Detection of faulty beam position monitors using unsupervised learning

E. Fol[1,2,*] R. Tomás,[1] J. Coello de Portugal,[3] and G. Franchetti[2,4]

[1]*CERN, 1211 Geneva 23, Switzerland*
[2]*Johann-Wolfgang Goethe University, 60438 Frankfurt am Main, Germany*
[3]*Paul Scherrer Institut, 5232 Villigen, Switzerland*
[4]*GSI Helmholtzzentrum fr Schwerionenforschung, 64291, Darmstadt, Germany*

Optics measurements at the LHC are mainly based on turn-by-turn signal from hundreds of beam position monitors (BPMs). Faulty BPMs produce erroneous signal causing unreliable computation of optics functions. Therefore, detection of faulty BPMs prior to optics computation is crucial for adequate optics analysis. Most of the faults can be removed by applying traditional cleaning techniques. However, optics functions reconstructed from the cleaned turn-by-turn data systematically exhibit a few nonphysical values which indicate the presence of remaining faulty BPMs. A novel method based on the Isolation Forest algorithm has been developed and applied in LHC operation, allowing to significantly reduce the number of undetected faulty BPMs, thus improving the optics measurements. This report summarizes the operational results and discusses the evaluation of the developed method on simulations, including extensive studies and optimization of the preexisting cleaning technique and verification of a new method in terms of coupling measurement. The advantages of the chosen algorithm compared to some other unsupervised learning techniques are also discussed.

I. INTRODUCTION

LHC optics measurements are mainly based on the transverse beam motion recorded by the beam position monitors (BPMs) around the ring. The optics functions are computed from the properties of this signal as inferred from a harmonic analysis [1–3]. The appearance of faulty signals can have significant impact on the optics computation and corrections. Several numerical thresholds as well as a cleaning technique based on singular value decomposition (SVD) are used in order to remove faulty signals. Since nonphysical values still can be observed in the optics functions computed from the cleaned data, it is not uncommon that additional manual cleaning of harmonic analysis data is required, followed by repeating the optics computation. Therefore, the preexisting techniques appear to be insufficient and alternatives for automatic identification of faulty BPMs are required to avoid human intervention and ensure that faulty signal does not corrupt the optics analysis.

Considering Machine Learning-based identification of faulty BPMs, application of supervised learning to classify the BPM signal appears not appropriate. For the supervised classification, labeled training data is required. Since no ground truth is available, only the cleaning results from the past measurements can be used as labels for good and bad BPMs. However, this would lead to reproducing the results of existing techniques, instead of improving the optics analysis. Since the reasons of BPM failures are partially unknown, we cannot define rules which would indicate faulty BPMs that actually cause the erroneous optics computation. It has to be noted that there is not necessarily a direct relation between the location where the error is observed and the actual bad BPMs. Due to the way the optics is calculated [4,5], a single faulty BPM may cause erroneous optics calculation at multiple locations, i.e., produced errors might appear not directly at the position of the bad BPM, but propagate to the locations of adjacent BPMs. The new method should automatically recognize bad signals in the online provided turn-by-turn data prior to optics computation, without requiring rules or thresholds which depend on optics settings or signal properties and have to be adjusted during operation. In this work we demonstrate that unsupervised learning successfully meets these requirements.

A. Unsupervised learning

Unsupervised learning deals with tasks where only input data is available and the target is to find patterns in the given data or to extract new information. The ability of

*elena.fol@cern.ch

unsupervised techniques to find hidden patterns in the data is a powerful tool for faulty BPM identification. The signal produced by a faulty BPM has different properties compared to normal functioning BPMs. Unsupervised methods can automatically find relevant data properties and the differences which indicate anomalous points in the data provided to the algorithm.

Most common clustering techniques based on centroid search such as K-means [6] appear not to be suitable for faulty signal detection, since the appearance of outliers (data points that significantly differ from other observations) affects the computation of the mean of parameters. A possible solution is to apply density-based algorithms which is discussed in Sec. III. Several unsupervised learning algorithms have been explored to improve the quality of the optics measurements at LHC [7,8], among others the Isolation Forest (IF) algorithm [9]. It detects data anomalies using an ensemble of randomized decision trees. Such decision tree represents a sequence of random splits which are performed until each single data point is "isolated." The principle is illustrated in Fig. 1. The split is selected randomly between maximum and minimum values of randomly selected data feature. On average fewer random splits will be needed to isolate an anomalous data point. The number of splits represents the path length from the root to the leaf of a decision tree. Using a single tree might lead to biased results [10], hence the path lengths are averaged over a number of trees (forest) in order to measure the normality of each isolated point. A significant advantage of the Scikit-Learn [11] implementation of the algorithm used in this work is that it requires only the number of trees and the expected proportion of outliers in the provided dataset (contamination rate) as tuning parameters. IF does not require any data normalization or rescaling and can be applied directly on harmonic analysis of BPM signals before computing the optics functions. This method is fully integrated into optics measurements at LHC and has been successfully used during commissioning and machine developments under different optics settings in 2018.

B. Traditional techniques

Statistical techniques capable of finding independent physical signal components have found their application in
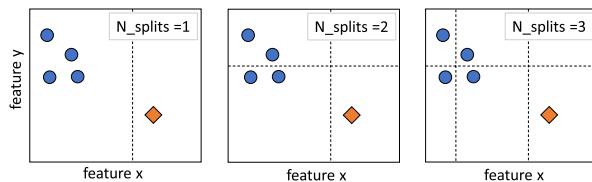

FIG. 1. A conceptual illustration of the IF algorithm. An anomalous point is more likely to be isolated by just one random split on a randomly selected input feature, compared to three splits required to isolate each of the core data points.

beam diagnostics and in particular in BPM data analysis in several facilities [12,13]. Traditional techniques for the cleaning of turn-by-turn data at the LHC include thresholds to define anomalously large and low values of the measured beam position, identification of zero values which replace unphysical BPM readings, as well as manual cleaning. A special list of systematically failing BPMs is used in the dedicated analysis software to exclude these BPMs from the actual data processing. An approach to identify these BPMs dynamically is discussed in Sec. V.

