

Using a Novel Change Detection Algorithm to Predict Daily Linear Accelerator Output Changes in a Radiation Department.

Mark Ashburner (✉ ashburnermark@gmail.com)

Waikato District Health Board <https://orcid.org/0000-0002-1975-1109>

Koki Mugabe

Waikato District Health Board

Yathu Mohanathas

University of Auckland

Ahmad Umar

University of Auckland

David Huang

University of Auckland

Gill Dobbie

University of Auckland

Research Article

Keywords: Artificial Intelligence, Machine Learning, Quality Assurance, Radiotherapy, Linac

Posted Date: August 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-759145/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Machine learning tools and techniques were utilised to create a Novel Change Detection Algorithm (NCDA) as a supplementary quality assurance tool to alert users to potential significant changes in linear accelerator output using daily measurements.

A prototype model was developed and validated that provides a forecast for the daily dose and indicated when there is the potential for the output to change beyond what is considered normal daily drift. The model provided a good fit with the validation dataset used, meaning that trends in daily output were easy to identify.

The NCDA can be used either daily whereby output data is manually entered from the QA3 user interface, or on a weekly basis allowing input of the other data which is not traditionally presented on the interface. The NCDA should serve as a tool to alert users when data is trending out of tolerance and inform of possible output adjustments, therefore enabling effective allocation of time and resources.

1. Introduction

The International Commission on Radiation Units and Measurements (ICRU) [1] recommends that treatments using radiotherapy should deliver dose to patients within $\pm 5\%$ of prescription, and to achieve this, each step in the dosimetry chain must be performed to high accuracy. A robust quality assurance (QA) program for linear accelerators (Linacs) should be designed in such a way that users are assured that the machine characteristics have not deviated from their baseline values to such an extent that treatments are delivered above recommended thresholds. Further recommendations by AAPM in TG142 [2] that (Linac) outputs should be measured daily with suggested tolerances of $\pm 3\%$ of nominal output.

Daily measurements should be checked and monitored by a Medical Physicist, and when out of tolerance further action taken to ensure that the result is 'true', meaning that there has been a change in machine output not related to set up issues or daily device calibration issues, as opposed to 'false', whereby a negative result is due to equipment failure or user error. In an instance where a Linac output is found to be out of tolerance it should be returned into acceptable limits, followed by independent verification measurements before clinical treatments can resume. Regardless of whether or not the Linac output requires adjustment, this process of verifying a negative daily result is time consuming, in particular when it comes to the setting up of equipment accurately for verification and measurement, resulting in the Linac being out of clinical use whilst the issue is resolved.

Waikato Regional Cancer Centre (WRCC) has 4 beam matched Linacs (Varian TrueBeams, Silhouette and C-Series) used for delivering therapeutic radiation. Daily output is measured using the Sun Nuclear Daily QA 3 (DQA3, Sun Nuclear Systems). DQA3 measures radiation using a combination of ionization chambers and diodes, arranged in a fixed pattern allowing for good reproducibility, whereby a deviation in daily measurements will reveal changes in beam characteristics. The parameters measured are dose output, beam symmetry, beam flatness, beam energy, and light/radiation field coincidence. These

parameter measurements are compared to a standard using trend graph analysis in the accompanying software, which stores measurement results in a database allowing trend analysis and report generation.

The device has been shown to consistently deliver good linearity, reproducibility and results similar in degree to ionisation chamber measurements, for a period of up to eight months, and routine DQA3 calibrations can improve consistency of the detector [3]. Accuracy of the DQA3 is verified alongside monthly output measurements done in water tank or solid water using an FC65 farmer or Roos parallel plate ionisation chambers, as part of routine QA. Measured values of the DQA3 and IC should be within +/- 0.5% of each other, with DQA3 recalibration done when this fails.

At WRCC a daily tolerance of +/-3% of nominal is set as the action level. Daily measurements can be carried out by a Physicist, Technician or Radiotherapist (RT), but data must be analysed and reviewed daily by the Medical Physicist. Large changes in output over the course of two or more days are flagged and brought to the attention of the team to determine whether machine output is drifting significantly or that the QA3 device calibration is drifting. This involves verification and clarification using more accurate and complex methods of ionisation chamber measurements in water or water like medium. If the output change is verified as true, machine recalibration is carried out and independently verified.

Daily output measurements have been observed to drift within +/-2% of nominal, the accuracy of which is dependent on techniques used for measurement and individual department processes. Published analysis of daily output data shows photon and electron beams tend to have a rising drift in output constancy and occasional sinusoidal variations in beam flatness and symmetry [4] [5].

For this project we examined the feasibility of using predictive modelling on daily QA measurements to anticipate drifts in Linac output on a day-to-day basis. The benefit of a system that enables identification of potential deviation beyond normal performance would allow for proper resource allocation and reduction in machine downtime and delays.

2. Methodology

2.1: Data modelling, cleaning and behaviour analysis

Historic measurements of daily data using the DQA3 was procured along with monthly output measurements, re-calibration data and any other recorded machine outputs dating back approximately six years. These daily measurements, when out of tolerance, can be matched with concurrent data measurements done using IC and water, furthermore the output of the daily device is directly compared to an absolute output on a monthly basis, meaning we have a cohesive labelled dataset. This along with the problem being a time series dataset interfaced at a front end by a user (Physicist) leads us to conclude that our data can best be presented by a supervised learning problem.

Since the analysis being carried out here is initially retrospective in nature, anomalous data points known to and identified by the Physicist shouldn't be included in the data set used for training and testing. Known excursions beyond control limits, such as those caused by incorrect equipment set-up or user error were omitted from the data as the reason for the error was clearly known and its recurrence limited. Results that were beyond limits, yet due to output drift, machine breakdown or failure were kept in the dataset and considered a true reflection of the Linacs output and behaviour. [6] [7]

The different supervised learning algorithms investigated fall into distinct groups of regression and classification. Regression models will use the data sets to predict the output changes on daily basis, and classification would be a simple binary answer of yes/ no to significant machine output change. The different supervised learning models considered and tested for this project are discussed below.

2.2 Cumulative Sum

A cumulative sum (CUSUM) chart plots the cumulative sum of the deviations of subgroup averages from the set target value. Shifts are detected by comparing them with previously occurred values. In an ideal case the CUSUM plot varies randomly in the vicinity of zero. CUSUM plots clearly show when the actual values are away from the set target value of the parameter, which in this case would be the output tolerance limits defined by the Physicist.

A tabular CUSUM chart accumulates deviations above and below the target value in two separate variables Upper and lower CUSUM, C+ and C- respectively, which are decided beforehand. When the upper or lower CUSUM line crosses these limits, the process is considered to have gone out of acceptable control limits [6]. For this project, they were initially set at $\pm 3\%$, based on the daily tolerance limits determined by the department and TG142, when the CUSUM line crosses this limit the machine would be determined to be out of daily tolerance.

The use of CUSUM charts and statistical process control techniques has been reported elsewhere, with moderate success in identifying instances of loss in process control in spite of the machine QA data being within allowable action limits. Concluding that the use of such methods can assist physicists in identifying early recognition of trends that are likely to lead to future breaches of tolerance [6].

2.3 Linear and Logical Regression and Support Vector Machines

Linear regression models the relationship between the response and one or more independent variables as a linear equation [8] but with some modifications we can use a linear regression model for our data points as well. The concept is that we can use the dates we have for the maintenances that took in the past to be used to calculate the number of days between breakdowns.

