Uncertainty Quantification in Radioactivity Level Estimation via Heteroscedastic Count Regression

1st Arthur Roblin
PSN-RES/SCA/LPMA
Autorité de Sûreté Nucléaire et de Radioprotection (ASNR)
Saclay, F-91400, France

2nd Jean Baccou
PSN-RES/SEMIA/LSMA
Autorité de Sûreté Nucléaire et de Radioprotection (ASNR)
Saint-Paul-lez-Durance, F-13115, France

3rd Grégoire Dougniaux
PSN-RES/SCA/LPMA
Autorité de Sûreté Nucléaire et de Radioprotection (ASNR)
Saclay, F-91400, France

4th Santiago Velasco-Forero
Centre for Statistics and Images (STIM)
Mines Paris, PSL University
Fontainebleau, 77300, France

***Abstract*—In nuclear facilities, dedicated instruments are used to monitor airborne radioactivity in real-time. Current nuclear measurement analyses rely on simple statistical methods that exhibit significant limitations in atypical or low-count situations. To address these shortcomings, we propose a deep neural network approach for count regression, enabling more accurate estimation of transuranic alpha emissions. Uncertainty quantification is integrated through parametric distributional prediction combined with the theoretical guarantees of conformal prediction. In addition, we introduce a novel optimization strategy for heteroscedastic predictive distributions, which mitigates the inflated variance behavior commonly observed in existing approaches.**

***Index Terms*—Deep Learning, Uncertainty Quantification, Conformal Prediction, Radioactivity.**

I. Introduction

Continuous Air Monitors (CAMs) are widely deployed in nuclear facilities to detect airborne transuranic contamination in real time. These systems operate by continuously collecting aerosols from ambient air onto a filter and analyzing the resulting alpha spectra to trigger alarms when contamination is detected. However, In complex environments such as decommissioning sites, CAMs are known to generate a high rate of false positives. This behavior is primarily caused by non-standard atmospheric conditions, including dust-laden aerosols and fluctuations in radon progeny activity, which significantly perturb the measured spectra [1]–[3]. These false alarms lead to frequent operational interruptions and delays in decommissioning activities. Current CAM decision algorithms mainly relying on fixed regions of interest (ROIs) and alarm thresholds [4]. While effective under nominal conditions, such rule-based approaches lack adaptability to atypical spectral distortions induced by non-radioactive aerosols or variable radon backgrounds. As a result, their performance degrades in the challenging environments encountered during dismantling operations, where atmospheric conditions often fall outside the assumptions underlying standard calibration procedures.

To address these limitations, recent work has explored the integration of statistical learning techniques into CAM data analysis. In particular, a neural network–based classification approach was proposed in [5], demonstrating improved adaptability to atypical conditions and a substantial reduction in false alarm rates. Despite these promising results, the proposed model suffers from a key limitation inherent to many deep learning approaches: its black-box nature. In safety-critical applications such as radiological monitoring, the absence of uncertainty quantification and interpretability significantly limits the model's suitability for operational decision-making.

The objective of this work is to overcome these limitations by developing a more trustworthy and interpretable data-driven approach for CAM analysis. Rather than formulating the problem as a classification task, we propose a regression-based framework that directly estimates transuranic alpha activity while explicitly modeling predictive uncertainty. Uncertainty quantification is integrated through parametric distributional prediction, calibrated using the conformal prediction framework, which provides strong theoretical guarantees on the validity of the resulting confidence intervals. In addition, we introduce a novel optimization strategy for heteroscedastic predictive distributions that mitigates the inflated variance behavior commonly observed in existing approaches, leading to improved uncertainty calibration and superior accuracy compared to state-of-the-art solutions.

II. Double Count Regression Network

This section is devoted to the construction of a deep learning model for count regression. After a brief description of the motivations and notations, the different elements of the construction are introduced. They concern a suitable choice of the type of network to model a count variable, an efficient learning process to estimate the network parameters and an uncertainty calibration step to derive reliable information for radioactivity level estimation.

A. Motivation of the regression framework

The model's objective is to determine whether transuranic activity is present. Although this is inherently a binary decision, we separate the measurement from the decision stage

and therefore adopt a regression framework instead of binary classification. The number of transuranic events in the measurement is directly predicted, *i.e.* the number of emissions that are due to a transuranic radioisotope. We motivate this choice with multiple reasons:

- The alarm threshold is controllable, thanks to the physical meaning behind the network's prediction.
- The decision, separated from analysis, can be taken by usual means from the estimated measurement. The interpretability of the decision process is thereby enhanced.
- Regression naturally accommodates the varying severity of false negatives (e.g., missing large vs. small contamination), eliminating any complex loss function tuning.
- As described in Section III, the database includes exact information on the number of transuranic-related events in each spectrum. The count regression framework thus allows to use all information available to train the model.

