Probabilistic graphical models for diagnosing defectivity patterns

Leonardo Barbini[1], Peter Kruizinga[2], Micha Lipplaa[3], and Álvaro Piedrafita[4]

[1,3,4] *TNO-ESI, Eindhoven, 5656 AE, The Netherlands*
leonardo.barbini@tno.nl
micha.lipplaa@tno.nl
alvaro.piedrafitapostigo@tno.nl

[2] *Canon Production Printing, Venlo, 5914 HH, The Netherlands*
peter.kruizinga@cpp.canon

ABSTRACT

In the high-tech sector, diagnosing performance issues often involves analysing a variety of defectivity patterns on products. High-tech systems perform numerous processes – e.g. product handling, light projection, jetted ink application, and thermal treatments – all of which affect the quality of the product itself. Their potential malfunctioning can contribute to defects with characteristic patterns. Often there is not a one-to-one mapping between root causes of these malfunctions and the resulting observable defectivity patterns. Consequently, identifying the root cause of these patterns is a challenging and an intrinsically probabilistic task. This paper proposes a framework to relate these patterns to the underlying root causes and employs Probabilistic Graphical Models (PGM) to reason about these relations. We find that PGMs ability to contain arbitrary graph topologies and jointly reason across all root causes empowers the modeller to adapt the models to the system at hand and include domain specific knowledge that would be hard to account for using more data-driven approaches. When provided with data from an operational system in the field, the PGM identifies the underlying root causes of product quality issues. We demonstrate the methodology with a real use case from the production printing industry.

1. INTRODUCTION

The overall output quality in modern high-tech systems is determined by the joint performance of several system components across various integrated processes. Directly monitoring each individual process within a high-tech system is often operationally infeasible. Instead, it is typically only possible to observe their overall effect by measuring the resulting quality at multiple locations on the product. Product quality below specification is measured as a distribution of defects on the product's surface, i.e. a defectivity pattern. Inferring the root causes from the available indirect and aggregated set of measurements, i.e. observations, results in a significant challenge. For example, consider the semiconductor lithography process. The precise alignment of different layers on each die is a major contributor to product quality. If a wafer shows misaligned regions after exposure, it is crucial to determine whether the issue originated from the wafer itself or the exposure process.

This paper introduces a framework to assist engineers in diagnosing product quality issues. Central to our framework is the observation that different root causes acting on the system's processes, components and product will induce characteristic correlations among measurements. We capture these correlations by defining different distance metrics between the localized observations. These metrics are used to reason about the root causes underlying the abnormal observations.

For example, in the exposure cycle of a lithography system the same pattern is repeatedly projected onto dies on a wafer following a particular path, i.e., exposure sequence. In this case our methodology establishes both a spatial distance metric and an exposure sequence metric between any two dies. As a result, two abnormal alignment measurements could be correlated based on their spatial proximity on the wafer or on their exposure sequence proximity. If the first is the case, we reason that the problem originates in the wafer, while the second indicates a problem with the exposure process, e.g. the robot that moves the wafer under the light source.

A different example of a high-tech system that can be modelled in a similar way is the printing process of an industrial printer. There, ink-jetting nozzles are arrayed in a 2D grid and follow a predefined jetting sequence. We will develop this example further in Section 3.

In this paper we employ Probabilistic Graphical Models (PGM) for modelling and reasoning over the system's observations, structure and hidden states. The application of PGMs in diagnostics, specifically in the medical domain, stems from a long line of research (Pearl, 1988), (Darwiche, 2009), with specific applications in high-tech diagnostics (Barbini, Bratosin, & Nägele, 2021; van Gerwen, Barbini, Borth, & Passmann, 2024). In the PGM framework, the diagnostic inference task is framed as a problem of statistical inference, specifically the computation of a diagnosis as the posterior probability given observed evidence. In this work, we use Conditional Markov Random Fields (CRF) to model root causes as the hidden states of the system's components, conditioned on the observed qualities measured at multiple points on the product.

The paper is organized as follows. Section 1.1 gives an overview of the relevant approaches to the diagnosis of product quality problems. Section 2 gives details on how we specify and use CRF for the diagnostic inference task. In Section 3, the proposed methodology is applied to a use case from the industrial production printing domain. Finally, Section 5 summarizes the main findings, the current limitations and outlines future research directions.

1.1. Current Approaches

The current approaches to diagnosing defectivity patterns in high-tech production systems primarily use data-driven methods, especially machine learning and deep learning (Bai et al., 2025; Henn, Zhou, & Barnes, 2019), alongside expert-based approaches. A well-known challenge for purely data-driven methods, particularly in performance diagnostics, is acquiring a sufficiently large dataset of labelled field data. This is because during both design and early usage, the full range of potential performance problems and their underlying root causes are often unknown, emerging only during extensive real-world operation. Consequently, it is difficult to systematically label data linking the observed defectivity patterns to the root causes. Furthermore, reliable ground truth is frequently unavailable or requires costly designed experiments, for example by inducing a specific failure and measuring its effects. As a result, a common practice is to rely on expert interpretation of complex dashboards displaying measured data in various visualizations, along with estimated features, leaving the identification of root causes to human inference.

