Physics and Imaging in Radiation Oncology 15 (2020) 72–76

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.elsevier.com/locate/phro

Multivariate log file analysis for multi-leaf collimator failure prediction in radiotherapy delivery

Arkadiusz Mariusz Wojtasik[a,*], Matthew Bolt[b], Catharine H. Clark[b,c,d], Andrew Nisbet[d], Tao Chen[a]

[a] Department of Chemical and Process Engineering, University of Surrey, Guildford, UK
[b] Radiotherapy Physics, University College London Hospitals, London, UK
[c] National Physical Laboratory, Teddington, UK
[d] Department of Medical Physics and Biomedical Engineering, University College London, London, UK

ABSTRACT

Background and Purpose: Motor failure in multi-leaf collimators (MLC) is a common reason for unscheduled accelerator maintenance, disrupting the workflow of a radiotherapy treatment centre. Predicting MLC replacement needs ahead of time would allow for proactive maintenance scheduling, reducing the impact MLC replacement has on treatment workflow. We propose a multivariate approach to analysis of trajectory log data, which can be used to predict upcoming MLC replacement needs.

Materials and methods: Trajectory log files from two accelerators, spanning six and seven months respectively, have been collected and analysed. The average error in each of the parameters for each log file was calculated and used for further analysis. A performance index (PI) was generated by applying moving window principal component analysis to the prepared data. Drops in the PI were thought to indicate an upcoming MLC replacement requirement; therefore, PI was tracked with exponentially weighted moving average (EWMA) control charts complete with a lower control limit.

Results: The best compromise of fault detection and minimising false alarm rate was achieved using a weighting parameter (λ) of 0.05 and a control limit based on three standard deviations and an 80 data point window. The approach identified eight out of thirteen logged MLC replacements, one to three working days in advance whilst, on average, raising a false alarm, on average, 1.1 times a month.

Conclusions: This approach to analysing trajectory log data has been shown to enable prediction of certain upcoming MLC failures, albeit at a cost of false alarms.

1. Introduction

Equipment breakdown and subsequent unscheduled maintenance is a common reason behind radiotherapy delivery interruption. Not only does such interruption put undue stress on the patients but it also disrupts the workflow within the treatment centre, impacting staff workload [1,2].

MLC replacement need is a common reason behind unscheduled linear accelerator (linac) maintenance and, therefore, treatment disruption. One linac considered in this project, for example, had nine MLC related maintenance events over a four-month span. Four of those occurring during patient treatment hours [3]. Development of a method for predicting MLC leaf failure ahead of time would allow for proactive maintenance scheduling, limiting the impact of unscheduled MLC leaf replacements on patient treatment.

For these reasons, MLC leaf failure prediction has been given attention over recent years. Many of the approaches tested relied on tracking individual parameters related leaf behaviour to detect abnormalities. Some parameters investigated were total distance travelled [4], leaf velocity [5] or frequency of positional discrepancies [6]. The values for these were recorded either via trajectory logs [4,6] or with additional testing during daily QA [3]. Each of these studies reported some degree of success, showing that certain MLC leaf failures can be anticipated in advance, based on changes in the leaf behaviour prior to failure. These studies however, considered the parameters in isolation.

Trajectory log files track a wide range of parameters automatically. A range of these have been shown to act as predictors of MLC leaf failure [6]. It was noted that the parameters tracked by trajectory logs displayed intercorrelation between each other, revealing potential for application of multivariate approaches. This study aimed to develop a multivariate approach to MLC failure prediction and asses its performance against current approaches.

2. Materials & methods

2.1. Log files

Trajectory log files from two Varian TrueBeam linacs at Royal

* Corresponding author at: The Department of Chemical and Process Engineering, University of Surrey, Guildford, UK.
E-mail address: a.m.wojtasik@surrey.ac.uk (A.M. Wojtasik).

https://doi.org/10.1016/j.phro.2020.07.011
Received 29 February 2020; Received in revised form 27 July 2020; Accepted 28 July 2020
Available online 10 August 2020

Surrey County Hospital (RSCH), were collected for a period of seven and six months for linacs one and two, respectively. These files correspond to all fields delivered by the linacs over the stated period. Each log file tracks 130 different variables during treatment, comparing real time measurements to the predetermined treatment plan, sampled every 20 ms. The log files were anonymised by RSCH staff before being made available for analysis and were part of a service evaluation study at RSCH.

