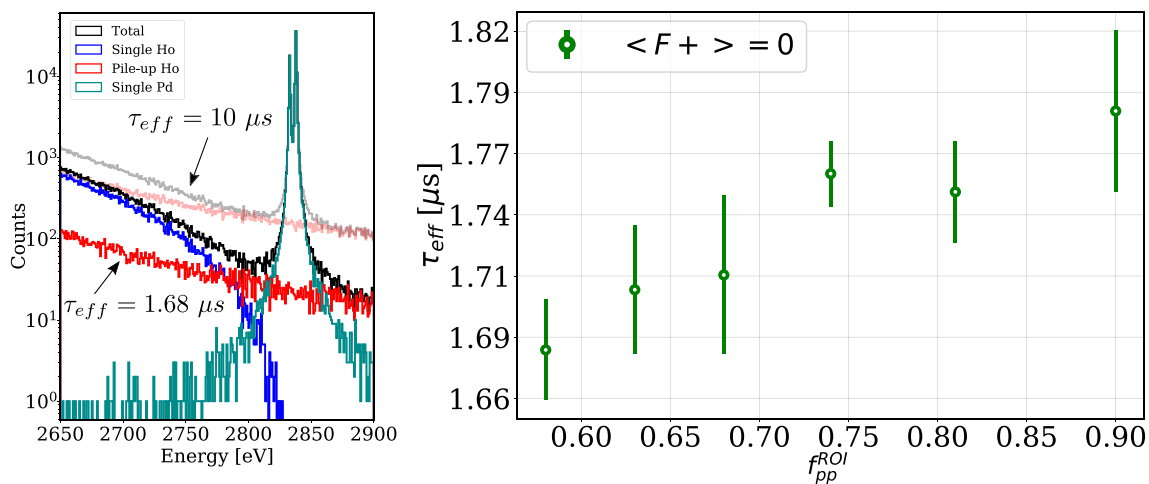


Fig. 7 Left panel: ROI energy spectrum with Pd $L\alpha$ peaks before and after the application of the DSVP technique. The initial f_{pp}^{ROI} was set to 0.58. Right panel: The dependence of τ_{eff} from the f_{pp}^{ROI} is shown

for the detector b . with a rise time of 17 μs and a sampling time of 2 μs . In this case, N_B was equal to the number of pile-up pulses in the ROI

closest to the ROI, thus reducing the non-linearity effects on the filters and on the discrimination algorithm and it fulfills the condition of $N_B \ll N_A$. The f_{pp} in this sector is expected to be of the order of $\sim 10^{-3}$, which can be further reduced with a *raw cleaning with PCA*, as described in Sect. 2.1.

4 Simulation results

To quantify the efficiency of the pile-up discrimination algorithms, we define an effective time resolution τ_{eff} as the ratio of the number of retained piled-up records to single-pulse records after the algorithm divided by the same ratio referred to raw data, times $10 \mu\text{s}$.

$$\tau_{eff} = \left(\frac{pup}{single} \right)_{final} \div \left(\frac{pup}{single} \right)_{initial} \times 10 \mu\text{s} \quad (8)$$

Through simulations similar to the ones in [9], [17], we have preliminarily estimated that even a small fraction of false negative modifies the single events spectrum and leads to a systematic error on the neutrino mass evaluation. We note that in our simulations no single pulse event was mistaken as pile-up. The DSVP technique described in Sect. 2 is designed to leave unaffected the A events.

In applications where a more robust discrimination of the B events is required, it is possible to adapt the algorithm toward this goal, for example by adjusting the threshold definition (Eq. 6), at the expenses of increasing the chance of deleting some A events.

The energy dependence of the method must be assessed for each specific application. In general, we would say that the events with the energies further away than the mean energy in the dataset \mathbf{M} are most likely to become false positive. In the end, the number of false positives is due to the threshold value (Eq. 6), while their nature is due to the mean ‘‘morphology’’ of the events present in the original dataset \mathbf{M} .

4.1 DSVP with Wiener filter and Volumetric cuts

We have estimated the τ_{eff} on the simulated data processed with the DSVP after lowering the initial f_{pp}^{ROI} using the ‘wiener filter’ or the volumetric cut techniques.

Furthermore, before being processed by the DSVP algorithm, the signals were also whitened, i.e. transformed to whiten noise by a fast Cholesky-factor backsolve procedure [18]. The results are reported in Table 1. All the simulations showed that the DSVP is able to reach a time resolution lower than the sampling time of the signal.

Table 1 shows that the time resolution is strongly dependent on the sampling time, the faster the better, but also on the rise time of the pulse. While the sampling frequency is constrained by the readout resources, there is more scope to

change the rise time of the detectors, acting on the electrical time constant of the biasing circuit. Changing the rise time by a factor two may be achieved reducing the inductance of the TES circuit by a similar factor. This change the noise spectrum too but usually it does not worsen the energy resolution.

Also, the non-linear detector response generally improves the efficiency of pile-up recognition algorithms. When two near-coincident energy depositions happen inside the TES, the detector will have different starting conditions. The shape of the pile-up pulse will be much more different from the single pulse for a non-linear TES than for a linear one, thus allowing the algorithms to recognize them more efficiently. As we stressed in Sect. 2, the only external parameter required by the DSVP algorithm is the number of events that it should discard at most, N_B . To quantify the influence of this parameter on the effectiveness of the algorithm, we fixed the dataset \mathbf{M} and varied N_B , computing the effective time resolution each time.