Another key limitation of currently used methods is their tendency to analyse individual defect contributors in isolation, rather than from a system-level perspective. This typically results in estimating a single system process or component's performance based on aggregated measurements. However, in complex systems, all processes and components influence these measurements. While effective for quality issues traceable to a single source, this isolated analysis fails when defects require simultaneous reasoning across multiple interacting factors. The challenges of the current approaches described above highlight the ongoing need for approaches that can incorporate domain knowledge across multiple aspects of a system, and can reason effectively even without large, perfectly labelled datasets.

2. METHODOLOGY

Our methodology aims to classify observations taken over time and at different positions of a product into a discrete set of possible root causes. The proposed approach is based on two key assumptions. First, we assume that observations of the product are discrete (e.g., OK, NOK) and indirectly reflect the hidden state of the hardware, processes, and the product itself. Second, we assume to have knowledge of failure behaviour, i.e. root causes, and of the system design, such as component layout or process geometry. This knowledge is used to determine the distance metrics to which the considered root causes adhere.

This classification task is reminiscent of the problem of image segmentation, where the goal is to label every pixel in a 2D image into a set of classes. Inspired by the literature on image segmentation (Kato & Zerubia, 2012), we use probabilistic graphical models (PGMs) in the form of Conditional Random Fields to model our system, and reason about it, in probabilistic terms. Our assumptions contain all the necessary ingredients to frame our problem in the language of PGMs (Koller & Friedman, 2009). In contrast with the square grids used in the image segmentation literature, we will make use of CRFs with arbitrary topology.

Let us take as an example a system where the observations are arrayed on the product in a 2D grid but also connected along a 1D path. Both the lithography and industrial printing processes mentioned in Section 1 are of this type. The construction of the PGM proposed by our framework is outlined in Figure 1.

First, we model each localized observation O_i as an observed random variable taking discrete values $\{OK, NOK\}$. The classification of observation O_i is then modelled as another discrete random variable H_i, this time hidden, taking values in $\{c_1, \dots c_k\}$, the set of root causes, of which we assume to have prior knowledge. Based on this description, we construct a graph, see Figure 1-a, for which the random variables are the nodes and edges connect each O_i and H_i pair (bottom and top, respectively). Edges in the graph represent our belief that the two variables are directly correlated. Second, we add edges between different H_i's. These edges encode proximity in one of the design-informed distance metrics. We call the set of edges induced by each distance metric a representation of the observations. In Figure 1-b we add the edges induced by the grid-like 2D structure on the product, while in Figure 1-c we add the edges induced by the 1D path.

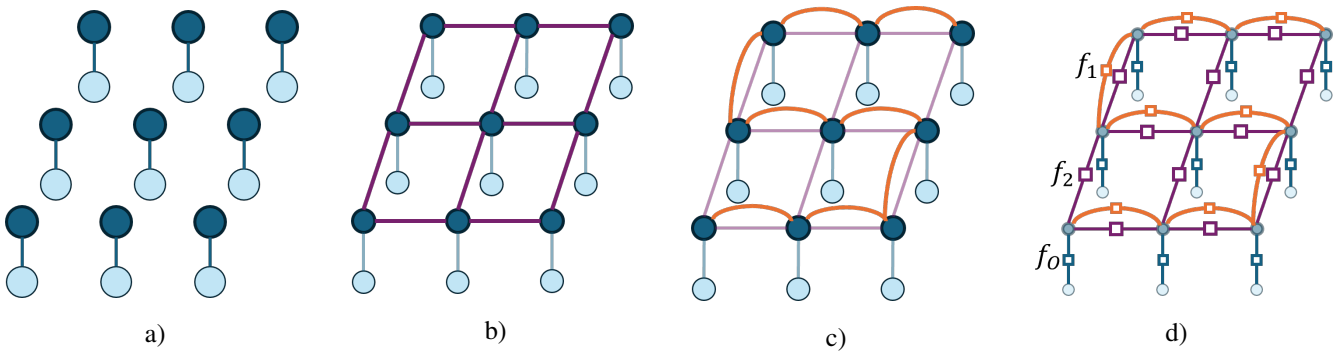


Figure 1. Construction of the CRF for a system with two representations.

To make a conditional random field, we place factors f_j on the edges between variables. These factors are numerical arrays quantifying the strength of correlation between the discrete states in the variables they connect. These factors are depicted in Figure 1-d with empty squares. In this work we use design knowledge and domain expertise to determine their values.

Equation (1) gives an example of how we use one-hot encodings to represent discrete random variables.

$$\begin{aligned} O_i \in \{OK, NOK\} &\cong \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, \\ H_i \in \{h = c_0, c_1, c_2\} &\cong \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} \end{aligned} \tag{1}$$

Specifically, it shows random variable O_i with two states $\{OK, NOK\}$, represented as orthogonal unit vectors in a 2-dimensional space; together with a random variable H_i with three states $\{h, c_1, c_2\}$ corresponding respectively to healthy state and root causes c_1 and c_2, represented as orthogonal unit vectors in a 3-dimensional space.