B. Notation

In this study, we begin by assuming access to a dataset, denoted as $\mathcal{D}$, which comprises a total of N distinct training examples. Each example within this dataset is represented as a pair $\{x_i, y_i\}_{i=1}^{N}$, where x_i corresponds to the input features taking values in $\mathcal{X} = {\mathbb{Z}_{\geq 0}}^d$, $d > 0$, and y_i, which takes values in the set of non-negative integers $\mathbb{Z}_{\geq 0}$, represents the associated label or target variable. These labels are assumed to be sampled from an underlying, yet unknown, count distribution $p(y_i \mid x_i)$, which describes the conditional probability of observing y_i given the input x_i. To formalize the problem setting, $\mathcal{P}$ is introduced to represent the space of parametric probability distributions $\{p_\psi, \psi \in \mathbb{R}^m\}$ defined over the set of non-negative integers $\mathbb{Z}_{\geq 0}$. Within this framework, a specific distribution $p \in \mathcal{P}$ can be uniquely identified and parameterized by a vector $\psi \in \mathbb{R}^m$, which encapsulates the essential characteristics of the distribution. Our primary goal is to model the space of distributions $\mathcal{P}$ using a neural network, denoted as f_Θ, which maps an input from $\mathcal{X}$ to a corresponding distribution in $\mathcal{P}$. The neural network f_Θ is fully parameterized by a set of learnable weights Θ, which are organized across L successive layers. Although the idealized network maps inputs directly to distributions, in practical implementations, we restrict the probability space to a family of parametric distributions and design f_Θ to output the parameter vector ψ, effectively defining the mapping $f_\Theta : \mathcal{X} \to \psi$. Once the network is trained, it enables us to derive a parametric predictive distribution, denoted as $\hat{p}(y_i \mid f_\Theta(x_i))$, for any given input x_i in the dataset. We introduce in next sections two families of parametric distributions suitable for our problem.

C. Double Poisson networks

When estimating a count variable, a natural first choice is to model the targeted count using a Poisson distribution. However, this distribution fails to adequately account for the uncertainty inherent in our heteroscedastic data. Specifically, the variance of y is influenced not only by its mean but also by external factors such as Radon activity and filter dustiness. To address this limitation, we adopt a Double Poisson (DP) distribution [6] to model the response variable y. This distribution generalizes the traditional Poisson model by incorporating a mean parameter $\mu > 0$ and an independent inverse dispersion parameter $\phi > 0$, enabling capture of over- and under-dispersion. Accordingly, we restrict the output space to $\mathcal{P}_{\text{DP}} \subset \mathcal{P}$, the family of double Poisson distribution over y. Any distribution $p \in \mathcal{P}_{\text{DP}}$ is uniquely parameterized by the pair $\psi := (\mu, \phi)$ and its mass function is defined for all $y \in \mathbb{Z}_{\geq 0}$ as:

$$p_{\text{DP}}(y \mid \mu, \phi) = \frac{\phi^{1/2} \exp(-\phi\mu)}{c(\mu, \phi)} \left(\frac{e^{-y} y^y}{y!}\right) \left(\frac{e^\mu}{y}\right)^{\phi y},$$

where $c(\mu, \phi)$ is approximated by:

$$c(\mu, \phi) \approx 1 + \frac{1-\phi}{12\phi\mu}\left(1 + \frac{1}{\mu\phi}\right).$$

This formulation allows the Double Poisson distribution to model under-dispersion ($\phi > 1$, where $\text{Var}(y) < \mu$) as well as over-dispersion ($\phi < 1$). When $\phi = 1$, the distribution closely approximates the standard Poisson distribution.

The Double Poisson distribution can be integrated into a generalized linear model (GLM; [7]) and, similarly, into a neural network framework [8]. In the neural network implementation, the two parameters of the distribution, μ and ϕ, are estimated simultaneously using a two-neuron output layer $f_\Theta : \mathcal{X} \to (\mu, \phi)$ with a Softplus[1] activation function to enforce non-negativity of the output, defined by $\text{Softplus}(x) = \log(1 + \exp(x))$.

