A Machine Learning Approach to Pattern Detection and Prediction for Environmental Monitoring and Water Sustainability

Michael Osborne[*] mosb@robots.ox.ac.uk
Roman Garnett[†] rgarnett@andrew.cmu.edu
Kevin Swersky[‡] kevin@aquaticinformatics.com
Nando de Freitas[§] nando@cs.ubc.ca

[*]Department of Engineering Science, University of Oxford
[†]Department of Computer Science, Carnegie Mellon University
[‡]Aquatic Informatics
[§]Department of Computer Science, University of British Columbia

Abstract

We describe one of the successful products of a research partnership among several academic institutions (CMU, Oxford and UBC) and a water monitoring company (Aquatic Informatics). Water monitoring sensors are very diverse and remotely distributed. They produce vast quantities of data. The data itself is nonlinear and nonstationary. In addition, unanticipated environmental conditions and limitations in the sensing and communications hardware cause the data to be corrupted by previously uncharacterized nonlinearities, missing observations, spikes and multiple discontinuities. To improve the quality of the data and the monitoring process, this paper introduces an approach that uses Gaussian processes and a general "fault bucket" to capture *a priori* uncharacterized faults, along with an approximate method for marginalizing the potential faultiness of all observations. This gives rise to an efficient, flexible algorithm for the detection and automatic correction of faults. The probabilistic nature of the method is ideal for reporting uncertainty estimates to human operators. The approach can also be applied to detect patterns, other than faults, which are of great environmental significance. We present a fish sustainability example, where specific patterns in water level need to be detected so that fish don't get trapped and die in shallow pools.

Appearing in the *ICML 2011 Workshop on Machine Learning for Global Challenges*, Bellevue, WA, USA, 2011.

1. Introduction

Water sustainability is one of the greatest challenges that humankind faces. It is also a problem to which the machine-learning community can make positive, significant contributions. Water sustainability begins with proper water monitoring, which requires the analysis and interpretation of vast amounts of environmental data (Wagner & US Geological Survey, 2006). We refer readers to the website of Aquatic Informatics[1] for a broad picture of water monitoring.

In this paper, we attack the problem of pattern detection, correction, and prediction in water monitoring signals. Here measurements are often corrupted in non-trivial ways by various intermittent faulty sensing and communication mechanisms, giving rise to outliers, telemetry spikes, missing data, drift, and multiple unanticipated exogenous disturbances (see Figure 1). Further, signals are not well-modelled by simple parametric approaches, such as linear or Markovian models. Despite the enormous importance of such monitoring, appropriate machine-learning techniques are yet to be deployed for this purpose. In particular, there is a clear need for flexible algorithms, able to cope with signals and faults of many different types without placing a significant model-building burden upon users. Such algorithms must also be able to run reliably in real-time on incoming data. These techniques will enable us to provide operators with high-level summaries for better decision support and, in the future, to increase the level of automation and efficiency in water-management systems.

The collection of literature on fault- (also known as novelty-, anomaly- and one-class-) detection is vast