Let $f_O(i) = f_O$ in Eq. (2) be the factor connecting the variables O_i and H_i in Figure 1-d for any index i.

$$f_O = \begin{matrix} H_i = h & H_i = c_1 & H_i = c_2 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{matrix} O_i = OK \\ O_i = NOK \end{matrix} \tag{2}$$

Rows and columns of the factor correspond to the states of the variables O_i and H_i respectively. The entries of the array $f_O[j,k]$ indicate the strength of correlation we attribute to the combination of variable's states indicated by the j-th row and k-th column. For this factor, the first row encodes the fact that $O_i = OK$ perfectly correlates with $H_i = h$. The second row encodes the fact that a NOK observation could be caused by either c_1 or c_2. The zeros in the array encode impossible combinations. If need be, false positives and negatives could be incorporated into the factor. Similarly, $f_1(i) = f_1$ shown in Eq. (3), is the factor connecting any two variables H_i, H_{i+1} consecutive along the 1D path.

$$f_1 = \begin{matrix} H_{i+1} = h & H_{i+1} = c_1 & H_{i+1} = c_2 \end{matrix} \\ \begin{bmatrix} 1 & \epsilon & \epsilon \\ \epsilon & 2 & \epsilon \\ \epsilon & \epsilon & 1 \end{bmatrix} \begin{matrix} H_i = h \\ H_i = c_1 \\ H_i = c_2 \end{matrix} \tag{3}$$

The rows of f_1 correspond to the different states of H_i and the columns correspond to the different states of H_{i+1}. As in the case of f_O, $f_1[j,k]$ is the weight given to the combination $H_i = c_j$ and $H_{i+1} = c_k$, for $j,k = 0,1,2$. Let us assume root-cause c_1 displays larger correlation along the 1D path than cause c_2. Factor f_1 enforces this by making $f_1[1,1]$ larger than the other entries. For simplicity we have given a small weight of ϵ to all non-diagonal entries of f_1. Depending on the application, these entries can be given distinct values or be functions parametrized by, e.g., time, or the factor's location on the graph.

Let $\mathbf{O} = \{O_i\}, \mathbf{H} = \{H_i\}$ be the observed and hidden variables, and $\mathbf{V} = \mathbf{O} \cup \mathbf{H}$ be their union. Let $\mathbf{E} \subset \mathbf{V} \times \mathbf{V}$ be the set of edges between variables. Let $\mathbf{F} = \{f_j\}$ be the set of factors and $\Lambda : \mathbf{E} \to \mathbf{F}$ be a map assigning a factor to each element in $\mathbf{E}$. Then, the probabilistic graphical model $\mathcal{M} = (\mathbf{V}, \mathbf{E}, \mathbf{F}, \Lambda)$ defines a Conditional Random Field (CRF).

A CRF encodes the conditional probability distribution $p(\mathbf{H}|\mathbf{O})$, meaning that it encodes the joint posterior probability of the hidden variables H conditioned on the observed variables O. We then obtain the marginal posterior probability $p(H_i|\mathbf{O})$ for every hidden variable H_i using loopy belief propagation as our inference algorithms, although others exist, see Section 4. This marginal posterior probability conditioned on the observations becomes our classification, i.e. our diagnosis for the root cause of the observation.

The strength of the CRF, and more generically of PGMs, as a framework for modelling defectivity patterns resides in its ability to contain arbitrary graph topologies, and to perform

inference, i.e., reasoning, across all representations forming the graph. This empowers the modeller to adapt the model to the system at hand, including domain specific knowledge that would be hard to account for using more data-driven forms of image segmentation. Furthermore, we remark that our methodology is not limited to 1 or 2-dimensional lattices embedded in a graph. Indeed one could conceive, e.g., of fan-like representations of electrical interconnections on solar panels, or completely irregular structures like the overlapping high-voltage grid and communication networks forming a modern electricity grid.

2.1. Multi-Resolution Spatial and Temporal Analysis

The factors in Figure 1 are all binary in the sense that they connect two variables. It is also possible to use unary (affecting just one variable) or higher order factors. Unary factors are useful because they act as weights on variables. In Bayesian terms, we can think of them as priors. These can come from historical data or from previous computations on the same data. In this work we will use unary factors to extend the models in two directions: multi-resolution and temporal analysis.

When one of the relevant distance metrics induces a 2D square lattice structure on the model graph, one may model the system at a lower resolution scale. At low resolutions, patterns that are too discontinuous for the high-resolution model with only nearest neighbour interactions, are smoothed over and become once again continuous, and thus, visible to the model. While multi-resolution analysis is not limited to 2D lattices, or indeed, lattices of any dimension (Z. Zhang et al., 2023; S. Zhang, Tong, Xu, & Maciejewski, 2019), in this work we limit ourselves to the 2-dimensional case.