Accordingly, for training the *Double-Poisson network*, we learn the weights Θ by minimizing the *negative log-likelihood* (NLL) of the Double Poisson distribution in average across all the triples $(\hat{\mu}_i, \hat{\phi}_i, y_i)$ using the following loss function:

$$\begin{aligned}\mathcal{L}_{\text{DP}}(\hat{\mu}_i, \hat{\phi}_i, y_i) = &-\frac{1}{2}\log(\hat{\phi}_i) + \hat{\phi}_i\hat{\mu}_i + \log(c(\hat{\mu}_i, \hat{\phi}_i)) \\ &- \hat{\phi}_i y_i\Big(1 + \log(\hat{\mu}_i) + \hat{\phi}_i\hat{\mu}_i - \log(y_i)\Big).\end{aligned}$$

D. Negative Binomial networks

An alternative for Double Poisson modeling is the Negative Binomial (NB) distribution [9], which generalizes the Poisson by introducing a dispersion parameter $\omega > 0$. This allows the variance to depend not only on the mean $\mu > 0$ but also on ω, accommodating the additional variability observed in heteroscedastic data by modeling an eventual over-dispersion. We restrict our output space to $\mathcal{P}_{\text{NB}} \subset \mathcal{P}$, the family of Negative Binomial distributions over y. Each distribution $p \in \mathcal{P}_{\text{NB}}$ is uniquely parameterized by the pair $\psi := (\mu, \omega)$, and can be formulated following two common parameterizations:

[1]Softplus predict positive values, is defined for all real numbers and is differentiable everywhere. Exponential activation would preserve canonical exponential-family structure and classical GLM theory, but might suffer from unstable gradients and curvature. Softplus reparameterization trades canonical form for smoother gradients, bounded curvature, and improved numerical behavior, while preserving the underlying exponential-family likelihood.

1) Negative Binomial Type I (NBI): The variance scales quadratically with the mean, $\text{Var}(y) = \mu + \omega\mu^2$. Its probability mass function is given by:

$$p_{\text{NBI}}(y \mid \mu, \omega) = \frac{\Gamma\left(y + \frac{1}{\omega}\right)(\omega\mu)^y}{\Gamma\left(\frac{1}{\omega}\right)\Gamma(y+1)(1+\omega\mu)^{y+1/\omega}}.$$

The corresponding negative log-likelihood used as loss function is easily derived.

2) Negative Binomial Type II (NBII): The variance scales linearly with the mean, $\text{Var}(y) = \mu + \omega\mu$. Its probability mass function is:

$$p_{\text{NBII}}(y \mid \mu, \omega) = \frac{\Gamma\left(y + \frac{\mu}{\omega}\right)\omega^y}{\Gamma\left(\frac{\mu}{\omega}\right)\Gamma(y+1)(1+\omega)^{y+\mu/\omega}}.$$

E. Improving optimization process: Stop-Gradient Scheduler

The dual nature of the prediction task of the Double-Poisson neural network may lead to premature convergence due to inflated predicted variance, in particular for the most difficult instances of the training data, which can lead to sub-optimal mean parameter fit [10]. We aim to re-balance the optimization process in order to achieve a better fit for μ, without sacrificing accuracy on ϕ. To mitigate this particular problem, [10] proposed to introduce a trade-off in the loss function via an hyper-parameter $\beta \in [0, 1]$ in a slightly modified loss, called *Double-Poisson β-NLL* :

$$\mathcal{L}(\hat{\mu}, \hat{\phi}, y) = \lfloor \phi^{-\beta} \rfloor \Bigg(-\frac{1}{2}\log(\hat{\phi}) + \hat{\phi}\hat{\mu} - \hat{\phi}y\big(1 + \log(\hat{\mu}) - \log(y)\big) + \log\left(c(\mu, \phi)\right) \Bigg),$$

with $\lfloor . \rfloor$ designating the stop-gradient function[2]. Taking $\beta = 0$ recovers the standard Double-Poisson NLL, and increasing β towards 1 progressively reduce the importance of the dispersion parameter to the profit of μ. $\beta = 0.5$ generally achieves the best trade-off between accuracy and log-likelihood [10]. This loss will be denoted $\beta_{0.5}$-Double-Poisson-NLL.