[1] www.aquaticinformatics.com

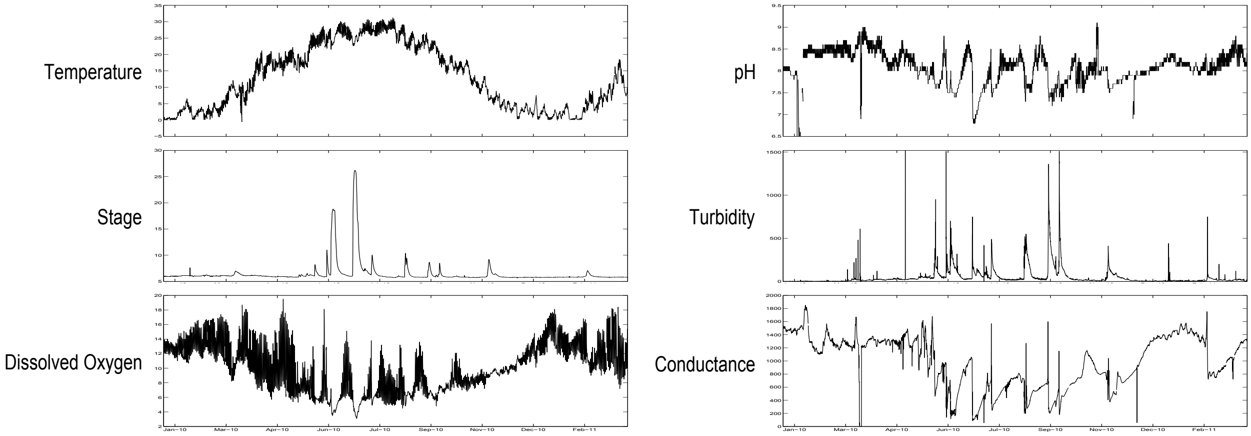

Figure 1. 16 months worth of data from six representative signals in water quality monitoring, which corresponds to approximately 11 000 measurements per series. These signals are highly nonlinear, demonstrating periodicity at different scales, intermittent pulses, and changes in dynamics. Not only do these signals exhibit a wide range of dynamics, but different signals of the same measurement type can also differ drastically if they are taken in different regions. The figure also depicts typical problems with the data, including missing observations, outliers and discontinuities.

(Isermann, 2005; Ding, 2008; Markou & Singh, 2003; Chandola et al., 2009; Khan & Madden, 2010; Dereszynski & Dietterich, 2011). Unfortunately, the problems solved by most of these techniques are of very different character to our own, rendering such techniques inapplicable. Further, after much experimentation with those methods that are applicable (some of the results appear in our experiments section) it became clear that off-the-shelf techniques could not satisfy our requirements for reliable water monitoring. This was predominately due to excessively restrictive assumptions (e.g., that signals were linear, Markov or Gaussian), and/or a failure to produce reasonable uncertainty estimates. Green-tech areas, including environmental monitoring and energy-demand prediction, are still far from full automation; the provision of uncertainty estimates is necessary to allow human operators to make appropriate decisions. For this reason, we focus on developing probabilistic nonlinear models of the signal. In addition to providing posterior probabilities of observation faultiness, we are able to perform effective prediction for the latent process even in the presence of faults.

Our proposed method will rely on Gaussian processes (GPs) due to their flexibility and widely demonstrated effectiveness at modeling nonlinear distributions. GPs have been used previously for fault detection in (Eciolaza et al., 2007), but in a very different context, unsuitable for our problem. Previous work along similar lines has approached this problem by creating observation models that specify the anticipated potential fault types *a priori* (Garnett et al., 2010), but this is usually an unreasonable assumption in highly variable or poorly understood environments. In our proposed "fault bucket" approach, we do not require the specification of precise fault models. In this way, our model can simultaneously identify anomalies and robustly make predictions in the presence of sensor faults. The result is a fast and efficient method for data-stream prediction that can manage a wide range of faults without requiring significant domain-specific knowledge.

2. Fault Bucket

Gaussian processes provide a simple, flexible framework for performing Bayesian inference about latent functions (Rasmussen & Williams, 2006). In environmental monitoring, exact measurements of the latent function are typically not available. Let $z(x)$ represent the value of an observation of the signal at x and $y(x)$ represent the value of the unknown true latent signal at that point. When the observation mechanism is not expected to experience faults, the usual noise model used is

$$p(z \mid y, x, \sigma_n^2) \triangleq \mathcal{N}(z; y, \sigma_n^2), \tag{1}$$

which represents additive i.i.d. Gaussian observation noise with variance σ_n^2. Note that this model is inappropriate when sensors can experience faults, which complicate the relationship between z and y.