A total of 4280 and 3205 log files were collected and analysed for Linac 1 and Linac 2 respectively [7].

Data sets with a wide range of parameters and inter-correlation lend themselves towards multivariate statistical approaches. In correct circumstances, a multivariate approach offers several advantages over its univariate counterpart. A single multivariate performance index (PI) would replace the need to track each parameter separately, highlighting an advantage this approach has over univariate methods.

Another main advantage of utilising multivariate approaches lies in detecting false negatives. A false negative is a case where erroneous behaviour of a single parameter looks normal in isolation and thus is assumed to be operating as intended [8]. Since the multivariate approach accounts for the behaviour of a parameter in relation to all other parameters in the data set, it performs better in highlighting unnatural behaviours even when no individual parameter breaches its control limits [9].

2.2. Data cleaning

To extract the necessary data from the log files provided, the files were first converted from binary to csv format to allow for them to be analysed. This was performed in python using the pylinac package [10]. The resultant csv files were then loaded into python for further manipulation. The data manipulation was performed with the NumPy [11] and pandas [12] packages. Since log files return data on expected and actual parameter measurements, the deviation of each parameter was found by subtracting the actual measurement from the expected value. The mean deviation in each parameter, for each treatment delivered, was then calculated, and written into a separate master database, indexed by the time and date of the treatment. This new master database, containing the mean deviation of every measured parameter for each field delivered was used to perform the analysis. This process can be visualised in Supplementary Material Two, Figure One.

2.3. Performance index generation

Principal component analysis (PCA) was used as a means for data exploration to determine what approach to take when developing a performance index. An explanation of PCA can be found in Supplementary Material One.

Following PCA the loading matrices of the analysed data windows were inspected. It was noted that the key variance contributions in the data sets originated from two sources. Natural process variation of all variables in principal component one (PC1) as well as the covariances between bank and leaves within the specified bank in principal component two (PC2). Based on that observation, the variance explained by the first two PCs was to be used as a performance index monitoring process behaviour.

To calculate the performance index firstly, the master data set for a linac was loaded into python and a window size to be employed in the analysis was selected. The size of the moving window corresponds to the number of log files under analysis and therefore the number of radiation fields delivered.

For the window undergoing analysis, containing a subsection of the mean deviations cohort as described in Section 2.3, PCA was performed using the scikit-learn package available in python [13]. The explained variance for the first two PCs based on the window was extracted into a separate list.

The window moved on, dropping the first data point from the dataset whilst adding the next available datapoint. This process was repeated until the window swept over the full set of data, collecting the explained variances throughout. The variances explained by the two PCs for each window were summed together, resulting in a list of PIs for each data window. This process has been summarised in Supplementary Material Two, Figure Two.

Window sizes of 40, 60, 80 and 100 were tested on datasets for both linacs in the study. 40 was chosen as the default starting point since during uninterrupted linac operation, an average of 40 fields per day per linac were delivered. The performance of each data window was assessed based on accurate failure prediction and false alarm rates.

2.4. Exponentially weighted moving average control chart

Having generated a single monitoring statistic of PI, a corresponding control chart with appropriate control limits was then developed. Since the changes in parameter behaviour in the lead up to failure are small and gradual, an EWMA chart was chosen for this purpose since it performs well in such scenarios [14].

The EWMA statistic z is calculated using the following equation:

$$z_n = \lambda x_n + (1 - \lambda) z_{n-1} \tag{1}$$

where z is a weighted running average of the observations, λ is a weighting factor where $0 < \lambda < 1$ and × is the mean of the current subgroup being considered [15]. Adjusting the value of λ, adjusts the extent to which older data affects the value of the EWMA statistic.