Most of the noise and faulty signals can be removed using these methods, as well as through applying advanced signal improvements techniques based on SVD. The singular vectors of a turn-by-turn data matrix containing the signal of several BPMs correspond to temporal and spatial modes variations describing the beam motion [14]. SVD-based cleaning allows to identify faulty BPMs and to reduce the noise in recorded signal. SVD modes with localized spikes in their spatial vectors indicate faulty BPMs using so-called *SVD cut* as threshold to find such spikes. To globally reduce the noise on all BPM readings, only a predefined number of strongest singular modes, defined by the *SVD mode* setting, remain in the turn-by-turn data. While the SVD cut value has a direct influence on the number of BPMs identified as faulty, the SVD mode setting affects the overall noise level in turn-by-turn signal. In Sec. III we demonstrate the importance of choosing appropriate SVD settings and the interplay between IF which is the main subject of this work, and the SVD technique with respect to the overall cleaning performance.

The default settings used in LHC operation in 2018 were originally defined from Relativistic Heavy Ion Collider (RHIC) turn-by-turn data statistical analysis [14]. Table I summarizes the SVD settings and numerical thresholds used as defaults in 2018. Peak thresholds choose the minimum peak-to-peak signal and maximum value of the signal to be considered physical. Recording exact 0 in one or several turns indicates a bad BPM since this value is used to replace unphysical BPM readings. The betatron tune obtained as the main frequency of the BPM signal can serve as another indicator for faulty BPMs. The threshold for the tune deviation from the average value over the entire set of BPMs defines the limit of tune variability for correct BPM signals. The operational results of the preexisting cleaning tools presented in the following section are obtained using default settings.

TABLE I. Default SVD settings and signal cuts used in 2018.

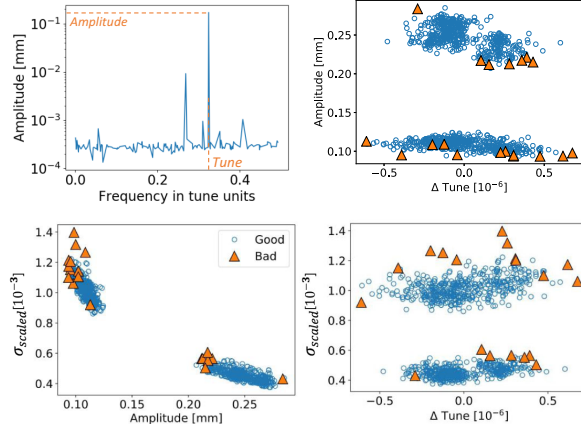| SVD settings		Signal-based cuts	
SVD cut	0.925	Min peak-to-peak	10^2 mm
SVD modes	12	Max peak value	20 mm
		Max tune deviation	10^{-5}
		Unphysical signal	Exact 0

FIG. 2.   The upper left plot represents, for a single BPM, the tune and amplitude from harmonic analysis that are used together with noise to amplitude ratio σ_{scaled} as input features for the IF algorithm. 2D projections of the feature space illustrate the detection of bad BPMs.

II. OPERATIONAL RESULTS OF ISOLATION FOREST APPLICATION

In the LHC there are 523 BPMs per plane and per beam. The analysis of the results obtained by traditional cleaning tools from the measurements before 2018 has shown that around 10% of all BPMs are identified as faulty using these tools [7]. Due to experience with the systematical observation of few nonphysical outliers in the reconstructed optics functions, we assume that only a small fraction of bad BPMs is remaining in the data, so most of the bad BPMs are eliminated by existing techniques.

As discussed in the Introduction, IF requires the contamination (expected fraction of anomalous samples) as input of the algorithm. During beam commissioning and Machine Development sessions (MD) in 2018 the contamination was set to 1% in the arcs and 2.5% in the interaction regions (IRs). A higher contamination rate in the IRs is assumed based on the analysis of previous measurements in [7]. The separation into IRs and arc BPMs is needed due to the fact that BPM hardware installed in IRs is

different from the arc sections [15], therefore they have to be treated separately.

The target is to apply the new cleaning method on the properties of BPM readings computed by a special harmonic analysis [16,17]. The parameters that are considered as significant for bad BPMs identification are the betatron tune, the amplitude obtained from the FFT scaled by factor 2 with respect to the oscillation amplitude, A, and the noise to amplitude ratio, σ_{scaled}. The latter is defined as

$$\sigma_{\text{scaled}} = \frac{\text{std}(x_{\text{raw}} - x_{\text{clean}})}{A\sqrt{N}}, \qquad (1)$$

where x_{raw} and x_{clean} stand for the BPM data before and after applying the SVD cleaning, respectively. The standard deviation of the difference of these two signals, $\text{std}(x_{\text{raw}} - x_{\text{clean}})$, represents the BPM noise. N is the number of turns. Figure 2 illustrates detection of faulty signals by IF, using the described features as input data. During commissioning and MDs in 2018, the IF algorithm was used complementary to the preexisting techniques. During the evaluation phase of the new introduced technique, the optics was first computed using the data cleaned with traditional tools only and then repeated with turn-by-turned data additionally cleaned with IF. This allowed to observe the positive impact of the new method on the quality of optics analysis. Figure 3 shows a comparison between the β beating reconstructed from the measurements cleaned with the traditional tools only and the measurements additionally cleaned with IF. It demonstrates that most of the observed outliers in the optics obtained from SVD-cleaned data can be prevented by using IF. The removal of BPMs at the locations where no spike has been observed did not cause significant data loss.