Rather than specifying explicit parameters for every possible fault type, we propose a single catch-all "fault bucket" that can identify and treat appropriately measurements that are suspected of being faulty. The basic idea is to model faulty observations as being generated from a Gaussian distribution with a very wide variance;

points that are more likely under this model than under the normal predictive model of the Gaussian process can reasonably be assumed to be corrupted in some way, assuming we have a good understanding of the latent process. It is hoped that a very broad class of faults can be captured in this way. To formalize this idea, we choose an observation noise distribution to replace that in (1) that models the noise as independent but not identically distributed with separate variances σ_n^2 and σ_f^2 for the non-fault and fault cases.

Of course, *a priori*, we do not know whether any given observation will be faulty. Unfortunately, managing our uncertainty about the "faultiness" of all available observations is a challenging task. With N observations available, there are 2^N possible assignments of faultiness. It quickly becomes computationally infeasible to marginalize over all these possible values. Instead, we approximate our marginal predictions, a sum of Gaussians weighted by the posterior probabilities of faultiness of old data, as a single moment-matched Gaussian. In order to effect this approximation, we adopt a sequential approach, applicable for ordered data such as time series. For time series, the value to be predicted $y_\star$ typically lies in the future; we can hence assume that the faultiness of old observations is less pertinent than that of newer observations. At any point in time, then, we approximately resolve our sum over faultiness by representing each observation as having a known variance lying between between σ_n^2 and σ_f^2. The more likely an observation's faultiness, the closer its assigned variance will be to the (large) fault variance and the less relevant it will become for inference about the latent process. This approximate observation is then used for future predictions; we need never consider the full sum over all observations. Nonetheless, this approximate marginalization over faultiness is preferable to heuristics that would designate all observations as either faulty or not; our method acknowledges the uncertainty that may exist in our belief about faultiness.

The further mathematical details of our approximations are not reproduced here for brevity.

3. Results

Figure 2 shows the performance of the fault-bucket algorithm on the two real datasets. The first dataset contains a fault type called "painting," which is an error that occurs when ice builds on a sensor obscuring some of the readings. It is characterized by frequent sensor spikes interlaced with the original, and still accurate, signal. Our second dataset, which we dub "fishkiller", comes from a sensor near a dam on a river in British Columbia, Canada. It contains an otherwise normal water level-reading that is occasionally interrupted by a short period of rapid oscillation. This occurs when dam operators open and close the floodgates too quickly, leading to rapid water level drops followed by salmon becoming stranded and suffocating. Detecting these events is critical to proper regulation of dams. It is however a difficult problem as often these events occur during other transitions.

Figure 2 shows that, in both cases, the fault-bucket algorithm is able to detect these markedly different types of pattern. It is also capable of making accurate predictions.

Figure 3 shows the results of the fault-bucket algorithm on two additional sustained faults that were created artificially from real data, allowing the predictions made by the algorithm throughout the fault period to be compared with the now-known (but unobserved by the algorithm) ground truth. The fault-bucket algorithm again performs well, despite the very different nature of the faults.

We also tested against a number of different methods in order to establish the efficacy of the fault bucket algorithm. All GP-based approaches used the same hyperparameters employed by our algorithm. The training set used to learn those hyperparameters was also supplied to other methods for their respective model learning phases. Several methods identify a new observation y as a fault if

$$\left|y - m(y \mid \mathbf{y})\right| > 3\sigma_T\,, \qquad (2)$$

where $m(y \mid \mathbf{y})$ is the method's *a priori* prediction for y, and σ_T is the noise standard deviation on the faultless training set. Of course, methods using (2) or similar can not provide the posterior probability of a point's faultiness, as our algorithm can. Methods tested include:

XGP: A GP in which we exhaustively search over the faultiness of the last 10 points, and approximate the noise variance of all previous points in the window as having the value $\sigma_f^2 p(\text{fault} \mid \mathbf{y}) + \sigma_n^2 p(\neg\,\text{fault} \mid \mathbf{y})$, fixed at the time the point was observed (when data $\mathcal{D}$ was available). Clearly, this method is very much more computationally expensive than the fault bucket algorithm (roughly 2^9 times more), but offers a useful way to quantify the influence of our approximations.