The process starts with grouping and fusing the 2D observations. In this work we group them into blocks of $l \times l$. We use $MaxPool(O_i, \ldots, O_{i+l^2-1})$ to determine the value of the effective observation $O_k^{'}$ representing the block, and assign also an effective hidden variable to each block $H_k^{'}$. Then we create a model by connecting the pooled observations to the new hidden root causes for the blocks and connect these to each other following the 2D grid pattern inherited from the base model. This is shown in Figure 2-a. Although some features are shared, low- and high-resolution models might have different topologies and might have different factors. Upon usage, we use the low-resolution model to infer the marginal distributions of the blocks' hidden variables and subsequently we insert them as priors on the full-resolution model (solid light-blue squares), see Figure 2-b.

The second use of unary factors or priors in this work is in modelling the evolution of the hidden states over time. We assume the data consists of a time series of 2D observations. Our models so far have been geared towards representing and reasoning on one timestep of the series of observations. To allow the inferred hidden states in the previous timestep $H_{i,t-1}$ to affect the current timestep $H_{i,t}$, we insert them as unary factors containing the posterior distribution of $H_{i,t-1}$ computed in the previous iteration (solid green squares in Figure 2-b) and connect them to the variables $H_{i,t}$ via a factor f_T encoding the time evolution of the hidden root-causes (empty green squares). This carrying of information forwards in time is a form of Bayesian filtering (Särkkä, 2013).

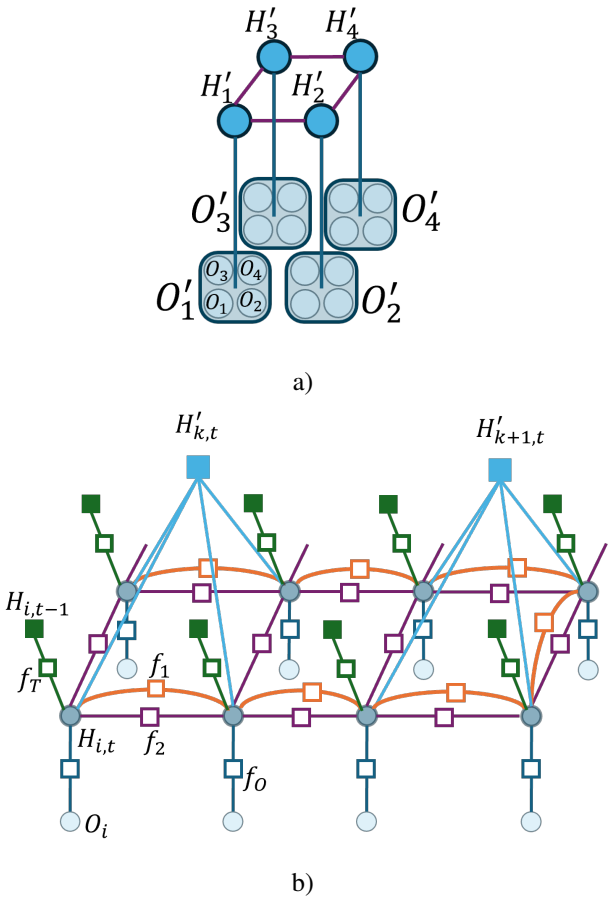


Figure 2. Construction of a low-resolution model for the system in Figure 1. a) Observations are pooled into blocks of size 2×2 and the new hidden states $H_k^{'}$ are connected. b) Depicts the full model with multi-resolution and temporal information. The inference results of the low-resolution model inserted as unary factors $H_{k,t}^{'}$ (solid blue), and the posteriors of the previous time-step inserted as unary factors $H_{i,t-1}$ (solid green), connected to the current variables $H_{i,t}$ via another factor accounting for time evolution of the states.

2.2. Workflow

The full workflow of the methodology to perform classification at multiple resolutions and over time, is shown in Figure 3. At discrete points in time, the system is measured providing raw data D_t. Depending on the application, this data can be a float, a vector, or even a nan. Therefore, it

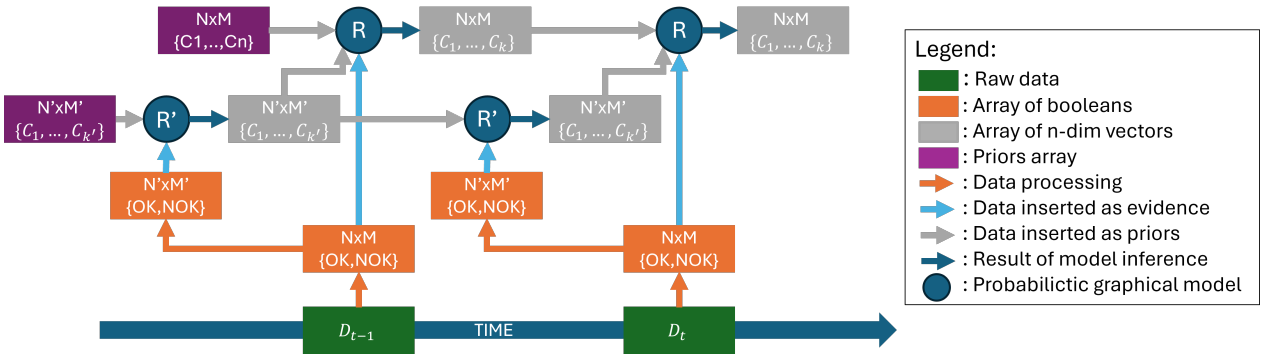


Figure 3. Graphical representation of the workflow for multi-resolution and temporal inference.