We introduce an alternative progressive learning strategy, which we refer to as the *Stop-Gradient Scheduler* (SGS). This method can be applied to any two-parameter distribution characterized by a mean and a dispersion parameter, such as the Double Poisson or the Negative Binomial (and thus also including the Gaussian distribution). A brief description is provided below. At the beginning of training, only μ is optimized, while the gradient of the dispersion parameter ϕ (resp. ω) is frozen and this parameter is fixed to a default value selected via cross-validation. Subsequently, ϕ (resp. ω) is progressively incorporated into the loss function, with the probability of including it increasing from 0 to 1 over the course of training. This probability is set to increase linearly, as illustrated in Figure 1, although alternative schedules could be explored. Consequently, the proportion of batches used to optimize ϕ rises from 0% in the first epoch to 100% in the final epoch. At the end of the training, it is reasonable to consider that every data point contributed multiple times to fit the dispersion parameter, especially if numerous epochs are used. This heuristic does not require any trade-off between the two parameters of the distribution.

[2]In a computational graph, any path passing through z will have a gradient contribution of $\frac{\partial \lfloor (z) \rfloor}{\partial x} = 0$, meaning it does *not* affect the updates of the parameters preceding z.

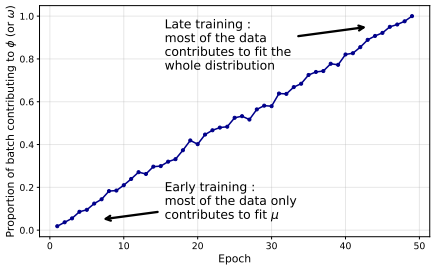

Fig. 1. Empirical proportion of training batches contributing to the dispersion parameter ϕ (or ω) at each epoch of the training.

F. Uncertainties calibration using conformal prediction

The predictive probability distributions obtained from the previous networks allows to get estimated confidence intervals but does not achieve theoretical guaranties on the calibration of these confidence intervals, beyond the behavior we could empirically observe on a test dataset. To tackle this limitation, it is proposed to perform a calibration using the conformal prediction framework.

Conformal prediction is a recent uncertainty quantification framework originally introduced by Vladimir Vovk [11]–[13]. It allows to generate prediction sets for any model f, associated with a chosen confidence level. In the regression framework, this will naturally corresponds to confidence intervals. The main idea is to separate the learning data into two distinct sets: the proper training set $\{(x_i, y_i), i \in \mathcal{I}_{train}\}$, used to fit the model's parameters Θ, and a *calibration* set $\{(x_i, y_i), i \in \mathcal{I}_{cal}\}$, used to build the prediction sets, with $\mathcal{I}_{train} \cap \mathcal{I}_{cal} = \emptyset$ and $|\mathcal{I}_{train}| + |\mathcal{I}_{cal}| = N$.

The *locally adaptive split conformal prediction* framework allows to turn the uncertainties estimated by our Double Poisson (or Negative Binomial) network into rigorous prediction intervals $\hat{C}_{\mathcal{I}_{train} \cup \mathcal{I}_{cal}}$, benefiting from strong theoretical guaranties. The results of conformal prediction hold under the assumption that the new test point (x_{N+1}, y_{N+1}) is *exchangeable* with the calibration data, i.e. their joint distribution remains unchanged under permutations. This assumption is weaker that the i.i.d. assumption. The validity of this hypothesis for our data is discussed Section III-F. Under this hypothesis, the prediction band $\hat{C}_N = \hat{C}_{\mathcal{I}_{train} \cup \mathcal{I}_{cal}}$ built with the conformal approach satisfies

$$\mathbb{P}\big(y_{N+1} \in \hat{C}_N(x_{N+1})\big) \geq 1 - \alpha \tag{1}$$

In particular, the use of scaled residuals ordination [14], [15] is a simple way to build $\hat{C}_N(x_{N+1})$ for any new test point (x_{N+1}, y_{N+1}) [14] following four steps.

