Towards Gaussian Processes for Automatic and Interpretable Anomaly Detection in Industry 4.0

Fabian Berns[1], Markus Lange-Hegermann[2] and Christian Beecks[1]
[1]*Department of Computer Science, University of Münster, Germany*
[2]*Department of Electrical Engineering and Computer Science,*
OWL University of Applied Sciences and Arts, Lemgo, Germany

Keywords: Anomaly Detection, Gaussian Processes, Explainable Machine Learning, Industry 4.0.

Abstract: Discerning unexpected from expected data patterns is the key challenge of anomaly detection. Although a multitude of solutions has been applied to this modern Industry 4.0 problem, it remains an open research issue to identify the key characteristics subjacent to an anomaly, sc. generate hypothesis as to why they appear. In recent years, machine learning models have been regarded as universal solution for a wide range of problems. While most of them suffer from non-self-explanatory representations, Gaussian Processes (GPs) deliver interpretable and robust statistical data models, which are able to cope with unreliable, noisy, or partially missing data. Thus, we regard them as a suitable solution for detecting and appropriately representing anomalies and their respective characteristics. In this position paper, we discuss the problem of automatic and interpretable anomaly detection by means of GPs. That is, we elaborate on why GPs are well suited for anomaly detection and what the current challenges are when applying these probabilistic models to large-scale production data.

1 INTRODUCTION

Anomaly detection is an important data mining process to distinguish expected from unexpected data patterns. It enables researchers as well as practitioners to assess the condition and current state of a system of interest (Chandola et al., 2009). Applications are manifold, e.g. in medicine (Rajpurkar et al., 2017), credit card fraud (Sorournejad et al., 2016), or network intrusion detection (Ioannou et al., 2017). Especially with regards to industrial applications, monitoring sensor data from complex processes in order to detect outliers or low-performing production behavior caused by undesired patterns and trends, which we summarize as *anomalies*, is a challenging task (Beecks et al., 2019). Not only due to the massive amount of sensor data but also due to different types of anomalies, manual or automatic inspection systems are frequently supported by anomaly detection algorithms (Stojanovic et al., 2016). It is often stressed that anomaly detection should be part of understanding the production process as a whole (Niggemann and Lohweg, 2015), i.e. that it is not sufficient to detect anomalies but also to generate hypothesis as to why they appear.

While the last years have witnessed the development of different anomaly detection algorithms (cf. the work of Renaudie et al. (2018) for a recent performance evaluation in an industrial context) only less effort has been spent on the investigation of the inherent structure of an anomaly. Utilizing techniques of machine learning and Artificial Intelligence (AI) for anomaly detection seems like a natural choice to not only detect unexpected patterns, but to understand them. Still, achieving the level of trustworthiness required for sensitive industrial applications is a nontrivial task (cf. High-Level Expert Group on Artificial Intelligence, 2019; Kwon et al., 2019).

With regards to the key requirements for trustworthy AI defined by the High-Level Expert Group on Artificial Intelligence (2019) of the European Commission, we consider GPs (Rasmussen and Williams, 2006) as an appropriate machine learning model to fulfill those demands. GPs deliver robust and reliable statistical data models, which are able to cope with unreliable, noisy, or partially missing data (Boškoski et al., 2012). Since these models are composed from simple components defined by explainable hyperparameters, they are interpretable by nature and deliver the capabilities to trace their predictions back to those

Berns, F., Lange-Hegermann, M. and Beecks, C.
Towards Gaussian Processes for Automatic and Interpretable Anomaly Detection in Industry 4.0.
DOI: 10.5220/0010130300870092
In *Proceedings of the International Conference on Innovative Intelligent Industrial Production and Logistics (IN4PL 2020)*, pages 87-92
ISBN: 978-989-758-476-3

components and the according training data (Lloyd et al., 2014; Duvenaud et al., 2013).