must be interpreted, i.e. processed, into an $N \times M$ array of binary observations. This processed data is then down sampled to size $N^{'} \times M^{'}$ and used as evidence for a low-resolution model R' that possibly admits information from prior timesteps. Inference on model R' produces an $N^{'} \times M^{'}$ array of k'-dimensional vectors representing the posterior probability that each observation is caused by each one of the k' effective root causes. These values are then used as priors in the full-resolution model R together with the $N \times M$ vectors of posteriors from the previous timestep and the $N \times M$ observations of the current timestep. The outcome of this process is an $N \times M$ array of k-dimensional vectors containing the probability of the root causes of each observation.

3. APPLICATION

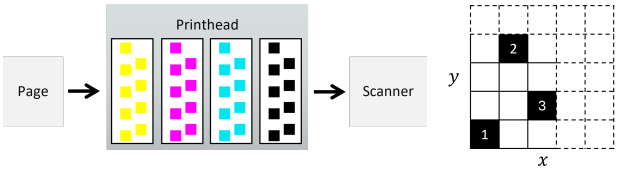
In this section, we discuss the application of the proposed methodology to an industrial use-case. Specifically, the diagnosis of print quality artefacts manifesting in an industrial inkjet printer manufactured by Canon Production Printing. Industrial inkjet printers are complex machines, featuring thousands of tiny nozzles (just a few micrometres in diameter) that jet billions of ink droplets every second. To monitor performance, the machine occasionally prints, scans, and analyses a test page. This process provides an indirect measurement of the printing accuracy of each individual nozzle. These measurements are then further analysed to determine the causes of suboptimal nozzle performance, allowing for optimization of the printing process and ensuring high print quality. The main challenge in this diagnostic process lies in the second analysis step: inferring the root causes of underperforming nozzles from their observations. A high-level schematic representation of the printing and scanning process is given in Figure 4.

Our methodology is well suited for this use-case as different distance metrics can be defined for the nozzles. Here we will consider two metrics: the spatial location of nozzles on the printhead and their jetting order. Figure 4-b illustrates the spatial arrangement of three nozzles on the printhead, where empty cells denote locations without a nozzle and black cells indicate nozzle presence. Notice that nozzles in close spatial proximity, such as nozzles 1 and 3, are not consecutive in the jetting order.

Given these two metrics we construct a CRF like the one shown in Figure 2-b. Purple edges and factors correspond to the spatial location metrics while orange edges and factors correspond to the jetting order metric. Specifically, in our implementation we define $|O| = |H| = 2^{10}$ random variables for the Observed and Hidden variables, respectively. These random variables are arrayed in an $N \times M$ lattice with $M = N = 2^5$, following the layout of the nozzles on the printhead. In the context of print quality, we consider four possible root causes for nozzle underperformance. Hence, H_i can take a value between $\{h, RC1, RC2, RC3, RC4\}$ where h stands for $healthy$ and RC for root cause. For the multi-resolution, we set $l = 2^3$ so we have$|O'| = |H'| = 2^4$. All the factors in the CRF, which we do not show for confidentiality reasons, are specified using experts' knowledge.

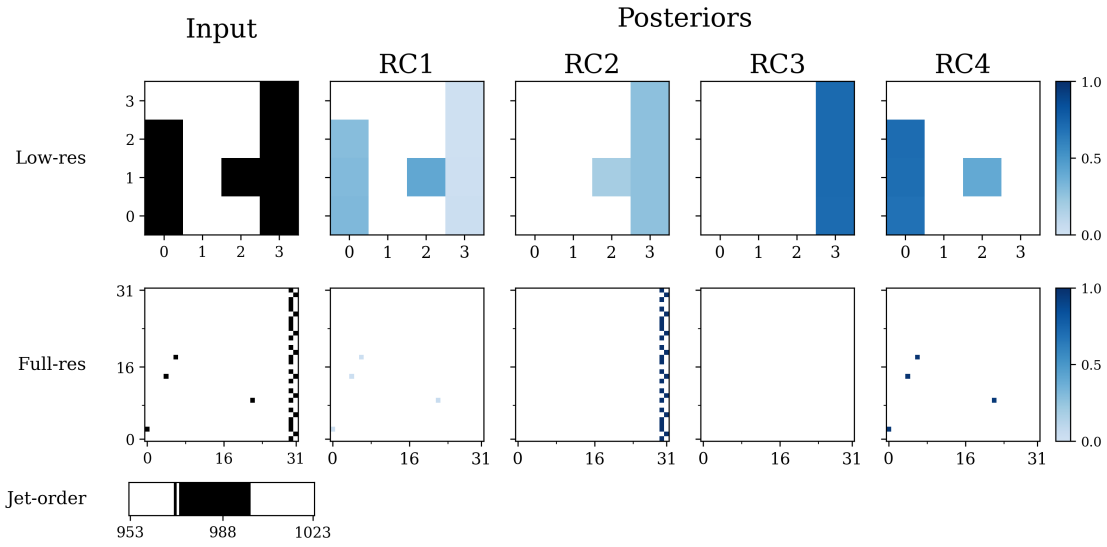
3.1. Results

To demonstrate the application of our methodology, we present the results of the CRF using nozzle performance data from two test sheets of an industrial printer operating in the field. As described in Section 2.2, the input to the CRF is an $N \times M$ array of binary observations. When visualized as an image, this forms a binary map where black pixels represent