The control limits for EWMA charts are based on standard deviation of the data, similarly to Shewhart control charts. in the case of EWMA charts, however, the standard deviation needs to be adjusted by the weighting factor, such as:

$$\sigma_{EWMA} = \sigma\left(\frac{\lambda}{2 - \lambda}\right) \tag{2}$$

where σ is the standard deviation of the historical data. The lower control limit (LCL) would then be calculated as:

$$LCL = \mu - 3 * \sigma_{EWMA} \tag{3}$$

where μ is the mean of the historical data. The factor of three is introduced to match other commonly encountered control charts [14].

Using PI as the historical data, EWMA control charts for the two linacs have been generated along with a lower control limit. The control limit was based on PI readings corresponding to a period of normal, faultless linac behaviour as commonly seen in control chart generation. This ensures that the control limit is designed independently of the faults it is trying to detect and describes the natural deviation of the process only [14]. To test the performance of this approach, the control charts were compared against linac service logs to see if a drop below the control limit was visible ahead of an MLC motor replacement event. The selection of window size and λ in Eq. (1) was done by comparing the plots of the performance index over time against each other for all window sizes and assessing their false alarm rate as well as failure prediction ability.

The sensitivity and specificity of the final model was calculated to assess the performance of the approach. The sensitivity of the model for each linac was determined by dividing the number of correctly predicted MLC replacement events by the number of total MLC replacement events recorded in the maintenance logs. To calculate the specificity of the model it was first necessary to note down the number of datapoints breaching the control limit due to false alarms and the number of datapoints correctly within the control limit (i.e. not directly preceding an undetected MLC replacement event). The sum of these values results in the total number of datapoints relating to controlled linac operation. The specificity of the model for each linac was then found by dividing the number of datapoints correctly within the control limits by the total number of data points related to controlled linac operation. This approach assumed that every single point on the EWMA charts breaching the control limit was to be treated as a potential failure. This assumption was deemed appropriate due to the aggregatory nature of the PI and EWMA statistic. Each value of PI was dependant on the window size selected (40–100 log files) with the EWMA statistic being calculated based on cohorts of PI. In all cases considered, data from at least 49 log files was used to generate a single EWMA statistic value which was then to be plotted. This value was a significantly large enough sample to signal a fault.

3. Results

Analysis of the log file data showed MLC positioning errors for leaves in each bank correlated closely (correlation coefficients in the range of 0.55–0.85) with the positioning errors of the bank they were in. Strong correlations (correlation coefficients between 0.5 and 0.9) were also seen between leaves in the same bank of the collimator. No strong correlations or covariances were seen outside the MLC parameters, therefore the continued focus was on MLC related data only in further analysis.

Another key result from the preliminary analysis was that the erroneous MLC behaviour in the lead up to a subsequent leaf replacement was short lived in comparison to the size of the dataset. This finding led to the development of a moving window approach towards PCA, scanning through the data set treatment by treatment.

In the case of the linacs investigated, an 80 datapoint moving window achieved the best compromise between fault prediction and false alarm rate. The summary of these results is shown in Table 1. The PI based on this window size was used for control chart generation.

λ values of 0.05, 0.1, 0.2 and 0.3 were tested on the PI values for both linacs based on 80 data point windows. All values of λ resulted in the same fault detection rates, however, λ of 0.05 resulted in lower false alarm rates and thus achieved the most desirable performance.

Over the combined thirteen months of linac data, a total of 23 control limit breaches occurred. Fourteen of those were false alarms whilst the other nine corresponded to upcoming leaf replacement events. Eight of the thirteen recorded MLC maintenance events were preceded by a control limit breach. This approach, therefore, yielded a sensitivity of 62% and raised a false alarm 1.1 times a month. The specificity of the model was found to be 76% for the combined data of the two linacs based. A breakdown of the results is given in Table 2. EWMA control charts for the two linacs can be seen in Fig. 1 and Fig. 2.

4. Discussion

Based on the preliminary analysis of the master data set, the proportion of variance, as explained by the first two PCs, was used as the performance index of each linac. As previously mentioned, the first two PCs encompass the variance caused by the natural variation in MLC positioning as well as the covariances of the leaves with their respective banks. From a theoretical standpoint, these two PCs would sufficiently explain the variance/covariance within the MLC. Any sharp drops in the performance index, therefore, would indicate the presence of a new, unexpected cause of variance elsewhere in the MLC identifying potential upcoming failure.