1) After training the model f on the training set, the predictions are performed on the calibration set $\left\{f(x_i) = (\hat{\mu}_i, \hat{\phi}_i), i \in \mathcal{I}_{\text{cal}}\right\}$.
2) The estimated standard errors $\left\{\hat{\sigma}_i = \hat{\sigma}(x_i), i \in \mathcal{I}_{\text{cal}}\right\}$ are obtained from the estimated Double Poisson density $\left\{p_{\text{DP}}(\hat{\mu}_i, \hat{\phi}_i), i \in \mathcal{I}_{\text{cal}}\right\}$, or the Negative Binomial density respectively.
3) The *scaled residuals* of the predictions are then defined as: $R_{\text{cal}} = \left\{\frac{|y_i - \hat{\mu}_i|}{\hat{\sigma}_i}, i \in \mathcal{I}_{\text{cal}}\right\}$.
4) Finally, define $\hat{q}_{\text{cal}}^{1-\alpha}$ as the empirical $(1-\alpha)$-quantile of R_{cal}, that is to say the $\lceil (1-\alpha)(|\mathcal{I}_{\text{cal}}|+1) \rceil$ smallest residual among R_{cal}. After performing the predictions $f(x_{N+1})$ and computing $\hat{\sigma}_{N+1}$, a conformal prediction band associated to x_{N+1} can be written as:

$$\hat{C}_n(x_{N+1}) = \left[\hat{\mu}_{N+1} \pm \hat{\sigma}_{N+1} \hat{q}_{\text{cal}}^{1-\alpha}\right]. \tag{2}$$

An alternative to scaled residuals ordination relies on quantile regression. However, in our setting the confidence level must remain fully controllable at any time by the monitoring operator. We therefore exclude quantile regression methods, as they require specifying the confidence level prior to deployment, which is incompatible with practical usage.

III. Results

This section presents a performance evaluation of the proposed deep learning approaches on alpha-spectroscopy case. Following a description of the dataset, the evaluated models and performance metrics are outlined. The regression results are then reported, and the section concludes with a brief discussion of the classification task.

A. Dataset

A simulation code based on classical equations has been developed in [5]. This code allows the simulation of a synthetic database containing 200,000 spectra (an example is shown in Figure 2), representing a large variety of atmospheric situations. This database includes extreme conditions, such as high dust levels (up to 8 mg on the filter), high Radon activity (up to 100 Bq/m³) and instrument miscalibration.

B. Tested models

For our predictive distribution approach, the three distributions presented in II-C and II-D were first tested using their negative log-likelihood as loss function. The improved learning process with SGS introduced in II-E was then conducted on these likelihoods. Finally, a third variant was obtained by applying the conformal-calibration of the predicted confidence intervals on trained models. For the case of Double Poisson likelihood, the state-of-the-art optimization process $\beta_{0.5}$-Double Poisson NLL is also added to the benchmark [8], [10]. For the sake of completeness, natural choices for the standard loss function were also used to train two baseline models: Mean Absolute Error (MAE) and Mean Squared Error (MSE). The same one-dimensional convolutional neural

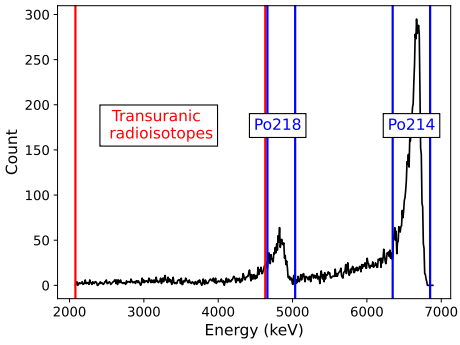

Fig. 2. A simulated spectrum, with the ^{218}Po peak between 4.5 and 5 MeV, and the ^{214}Po peak between 6.5 and 7 MeV. The tails of these two Radon-descendants create background noise in the transuranic region of the spectrum.

network was employed for all loss functions. Its architecture closely follows that of the classification network described in [5]. For the distributional variant, an additional output neuron was introduced to estimate the dispersion parameter alongside the mean of the predicted distribution. Training was conducted using a batch size of 100 spectra and the AdamW optimizer [16]. The network was trained for 50 epochs with an initial learning rate of 10^{-3}. A learning rate scheduler was also employed to facilitate faster convergence.

The training set is composed of 80% of the total database, *i.e.*, 160K spectra, leaving 20% for the test set. 25% of the training set is kept for a calibration set, used for the conformal calibration of the uncertainties, leaving 120K spectra for the proper optimization of the network weights. Ten models were trained each time with different random initialization seed, in order to quantify the dispersion of the results. Standard errors of the ten models outputs will be indicated and represented when presenting the results.