In this position paper, we discuss the problem of automatic and interpretable anomaly detection by means of GPs. That is, we elaborate on why GPs are well suited for anomaly detection and what the current challenges are when applying these probabilistic models to large-scale production data. To this end, our contributions are two-fold:

- We introduce GPs as an interpretable model for anomaly detection.
- We outline challenges that have to be addressed in order to scale GPs to large-scale production data

This position paper is structured as follows. We outline related work in Section 2. We introduce GPs in Section 3, while the concept and rationale for anomaly detection by means of these statistical data models and the remaining challenges in that field are explained in detail in Section 4. We conclude our paper with an outlook on future work in Section 5.

2 RELATED WORK

In the era of Industry 4.0, the field of anomaly detection has become crucially important. As a result, there is a plethora of classical anomaly detection algorithms that have been proposed in recent years such as Z-Score (Domingues et al., 2016), Mahalanobis Distance-Based, Empirical Covariance Estimation (Pedregosa et al., 2011; Chandola et al., 2009), Robust Covariance Estimation (Rousseeuw, 1984; Chandola et al., 2009), Subspace-based PCA Anomaly Detector (Chandola et al., 2009), One-Class SVM (Schölkopf et al., 2001; Pedregosa et al., 2011; Chandola et al., 2009; Eskin et al., 2002), Isolation Forest (I-Forest) (Liu et al., 2008; Pedregosa et al., 2011), Gaussian Mixture Model (Pedregosa et al., 2011; Chandola et al., 2009; Phua et al., 2010), Deep Auto-Encoder (Candel et al., 2018; Gong et al., 2019), Local Outlier Factor (Breunig et al., 2000; Pedregosa et al., 2011; Chandola et al., 2009; Auslander et al., 2011), Self-Organizing Maps (Von Birgelen et al., 2018), Least Squares Anomaly Detector (Tavallaee et al., 2010), GADPL (Graß et al., 2019), Automata (Vodenčarević et al., 2011), and k-Nearest Neighbor (Goldstein and Uchida, 2016; Auslander et al., 2011; Eskin et al., 2002).

Current approaches (Chalapathy and Chawla, 2019; Zenati et al., 2018; An and Cho, 2015; Zhang and Chen, 2019; Sabokrou et al., 2018; Suh et al., 2016; Berkhahn et al., 2019; Li et al., 2019; Kawachi et al., 2018; Guo et al., 2018; Wang et al., 2019; Dias et al., 2020) frequently make use of generative models for anomaly detection, e.g. Variational Autoencoders (Kingma and Welling, 2014), Generative Adversarial Networks (Goodfellow et al., 2014), GP Latent Variable Models (Damianou et al., 2016), or Normalizing Flows (Rezende and Mohamed, 2015), in particular for sequence data (Bowman et al., 2016). These models can be trained automatically for the usage of anomaly detection (Müller et al., 2020).

While these algorithms are all possible approaches for anomaly detection, as shown in different surveys (Goldstein and Uchida, 2016; Phua et al., 2010; Chandola et al., 2009), they are not directly suited for describing the inherent structure of anomalies, which is the major focus of this position paper. We choose GPs (Rasmussen and Williams, 2006) for anomaly description due to their capability to not only gather statistical indicators, but deliver the very characteristics of specific anomalous behavior from the data.

For automatically describing the underlying data characteristics, Lloyd et al. (2014) have proposed the Automatic Bayesian Covariance Discovery System that adapts the Compositional Kernel Search Algorithm (Duvenaud et al., 2013) by adding intuitive natural language descriptions of the function classes described by their models. Hwang et al. (2016) further expand on those concepts by expanding these models to discover kernel structures which are able to explain multiple time series at once. Recently, the 3CS (Berns et al., 2020) and LARGe algorithms (Berns and Beecks, 2020) have shown to outperform the aforementioned approaches in terms of efficiency. As these methods are all based on GPs, we give a short introduction of GPs in the following section.