Table 1
Summary of window size selection results.

Linac 1			Linac 2		
Window Size	Faults Detected	False Alarms	Window Size	Faults Detected	False Alarms
40	6	9	40	3	11
60	6	7	60	2	10
80	6	6	80	2	8
100	3	5	100	2	8

The study assumes that trajectory log files are a suitable source of data to base a predictive model on. This assumption can seem dubious at first given that the reliability of MLC positioning data has been questioned in recent studies. MLC positioning accuracy has been shown to exhibit high levels of variability and root mean square (RMS) error in cases where leaves where accelerating or decelerating [16]. The high mid-treatment standard deviation in the leaf positioning have been noted in this project as well, manifesting itself as the key variance contributor in PC1. This variability could be a contributor to the somewhat erratic behaviour of the PI visible on the control chart. On the other hand, the same study states that the average leaf positioning errors were negligible over all treatment sites. The PI in this study is based on average deviations in MLC parameters across the treatments delivered and as such should benefit greatly from the negligible average errors. Log files also have the benefit of being generated automatically and collecting data with every radiation field delivered. This upside of bulk data collection without introducing linac downtime cannot be understated.

Given the doubts surrounding the limitations trajectory logs, a model based purely on log data would not be able to replace current QA protocols. It is deemed, however, that it can be a very useful supplementary tool which fulfils the purpose of this project [17].

A potential limitation of this approach to log file analysis has been the fact that it has not been tested on other linac models or linacs from other vendors. Given the nature of the performance index, different calculations of leaf position relative to the respective bank would obsolete this approach in the current iteration. As of right now, this approach should be applicable to Varian TrueBeam linacs whilst a different approach would be necessary for other linac models, if the parameters in the log files are measured in a different manner. An avenue of future work would be to transfer learning from this multivariate approach designed for Varian TrueBeam linac onto other linac models, increasing its practicality.

By iterating over a small selection of the datapoints during PCA, smaller, shorter-lived erroneous deviations become more visible in comparison to when the full dataset was considered, supporting the decision to utilise a moving window approach. Linac service logs also revealed that MLC motor replacement is commonly followed up by calibration. It was uncertain whether MLC behaviour pre and post calibration would be comparable enough to ensure accurate model development, further supporting the decision to use a moving window approach. Different window sizes were tested to achieve the best compromise between failure prediction and false alarm rate.

Of the four window sizes tested, a set of 80 datapoints was the most optimal. Window sizes of 40 and 60 displayed erratic performance index behaviour, resulting in a much higher rate of false alarms than window sizes 80 & 100. Window size 100 handling greater quantities of data drowned out the subtle changes in the MLC behaviour which need to be detected for this model to detect upcoming failure. This has been summarised in Table 1.

The resultant model achieved an MLC failure prediction sensitivity of 62%, accurately predicting eight out of the thirteen logged failures. Another study based on tracking MLC positioning errors via log file analysis reported a failure prediction rate of 59% for a linac over a three-year period which is very much comparable to the performance of our model [6]. The same study reports nine flagged failures which did not correspond to a logged MLC failure event, resulting in a false alarm rate of around 0.25 per month. This is significantly lower than the 1.1 false alarms per month reported in this study. This may be explained by the larger historical backlog available to the other research team, which allowed for more comprehensive parameter finetuning.

The specificity of the prediction model was calculated to be 76%, based on the approach described in section 2.4. Approaches described in other publications dealing with the issue of MLC failure prediction have not stated their specificity rates so no parallels between this and

Table 2
Performance summary of the EWMA control charts for the two linacs analysed.