In order to conclude on the effectiveness of the new method for online optics analysis, we collected a summary of the measurements where IF has been used. Figure 4 shows the summary of cleaning results on several measurements in 2018 that are listed in Table II. The BPMs at the locations of outliers usually need to be manually analyzed and removed from harmonic analysis of turn-by-turn data before recomputing the optics. This procedure requires additional human intervention and still does not
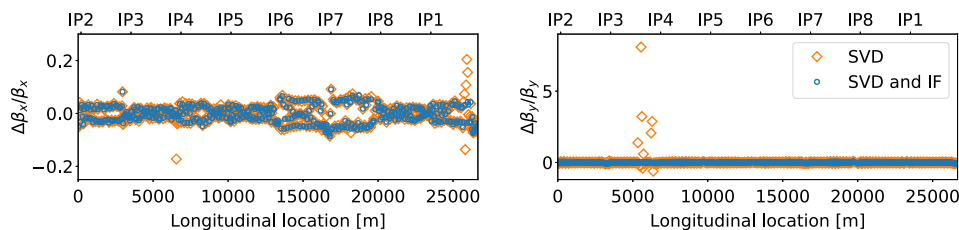

FIG. 3.   Comparison between β beating measured from SVD-cleaned data and the data additionally cleaned with IF. The data is obtained during ion commissioning in 2018. Without applying IF nonphysical spikes appear in β beating reconstructed from the BPM signal.
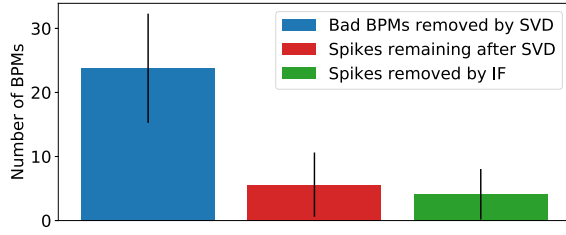
FIG. 4. Summary on cleaning results based on the number of outliers (spikes) per plane per beam appearing in computed β beating and phase advance averaged over ten measurements listed in Table II.

guarantee that manually removed BPMs are correctly detected and erroneous measurements do not appear in the recomputed optics. The statistics on operations data shows that most of the spikes remaining after SVD and thresholds-based cleaning could be successfully removed by IF, without requiring any manual cleaning.

III. SIMULATING FAULTY BPM SIGNAL IDENTIFICATION

Since the definition of erroneous values is exclusively based on the observation of computed optics functions, it is not possible to conclude about the exact amount of actual faulty BPMs which are removed and the number of good BPMs which are wrongly recognized as faulty. The knowledge about actual malfunctioning BPMs currently present in the machine is not available during the operation. Therefore, it has to be studied on simulated measurements with artificially introduced BPM faults, such that cleaning results can be verified against the ground truth. This avoids using the number of manually identified outliers remaining

TABLE II. List of measurements where IF has been used during online data analysis in 2018. Beam 1 appears more often in the statistics just due to the human intervention in the optics analysis and subjective decision about the need for additional cleaning with IF.

Date	Optics, β^*	Beam	Corrected
29.05.2018	Commissioning, 90 m	1	No
29.05.2018	Commissioning, 90 m	1	Yes
29.05.2018	Commissioning, 90 m	2	No
12.06.2018	Flat, 60/15 cm	1	No
12.06.2018	Flat, 60/15 cm	1	Yes
22.08.2018	Ion commissioning, 1 m	1	Yes
08.10.2018	Ion commissioning, 50 cm	1	No
08.10.2018	ion commissioning, 50 cm	1	Yes
08.10.2018	Ion commissioning, 50 cm	2	No
08.10.2018	Ion commissioning, 50 cm	2	Yes

in the computed optics functions as a figure of merit for the validation of applied cleaning methods.

A. Model of faulty BPMs

First, turn-by-turn BPM signals are generated for 6600 turns using ion optics model with $\beta^* = 50$ cm in IP1, 2 and 5 using MAD-X and a dedicated PYTHON script. Every BPM is given 0.1 mm Gaussian noise. In the second step, the signal of some randomly chosen BPMs is artificially perturbed—these BPMs have to be identified as bad. In real measurements, the reasons for the appearance of faulty signal are unknown, but there are specific artifacts which are known to be related to faulty BPMs. Therefore, we use these known properties to model a presence of faulty BPMs in the LHC. It has to be noted that these artifacts do not describe all possible failures, however in order to verify the effectiveness of cleaning methods, it should be sufficient to assign the known failures to a realistic number of BPMs in the simulations according to the number of observed anomalies in the measurements data.

The following failure modes are used to introduce BPM faults: (i) Gaussian noise added to the signal is 0.3 mm, 3 times higher compared to good BPMs; (ii) signal in one turn is replaced by a random value in range $[-20, 20]$ mm, such that produced local spike is smaller than the threshold for the maximal absolute peak value used for simple cuts; (iii) tune computed from the signal deviates by 10^{-5} from the rest of the BPMs; (iv) all described failures are present.