TGP: A GP in which a point was flagged as a fault using (2); if faulty, a point was treated as having noise variance σ_f^2.

MLH: The most likely heteroscedastic GP (Kersting et al., 2007). Note that for this method we perform retrospective prediction (so that all data is available

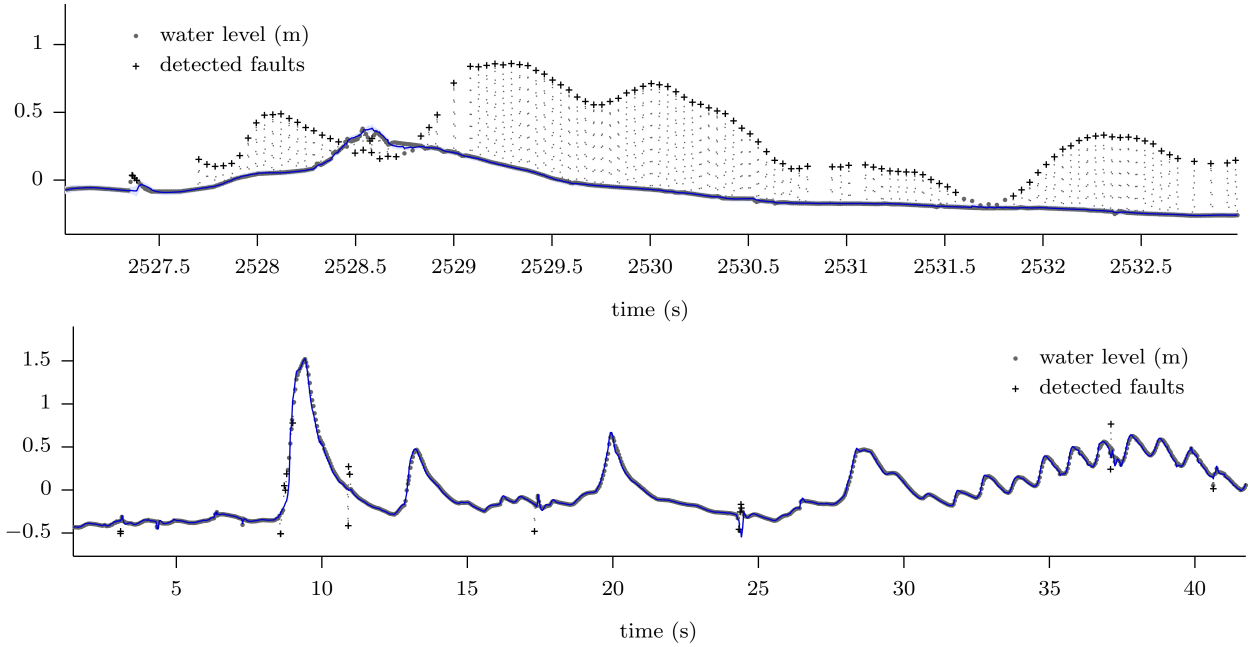

Figure 2. Mean and ±3σ standard-deviation bounds for the predictions of the fault-bucket algorithm on (top), the painting dataset and (bottom), the "fishkiller" dataset.