Figure 6 shows that no false positive was detected even if we get the number of event to eliminate wrong up to 50%.

4.2 DSVP with additional calibration source

We have also tested if the performance of the DSVP technique remains unchanged reducing the f_{pp}^{ROI} by adding an external source of single events with energy inside the region of interest instead of using preliminary filters. We added a source from $L\alpha$ x-ray emission lines of Pd (2.833, 2.839 keV). Figure 7 shows that increasing the number of photons of the Pd source (thus decreasing f_{pp}^{ROI}) the effective time resolution of the DSVP improves. Moreover, τ_{eff} always remains below the sampling time even for a pile-up fraction up to 0.9.

5 Conclusions

The DSVP algorithm represents a very powerful technique to decrease the number of undesirable events in a dataset.

In this work we have applied this algorithm for pile-up discrimination, which can lead to major improvement in experimental sensitivity for experiments such as HOLMES (neutrino mass measurement) or CUPID ($0\nu\beta\beta$) [19]. It can also be useful to recognize single-site events of the $0\nu\beta\beta$ interactions from multi-site background events in GERDA [20]. We tested the DSVP technique for the HOLMES application and we compared its efficiency, represented in this case by the effective time resolution τ_{eff} , to more ‘classical’ discrimination techniques, resulting in a better time resolution.

With the target detector of HOLMES, the DSVP techniques allows us to reduce the total fraction of pile-up events from 10^{-3} ($\sim \tau_{eff} 3 \mu\text{s}$) to 10^{-4} ($\sim \tau_{eff} 1.5 \mu\text{s}$), thus improving

the neutrino mass sensitivity from 2 eV to about 1.4 eV. To put this result in perspective, achieving the same improvement would require to increase the acquisition time by a factor 4: from 3 to 12 years.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: All the relevant data and the necessary informations are reported in the figures.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.
Funded by SCOAP³.

References

1. A. Nucciotti, Adv. High Energy Phys. **2016** (2016). <https://doi.org/10.1155/2016/9153024>
2. B. Alpert et al., J. Low Temp. Phys. **184**(1–2), 263 (2016). <https://doi.org/10.1007/s10909-015-1402-y>
3. B. Alpert, M. Balata, D. Bennett, M. Biasotti, C. Boragno, C. Brofferio, V. Ceriale, D. Corsini, P.K. Day, M. De Gerone et al., Eur. Phys. J. C **75**(3), 112 (2015)
4. I. Jolliffe, *Principal Component Analysis* (Springer, Berlin, Heidelberg, 2011), pp. 1094–1096. https://doi.org/10.1007/978-3-642-04898-2_455
5. G.H. Golub, C.F. Van Loan, *Matrix Computations*, 3rd edn. (Johns Hopkins University Press, Baltimore, 1996)
6. Ransac sklearn (2020). https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RANSACRegressor.html
7. A. De Rújula, M. Lusignoli, Phys. Lett. B **118**(4–6), 429 (1982)
8. M. Aker, K. Altenmüller, M. Arenz, M. Babutzka, J. Barrett, S. Bauer, M. Beck, A. Beglarian, J. Behrens, T. Bergmann et al., Phys. Rev. Lett. **123**(22), 221802 (2019)
9. A. Nucciotti, Eur. Phys. J. C **74**(11), 3161 (2014)
10. S. Eliseev, K. Blaum, M. Block, S. Chenmarev, H. Dorrer, C.E. Düllmann, C. Enss, P. Filianin, L. Gastaldo, M. Goncharov et al., Phys. Rev. Lett. **115**(6), 062501 (2015)
11. L. Gastaldo, K. Blaum, K. Chrysalidis, T.D. Goodacre, A. Domula, M. Door, H. Dorrer, C.E. Düllmann, K. Eberhardt, S. Eliseev et al., Eur. Phys. J. Spec Top. **226**(8), 1623 (2017)
12. K.D. Irwin, G.C. Hilton, in *Cryogenic Particle Detection* (Springer, New York, 2005), pp. 63–150
13. I.J. Maasilta, AIP Adv. **2**(4), 042110 (2012)
14. B. Cabrera, J. Low Temp. Phys. **151**(1–2), 82 (2008)
15. B. Alpert, D. Becker, D. Bennet, M. Biasotti, M. Borghesi, G. Gallucci, M. De Gerone, M. Faverzani, E. Ferri, J. Fowler et al., Eur. Phys. J. C **79**(4), 304 (2019)
16. E. Ferri et al., J. Low Temp. Phys. **184**(1–2), 405 (2016). <https://doi.org/10.1007/s10909-015-1466-8>
17. A. Nucciotti, E. Ferri, O. Cremonesi, Astropart. Phys. **34**(2), 80 (2010)
18. J.W. Fowler, B.K. Alpert, W.B. Doriese, D. Fischer, C. Jaye, Y.I. Joe, G. O'Neil, D. Swetz, J. Ullom, Astrophys. J. Suppl. Ser. **219**(2), 35 (2015)
19. D. Chernyak, F. Danevich, A. Giuliani, E. Olivieri, M. Tenconi, V. Tretyak, Eur. Phys. J. C **72**, 1989 (2012). <https://doi.org/10.1140/epjc/s10052-012-1989-y>
20. M. Agostini et al., Eur. Phys. J. C **73**(10), 2583 (2013). <https://doi.org/10.1140/epjc/s10052-013-2583-7>