The examples of perturbed turn-by-turn signal produced by the introduction of the listed failure modes are shown in Fig. 5. In addition to the described failure modes, we also introduce BPMs with flat zero signal in all turns and zero signal in ten randomly chosen turns. These two failure types are trivial to detect and hence not relevant for the method verification. Nevertheless, we need to include them to produce more realistic turn-by-turn data simulations. Considering the number of bad BPMs found by traditional
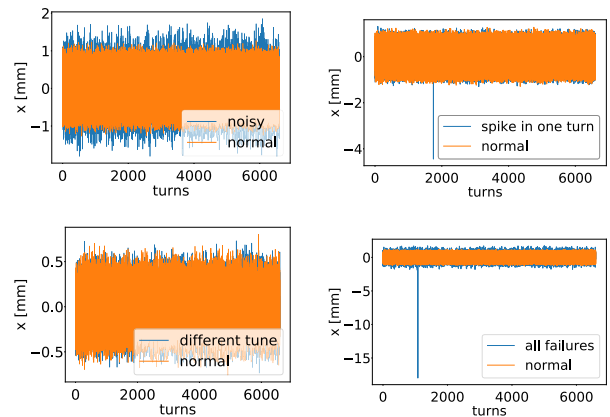

FIG. 5. Simulated turn-by-turn data of the bad BPMs compared to normal unperturbed signal over 6600 turns.

tools and remaining spikes before applying IF ($\approx$30 per plane per beam as shown in Fig. 4), the ratio of bad BPMs over the total number of BPMs per plane per beam in 2018 was $\approx$5.5%. Hence, we perturb 5.5% of BPMs in original simulated turn-by-turn data. All failures, apart from flat zero signal, are equally distributed over generated BPM signal with five occurrences and flat zero signal is simulated at two BPMs, producing 27 bad BPMs per plane in total.

B. Isolation Forest results

First, we perform harmonic analysis on the generated turn-by-turn data using the traditional cleaning techniques without changing the default settings. Knowing the actual bad BPMs and their faults, the unsupervised method can be evaluated in the combination with traditional tools, applying IF on harmonic analysis of the SVD cleaned BPMs data.

Statistical analysis of operational data presented in II shows that around 15 bad BPMs remain after using traditional cleaning tools and 12 bad BPMs are removed, thus the contamination should be set to $15/(523-12) \approx 0.029$. To study the influence of the contamination parameter on the optics computation, we use simulations to run IF multiple times increasing the contamination number from 0 to 0.15 stepwise. Figure 6 illustrates the trade-off between eliminating bad BPMs and removing good BPMs as a side effect. In order to find an optimal contamination factor, we have to define an acceptable maximum number of missing good BPMs which does not cause negative effects on optics analysis. In Fig. 6 we observe a sharp increase of the number of detected bad BPMs and slow increase of removed good BPMs. After the contamination factor reaches 0.02, the trend changes and the rise of the number of removed good BPMs becomes steeper than the increase of detected bad BPMs. After the contamination number reaches 0.04, there is no significant increase in the number of removed bad BPMs anymore. Based on the obtained results we conclude that the optimal contamination factor lies between 0.02 and 0.04 as expected, if the data is

TABLE III. Number of faulty BPMs detected by SVD together with combination of thresholds compared to number of BPMs detected after additionally applying IF on cleaned data. The results are obtained using SVD default cut value 0.925 and IF contamination factor set to 0.02. The numbers are averaged over 20 simulated measurements. The basic zero cut failures are not included into comparison.

Failure	SVD	IF	Not detected
Higher noise	0	3 ± 1	1 ± 1
Random values	1 ± 1	2 ± 0	1 ± 1
Tune deviation	0	2 ± 1	2 ± 1
All failures present	1 ± 1	3 ± 1	0

previously cleaned by SVD using the default parameter settings. Considering the different failure modes, Table III shows that under the default settings, IF complements the traditional techniques exactly in the cases where they are insufficient.

C. Exploring SVD settings

The simulated data with artificially introduced bad BPMs has also been used to explore if the change of SVD cut threshold can improve the detection of faulty signals. The change in cleaning results with respect to SVD cut values presented in Fig. 7 shows that the optimal SVD cut threshold range is [0.3, 0.5], noting equal results for 0.4 and 0.3 values. Bigger values lead to an increase in the number of remaining bad BPMs, which has been demonstrated not only in simulations, but also in LHC measurements in relation to the new failure mode described in V. Lowering the threshold to values smaller than 0.3 leads to an increase in the number of good BPMs wrongly identified as bad. The change of SVD cut from its default value 0.925 requires also the change of the IF contamination factor. Since less faulty BPMs will appear in the data, lower contamination factor should be needed. So, we repeat the scan of IF contamination factor on the data cleaned with SVD using the lowest optimal cut value 0.3 in order to compare it with IF performance under the default SVD
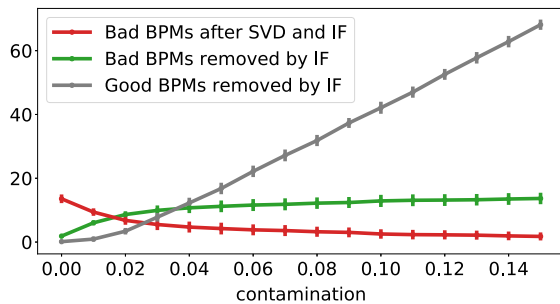

FIG. 6. Adjustment of contamination factor of the IF algorithm. The target is to keep the number of bad BPMs remaining after applying SVD and IF low, while avoiding a significant number of good BPMs to be wrongly identified as bad.
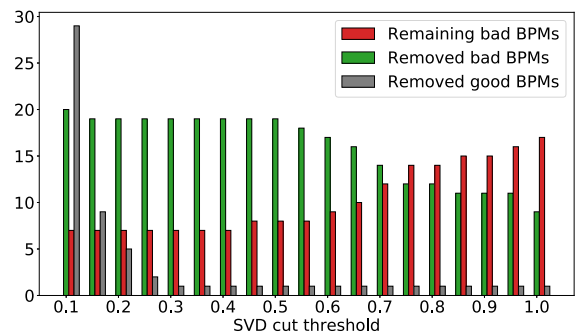