	Months of operation	Control limit breaches	False alarms	Readouts below control limit due to false alarms	MLC fails predicted (sensitivity)	Readouts correctly above control limit	Specificity
Linac 1	7	13	6	477	6 out of 7 (86%)	2800	85%
Linac 2	6	10	8	1046	2 out of 5 (40%)	2000	66%
Total	13	23	14	1523	8 out of 13 (62%)	4800	76%

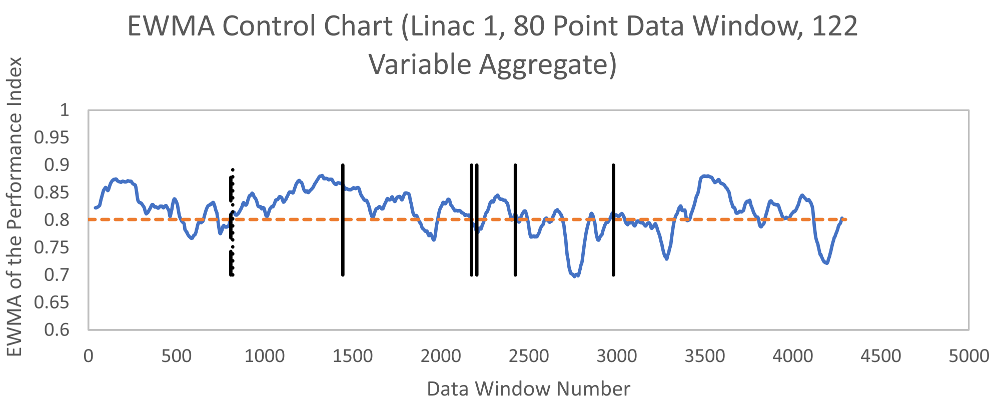

Fig. 1. An EWMA chart of the performance index for linac 1 encapsulating all 122 MLC variables found in log files, vertical lines indicated motor replacement events. The dashed line around the 800 represents two failures in proximity of each other. Drops of the index value below the control limit in the lead-up to a replacement even quantified a successful MLC failure prediction.

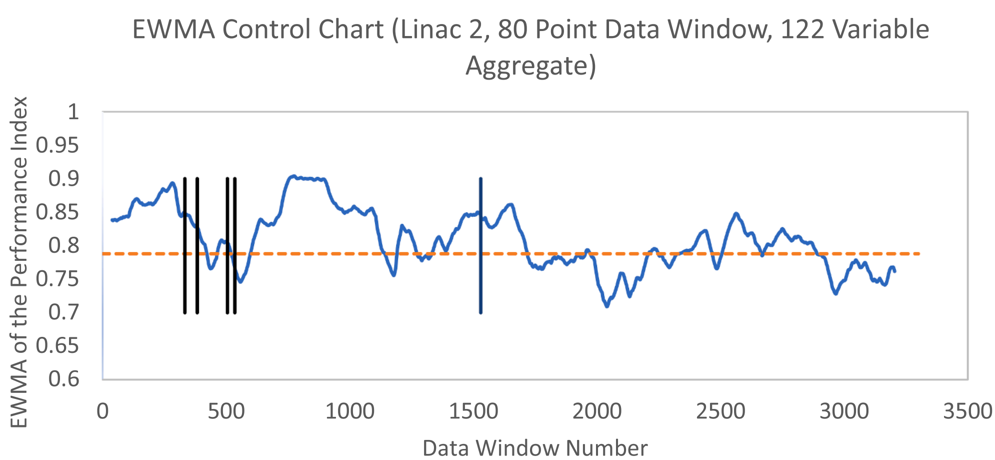

Fig. 2. An EWMA chart of the performance index for linac 2 encapsulating all 122 MLC variables found in log files, vertical lines indicated motor replacement events. Drops of the index value below the control limit in the lead-up to a replacement even quantified a successful MLC failure prediction.

other studies can be made.

In its current iteration, the model flags possible upcoming MLC leaf replacement events but it does not directly identify which leaf will cause the failure. Future work on this project will focus on expanding the model to not only alert of an upcoming fault but also isolate the variables responsible for it.

Currently the performance index has not been compared to any other QA measurements, so it is impossible to comment on whether any of the false alarms correlate with potential issues elsewhere in the linac. An extension of this project does intend on comparing/contrasting log file data with data collected in daily QA and via Varian's Machine Performance Check to widen the parameter pool in which faults can be predicted. The various data sources would also act as cross-validation checkpoints for flagging potential faults hopefully lowering the occurrence of false alarms.