3 GAUSSIAN PROCESSES

A Gaussian Process (GP) (Rasmussen and Williams, 2006) is a stochastic process over random variables $\{f(x) \mid x \in \mathcal{X}\}$, indexed by a set $\mathcal{X}$, where every finite subset of random variables follows a multivariate normal distribution. The distribution of a GP is the joint distribution of all of these random variables and it is thus a probability distribution over the space of functions $\{f : \mathcal{X} \to \mathbb{R}\}$. A GP is formalized as

$$f(\cdot) \sim GP\big(m(\cdot), k(\cdot,\cdot)\big), \tag{1}$$

where the mean function $m : \mathcal{X} \to \mathbb{R}$ and the covariance function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ are defined $\forall x, x' \in \mathcal{X}$ via

$$m(x) = \mathbb{E}\left[f(x)\right] \tag{2}$$

$$k(x,x') = \mathbb{E}\left[\big(f(x) - m(x)\big) \cdot \big(f(x') - m(x')\big)\right] \tag{3}$$

Given a finite dataset $D = \{X, Y\}$ with input $X = \{x_i \mid x_i \in \mathcal{X} \wedge 1 \leq i \leq n\}$ representing the underlying index values, such as timestamps or locations, and target $Y = \{y_i \mid y_i = f(x_i), x_i \in \mathcal{X}\}$ representing the actual data values, such as sensor values or other complex measurements, a GP can be used to statistically represent the dataset D by optimizing the hyperparameters θ of both mean and covariance function. This optimization is frequently carried out by maximizing the log marginal likelihood $\mathcal{L}$ (Rasmussen and Williams, 2006; Kim and Teh, 2018) of the GP:

$$\mathcal{L}(m,k,\theta \mid D) = -\frac{1}{2} \cdot \left[(y-\mu)^T \Sigma^{-1}(y-\mu) + \log |\Sigma| + n \log(2\pi)\right] \quad (4)$$

As can be seen in Equation 4, the marginalization of a GP for a given dataset D of n records results in a mean vector $\mu \in \mathbb{R}^n$, and a covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ which are defined as $y[i] = f(x_i)$, $\mu[i] = m(x_i)$, and $\Sigma[i,j] = k(x_i, x_j)$ for $1 \leq i, j \leq n$, respectively.

How GPs are utilized in the context of anomaly detection and how the resulting challenges can be solved are described in the following section.

4 GPs FOR ANOMALY DETECTION

GP models are robust Bayesian models that intrinsically prevent overfitting. This sturdiness is a clear advantage of GP models in dynamic and noisy production environments with always changing circumstances like evolving processes or sensors modifications. GP models yield smooth models even when only a few data points are given, which is particular important in production, where the batch size is steadily decreasing.

In addition to recognizing an anomaly, it is important to understand and interpret its underlying structure. Only this makes it possible to rectify the underlying problem. Being interpretable is immanent in GPs, due to their rigid mathematical structure. Their covariance functions can induce combinations of various physically motivated behaviors like trends, periodicity, differential equations (Lange-Hegermann, 2018, 2020), and change points, and each of these covariance functions comes with hyperparameters, which are often interpretable physical constants like periods. The model prediction can even be disassembled into parts stemming from the individual interpretable parts of the covariance function. All this physically interpretable structure can be learned from data, and additionally domain knowledge can be pre-encoded into the models by experts.

What makes the application of GPs difficult to large-scale, streaming production data is the super-cubic computation time complexity of optimizing GPs. Though efficient solutions exist (Snelson and Ghahramani, 2007), they do not provide a fast solution for *automatic* and *interpretable* anomaly detection. For this reason, we identified the following major challenges.

4.1 Efficient Algorithms for Large-scale GP Models

The first challenge is concerned with the development of efficient retrieval algorithms for GP models. This includes the central question of how to efficiently search and determine suitable covariance functions for anomaly detection. Instead of applying a greedy search through the space of possible covariance functions, as done by state-of-the-art algorithms (Duvenaud et al., 2013; Lloyd et al., 2014; Steinruecken et al., 2019), one could learn the impact of individual hyperparameters in order to specifically derive new covariance functions. In addition to the development of such intelligent covariance search heuristics, another challenge is to process and infer GP models for multivariate, event-based sensor data in (near) real time. This demands for resource-efficient streaming GP retrieval algorithms that scale to state of the art big data processing frameworks such as Apache Hadoop, Spark, and Storm.