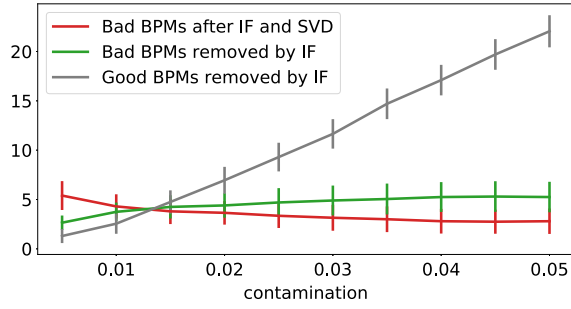FIG. 7. Number of removed and remaining bad BPMs after applying SVD depending on SVD cut value.

FIG. 8. Adjustment of contamination factor of the IF algorithm after changing the SVD cut to its lowest optimal value 0.3.

FIG. 9. Averaged results of faulty BPMs detection on 20 simulated measurements using IF contamination factor 0.01 and SVD cut 0.3.

settings. Figure 8 demonstrates how the cleaning result changes with the increase of contamination factor. In case the SVD cut is set to 0.3, the IF contamination factor should be within the range [0.01, 0.015] since further increase results in a bigger number of good BPMs wrongly removed by the algorithm. We repeat the study on the detection of particular failures using the obtained optimal settings for both, SVD and IF, namely SVD cut of 0.3 and contamination of 0.01. The averaged results for each of the introduced failure modes are summarized in Table IV. The final result considering the full set of failures is presented in Fig. 9.

The presented result shows that the simulations qualitatively reproduce experiment observations, such that we could obtain an understanding of unsupervised learning as a method for faulty BPMs identification. We also investigated the interplay between previously available cleaning techniques and the introduced IF algorithm, demonstrating the advantage of applying IF prior to optics computation, instead of using SVD only to identify BPM faults. The number of bad BPMs remaining in the data can be reduced by factor 2 to less than 1% performing anomaly detection with IF on SVD-cleaned data. Both methods, SVD and IF, can be used complementary noting the importance of adapting the thresholds accordingly.

D. Comparison to clustering

Prior to the integration of the IF algorithm into optics measurements, we have considered several clustering

TABLE IV. Cleaning results applying the SVD cut optimal value 0.3 and adjusting the IF contamination factor to 0.01. The results differ from the cleaning using the default setting for all introduced failure modes.

Failure	SVD	IF	Not detected
Higher noise	5	0	0
Random values	2 ± 1	1 ± 1	2 ± 1
Tune deviation	0	3 ± 1	2 ± 1
All failures present	5	0	0
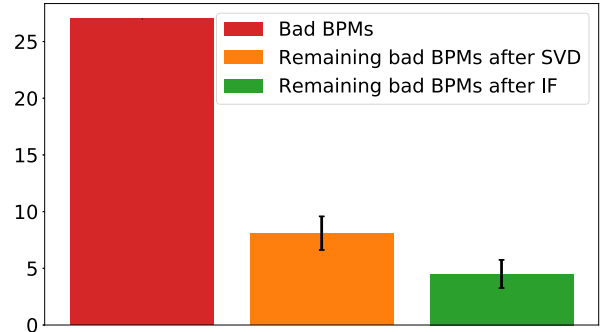
techniques as possible solutions to improve the cleaning results. These techniques have been tested together with IF on the harmonic analysis of LHC turn-by-turn measurements obtained in the past. One of the investigated approaches is a density-based DBSCAN algorithm [18] which views clusters as areas of high density separated by areas of low density. The method finds core points which build a cluster center, assigns neighboring points to this cluster and considers as anomalies the points which do not belong to any cluster. A core point is defined by the minimum number of points within a distance. Hence, there are the following parameters to be defined: minimum number of samples in the neighborhood of a core point, the distance to the neighbors and the metric to be used to compute the distance. The results of applying DBSCAN to bad BPM detection demonstrated improvements on data cleaning [19], however a significant amount of outliers remained present in the measured optics functions. Another technique which has been applied to the LHC turn-by-turn data is the local outlier factor (LOF). This factor indicates the local deviation of density of a given sample with respect to its nearest neighbors [20]. LOF measures how isolated an object is with respect to the surrounding neighborhood, which is very similar to the IF algorithm. However, apart from the contamination rate, LOF requires the number of neighboring data points and the definition of a metric to compute the distance between the points in order to find the nearest neighbors. Since the structure of the measurements data can vary significantly depending on the BPM location and machine settings, a general valid definition of the distance and number of neighbors to build a cluster becomes problematic.

In order to examine the performance and suitability of each method for faulty BPM detection, we carry out the previously described simulation procedure. The parameters of clustering techniques have been defined empirically during the application on LHC data. The distance between the points to define a cluster in DBSCAN is set to 0.7, LOF contamination is 0.05. Both methods use the Euclidean
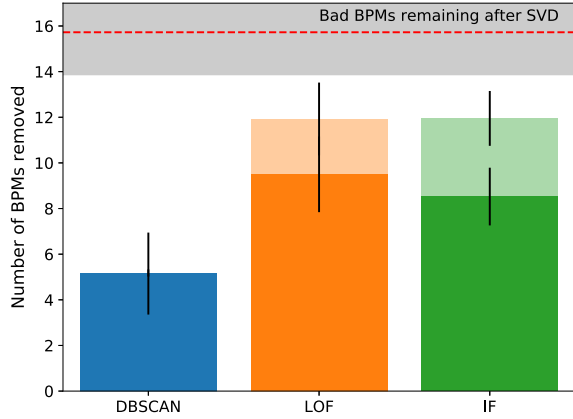
FIG. 10. The comparison is carried out on 20 simulations for each plane, the results are averaged. Each bar represents the number of BPMs removed by the method. The dark fraction corresponds to the number of removed BPMs that are actually bad.