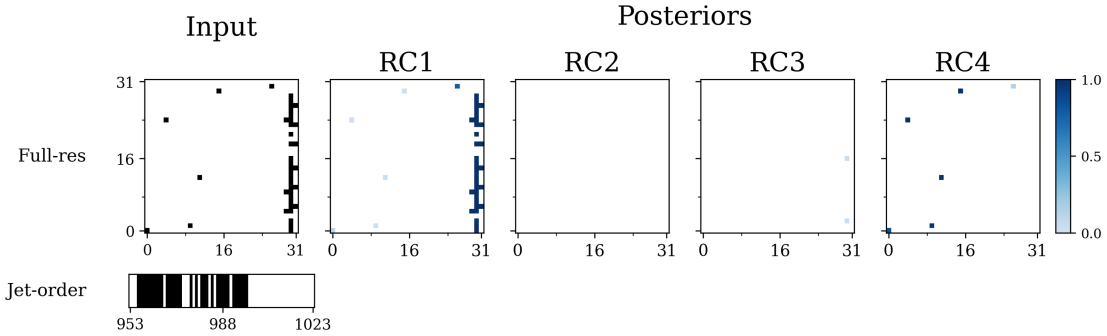


a) b)

Figure 4. a) High level schematics of ink-jet printing process. Only test pages are scanned. b) Example position of 3 nozzles on the printhead. Nozzle numbers (1,2,3) correspond to jetting order.



a)



b)

Figure 5. Results for two test sheets a) and b) respectively. The first column displays the binary nozzle performance data used as input for the model. Columns $RC1$ to $RC4$ display the posterior probabilities for each root cause in the spatial representation. The first row of a) shows the down sampled input (obtained by max-pooling), the second row shows the full-resolution spatial representation, the third row shows the jetting order representation (only the last 70 nozzles are shown for clarity). In b), only the full-resolution input and jetting order representation are shown.

underperforming nozzles and white pixels indicate normally functioning nozzles, see Figure 5. The down sampled evidence required for the multi-resolution analysis is obtained through a two-dimensional max-pooling operation applied to this binary image. Alternatively, when the same binary data is arranged according to the jetting order, it results in a one-dimensional array, see last row of Figure 5.

The first row of Figure 5-a shows the posterior probabilities of the low-resolution model for the first test sheet. For clarity, the posteriors corresponding to the healthy state have been omitted. Because the input is down sampled, underperforming nozzles merge into a continuous vertical band, enabling the model to capture broad, global patterns. $RC1$, $RC2$, and $RC3$ are associated with root causes that appear as vertical structures in the spatial domain. Whereas $RC4$ is associated with isolated underperforming nozzles, which can appear as clusters of any shape when the input is downsampled. By design, the low-resolution model allows for greater uncertainty between the different root causes, resulting in probabilities spread across $RC1-4$ for this test sheet. Figure 5 shows probabilities in gray scale, with white corresponding to a probability of 0 and black corresponding to a probability of 1. The actual diagnosis is performed at full resolution, where the model reasons jointly over both the jetting order and spatial representation. The second row of Figure 5-a presents the full-resolution posterior probabilities for the first sheet, computed using the low-resolution model outputs as priors. Figure 5-b displays the full-resolution results for the second test sheet (low resolution results omitted).

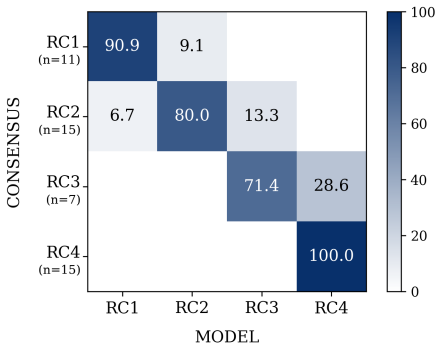


Figure 6. Row-normalized confusion matrix comparing the CRF-predicted classes with the corresponding expert labels.

Both test sheets show a similar vertical band of underperforming nozzles along the right edge of the paper. However, as identified by the CRF, the underlying causes differ. In the first sheet, the pattern originates from an issue in the jetting order. In the second sheet, the root cause manifests spatially on the printhead. This distinction becomes apparent when examining the jetting order representation: the first sheet exhibits a continuous sequence of underperforming nozzles, whereas the second shows a more intermittent pattern. Additionally, the few isolated NOK observations scattered throughout the test sheets that were not clearly identified in the low-resolution inference are correctly identified as $RC4$ when considered by the high-resolution model. By incorporating and reasoning upon both the jetting order and spatial representations, the CRF can effectively distinguish between these different root causes on a nozzle-by-nozzle basis. These results were confirmed by domain experts at Canon Production Printing.