4.2 Model Selection for GPs in Anomaly Detection

Besides the development of efficient algorithms, the second major challenge lies in the development of suitable GP model selection approaches. While prominent model quality estimators, such as the likelihood function, tend to prefer complex models containing many hyperparameters, Laplace approximations enable to capture the model evidence of GP models more properly. The particular challenge is thus to couple Laplace approximations with unsupervised GP Latent Variable Models (Damianou et al., 2016) in order to enable conclusions on the underlying physical processes and individual sensor dimensions.

To sum up, the development of real-time GP model retrieval and selection algorithms for multivariate data streams that are open-source and are based on open industry standards are necessary for detecting, analyzing, and understanding anomalies in an domain-agnostic, automatic, and interpretable manner.

5 CONCLUSION

In this paper, we have argued for the application of GPs for automatic and interpretable anomaly detection. GPs are robust bayesian machine learning tools, which enable inference for noisy as well as unreliable data. GP models yield smooth models even when only a few data points are given, which is particular useful in industrial scenarios, where creating those records is potentially expensive.

Along with the advantages, several challenges have to be met. In particular, concepts and algorithms are needed to facilitate efficient GP model retrieval and selection on large-scale streaming data. We aim to address these challenges in our future work and to elaborate our developments in various application domains.

REFERENCES

An, J. and Cho, S. (2015). Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, 2(1):1–18.

Auslander, B., Gupta, K. M., and Aha, D. W. (2011). A comparative evaluation of anomaly detection algorithms for maritime video surveillance. In Carapezza, E. M., editor, *Proc. SPIE 8019, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X*, SPIE Proceedings, page 801907. SPIE.

Beecks, C., Schmidt, K. W., Berns, F., and Graß, A. (2019). Gaussian processes for anomaly description in production environments. In *EDBT/ICDT Workshops*, volume 2322 of *CEUR Workshop Proceedings*.

Berkhahn, F., Keys, R., Ouertani, W., Shetty, N., and Geißler, D. (2019). Augmenting variational autoencoders with sparse labels: A unified framework for unsupervised, semi-(un) supervised, and supervised learning. *arXiv preprint arXiv:1908.03015*.

Berns, F. and Beecks, C. (2020). Automatic gaussian process model retrieval for big data. In *CIKM*. ACM.

Berns, F., Schmidt, K., Bracht, I., and Beecks, C. (2020). 3CS Algorithm for Efficient Gaussian Process Model Retrieval [accepted]. *25th International Conference on Pattern Recognition (ICPR)*.

Boškoski, P., Gašperin, M., and Petelin, D. (2012). Bearing fault prognostics based on signal complexity and gaussian process models. In *2012 IEEE Conference on Prognostics and Health Management*, pages 1–8.

Bowman, S. R., Vilnis, L., Vinyals, O., Dai, A. M., Józefowicz, R., and Bengio, S. (2016). Generating sentences from a continuous space. In *CoNLL*, pages 10–21. ACL.

Breunig, M., Kriegel, H.-P., Ng, R. T., and Sander, J. (2000). Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pages 93–104. ACM.

Candel, A., LeDell, E., Parmar, V., and Arora, A. (2018). Deep learning with h2o. https://www.h2o.ai/wp-content/themes/h2o2016/images/resources/DeepLearningBooklet.pdf. (Accessed on 09/28/2020).

Chalapathy, R. and Chawla, S. (2019). Deep learning for anomaly detection: A survey. *CoRR*, abs/1901.03407.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58.

Damianou, A. C., Titsias, M. K., and Lawrence, N. D. (2016). Variational inference for latent variables and uncertain inputs in gaussian processes. *J. Mach. Learn. Res.*, 17:42:1–42:62.