In conclusion, this study showed that a multivariate statistical model can predict an MLC replacement requirement, removing the need to track all the parameters individually. The exact leaf responsible for the failure, however, cannot be directly identified. Instead, a singular performance index based on PCA has been shown to flag upcoming failures in advance, allowing the treatment centre staff to schedule linac maintenance in a proactive manner.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Firstly, we would like to thank Matthew Jones, the Head of Dosimetry, as well as the technical services team at RSCH for enabling the data collection and anonymisation necessary for the development of the predicted model presented here. We would also like to thank both the National Physical Laboratory and the University of Surrey for funding the research, via a strategic research studentship, which led to the development of the model.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phro.2020.07.011.

References

[1] The Royal College of Radiologists. Timely delivery of radical radiotherapy: guidelines for the management of unscheduled treatment interruptions. 4th ed. London: The Royal College of Radiologists; 2019.

[2] Wroe LM, Ige TA, Asogwa OC, Aruah SC, Grover S, Makufa R, et al. Comparative analysis of radiotherapy linear accelerator downtime and failure modes in the UK, Nigeria and Botswana. Clin Oncol 2020;32:111–8. https://doi.org/10.1016/j.clon.2019.10.010.

[3] Hospital RSC. MLC Motor failures 04Feb20. Guildford: Royal Surrey County Hospital; 2020.

[4] Stathakis S, Defoor D, Linden P, Kirby N, Mavroidis P, Papanikolaou N. SU-E-T-545: MLC distance travelled as a predictor for motor failure. Med Phys 2015;42:3460. https://doi.org/10.1118/1.4924907.

[5] Able C, Hampton C, Baydush A, Bright M. SU-E-T-205: MLC predictive maintenance using statistical process control analysis. Med Phys 2012;39:3750. https://doi.org/10.1118/1.4735265.

[6] Wu B, Zhang P, Tsirakis B, Kanchaveli D, LoSasso T. Utilizing historical MLC performance data from trajectory logs and service reports to establish a proactive maintenance model for minimizing treatment disruptions. Med Phys 2019;46:475–83. https://doi.org/10.1002/mp.13363.

[7] Varian Medical Systems. TrueBeam trajectory log file specifications. s.l. : Varian Medical Systems, 2011.

[8] Banerjee A, Chitnis UB, Jadhav SL, Bhawalkar JS, Chaudhury S. Hypothesis testing, type I and type II errors. Ind Psychiatry J 2009;18:127–31. https://doi.org/10.4103/0972-6748.62274.

[9] Vargason T, Howsmon DP, McGuinness DL, Hahn J. On the use of multivariate methods for analysis of data from biological networks. Process Basel Switz 2017;5:36. https://doi.org/10.3390/pr5030036.

[10] Kerns, J.R. A TG-142 toolkit for doing routine linear accelerator quality assurance, https://github.com/jrkerns/pylinac/tree/5702bb0b67acb8216ec2cf5ea78db-b5e20fa3069; 2019. [Accessed: 18 02 2020].

[11] van der Walt S, Colbert SC, Varoquaux G. The NumPy array: A structure for efficient numerical computation. Comput Sci Eng 2011;13:22–30. https://doi.org/10.1109/MCSE.2011.37.

[12] W. McKinney, Data Structures for Statistical Computing in Python, in: S. van der Walt, J. Millman (Eds.), Proc. 9th Python Sci. Conf. 2010; p. 51–56.

[13] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825–30.

[14] Montgomery DC. Introduction to statistical quality control. 6th. Hoboken, New Jersey: John Wiley and Sons, Inc.; 2009. p. 400–32.

[15] Roberts SW. Control chart tests based on geometric moving averages. Technometrics 1959;1:239–50. https://doi.org/10.1080/00401706.1959.10489860.

[16] Agnew CE, Irvine DM, McGarry CK. Correlation of phantom-based and log file patient-specific QA with complexity scores for VMAT. J Appl Clin Med Phys 2014;15:204–16. https://doi.org/10.1120/jacmp.v15i6.4994.

[17] Legrand C. 49. Current uses of log files in the radiotherapy quality assurance workflow for IMRT and VMAT techniques. Phys Medica Eur J Med Phys 2017;44:24. https://doi.org/10.1016/j.ejmp.2017.10.074.