To quantitatively assess model performance, the posteriors of the CRF were compared against a set of 48 test sheets independently annotated by two domain experts according to one of the four root causes ($RC1-4$). Only sheets for which both experts reached consensus on the assigned label were retained, ensuring a high-confidence ground truth. For each sheet, the predicted root cause was determined by summing the posterior probabilities per root cause and selecting the one with the highest sum. This prediction was then compared against the expert-assigned label. Across this consensus set, the model achieved an overall accuracy of 0.88 (42 correctly classified sheets out of 48) and a macro-averaged F_1 score of 0.85. The confusion matrix is shown in Figure 6. Further validation on a larger dataset is ongoing to confirm the model's performance.

4. DISCUSSION

The proposed CRF framework can identify different root causes of system underperformance using real data from an operational high-tech system. The model captures both spatially localized and process-related defectivity patterns by jointly reasoning over multiple representations of the observations. Quantitative validation on a consensus-labelled dataset further confirmed the method's diagnostic capability, with an overall accuracy of 0.88 and a macro-averaged F_1 score of 0.85. These findings indicate that combining multi-resolution modelling with probabilistic reasoning enables interpretable diagnostics in complex high-tech systems. Although these results are promising, the current model can only diagnose root causes that correspond to predefined hidden states in the CRF. In practice, the range of failure modes that can lead to defectivity patterns is extensive, and it is not feasible to model all of them explicitly. The present validation therefore focused on the most common and well-understood root causes. Extending the framework to account for previously unseen causes is a subject of ongoing research.

In parallel with the development of the CRF framework, we explored unsupervised data-driven approaches for clustering the two-dimensional observation data. The choice of an unsupervised approach is motivated by the absence of a large labelled dataset, as explained in Section 1.1. Although this approach could capture broad similarities between patterns, it was considered unable to distinguish between the different root causes, as it relied solely on the spatial representation.

Below we discuss some computational aspects of the methodology, namely the choice and rationale for loopy belief propagation as an inference algorithm, as well as limitations of this approach and possible extensions. In this work we used the Python library $Quimb$ (Gray, 2018) to specify the models, and loopy belief propagation (K. P. Murphy, 2023) as the inference engine. The computational complexity of loopy belief propagation is $O(I \cdot E \cdot D^2)$, where I is the total number of iterations, E is the number of edges, in this case $E = O(N)$, and $D = O(1)$ is the number of distinct root causes. Although the computation time varies per case, it stays in the order of seconds on a standard laptop for models with 10^3 hidden variables, acceptable for the scope of this study. Loopy belief propagation is not guaranteed to converge to the true posteriors. In particular, when two or more joint configurations are equally likely, belief propagation will oscillate between them (K. Murphy, Weiss, & Jordan, 2013). This situation could arise in the diagnostic setting when an observed pattern is equally compatible with multiple root causes. For such cases, we are currently investigating the use of measures of computational hardness, e.g. number of iterations of belief propagation, together with the likelihood of the model given data, to characterize our confidence in the model's output.

Scaling our methodology to tens of thousands of variables may require changes on the inference algorithm. A possible improvement is message vectorization, which greatly reduces the compute time of belief propagation in the regime of large number of small factors, as is the case for our di-

agnostic models. Hardware acceleration, such as PyTorch, JAX and CUDA, is another possibility offered by the Quimb library yet to be explored. Another alternative would be to use Gibbs sampling (Kato & Zerubia, 2012), a technique employed in image segmentation known to scale to thousands of variables.

Lastly, we discuss factor specification. In this work we opted for a manual approach. As a consequence, the model quality depends on the modeller's skill at translating root-cause behaviour into probabilistic weights. Manual factor specification is an iterative process that employs failure simulators in order to test a modeller's hypothesis on synthetic data. This iterative process is a qualitative form of optimization. There is no mathematical guarantee that the parameters are the correct ones, in the sense that they do not necessarily maximize the likelihood of the parameters given the model. However, the fact that our models with manually specified factors perform well on real-world data speaks to the robustness of the approach.

A natural extension to the methodology would be to algorithmically learn the factors from a training dataset. The most widely known algorithm for learning of factors in probabilistic graphical models is expectation-maximization (K. P. Murphy, 2023). Deep learning-based training is also possible, as demonstrated in (Pastorino, Moser, Serpico, & Zerubia, 2023; Liu, Lin, Qiao, & Shen, 2017). Given the availability of some labelled data, semi-supervised learning is also an interesting possibility, as is treating the factors as random variables with distributions learnable through Bayesian inference (Cox, van de Laar, & de Vries, 2019). These investigations are left for future work.