Logistic Regression (LR) is used when a dependent variable is binary to describe a relationship between a discrete dependent variable and one or more independent variables by mapping an underlying function. The advantages of its use being that it is quite simple to implement, interpret and train compared to some other more complex algorithms [5]. The use of logistic regression was anticipated have limited use in

terms of developing a change detection algorithm, however, was used in this project to fit a logistic curve to the given dataset, and then give an output interpreted as a probability. Support Vector Machine (SVM) algorithms are generally used for classification modelling, each data point is plotted in n-dimensional space, n being the number of variables, and used to determine the separating hyperplane between two different classes or labels. SVMs have a high level of accuracy independent of the dimensionality of the data when compared to many other algorithms. The use of kernel functions allow addition of linear classifiers for data that have non-linear relationships, high levels of accuracy can be achieved for complex datasets, the downfall being that due to its complexity, the training speed is especially slow with larger datasets, such as the one being used for this study, however previous work found the use of SVMs for predictive maintenance had better performance than other algorithms such as k-Nearest Neighbour (kNN) [9] [10].

Work in logistic regression and support vector machines was completed using Jupyter Notebook and Google Colab, and the sci-kit learn library in Python.

2.4 Artificial Neural Networks

Artificial Neural Networks (ANN) consist of several processing nodes (neurons) organised amongst hidden layers where data is trained to recognise patterns within the dataset, and subsequently sent to the output layer. The main advantage of the use of ANNs is that is not held back by assumptions of linearity or variable independence. The major disadvantage is that its performance can be sensitive to the chosen parameters and magnitude of the hidden layer where data is processed, and its complexity leads to longer training times [9].

3. Results

Initial review and analysis of the data showed sinusoidal variations, drifting within +/- 2% of baseline for daily measurements, and monthly measurements done using water and ion chamber produces much more consistent doses within 1.5% of baseline, which is similar to results in the literature. Different energy levels, although acting in similar matter, with daily variations being sinusoidal overtime, never the less have different variations in daily output which means that being able to build a generalised model based off one energy level is not possible, and instead training data sets would be needed for each individual energy separately for our machine learning algorithm. These results provided us with an indication of the tolerances that should be applied to the final model in order to account for the random variations in daily measurements due to a myriad of factors which include: inter user variability, set up variability, actual output drift and daily device calibration drift, with our final implemented software able to distinguish when an output change is due to these normal fluctuations or when it is indicative of something more.

Daily QA measurements presents us with sequential time-series data. Furthermore, predictive modelling used on this data should help predict future drifts in LINACs, and the trends observed allowed for decisions to be made on the best models to use, the results of which are discussed below.

3.1: Cumulative Sum

The CUSUM model was developed and implemented using a classic algorithm in Python, accounting for both positive and negative deflections and programmed to give an alert when the mean of the input significantly differs from 0 (Equation 1).

$$\begin{cases} s[t] = x[t] - x[t - 1] \\ g^+[t] = \max(g^+[t - 1] + s[t] - drift, 0) \\ g^-[t] = \max(g^-[t - 1] - s[t] - drift, 0) \end{cases}$$

if $g^+[t]$ or $g^-[t] > threshold$:

$$\begin{cases} t_{talarm} = t \\ g^+[t] = 0 \\ g^-[t] = 0 \end{cases}$$

Equation 1: Example of a classic CUSUM algorithm to detect positive and negative deviations

Thresholds were adjusted until the algorithm was able to successfully predict an upcoming output change of significance. Initial threshold used was +/-3%, this being the daily tolerance in use, and decreased in value until a value was found that best corresponded to an impending significant output deviation as confirmed by a corresponding machine output adjustment or DQA3 recalibration.

3.2: Linear Extrapolation

Linear extrapolation was used to forecast output values for small increments with the aim of predicting small jumps in the dataset. When the jumps between the points were small, taking an average of the change and doing a linear extrapolation would give a close prediction. This was plotted, and linear regression used to forecast the number of days before the next output error would occur. In this case, we used the basic straight-line equation, $y=mx+c$, which meant we did not expect a highly accurate model at predicting drifts out of tolerance, however, was anticipated to be used as a starting point, whereby trends observed, if any, can be used in an ongoing machine learning algorithm.

3.3 Logistic Regression and Support Vector Machines

Treating the problem as a binary classification, where the primary outcomes were an indication of the output being out of tolerance and assigned a '1' in binary, and '0' was considered all other instances. Its use was anticipated to be limited for our primary goal; however, was utilised to aid in training the dataset to classify different labels. Modelling using logistic regression has a low computational training time making it easier to test using different parameters that are available within the sci-kit learn library. The

benefit of Logistic Regression is that is easy to implement and interpret, it does not assume linearity and gives an output values interpreted as a probability. These models require large datasets for training and all data points must be independent from one another, and care needed to be taken as dependency can lead to inaccurately overfitting of the data.

A SVM was also developed in parallel for binary classification. Different training-test splits were used to check if there would be any improvements in the accuracy for both models. SVM models reduce the risk of over-fitting, with high level accuracy for classification, even with datasets with high dimensionality. It was anticipated that an SVM model could be used to detect anomalies in daily outputs. SVM is good for outlier detection and cluster problems making it a good potential candidate for a solution to our problem. The data within the tolerance level can be placed in one region thus data outside tolerance level can be detected as anomalies. SVMs tend to be more complex algorithms, more difficult to implement with more computation required, especially for larger datasets such as ours.

3.4 ANN: LSTM Neural Network

A Long Short-Term Memory (LSTM) is a form of artificial recurrent neural network used to process long data sequences making it ideal for the time-series data such as used for this project. These networks apply a linear matrix operator to a current observation which in our study is the daily output, and hidden units from a previous time steps which give a resulting linear term serving as an activation function. Recurrent networks use the same matrix at each time step over multiple steps this feedback loop results in 'memory' in the model, which allows a decision to be made on whether the observed daily drift is normal or beyond what is expected, indicating underlying issues requiring investigation [11]. A particular advantage of using the ANN here is the ability to model the erratic changes that occur in the time-series data. The model was built using Keras instead of sci-kit learn which was previously used, however, the model still used Google Colab.

4. Discussion

4.1: CUSUM

The CUSUM algorithm was tested on all other energy levels available and did not show any obvious trends observed before intervention was required. Results suggest that Medical Physicists intervention would be required once the dose change was close to 1.6%, despite the tolerance of the DQA3 machine being 3%.

Figure 5 shows time series data for 6X. The alarms shown are for when tolerance threshold was exceeded, set to 2% for this example. After an alarm was set, there was a reset in the cumulative sum that was monitored noted by the start button. The y-axis shows the amplitude of the dose tolerance

Figure 6 shows the cumulative sum of all positive and negative changes in the data which depicted separately. The x-axis shows points labelled in order dating from the start of January 2014 to the end of

December 2019. The y-axis shows the change between data points on a daily basis.

The CUSUM model required thresholds to be manually determined, then retrospectively analysed to determine its accuracy, defeating the purpose to develop an algorithm able to provide a prediction based on the training dataset. CUSUM does have properties making it ideal for its use in change detection however its application in our final model that was limited. The algorithm shows promise as a graphical tool to help gain insight into the type of trends seen in the changes of dose tolerance levels.

CUSUM has been used in radiotherapy statistical control processes with moderate success elsewhere [6] [12] [13], and although it didn't meet the primary goal of this study, it has been shown the potential to be a valuable tool in assisting physicists in a radiotherapy department if developed further, namely helping to strike balance between two competing mistakes made when allocating time to quality assurance procedures being acting when a problem does not exist and 2: not acting when a problem exists. [12] A potential procedure for using CUSUM in daily QA management is detailed in Figure 7 adapted from [12]

4.2: Linear Extrapolation

The sinusoidal nature of daily outputs over time, as discussed previously, means a linear representation as a standalone feature will never easily be able to account for output drift changes on a daily basis, particularly when there is a continual change in the direction of the output shift (being from positive to negative difference from nominal). Simple linear modelling was not an ideal tool for predicting daily output trends on a linear accelerator, one such example of where failure occurs is illustrated in Figure 8 which shows a linear forecast, determined from a sample of 10 points, for 15MeV electrons. The forecast predicts a value of approximately +1% for the 5-05-2015, if the trend calculated from 14-04-2015 to 28-04-2015 was linear.

