Showcase #9

An NSF EPSCoR Track-2 RII Program

A Physics-informed Latent Variables of Corrosion Growth in Oil and Gas Pipelines

by Phat K. Huynh, Abdulsalam A. Alqarni, Om P. Yadav, Trung Q. Le

Background

Corrosion defects in pipeline systems is one of the significant threats to the structural integrity of buried pipelines worldwide. Therefore, the pipeline corrosion management is a crucial task, which includes an engineering assessment of corrosion defects called high-resolution inline inspections (ILI) to inspect the pipeline conditions and mitigate the effect of defects. To determine a re-inspection plan and a sound defect mitigation strategy, modeling the corrosion growth is substantially important. In the literature, methods to model corrosion growth in oil and gas pipelines can be categorized into three groups: deterministic, stochastic and machine-learning-based models. Power law model – one of the widely accepted deterministic models – has been used to estimate the time-dependent atmospheric corrosion. For predicting internal corrosion rate, the de Waard-Milliams model was proposed, which relates the corrosion rate with the CO_2 partial pressure and the temperature. However, those deterministic models cannot capture the stochastic nature of the corrosion process. To overcome it, various stochastic-process-based models have also been developed to capture the growth of corrosion defects on pipelines, such as inverse Gaussian processes, gamma processes , Markov chain models , and hierarchical Bayesian models. In recent years, there has been considerable interest in employing machine learning models to predict the corrosion growth of pipelines. However, the inherently stochastic and hidden-physics-driven nature of the corrosion growth process were not fully addressed by the previously developed models. This paper proposes a physics-informed latent variable model for nonstationary time series data.

Methodology

Consider a multivariate time-series data matrix $\boldsymbol{X}_{n\times T}$ that consists of n time series obtained from the sensors and uniformly sampled from time $t = 1$ to $t = T$, where X_{it} is the value of the i^{th} time series at time t, each time series $\{X_{i,t}\}_{t=1}^{T}$ represents the measurement of the same variable at n locations. A discrete-time multivariate nonstationary time series model was assumed to have the following form: $X_{i,t} = \underbrace{f\left(\boldsymbol{\xi}_i, t; \boldsymbol{\theta}_f\right)}_{deterministic} + \underbrace{g\left(\boldsymbol{\xi}_i, t; \boldsymbol{\theta}_g\right)}_{stochastic} + \underbrace{\varepsilon_{i,t}}_{measurement\ error}$ where $f(\cdot)$ and $g(\cdot)$ are the deterministic and the stochastic components of $X_{i,t}$ parametrized by $\boldsymbol{\theta}_f$ and $\boldsymbol{\theta}_g$ respectively $\boldsymbol{\xi}_i = \left[\xi_{i,1}, \xi_{i,2}, \dots \xi_{i,L}\right]$ represents the vector of latent variables for the i^{th} time series and $\varepsilon_{i,t} \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2, is the measurement error. For simplicity of the modeling process, we assumed $g(\cdot) \equiv \mathcal{N}\left(\mu_{i,t}(\boldsymbol{\xi}_i, , t; \boldsymbol{\nu}), \sigma_{i,t}^2(\boldsymbol{\xi}_i, , t; \boldsymbol{\varphi})\right)$, which models the non-stationarity of the time series $X_{i,t}$ (i.e., the change of stochastic process parameters). The true underlying model structure was depicted in *Figure 1*. To identify the latent variables of $\boldsymbol{\xi}_i$'s and discover the underling relationships between the time series and the latent variables, the agglomerative hierarchical clustering and physics-informed regression models were used.

Results and Discussion

In this case study, we simulated noisy high-resolution ILI data using a well-known corrosion growth model — power law function with initiation time coupled with a nonstationary Gaussian stochastic process. Here, we had a total of 150 time series for training the latent variable model, in which each soil type had 50 time series. Next, the deterministic component was separated from the stochastic component and noise by using the moving average with window length $k = 50$, the value of k was selected to optimally separate the deterministic and stochastic components. The implementation of physics-informed regression models with latent variables and long-term forecasting of time series using the latent variable model was demonstrated in *Figure 2* and *Figure 3*. For each time series in the training set, we performed moving average method to extract the deterministic component and computed the root-mean-square envelope of the stochastic component, which was demonstrated in *Figure 2*. Agglomerative hierarchical clustering was performed with the distance metric "chebychev" and the linkage method "weighted".