C. How to assess our models ?

The evaluation of a regression model is straightforward, as the MAE and the MSE computed on the test set are sufficient to quantify the relative accuracy of the predictions. For the likelihood-trained networks, the label y is compared with the predicted mean parameter, which exactly corresponds to the mean of the distribution for the Negative Binomial ($\mu = \mathbb{E}[Y]$), and approximately corresponds to the mean for the Double Poisson ($\mu \approx \mathbb{E}[Y]$). The mean parameter μ will be assimilated to the prediction $\hat{y}$. The evaluation of the uncertainty estimations quality lies on two criteria:

- Calibration of the estimated confidence intervals, where the empirical proportion of true values contained within the intervals should match the prescribed confidence level on a validation set.
- Width of the confidence intervals, with narrower intervals indicating better uncertainty quantification.

The *continuous ranked probability score* (CRPS) [17] provides a unified measure of the desirable properties of a predictive distribution. It attains its optimal value of zero when the predictive distribution collapses to a Dirac mass at the

true label. The CRPS increases as the distribution becomes more dispersed or as its central tendency moves away from the observed value.

D. Regression results

The results obtained with the different losses are detailed in the Table I. One hyper-parameter selection had to be done in the Stop-Gradient Scheduler optimization process: the choice of the default dispersion parameter, used when its gradient is deactivated in the loss. For the Double Poisson, cross-validation results clearly point toward the choice of $\phi = 1$, both in terms of MSE and CI width. Analogously for the negative binomial, the default ω value is chosen close to equi-dispersion, *i.e.* close to zero ($\omega = 0.001$). It is the best trade-off between a small default dispersion, which increases over-fitting tendencies, and a large dispersion, which dilutes the accuracy of the mean parameter fit.

In terms of pure regression performances, the MAE and MSE loss functions logically obtain among the best absolute and squared errors. However, once trained with SGS, the networks trained to maximize the three different distribution likelihood achieve comparable precision levels. The confidence intervals mean width obtained after training with the different losses are showed in Figure 3. It is clear that depending on the confidence level targeted, the best distribution choice might not be the same. For lower confidence requirements (below 90%), the two negative binomial distributions allows narrower confidence intervals. For higher confidence levels, intervals width becomes similar. Once trained with SGS, the two negative binomial distributions get very close results.

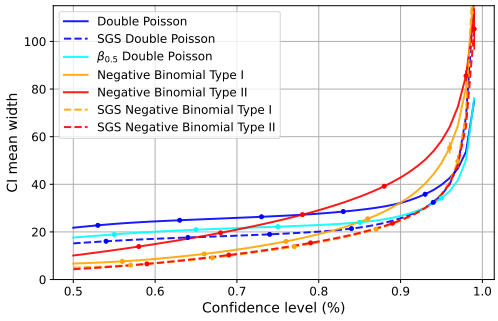

Fig. 3. Mean of the predicted confidence-interval widths versus confidence level for the different losses, obtained on the test set. For a fair comparison, all CI are conformal-calibrated, achieving the same calibration level.

The results from Table I and Figure 3 highlight the usefulness of the proposed Stop-Gradient Scheduler optimization process. The MSE of the predictions are significantly improved, except for the Type I NB. In the meantime, the width of the confidence intervals are considerably reduced, even for a same calibration rate. The CRPS are also greatly improved for the three distributions, in particular the two NB. For the Double Poisson, the SGS procedure outperforms the state-of-the-art $\beta_{0.5}$ Double Poisson NLL, both in MSE and in confidence interval width, at least until the 93% confidence level.

E. Classification results

In general, from the predicted count and its confidence interval, any binary decision can be applied depending on the specialist preference. We may apply a CI-based alarm method through the standard statistical test, i.e. $\mathbb{I}\{0 \notin \hat{\text{CI}}(x)\}$.

In our case, we choose to apply the traditional sigma-factor alarm method, used in many continuous air monitors [4]. It consists in triggering an alarm if $\hat{y} > k \times \sigma_{\hat{y}}$, with $\hat{y}$ the estimated transuranic count, and k the so-called *sigma-factor*. $\hat{y}$ may be estimated by various means, and is generally approached by a rudimentary statistical heuristic, such as the 4-Regions-of-Interests method [4]. The sigma-factor is typically chosen between 4 and 6 depending on the targeted false alarm rate, usually very low ($< 0.001\%$). The instrument takes indeed multiple independent decisions per hour and is often working 24/7. Furthermore, one facility may use dozens of such instruments. For instance, aiming a false alarm rate of one per year for ten instruments working on 30-minutes acquisition windows, corresponds to a false alarm rate of 5.7×10^{-6}. Estimating y by the neural network, our applicable alarm equation becomes : $\hat{y} > k \times \hat{\sigma}(\hat{y})$. The unusual precision of the estimation of y and of $\sigma(\hat{y})$ allows to chose a sigma-factor lower than usual: empirically, the choice of k equal to three already satisfies high conservatism standards. This means that the sensitivity of our approach is also improved.