The actual output data is presented in Figure 9 which shows some initial agreement with the linear trend prediction, however there is a significant decrease (>2%) in output from 4/05/2015 to 5/05/2015. In this instance there was a significant issue with the machine requiring intervention due to breakdown, which would be unable to have been forecast using simple linear regression methods. Unpredictable events such as this lead to substantial under-fitting using linear extrapolation.

3.3 Logistic Regression and Support Vector Machines

The use of a logistic regression algorithm requires independence between data points, which meant that when training the model, our data was not treated as a time-series dataset, leading to errors in its predicted value. There was significant imbalance in the classes used since the data rarely requires any 'reset' points. For data in each different energy level, less than 1% of the data points were 'reset' points which meant that the model nearly always predicted that there was no reset required from the validation sets used. This bias introduced in the training set meant the model always predicted zero, which also resulted in a false high accuracy as displayed in the confusion matrices below:

Figure 10 shows the confusion matrix when the class weighting is not changed. The class weighting parameter penalises mistakes in samples in order to put more emphasis on a particular class. In this case, there is a much higher frequency of '0' when there is not a 'reset' for a data point compared to a '1' when there is a 'reset'.

The confusion matrix in Figure 11 shows when the class weighting has been changed to be more balanced, meaning that the smaller class, which is the 'reset' (or '1') class, is replicated until there are as many samples as there are in the larger class in an implicit manner. When the weighting class is more balanced, it results in more false negatives, meaning that assigning more importance to the lower smaller class results in less accuracy. IN both cases the true negatives, ie true predictions are relatively high, which in turn is due to the bias inherent in the training data.

The SVM developed for binary classification yielded very similar results to Logistic Regression since there was a large imbalance in the classes being predicted. Different training-test splits were used to check if there would be any improvements in the accuracy for both models; however this did not lead to any further promising results.

There was limited success using either algorithm but using them helped the authors recognise that any future modelling must be able to support time-series data.

4.4 ANN: LSTM Neural Network

The use of an LSTM neural network algorithm provided the best results from the research conducted, which can be seen from the graph in Figure 12 which shows that the predicted values follow the trend of the validation set very closely even though there is slight under fitting in the model. This model used 200 epochs, and this means that the entire training dataset is passed forward and backwards through the neural network 200 times. As the number of epochs for the algorithm increased, the root mean square error (RMSE) would exponentially decrease. It was found that the RMSE decreased at a much lower rate at approximately 200 epochs, and there was no point in running it for any longer.

The root mean square error is a measure of the difference between the predicted value and the actual value in the dataset, and it should be minimised as much as possible.

One of the main drawbacks of the use of the LSTM neural network is that due to its complexity and the number of epochs that are run, there is a high computational training time required for the algorithm to build a model. In the future, if re-training is required of the dataset, there must be at least an hour dedicated to doing so, which is how long it initially took to train the model. The forecast in Figure 12 shows a plausible result for the prediction over the next 30 days however, further ongoing testing will need to be performed to check the accuracy of this prediction.

5. Conclusion

Predictive maintenance, using Neural Networks such as LSTM as used in this study, has the potential to help machine users to avoid unscheduled Linac downtime and unscheduled maintenance.

The authors present a prototype model (NCDA) that was developed and validated which provides a forecast for the daily output and alerts the user when there is the potential for the output to change beyond what is considered normal daily drift. The model provided a good fit with the validation dataset used, meaning that trends in daily outputs were able to be identified using the LSTM method.

Although we work with a monthly output tolerance of 2% and action level of 3% anything below this falls to the call and judgement of the Physicist on duty as to whether outputs need to be adjusted based on individual experience and intuition. Therefore, to predict re-calibration of the LINAC was a mixture of machine data and human behaviour, and this inconsistency in terms of what the tolerance level for re-calibration was one of the main challenges faced. Despite this the project has a huge scope in helping personnel make clinical decisions on whether to proceed with the more resource heavy output calibration and can reduce unscheduled maintenance.

The NCDA as described here, using a LSTM model as its basis shows promise as a tool for detecting changes in our daily Linac output measurements beyond that considered normal. Further work and development could include using more features provided in the original data such as breakdown occurrences and data available from the daily device such as flatness, symmetry and energy fluctuations. However, we would need to account for the fact that increasing the number of features we decide to use increases the layers in the LSTM model, making for a more complex system than the current NCDA tested here, leading to longer calculation times. Carrying out more research into and combining CUSUM with other change detection and time series algorithms to detect change, such as a moving average technique, may or may not give more fruitful results than seen here, however the authors consider the results presented by NCDA using LSTM to show promise, and will move forward in testing this using other daily output measurements devices, namely the Machine Performance Check (MPC) by Varian. Furthermore, automation of data processing and data storage so we don't have to manually add data points and making use of JSON files for storage can be used in creating dashboards with visual representation of the data and machine behaviour which has the potential to make for a more user-friendly experience.

Declarations

Funding N/A

Conflicts of interest/Competing interests N/A

Availability of data and material Available upon request

Code availability Upon Request

Consent for publication All Authors Consent to publication of the article in its current form

References

- [1] ICRU, "Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT). ICRU Report 83," *ICRU*, vol. 10, no. 1, 2010.

- [2] A. TG142, "Task Group 142 report: Quality assurance of medical accelerators," *AAPM*, vol. 36, no. 9, pp. 4917-4212, 2009.

- [3] D. Binny, C. Lancaster, T. Kairn, J. Trapp and S. Crowe, "Technical Note: Monitoring Daily QA 3 constancy for routine quality assurance," *Phys Med*, p. XX, 2016.

- [4] M. F. Chan, "Visual Analysis of the Daily QA Results of Photon and Electron Beams of a Trilogy Linac over a Five-Year Period," *International Journal of Medical Physics and Clinical Engineering in Radiation Oncology*, 2015.

- [5] M. Hossain, "Output Trends, characteristics, and measurements of three megavoltage radiotherapy linear accelerators," *Journal of Applied Clinical and Medical Physics*, 2014.

- [6] B. Pal, A. Pal, S. Das, S. Palit, P. Sarkar, S. Mondal, S. Mallik, J. Goswami, S. Das, A. Sen and M. Mondol, "Retrospective study on performance of constancy check device in Linac beam monitoring using Statistical Process Control," *Reports of Practical Oncology and Radiotherapy*, vol. 25, no. 1, pp. 91-99, 01 01 2020.

- [7] T. Sanghangthum, S. Suriyapee, S. Srisatit and T. Pawlicki, "Retrospective analysis of linear accelerator output constancy checks using process control techniques," *Journal of Applied Clinical Medical Physics*, vol. 14, no. 1, pp. 147-160, 2013.

- [8] Y. Luo, C. Shifeng and G. Valdes, "Machine Learning for Radiation Outcome Modelling and Prediction," *Medical Physics*, vol. 47, no. 5, May 202.

- [9] A. Singh, N. Thakur and A. Sharma, "A review of supervised machine machine learning algorithms. Proceedings of the 10th INDIACom," in *3rd International Conference on Computing for Sustainable Global Development, INDIACom*, 2016.

- [10] G. Susto, A. Schirru, S. Pampuri, S. McLoone and A. Beghi, "Machine Learning for predictive maintenance: A multiple classifier approach," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 812-820, 2015.

- [11] I. Witten, E. Frank, M. Hall and C. Pal, *Data Mining: Practical Machine Learning Tools and Techniques 4th Ed*, Morgan Kaufman, 2017, pp. 456-458.