metric to compute the distance and the minimum of 70 neighbors to build a cluster. SVD and thresholds based cleaning have been applied prior to clustering using the default values described in Sec. I B. Figure 10 summarizes the obtained result. The comparison shows nearly identical performance of LOF and IF algorithms on simulated BPM signal. Applying different methods on LHC measurements, we observe that the optics reconstructed from the data cleaned with IF contains fewer unphysical outliers compared to the other two methods. Moreover, due to a smaller number of settings, IF allows simpler tuning of the cleaning algorithm and more general application. Therefore, it is preferred as an alternative cleaning tool for faulty BPMs detection at the LHC.

IV. FAULTY BPM DETECTION IN THE PRESENCE OF LOCAL COUPLING

Betatron coupling drives the appearance of oscillations in the horizontal plane with the vertical tune and vice versa. Strong local coupling sources must be corrected to avoid luminosity loss [21] or its propagation to the rest of the machine, therefore it has to be ensured that cleaning tools do not have a negative impact on the computation of local coupling. The coupling is measured in terms of its resonance driving terms (RDTs), f_{1001} and f_{1010} [22,23]. Since these RDTs are calculated from the spectrum of cleaned BPM signal, the settings of the cleaning tools affect the coupling computation. As the local coupling information is contained in secondary lines of only few BPM spectra, using a small number of modes in SVD cleaning causes coupling information to be discarded as noise. Considering IF cleaning, the BPMs whose signal contains the information about local coupling might be removed completely based on the difference from the rest of BPMs. In the following we present the ability of IF to distinguish

local coupling related signal from faulty BPMs along with the optimization of SVD mode number with respect to local coupling. To simulate faulty BPM detection in the presence of local coupling, we first perform MAD-X tracking using ion optics from 2018 with $\beta^* = 50$ cm. In order to introduce a local coupling bump, the integrated field strength of the skew quadrupoles around IP2 are changed by ± 0.001 m^{-1}. We also introduce a small global β beating (1%) in order to get more realistic optics from the simulated signal. Produced tracking simulations are then converted into turn-by-turn measurements in order to introduce BPM faults as described in III. In this case we ensure that no fault is assigned to the BPMs at the location of coupling bump (BPMSW.1R2.B1, BPMSW.1L2.B1). In the following, we address the influence of SVD mode setting on coupling computation and keep SVD cut and IF at default values previously used in operation. The default SVD mode number used for the LHC in 2018 is 12. However, when reconstructing f_{1001} and f_{1010} from SVD-cleaned data, the presence of local coupling can be observed only by increasing SVD mode to 18 as shown in Fig. 11. To reproduce the actual expected value of simulated coupling around IP2 at least 22 SVD modes are needed. Further increase of SVD modes still produces correct f_{1001} and f_{1010} values, however it affects the phase computation negatively since the measurement becomes more noisy. Therefore, increasing the default value of SVD modes up to 22 could help to gain more information about local coupling without significant increase of noise in the computed phase. The scan of this parameter with respect to the number of removed BPMs performed on simulations is shown in Fig. 12. Starting from 16 modes, the number of identified bad BPMs increases, while keeping the number of wrongly removed good BPMs constant. Conclusively, the number of SVD modes used in turn-by-turn data analysis should be increased in order to achieve more reliable coupling computation without causing negative
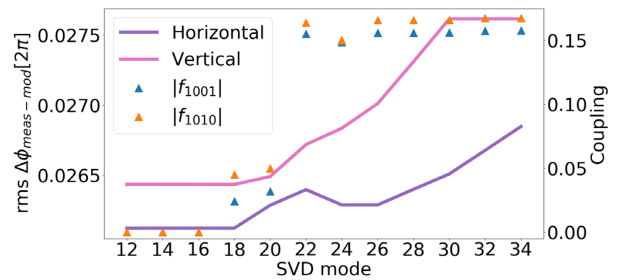

FIG. 11. Trade-off between correct computation of the expected coupling at the location of introduced skew quadrupole field error and increase of the noise in the measured phase advance errors $\Delta\phi_{\text{meas-mod}}$ defined as rms between measured and model phase advances. The right scale shows the computed values of f_{1001} and f_{1010} depending on the number of SVD modes used in SVD cleaning.
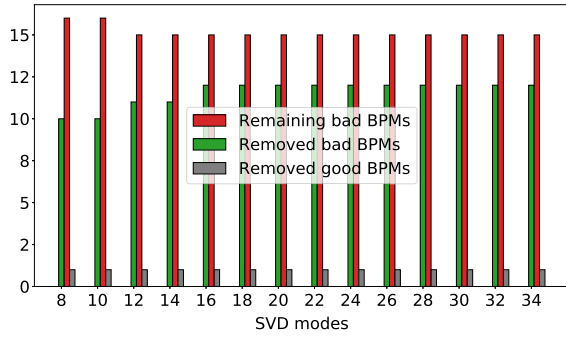
FIG. 12.   The relation between BPM cleaning results and the change of the number of SVD modes used in analysis.
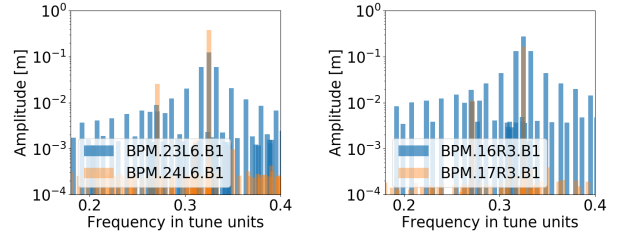