Dias, M. L. D., Mattos, C. L. C., da Silva, T. L. C., de Macêdo, J. A. F., and Silva, W. C. P. (2020). Anomaly detection in trajectory data with normalizing flows. *CoRR*, abs/2004.05958.

Domingues, R., Buonora, F., Senesi, R., and Thonnard, O. (2016). An application of unsupervised fraud detection to passenger name records. In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshop (DSN-W)*, pages 54–59.

Duvenaud, D., Lloyd, J. R., Grosse, R. B., Tenenbaum, J. B., and Ghahramani, Z. (2013). Structure discovery in nonparametric regression through compositional kernel search. In *ICML (3)*, volume 28 of *JMLR Workshop and Conference Proceedings*, pages 1166–1174. JMLR.org.

Eskin, E., Arnold, A., Prerau, M., Portnoy, L., and Stolfo, S. (2002). A geometric framework for unsupervised anomaly detection. In Barbará, D. and Jajodia, S., editors, *Applications of Data Mining in Computer Security*, volume 6 of *Advances in Information Security, 1568-2633*, pages 77–101. Springer US and Imprint and Springer, Boston, MA.

Goldstein, M. and Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PloS one*, 11(4):152–173.

Gong, D., Liu, L., Le, V., Saha, B., Mansour, M. R., Venkatesh, S., and Hengel, A. v. d. (2019). Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1705–1714.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *NeurIPS*.

Graß, A., Beecks, C., and Soto, J. A. C. (2019). Unsupervised anomaly detection in production lines. In Beyerer, J., Kühnert, C., and Niggemann, O., editors, *Machine Learning for Cyber Physical Systems*, pages 18–25, Berlin, Heidelberg. Springer Berlin Heidelberg.

Guo, Y., Liao, W., Wang, Q., Yu, L., Ji, T., and Li, P. (2018). Multidimensional time series anomaly detection: A gru-based gaussian mixture variational autoencoder approach. In *Asian Conference on Machine Learning*, pages 97–112.

High-Level Expert Group on Artificial Intelligence (2019). Ethics guidelines for trustworthy ai. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. (Accessed on 09/28/2020).

Hwang, Y., Tong, A., and Choi, J. (2016). Automatic construction of nonparametric relational regression models for multiple time series. In Balcan, M. F. and Weinberger, K. Q., editors, *ICML 2016: Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 3030–3039. PLMR.

Ioannou, C., Vassiliou, V., and Sergiou, C. (2017). An intrusion detection system for wireless sensor networks. In *2017 24th International Conference on Telecommunications (ICT)*, pages 1–5. IEEE.

Kawachi, Y., Koizumi, Y., and Harada, N. (2018). Complementary set variational autoencoder for supervised anomaly detection. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2366–2370. IEEE.

Kim, H. and Teh, Y. W. (2018). Scaling up the Automatic Statistician: Scalable structure discovery using Gaussian processes. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, volume 84.

Kingma, D. P. and Welling, M. (2014). Auto-encoding variational bayes. In *ICLR*.

Kwon, D., Kim, H., Kim, J., Suh, S. C., Kim, I., and Kim, K. J. (2019). A survey of deep learning-based network anomaly detection. *Cluster Computing*, pages 1–13.

Lange-Hegermann, M. (2018). Algorithmic linearly constrained gaussian processes. In *NeurIPS*, pages 2141–2152.

Lange-Hegermann, M. (2020). Linearly constrained gaussian processes with boundary conditions. *CoRR*, abs/2002.00818.

Li, D., Chen, D., Jin, B., Shi, L., Goh, J., and Ng, S.-K. (2019). Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks. In *International Conference on Artificial Neural Networks*, pages 703–716. Springer.

Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In Giannotti, F., editor, *Eighth IEEE International Conference on Data Mining, 2008*, pages 413–422, Piscataway, NJ. IEEE.