- [12] T. Pawlicki, M. Whitaker and A. Boyer, "Statistical Process control for radiotherapy quality assurance," *Medical Physics*, vol. 32, no. 9, pp. 2777-2786, September 2005.

- [13] J. U. Rah, D. Shin, D. H. Oh, T. H. Kim and G.-Y. Kim, "Feasibility study using statistical process control to customised quality assurance in proton therapy," *Medical Physics*, vol. 41, no. 9, pp. 092105-1-092105-8, September 2014.

- [14] C. Able, A. Baydush, C. Nguyen, J. Gersh, A. Ndlovu, I. Rebo, J. Booth, M. Perez, B. Sintay and M. Munley, "A model for preemptive maintenance of medical linear accelerators-predictive maintenance," *Radiation Oncology*, vol. 11, no. 1, 2016.

Figures

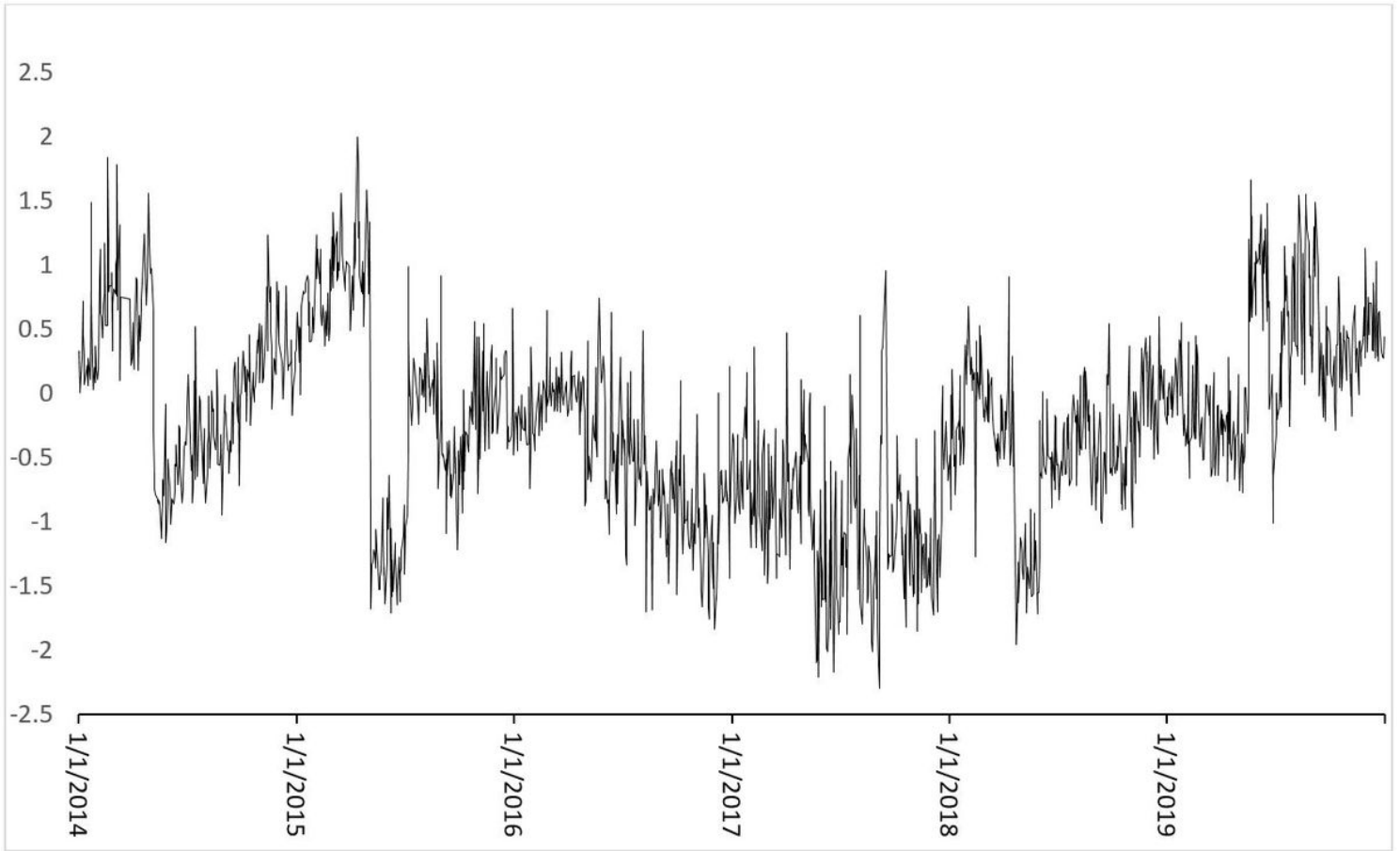


Figure 1

Daily Measurements of 6MV showing % difference from nominal

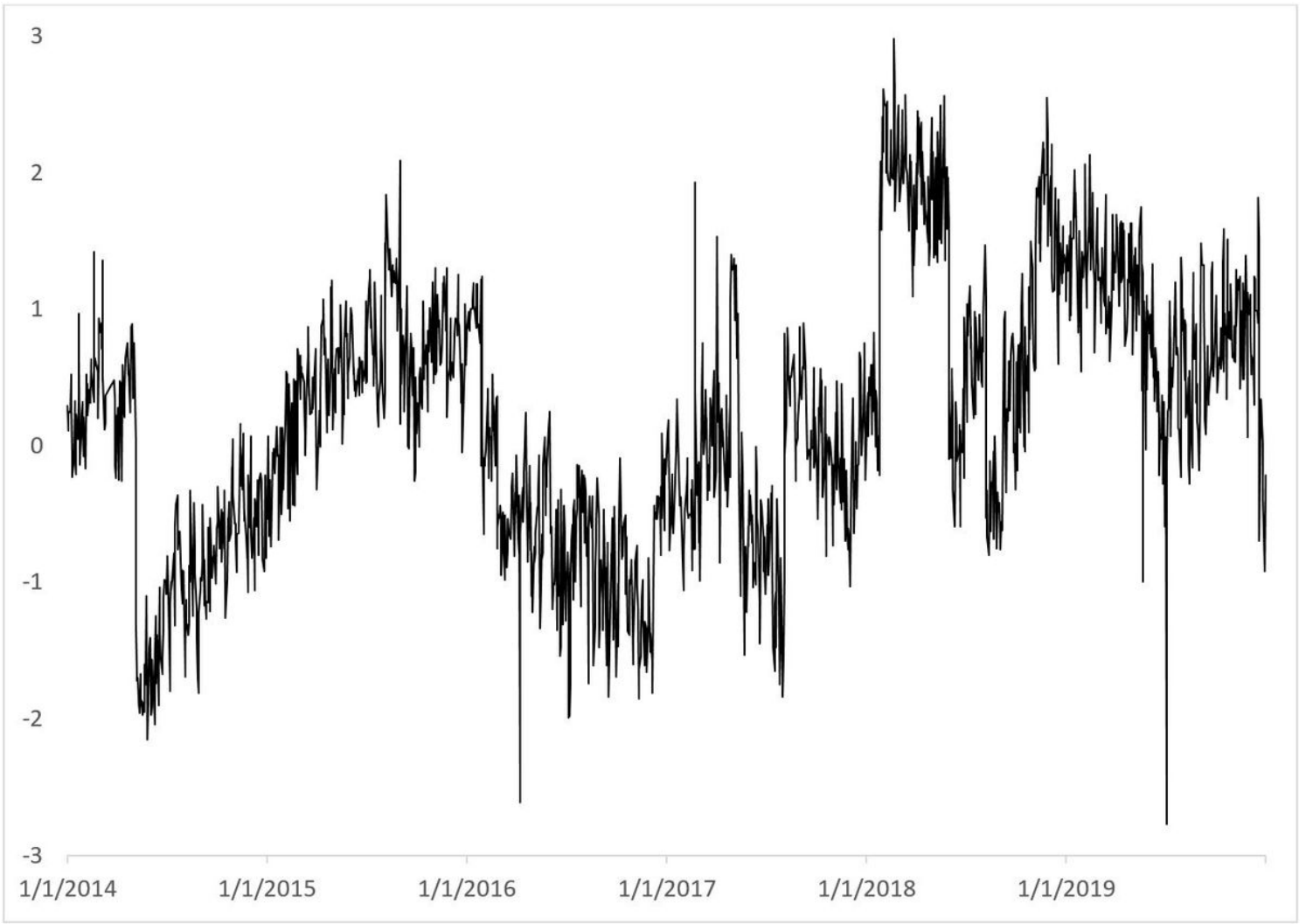


Figure 2

6MeV electron daily output variations