FIG. 14.   Examples of the faulty BPM signal spectra in the vertical plane (blue) compared to the spectra of normal functioning BPMs (orange). The measurement is performed on beam 1, $\beta^* = 65$ cm.

effects on signal noise cleaning and faulty BPMs detection. Figure 13 demonstrates the result of additional cleaning with IF compared to the optics obtained from data cleaned with traditional techniques only. The local coupling can be observed after application of IF since BPMSW.1R2.B1 and BPMSW.1L2.B1 are correctly identified as nonanomalies. At the same time, the outliers in β beating produced by the simulated faulty BPM signal are eliminated.

V. DETECTING UNKNOWN FAILURE MODE IN EXPERIMENTAL DATA

It has been observed in operational measurements that some BPMs repetitively caused erroneous optics calculations. These BPMs (23L6.B1, 16R3.B1, 22R8.B1, 15R8.B1) have been manually analyzed in regards of betatron tune domain and phase shifts aiming to identify signal properties which can serve as indicators for fault detection [24,25]. In the past, such indicators could not be found, and these BPMs were simply removed from the

data prior to data processing. Recently, these four BPMs together with three further BPMs (25L5.B1, 10R6.B1, YB.4R8.B1) [26] are found to exhibit an identical pattern in the spectra with sidebands around the tune frequency line [27]. Figure 14 shows examples of the signal related to the described pattern. The reason of the recently discovered failure mode still remains to be understood. We demonstrate that BPMs related to this failure mode can be identified using the combination of preexisting cleaning tools and IF, without observing the BPM signal spectra *a priori*. The seven listed BPMs can be detected using either specific settings of SVD clean or by applying IF in addition to SVD clean with arbitrary settings. This is crucial for successful automatic data cleaning since even a careful manual inspection did not guarantee full elimination of faulty BPMs in the computed optics functions.

In total there are seven BPMs found to have the described spectra pattern, however the presence of each of these BPMs causes multiple outliers in the computed optics as shown in Fig. 15. SVD is incapable of identifying this failure, unless reducing SVD cut to 0.6 which, however, cleans only 20% of affected BPMs. Removing
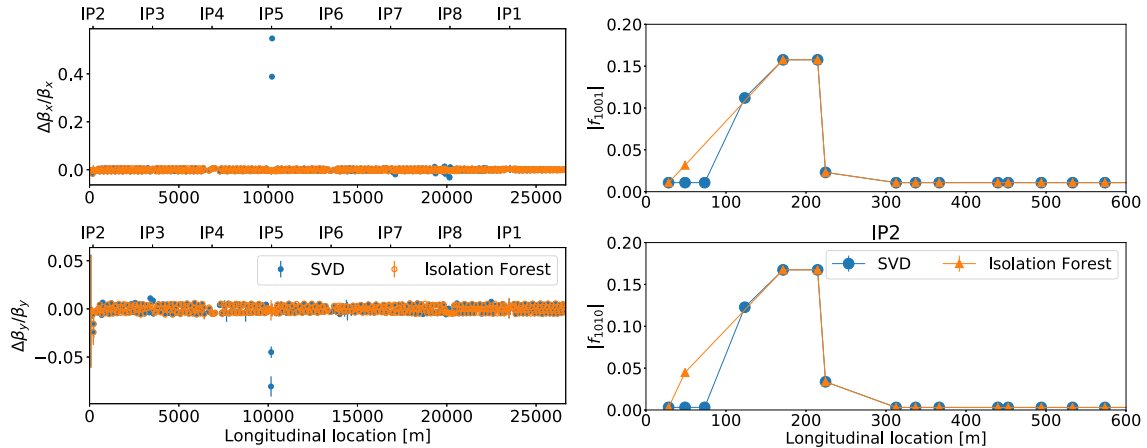

FIG. 13.   Application of IF on SVD cleaned data reduced the number of outliers in beta beating (left) so the simulated bad BPMs are detected correctly. At the same time the introduced coupling bump in IP2 is still present (right), showing that the signal at the location of the coupling bump (BPMSW.1R2.B1, BPMSW.1L2.B1) where f_{1001} and f_{1010} reach their maximum values is not classified as anomaly as expected.
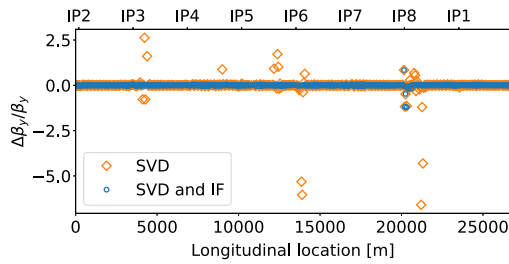
FIG. 15. In case the optics is computed from the data additionally cleaned by IF, the appearance of nearly all unphysical outliers is prevented.

all seven BPMs is possible only by using SVD cut 0.3 which agrees with the optimal range for this setting obtained from simulations described in the previous section. Independently of SVD settings, applying IF helps to clean 90% of these BPMs. The presented showcase demonstrates the advantage of combining traditional cleaning tools with unsupervised learning and the ability of presented techniques, given the optimal settings obtained from extensive simulation studies, to identify BPM faults without providing manually identified rules and to detect even *a priori* unseen failure pattern.