5. CONCLUSIONS

In this paper, we presented a novel application of Conditional Markov Random Fields (CRFs) for diagnosing root causes of defectivity patterns in high-tech manufactured products. Our methodology was preliminary validated on a real industrial scenario. To the best of the authors' knowledge, this is the first time CRFs have been employed for this task. Our methodology offers several advantages. First, it is model-based, allowing for the incorporation of domain expertise. This leads to a classification that is explainable, directly connected to known root causes, and grounded in the known design and operations of the system. Second, its inherent flexibility allows for time and spatial multi-resolution analysis, and reasoning across various representations. Third, it performs probabilistic diagnostics by computing the posterior probability of different root causes. This is important for those complex cases in which a diagnosis is not certain, e.g. a defect is due to concomitant root causes. Lastly, the approach demonstrates strong data robustness, as it is not dependent on high-quality data for model specification, and adapts effectively to diverse data shapes. Missing or uncertain data can be handled as well by treating them as unobserved (hidden) variables, with some prior distribution reflecting the uncertain beliefs of the observation's value.

Despite these strengths, the methodology in its current form has certain limitations. It is primarily limited to discrete classification on discrete data. Future extensions could explore alternative models such as hybrid CRFs to accommodate continuous data types. Image segmentation models with normally distributed observations and discrete labels have already been studied in the literature (Kato & Zerubia, 2012). Similar techniques could be applied to the case of non-Gaussian observations provided that the likelihoods of each label can be efficiently computed.

Furthermore, the manual specification of factors within the CRF model necessitates a good understanding of probabilistic graphical models alongside domain expertise. Future research could investigate the automatic learning of the factors. Additionally, while our models are Bayesian on variables, their parameters are currently fixed, even when learned. Future research should then focus on extending these models to incorporate Bayesian parameter estimation approaches. This will allow for a more complete probabilistic programming treatment of the problem.

ACKNOWLEDGMENT

The research is carried out as part of the Carefree program under the responsibility of TNO-ESI in cooperation with Canon Production Printing. The research activities are co-funded by TKI HTSM via the PPP Innovation Scheme (PPP-I) for public-private partnerships.

REFERENCES

Bai, J., Wu, D., Shelley, T., Schubel, P., Twine, D., Russell, J., ... Zhang, J. (2025, April). A comprehensive survey on machine learning driven material defect detection. *ACM Computing Surveys*. Retrieved from http://dx.doi.org/10.1145/3730576 doi: 10.1145/3730576

Barbini, L., Bratosin, C., & Nägele, T. (2021). Embedding diagnosability of complex industrial systems into the design process using a model-based methodology. *PHM Society European Conference*. doi: 10.36001/phme.2021.v6i1.2806

Cox, M., van de Laar, T., & de Vries, B. (2019). A factor graph approach to automated design of bayesian signal processing algorithms. *International Journal of Approximate Reasoning*, *104*, 185-204. doi: https://doi.org/10.1016/j.ijar.2018.11.002

Darwiche, P. A. (2009). *Modeling and reasoning with bayesian networks*. Cambridge University Press.

Gray, J. (2018). quimb: a python library for quantum information and many-body calculations. *Journal of Open Source Software*, *3*(29), 819. doi: 10.21105/joss.00819

Henn, M.-A., Zhou, H., & Barnes, B. M. (2019, Sep). Data-driven approaches to optical patterned defect detection. *OSA Continuum*, *2*(9), 2683–2693. doi: 10.1364/OSAC.2.002683

Kato, Z., & Zerubia, J. (2012). *Markov Random Fields in Image Segmentation.* Retrieved from https://inria.hal.science/hal-00737058

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. MIT Press.

Liu, F., Lin, G., Qiao, R., & Shen, C. (2017, 03). Structured learning of tree potentials in crf for image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, *PP*. doi: 10.1109/TNNLS.2017.2690453

Murphy, K., Weiss, Y., & Jordan, M. I. (2013). *Loopy belief propagation for approximate inference: An empirical study.* Retrieved from https://arxiv.org/abs/1301.6725

Murphy, K. P. (2023). *Probabilistic machine learning: Advanced topics*. MIT Press. Retrieved from http://probml.github.io/book2

Pastorino, M., Moser, G., Serpico, S., & Zerubia, J. (2023, November). Learning CRF potentials through fully convolutional networks for satellite image semantic segmentation. In *SITIS 2023 - 17th International Conference on Signal-Image Technology &Internet-Based Systems* (p. 93-98). Bangkok, Thailand: IEEE. Retrieved from https://inria.hal.science/hal-04255319 doi: 10.1109/SITIS61268.2023.00023

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann Publishers Inc.

Särkkä, S. (2013). *Bayesian filtering and smoothing*. Cambridge University Press.

van Gerwen, E., Barbini, L., Borth, M., & Passmann, R. (2024, 10). Efficient differential diagnosis using cost-aware active testing. *International Journal of Prognostics and Health Management*, *15*. doi: 10.36001/ijphm.2024.v15i3.3849

Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019, 11). Graph convolutional networks: a comprehensive review. *Computational Social Networks*, *6*. doi: 10.1186/s40649-019-0069-y

Zhang, Z., Bu, J., Ester, M., Zhang, J., Li, Z., Yao, C., ... Wang, C. (2023). Hierarchical multi-view graph pooling with structure learning. *IEEE Transactions on Knowledge and Data Engineering*, *35*(1), 545-559. doi: 10.1109/TKDE.2021.3090664