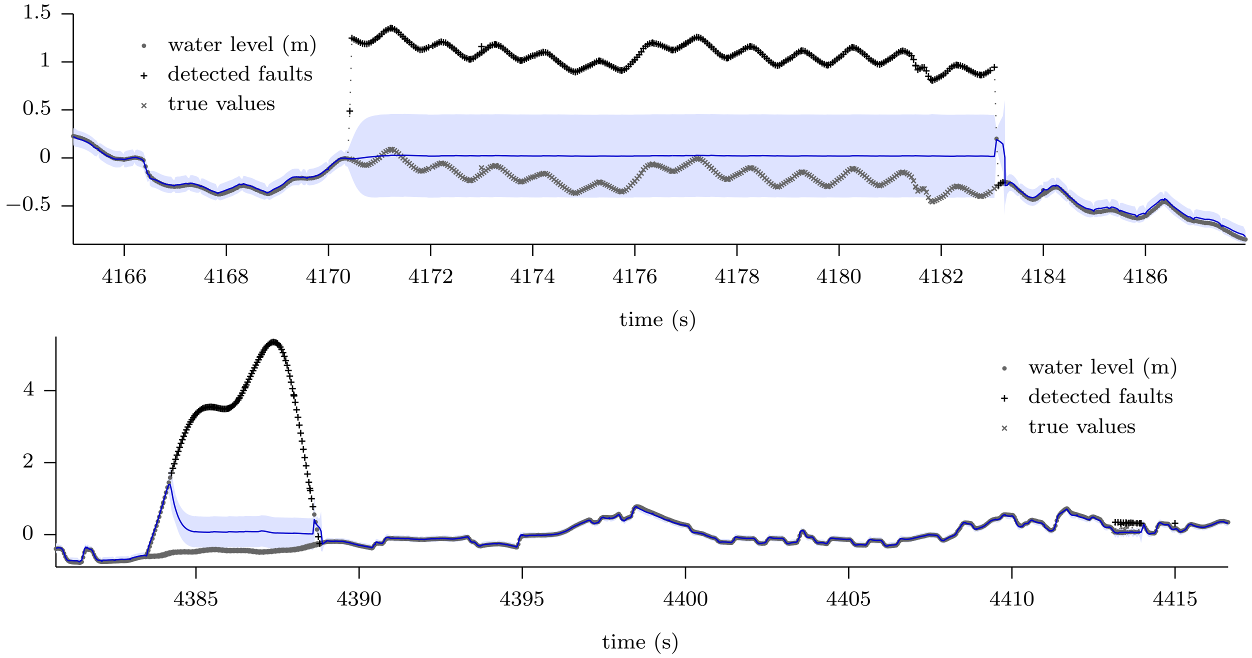

Figure 3. Mean and ±3σ standard-deviation bounds for the predictions of the fault-bucket algorithm on (top), a synthetic bias fault and (bottom), a synthetic change-in-dynamics fault. Detected faults are marked in black crosses, and the unobserved true values are marked in grey circles.

Table 1. Quantitative comparison of different algorithms on datasets with two simulated faults. For each dataset, we show the mean squared error (MSE), the log likelihood of the true data ($\log p(\mathbf{y} \mid \mathbf{x})$), and the true-positive and false-positive rates of detection for faulty points (TPR and FPR), respectively, with all methods permitted a 'burn-in' period of 50 points. The best value for each set of results is highlighted in bold.

	Bias dataset				"Flash-flood" dataset			
Method	MSE	$\log p(\mathbf{y} \mid \mathbf{x})$	TPR	FPR	MSE	$\log p(\mathbf{y} \mid \mathbf{x})$	TPR	FPR
FB	**0.024**	334	**0.997**	0.031	0.069	-5.77×10^3	**0.829**	0.016
XGP	0.037	**439**	0.982	**0.022**	**0.042**	$\mathbf{-1.52 \times 10^3}$	0.805	**0.012**
TGP	0.033	278	**0.997**	0.031	0.075	-8.29×10^3	**0.829**	0.083
MLH	0.940	-5.43×10^7	0.065	0.031	2.369	-2.27×10^7	0.045	0.262
EKF	0.060	-1.26×10^4	0.551	0.258	0.613	-1.81×10^4	0.169	0.768

to make predictions about even the first predictant), as the method is intended to be used. Clearly this allows the method a predictive advantage relative to sequential methods, and the multiple passes over the data effected by MLH cannot be readily applied to the sequential problem without requiring a great deal of expensive computation.

EKF: An autoregressive neural net trained with the extended Kalman filter to capture nonstationarity (de Freitas et al., 2000a;b). Again, (2) was used to identify and discard faulty data.