VI. CONCLUSION

IF successfully detected faulty BPMs during LHC operation and MDs in 2018 where this was previously done via tedious human intervention. Extensive studies and simulations presented here show that indeed the previously existing techniques cannot perform as efficiently as in combination with IF. Also, we demonstrated that measurement of important observables such as local coupling requires increasing the number of SVD modes above the default value of 12. IF does not influence negatively the local coupling measurement.

Considering a more general application, the developed cleaning approach can be potentially used in other accelerators, performing IF on a given set of signal properties suitable to a particular accelerator type. The choice of contamination rate has to be carefully estimated according to the available BPM infrastructure, by the means of statistical analysis of historical data or simulating BPM failures as demonstrated in this work.

The improved understanding of the combination of SVD cleaning and IF algorithm obtained in the presented study will further benefit future measurements during run III of the LHC.

ACKNOWLEDGMENTS

[1] T. Persson *et al.*, LHC optics commissioning: A journey towards 1% optics control, Phys. Rev. Accel. Beams **20,** 061002 (2017).

[2] J. C. de Portugal, R. Tomás, and M. Hofer, New local optics measurements and correction techniques for the LHC and its luminosity upgrade, Phys. Rev. Accel. Beams **23,** 041001 (2020).

[3] A. G.-T. Valdivieso and R. Tomás, Optics-measurement-based beam position monitor calibrations in the LHC insertion regions, Phys. Rev. Accel. Beams **23,** 042801 (2020).

[4] A. Langner, G. Benedetti, M. Carlà, U. Iriso, Z. Martí, J. C. de Portugal, and R. Tomás, Utilizing the N beam position monitor method for turn-by-turn optics measurements, Phys. Rev. Accel. Beams **19,** 092803 (2016).

[5] A. Wegscheider, A. Langner, R. Tomás, and A. Franchi, Analytical N beam position monitor method, Phys. Rev. Accel. Beams **20,** 111002 (2017).

[6] S. P. Lloyd, Least squares quantization in pcm, IEEE Trans. Inf. Theory **28,** 129 (1982).

[7] E. Fol, Evaluation of machine learning methods for LHC optics measurements and corrections software, Master's thesis, University of Applied Sciences, Karlsruhe, 2017.

[8] E. Fol, F. S. Carlier, J. M. Coello de Portugal, A. Garcia-Tabares, and R. Tomás, Machine learning methods for optics measurements and corrections at LHC, in *Proceedings of the 9th International Particle Accelerator Conference (IPAC'18), Vancouver, Canada, 2018*, pp. 1967–1970, https://doi.org/10.18429/JACoW-IPAC2018-WEPAF062.

[9] F. T. Liu, K. M. Ting, and Z.-H. Zhou, Isolation Forest, in *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM'08), Pisa, Italy* (IEEE Computer Society, 2008), pp. 413–422.

[10] T. G. Dietterich, Ensemble methods in machine learning, in *Proceedings of the First International Workshop on Multiple Classifier Systems (MCS'00)* (Springer-Verlag, Berlin, Heidelberg, 2000), pp. 1–15.

[11] F. Pedregosa *et al.*, Scikit-learn: Machine learning in PYTHON, J. Mach. Learning Res. **12,** 2825 (2011).

[12] X. Pang and S. Y. Lee, Independent component analysis for beam measurements, J. Appl. Phys. **106,** 074902 (2009).

[13] X. Shen, S. Y. Lee, and M. Bai, Uncovering beam position monitor noise at the relativistic heavy ion collider, Phys. Rev. ST Accel. Beams **18,** 014002 (2015).

[14] R. Calaga and R. Tomás, Statistical analysis of RHIC beam position monitors performance, Phys. Rev. ST Accel. Beams **7,** 042801 (2004).

[15] M. Gasior, J. Olexa, and R. Steinhagen, A high-resolution diode-based orbit measurement system—Prototype results from the LHC, CERN Technical Report No. CERN-ACC-NOTE-2017-0020, 2011.

[16] R. Bartolini and F. Schmidt, SUSSIX: A computer code for frequency analysis of nonlinear betatron motion, Technical Report, 1998.

[17] L. Malina *et al.*, HARPY: A fast, simple and accurate harmonic analysis with error propagation (to be published).

[18] M. Ester, H. Kriegel, J. Sander, and X. Xu, A density-based algorithm for discovering clusters in large spatial databases

with noise, in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD'96), Portland, OR, 1996*, pp. 226–231.

[19] E. Fol, *ICFA Beam Dynamics Mini-Workshop: Machine Learning Applications for Particle Accelerators*, 2018, https://indico.fnal.gov/event/16327/.

[20] M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, LOF: Identifying density-based local outliers, in *Proceeding of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, 2000*, pp. 93–104.

[21] R. Tomás *et al.*, LHC run 2 optics commissioning experience in view of HL-LHC, in *Proceedings of the 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 2019*, pp. 508–511, https://doi.org/10.18429/JACoW-IPAC2019-MOPMP033.

[22] T. Persson, Beam-based error identification and correction methods for particle accelerators, Ph.D. thesis, Chalmers University of Technology, 2014.

[23] T. Persson and R. Tomás, Improved control of the betatron coupling in the Large Hadron Collider, Phys. Rev. Accel. Beams **17**, 051004 (2014).

[24] A. Garcia-Tabares, Beta from amplitude: Ballistic and 2.5 km review, OMC Team meeting, https://indico.cern.ch/event/549409/.

[25] A. Garcia-Tabares, Bad BPMs, OMC Team meeting, https://indico.cern.ch/event/579332/.

[26] L. Malina, Noise and stabilities, OMC Team meeting, https://indico.cern.ch/event/859128/.

[27] D. W. Wolf, Analysis of tune modulations in the LHC, Bachelor's thesis, Karlsruhe Institute of Technology (KIT), 2018.