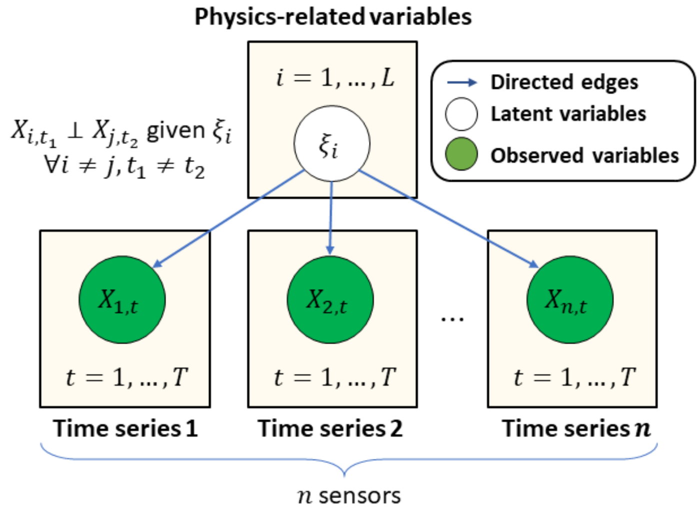

Figure 1. True underlying model structure with latent variables.

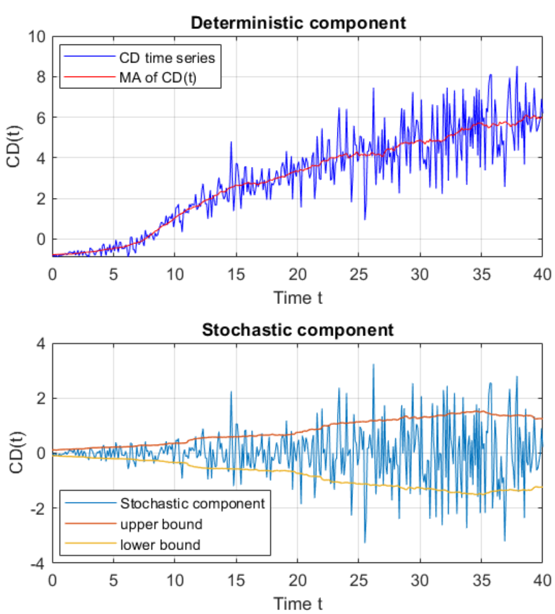

Figure 2. Decomposition of time series data into the deterministic component and stochastic component. The deterministic component was illustrated by red line in the first subplot. In the second plot, the root-mean-square envelope of the stochastic component was obtained.

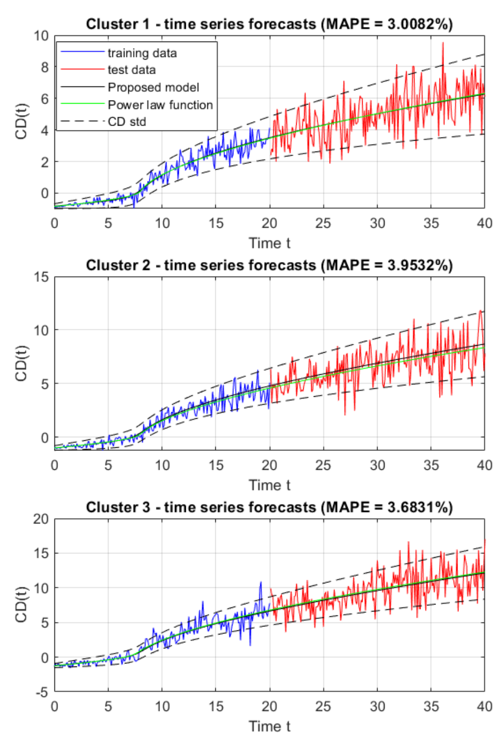

Figure 3. Long-term time series forecasts of 3 representative corrosion depth time series from three different clusters.

The showcase paper has been published in the 2023 Annual Reliability and Maintainability Symposium (RAMS) IEEE conference proceedings, please refer to this link to gain access to the paper:

https://ieeexplore.ieee.org/abstract/document/10088241