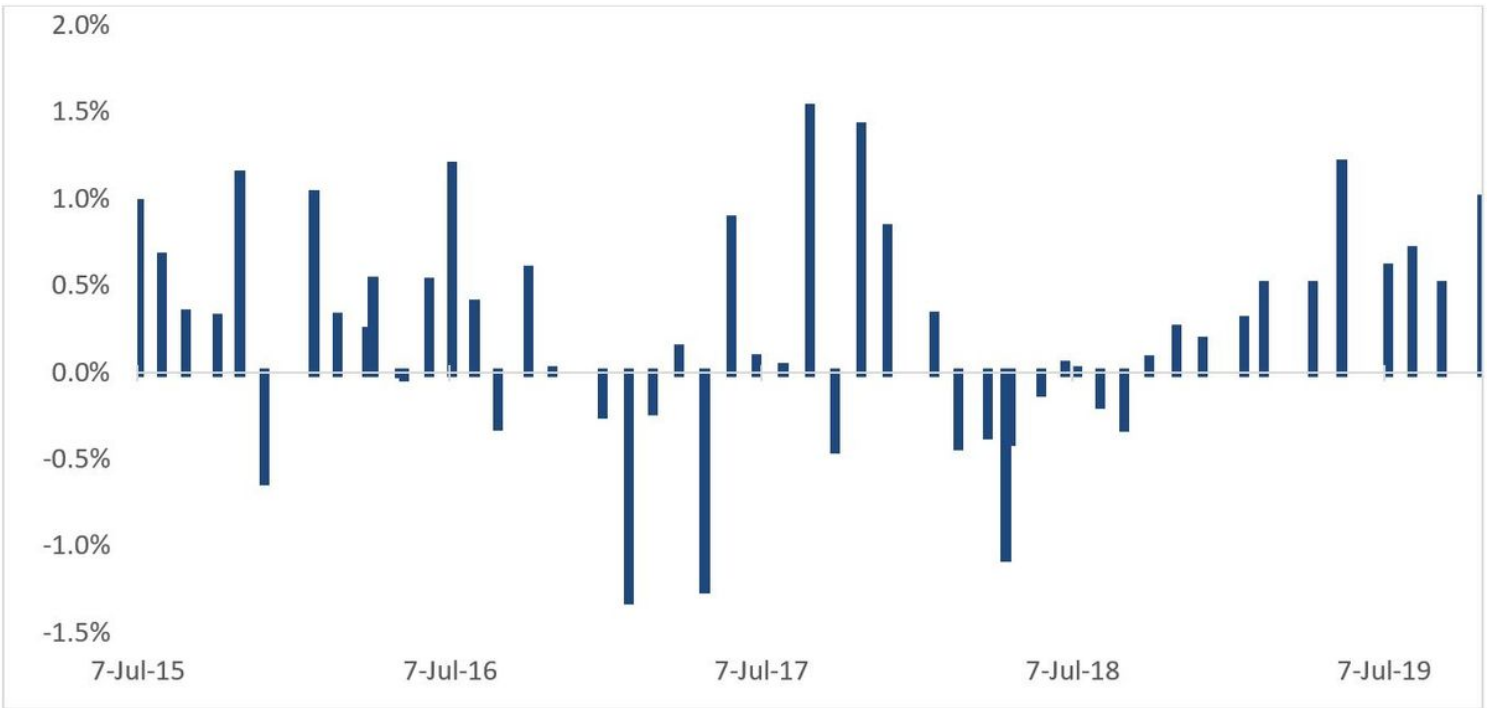


Figure 3

Monthly Output measurements for 6mv photons

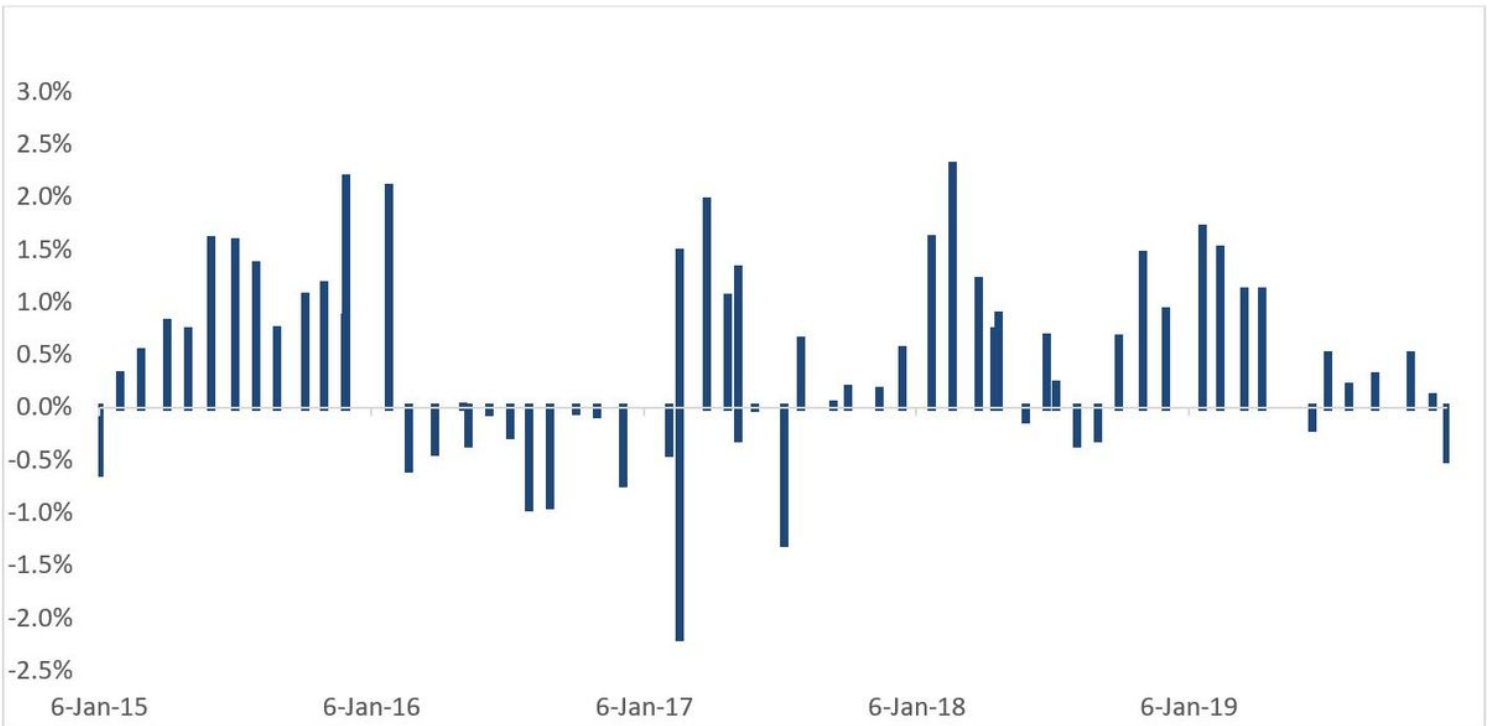


Figure 4

6MeV monthly output variation

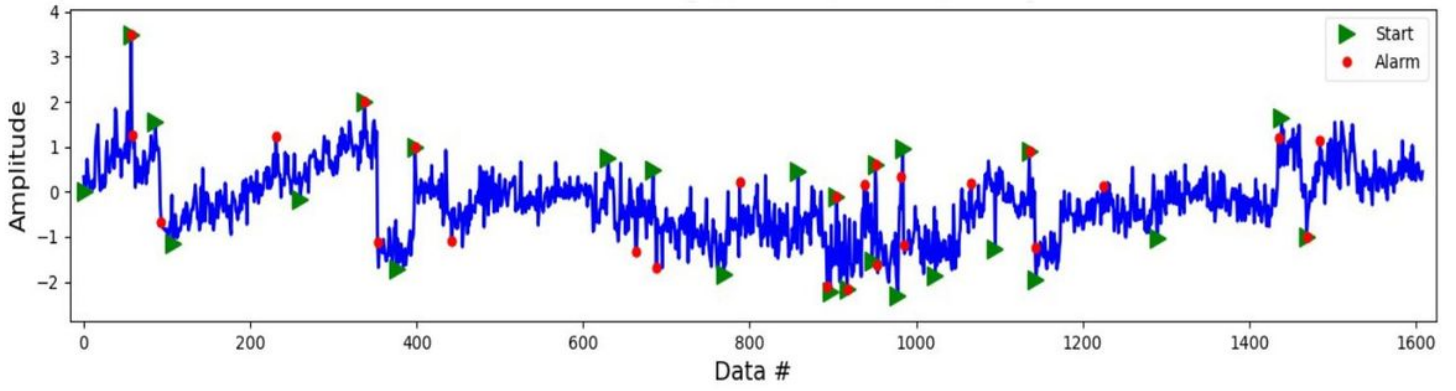


Figure 5

Time series data with detected changes (Alarms) and Resets (Start)

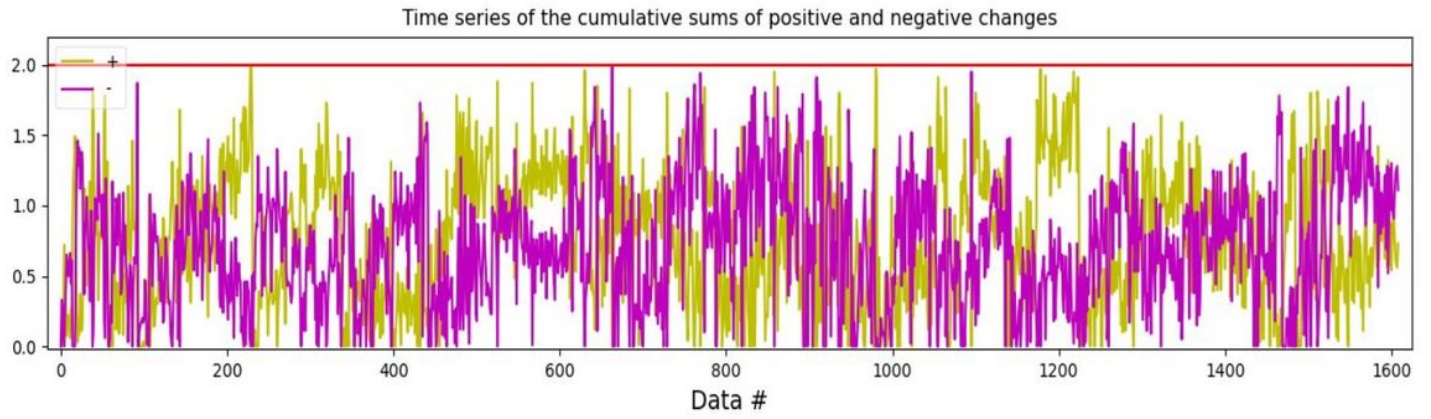


Figure 6

Time series of cumulative sums of positive and negative changes.