Table II summarizes the results obtained on the test dataset for the classification task. The distributional models were trained with SGS, but are not calibrated. For comparison, we added the performances of the CNN trained for classification with the cross-entropy, with an identical architecture except for the output layer. In moderate background noise conditions (background counts in transuranic region < 100), our distributional models achieve an excellent false positive rate (0.00%), already surpassing the 4-ROI. In high background noise (background > 100), the Double-Poisson network still achieves a strong 99.85% specificity. Note that the precision of the classification network may only be improved to the same standard by sharply raising the binary classification threshold from its default value 0.5 to 0.9995, obtaining hereby a comparable 80% sensitivity. We however understand that this trade-off approach is less rigorous, controllable and interpretable than the one proposed through the Double Count Regression networks followed by the sigma-factor algorithm. The CI-based alarm method achieves comparable performances.

F. Limitations

The very good coverage of the predicted confidence intervals obtained with conformal prediction is a powerful result, consistent for any confidence level. It however faces some limitations in real conditions. A monitor operating in real time will probably measure data that will not be exchangeable with our calibration set, because of the distribution drift. Furthermore, building a calibration set that would be representative of real world operations is not conceivable, since different monitors may face very different conditions during their lifetime. Future work will focus on an extension of these

TABLE I

REGRESSION PERFORMANCES OBTAINED WITH DIFFERENT LOSSES, WITH OR WITHOUT THE STOP-GRADIENT SCHEDULER (SGS), BEFORE AND AFTER CONFORMAL CALIBRATION. EACH MODEL IS TRAINED TEN TIMES WITH DIFFERENT RANDOM SEEDS.

Loss	MAE ↓	MSE ↓	$\hat{\text{CI}}_{90\%}$ cover	$\hat{\text{CI}}_{90\%}$ width ↓	CRPS ↓
MAE	**6.99** $\pm$ 0.05	253.3 ± 2.9	-	-	-
MSE	7.91 ± 0.07	239.3 ± 2.9	-	-	-
Poisson NLL	8.14 ± 0.05	282.4 ± 4.0	-	-	-
Double Poisson NLL	10.81 ± 0.32	372.8 ± 25.4	$95.62\% \pm 0.16$	46.22 ± 1.78	7.29 ± 0.21
$\beta_{0.5}$ Double Poisson NLL	9.00 ± 0.13	245.0 ± 7.6	$96.14\% \pm 0.17$	39.45 ± 1.05	6.20 ± 0.10
SGS Double Poisson NLL	8.28 ± 0.06	**230.9** $\pm$ 4.8	$95.04\% \pm 0.29$	36.73 ± 0.63	5.95 ± 0.04
Conformal SGS Double Poisson NLL	-	-	**90.10%** $\pm$ 0.12	25.73 ± 0.31	-
Negative Binomial Type I NLL	7.66 ± 0.09	237.0 ± 12.1	$94.24\% \pm 0.31$	56.77 ± 4.84	9.12 ± 1.20
SGS Negative Binomial Type I NLL	7.94 ± 0.08	255.6 ± 7.2	$93.52\% \pm 0.44$	31.84 ± 0.76	**5.71** $\pm$ 0.08
Conformal SGS Negative Binomial Type I NLL	-	-	**89.96%** $\pm$ 0.11	**24.69** $\pm$ 0.26	-
Negative Binomial Type II NLL	13.62 ± 0.42	765.5 ± 59.4	$83.41\% \pm 0.35$	38.60 ± 0.88	9.11 ± 0.29
SGS Negative Binomial Type II NLL	7.86 ± 0.04	255.3 ± 4.0	$93.62\% \pm 0.33$	31.00 ± 0.62	**5.67** $\pm$ 0.06
Conformal SGS Negative Binomial Type II NLL	-	-	**90.06%** $\pm$ 0.10	**24.95** $\pm$ 0.17	-

TABLE II

CLASSIFICATION PERFORMANCES OBTAINED WITH FIVE ALGORITHMS.