Table 1 displays quantitative measures of performance for the various algorithms on datasets with two simulated faults. We used simulated faults provided by Aquatic Informatics in order to have access to ground-truth. The two faults are common in water monitoring. The results in the table indicate that, in addition to superior predictive performance, our detection rates for the faulty points are generally excellent. The results reveal that our algorithm is competitive with exhaustive search. Our naïve approach to faults may, of course, suffer relative to better-informed models, but its probabilistic estimates provide a human operator with an indication as to whether more sophisticated analysis is necessary.

4. Conclusion

We have briefly presented an overview of some of the data challenges arising in water monitoring and sustainability. We demonstrated a novel algorithm, the "fault bucket," for managing time-series data corrupted by faults unknown ahead of time. The algorithm can also be used to detect other types of pattern, *e.g.* "fishkiller" events, and produces probabilistic predictions. This not only results in an increase in automation, but also in sensible uncertainty summaries to assist human operators. On the machine learning front, our chief contribution is a sequential method for marginalizing the faultiness of observations in a GP framework, allowing for fast, effective prediction in the presence of unknown faults and the simultaneous detection of faulty observations.

Acknowledgement

We would like to thank the management and employees of Aquatic Informatics, and in particular Touraj Farahmand, for their support in this project. All datasets have been provided by Aquatic Informatics. The data may be obtained by contacting Touraj Farahmand at tourajf@aquaticinformatics.com.

References

Chandola, V., Banerjee, A., and Kumar, V. Anomaly detection: A survey. ACM *Comput. Surv.*, 41:15:1–15:58, July 2009.

de Freitas, N., Niranjan, M., and Gee, A. H. Hierarchical Bayesian models for regularisation in sequential learning. *Neural Computation*, 12(4):933–953, 2000a.

de Freitas, N., Niranjan, M., and Gee, A. H. Dynamic learning with the EM algorithm for neural networks. *Journal of VLSI Signal Processing Systems*, 26(1/2): 119–131, 2000b.

Dereszynski, E. and Dietterich, T. G. Spatiotemporal models for anomaly detection in dynamic environmental monitoring campaigns. *ACM Transactions on Sensor Networks*, 2011.

Ding, S. X. *Model-based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools.* Springer, first edition, 2008.

Eciolaza, L., Alkarouri, M., Lawrence, N. D., Kadirkamanathan, V., and Fleming, P. J. Gaussian Process

Latent Variable Models for Fault Detection. In IEEE *Symposium on Computational Intelligence and Data Mining*, pp. 287—292, 2007.

Garnett, R., Osborne, M. A., Reece, S., Rogers, A., and Roberts, S. J. Sequential Bayesian Prediction in the Presence of Changepoints and Faults. *The Computer Journal*, 53, 2010.

Isermann, R. Model-based fault-detection and diagnosis – status and applications. *Annual Reviews in Control*, 29(1):71–85, 2005.

Kersting, K., Plagemann, C., Pfaff, P., and Burgard, W. Most likely heteroscedastic Gaussian process regression. In *Proceedings of the 24th international conference on Machine learning*, pp. 393–400. ACM, 2007.

Khan, S. and Madden, M. A survey of recent trends in one class classification. In Coyle, Lorcan and Freyne, Jill (eds.), *Artificial Intelligence and Cognitive Science*, volume 6206 of *Lecture Notes in Computer Science*, pp. 188–197. Springer Berlin / Heidelberg, 2010.

Markou, M. and Singh, S. Novelty detection: a review – Part 1: Statistical approaches. *Signal Processing*, 83(12):2481–2497, 2003.

Rasmussen, C. E. and Williams, C. K. I. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, USA, 2006.

Wagner, R. J. and US Geological Survey. *Guidelines and standard procedures for continuous water-quality monitors: Station operation, record computation, and data reporting*. US Department of the Interior, US Geological Survey, 2006.