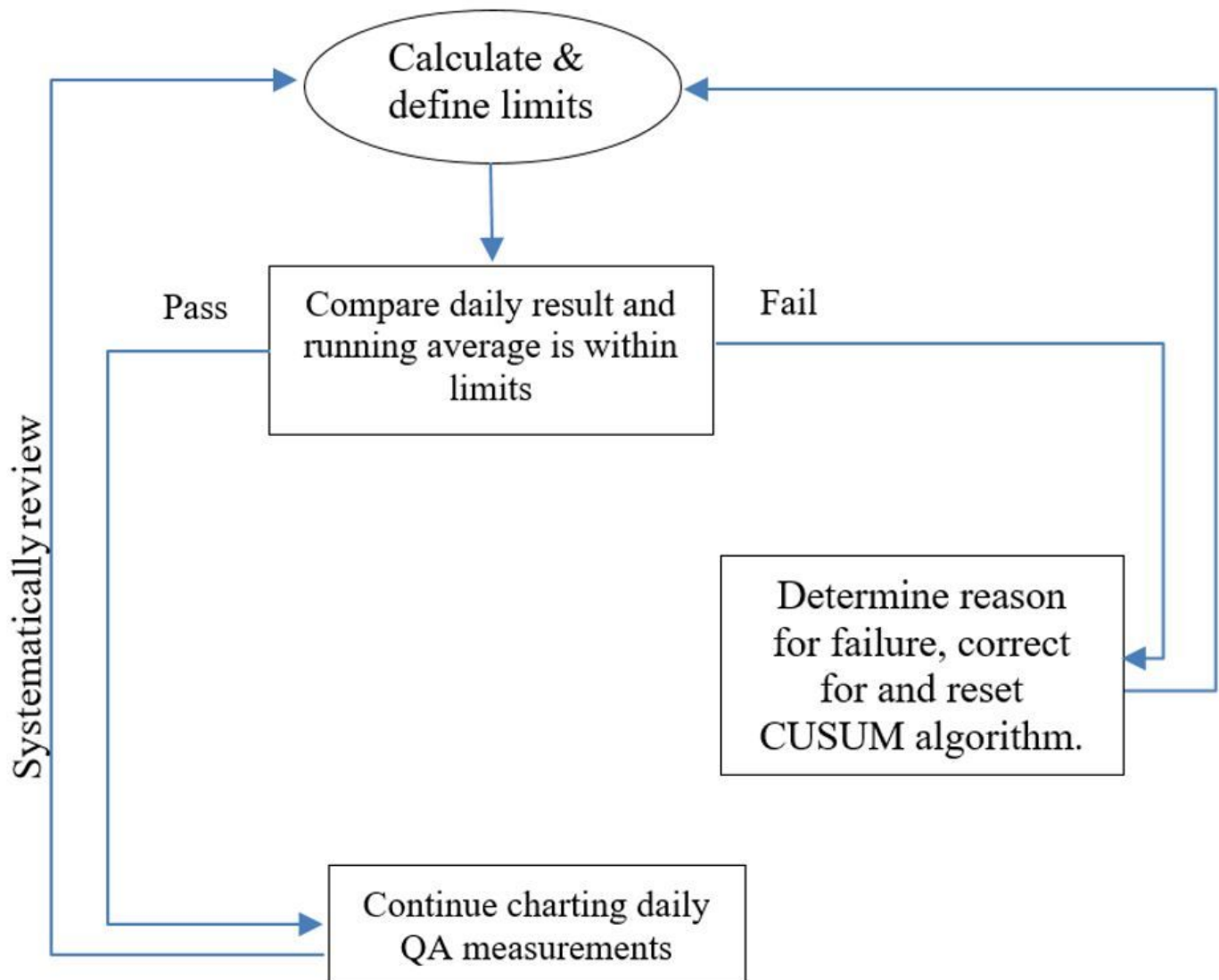


Figure 7

Potential for use of CUSUM to monitor daily performance processes of QA

Linear Extrapolation

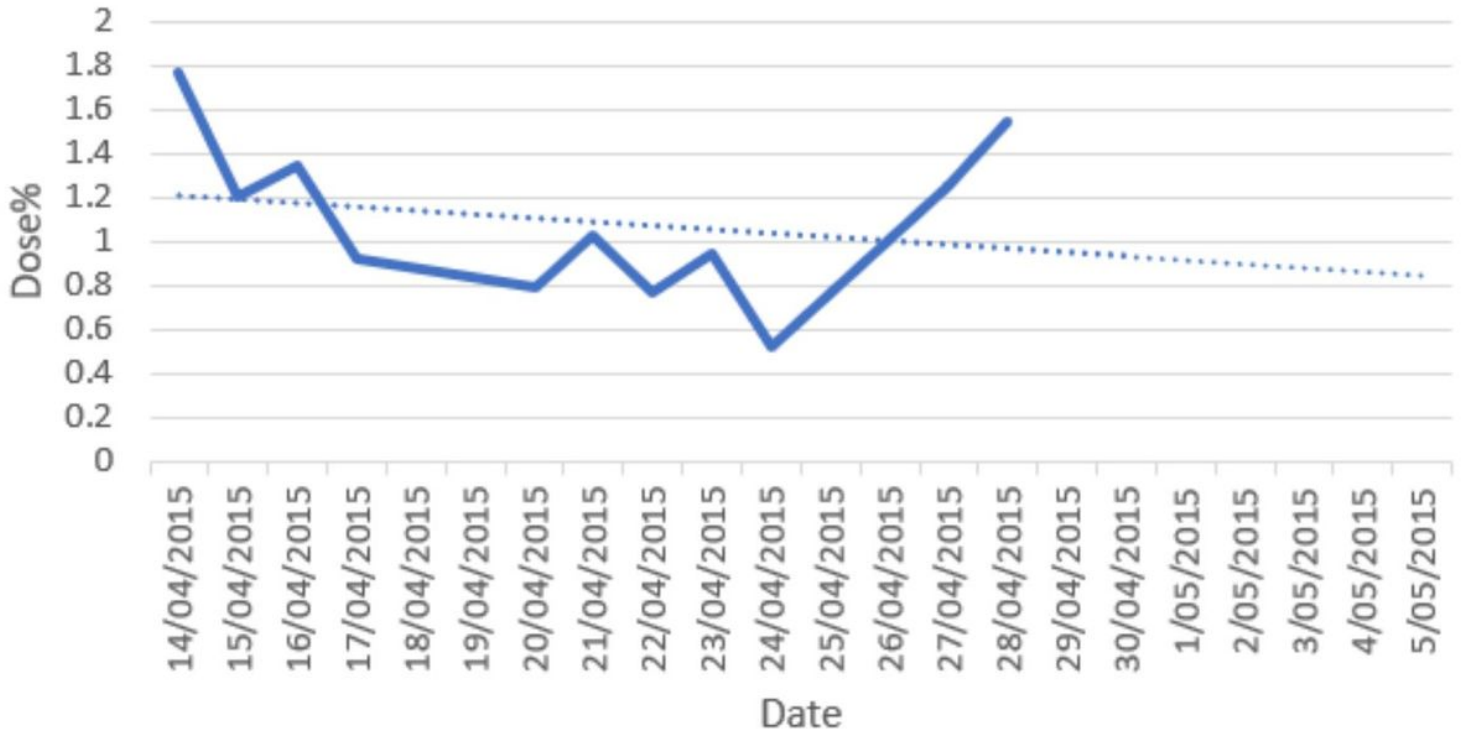


Figure 8

Linear Forecast of 15MeV (Dotted line) modelled from previous results (solid line).