Algorithm	Precision	Recall
Moderate background noise		
4-ROI	97.98 %	83.35 %
Direct classification network	98.16 %	96.97 %
DP count network	**100.00** %	97.78 %
NBI count network	**100.00** %	**98.18** %
NBII count network	**100.00** %	97.58 %
High background noise		
4-ROI	61.55 %	67.85 %
Direct classification network	96.06 %	**91.02** %
DP count network	**99.85** %	82.43 %
NBI count network	98.47 %	86.95 %
NBII count network	99.64 %	84.07 %

results to obtain stronger guaranties of confidence intervals coverage for real CAM operations. In particular, we may focus on conditional coverage to ensure a good confidence interval coverage for all spectrum configurations. We hope that this approach will mitigate drastically the lack of exchangeability.

Another limitation concerns the traditional sigma-factor inequality used in Section III-E. Even though it works very well, it is not based on solid statistical assumptions. Moreover, it does only use the vanilla distribution predicted by the model, but not the conformal-corrected confidence intervals.

Finally, the model's behavior on real data requires further study. The model may be deployed in various CAM devices.

IV. CONCLUSION

In this paper, we presented a new approach based on deep learning for the analysis of a raw nuclear measurement. The performances of the traditional algorithm used in this type of monitors were greatly improved by the neural network. Furthermore, the model was designed to estimate directly a concrete physical measure, rather than a direct binary decision, which increases the interpretability of the final decision algorithm. The integrated estimation of prediction dispersion, enhanced by a conformal prediction procedure, allows to efficiently capture the prediction uncertainty. The optimization heuristic proposed for the training of the two-parameters heteroscedastic distribution allows a wider use of this method for a versatile and reliable uncertainty quantification approach in deep learning in both Negative Binomial and Double Poisson distribution, without drawbacks on regression performances.

REFERENCES

[1] G. Dougniaux *et al.*, "Results from a measurement campaign in dismantling nuclear sites: a study of the false alarms emitted by CAM," 2016.

[2] G. Hoarau, "Étude de la limite de détection et des fausses alarmes émises par les moniteurs de mesure de la contamination radioactive atmosphérique dans les chantiers de démantèlement," Ph.D. dissertation, 2020.

[3] G. Dougniaux and G. Hoarau, "Detection limit variation against coarse aerosol during airborne contamination measurement with continuous air monitor," *Radiation Protection Dosimetry*, vol. 199, no. 18, pp. 2229–2232, 2023.

[4] A. Justus, "Technical Details of the Sigma Factor Alarm Method within Alpha CAMs," *Health Physics*, vol. 120, no. 4, pp. 442–453, 2021.

[5] A. Roblin *et al.*, "Deep learning approach for airborne alpha radioactivity monitoring in atypical atmospheric conditions," *Journal of Aerosol Science*, vol. 187, p. 106573, 2025.

[6] B. Efron, "Double Exponential Families and Their Use in Generalized Linear Regression," *Journal of the American Statistical Association*, vol. 81, no. 395, 1986.

[7] D. Toledo *et al.*, "Flexible models for non-equidispersed count data: comparative performance of parametric models to deal with underdispersion," *AStA Advances in Statistical Analysis*, vol. 106, no. 3, pp. 473–497, 2022.

[8] S. Young *et al.*, "Flexible Heteroscedastic Count Regression with Deep Double Poisson Networks," 2024.

[9] R. A. Rigby *et al.*, *Distributions for modeling location, scale, and shape: Using GAMLSS in R*. Chapman and Hall/CRC, 2019.

[10] M. Seitzer *et al.*, "On the pitfalls of heteroscedastic uncertainty estimation with probabilistic neural networks," in *International Conference on Learning Representations*, Apr. 2022.

[11] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic learning in a random world*. Springer, 2005.

[12] G. Shafer and V. Vovk, "A tutorial on conformal prediction." *Journal of Machine Learning Research*, vol. 9, no. 3, 2008.

[13] A. N. Angelopoulos *et al.*, "Conformal prediction: A gentle introduction," *Foundations and trends in ML*, vol. 16, no. 4, pp. 494–591, 2023.

[14] Y. Romano *et al.*, "Conformalized quantile regression," *Advances in neural information processing systems*, vol. 32, 2019.

[15] R. Tibshirani, "Conformal prediction under distribution shift," *Notes for Advanced Topics in Statistical Learning, Spring*, 2023.

[16] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representation*, 2019.

[17] H. Hersbach, "Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems," *Weather and Forecasting*, vol. 15, no. 5, 2000.