Lloyd, J. R., Duvenaud, D., Grosse, R. B., Tenenbaum, J. B., and Ghahramani, Z. (2014). Automatic construction and natural-language description of nonparametric regression models. In *AAAI*, pages 1242–1250. AAAI Press.

Müller, A., Lange-Hegermann, M., and von Birgelen, A. (2020). Automatisches training eines variational autoencoder für anomalieerkennung in zeitreihen. In *VDI Kongress Automation 2020*, volume VDI-Berichte 2375, pages 687–698, Baden-Baden. VDI Wissensforum GmbH, VDI Verlag GmbH.

Niggemann, O. and Lohweg, V. (2015). On the diagnosis of cyber-physical production systems. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12:2825–2830.

Phua, C., Lee, V. C. S., Smith-Miles, K., and Gayler, R. W. (2010). A comprehensive survey of data mining-based fraud detection research. *CoRR*, abs/1009.6119.

Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., and Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *CoRR*, abs/1707.01836.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning (Adaptive Computation And Machine Learning)*. The MIT Press.

Renaudie, D., Zuluaga, M. A., and Acuna-Agost, R. (2018). Benchmarking anomaly detection algorithms in an industrial context: Dealing with scarce labels and multiple positive types. In *IEEE International Conference on Big Data*, pages 1227–1236.

Rezende, D. J. and Mohamed, S. (2015). Variational inference with normalizing flows. In *ICML*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 1530–1538. JMLR.org.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American statistical association*, 79(388):871–880.

Sabokrou, M., Khalooei, M., Fathy, M., and Adeli, E. (2018). Adversarially learned one-class classifier for novelty detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3379–3388.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J. C., Smola, A. J., and Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Comput.*, 13(7):1443–1471.

Snelson, E. and Ghahramani, Z. (2007). Local and global sparse gaussian process approximations. In *AISTATS*, volume 2 of *JMLR Proceedings*, pages 524–531. JMLR.org.

Sorournejad, S., Zojaji, Z., Atani, R. E., and Monadjemi, A. H. (2016). A survey of credit card fraud detection techniques: Data and technique oriented perspective. *CoRR*, abs/1611.06439.

Steinruecken, C., Smith, E., Janz, D., Lloyd, J. R., and Ghahramani, Z. (2019). The automatic statistician. In *Automated Machine Learning*, The Springer Series on Challenges in Machine Learning, pages 161–173. Springer.

Stojanovic, L., Dinic, M., Stojanovic, N., and Stojadinovic, A. (2016). Big-data-driven anomaly detection in industry (4.0): An approach and a case study. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 1647–1652.

Suh, S., Chae, D. H., Kang, H.-G., and Choi, S. (2016). Echo-state conditional variational autoencoder for anomaly detection. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 1015–1022. IEEE.

Tavallaee, M., Stakhanova, N., and Ghorbani, A. A. (2010). Toward credible evaluation of anomaly-based intrusion-detection methods. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(5):516–524.

Vodenčarević, A., Büning, H. K., Niggemann, O., and Maier, A. (2011). Using behavior models for anomaly detection in hybrid systems. In *2011 XXIII International Symposium on Information, Communication and Automation Technologies*, pages 1–8. IEEE.

Von Birgelen, A., Buratti, D., Mager, J., and Niggemann, O. (2018). Self-organizing maps for anomaly localization and predictive maintenance in cyber-physical production systems. *Procedia CIRP*, 72:480–485.

Wang, X., Du, Y., Lin, S., Cui, P., and Yang, Y. (2019). Self-adversarial variational autoencoder with gaussian anomaly prior distribution for anomaly detection. *CoRR, abs/1903.00904*.

Zenati, H., Romain, M., Foo, C.-S., Lecouat, B., and Chandrasekhar, V. (2018). Adversarially learned anomaly detection. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 727–736. IEEE.

Zhang, C. and Chen, Y. (2019). Time series anomaly detection with variational autoencoders. *CoRR*, abs/1907.01702.