Actual Values

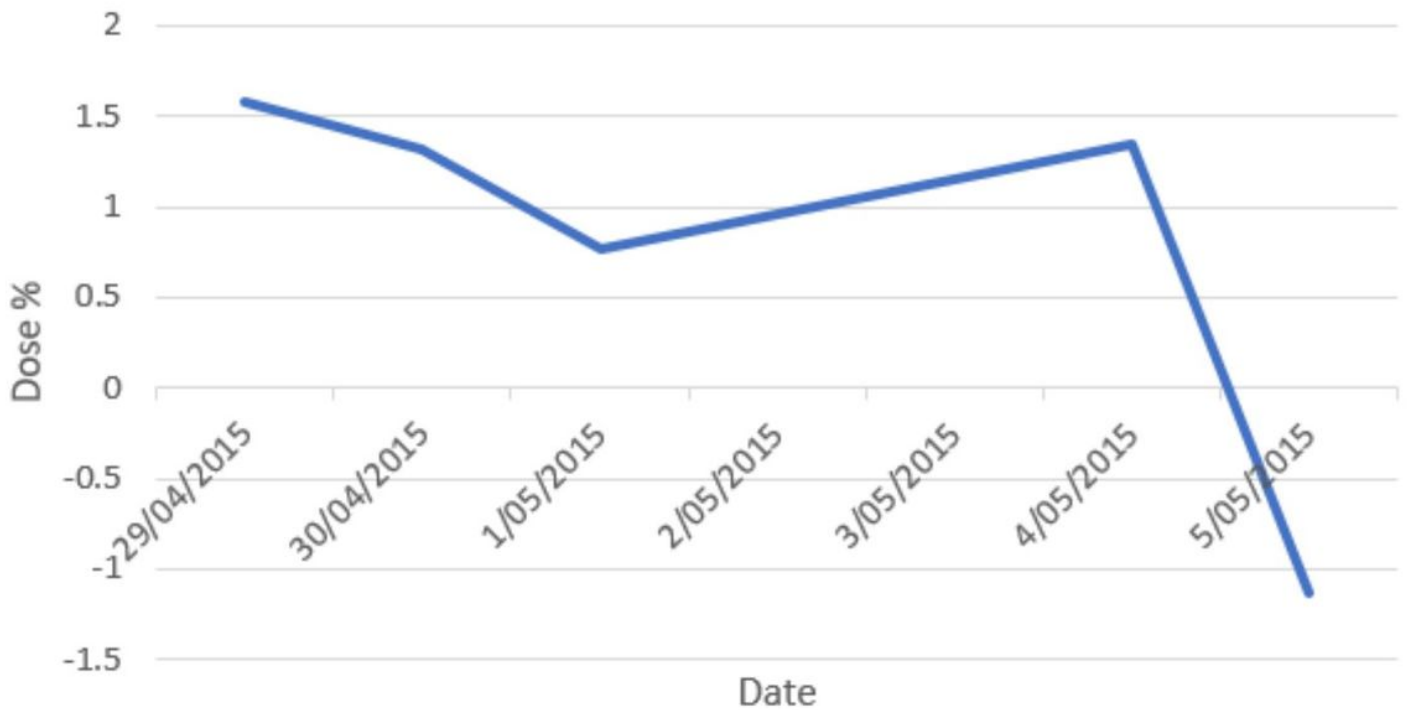


Figure 9

Actual daily results for 15MeV over a 7 day period

	Predicted 0	Predicted 1
Actual 0	313	0
Actual 1	1	0

Figure 10

Confusion matrix for 15MeV with no class weighting used

	Predicted 0	Predicted 1
Actual 0	218	95
Actual 1	0	1

Figure 11

Confusion matrix for 15 MeV with equal weighting

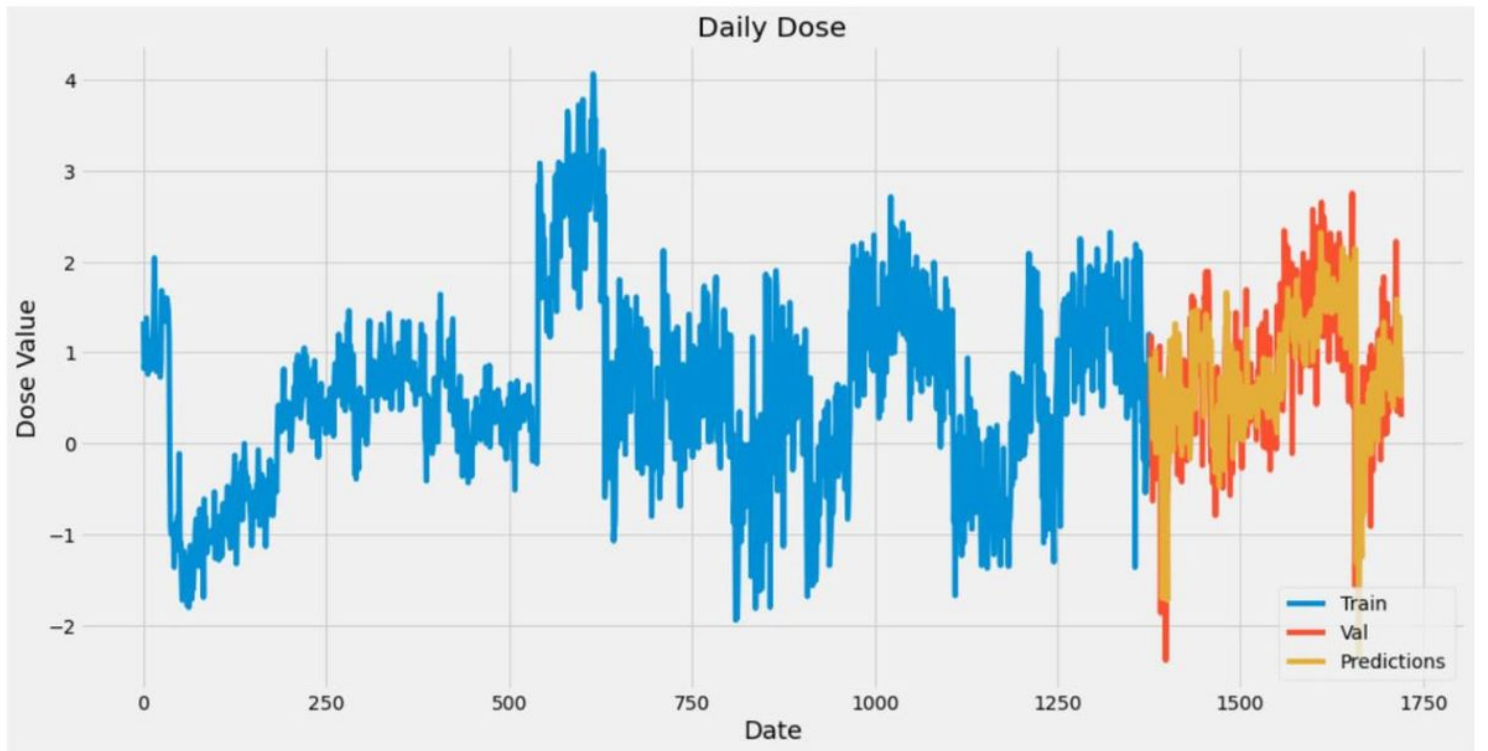


Figure 12

Daily dose prediction for 15MeV

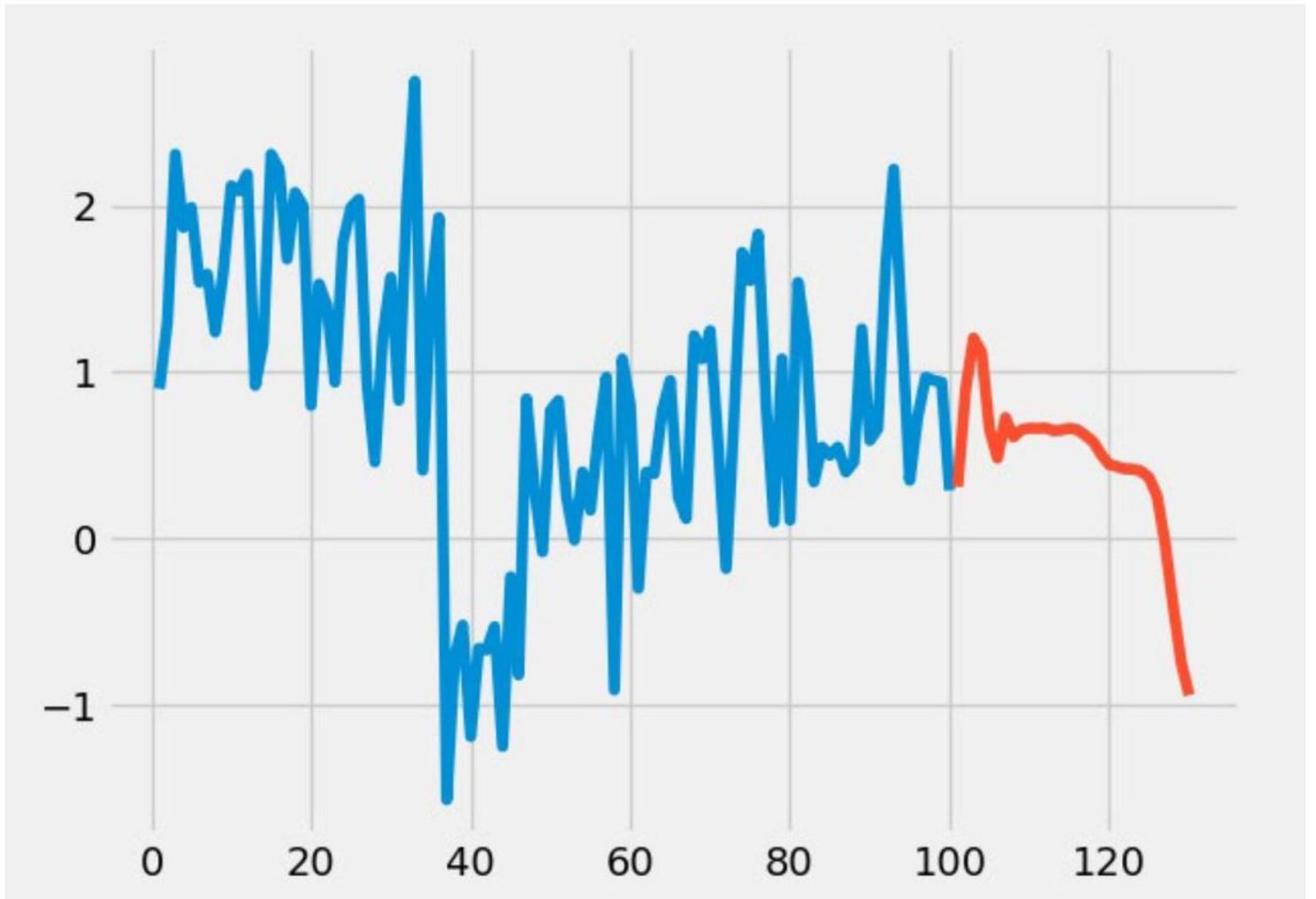


Figure 13

30 Day forecast for 15MeV using LSTM

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [scheme.jpg](#)