Reliability Monitoring Using Log Gaussian Process Regression

Martin Wayne

Mohammad Modarres

PSA 2013

Center for Risk and Reliability
University of Maryland
Department of Mechanical Engineering
College Park, MD 20742, U.S.A.
mwayne1@umd.edu; modarres@umd.edu

Traditional Regression and Bayesian Regression

- Traditional Regression assumes an underlying process (e.g., failure process described by the failure rate model λ(t)) which generates clean data. The goal is to describe the underlying model in the presence of noisy data.

- Bayesian Regression Specify the prior $P(H_\alpha)$ of a set of probabilistic models. The likelihood of H_α after observing data *D* is $P(D \mid H_\alpha)$

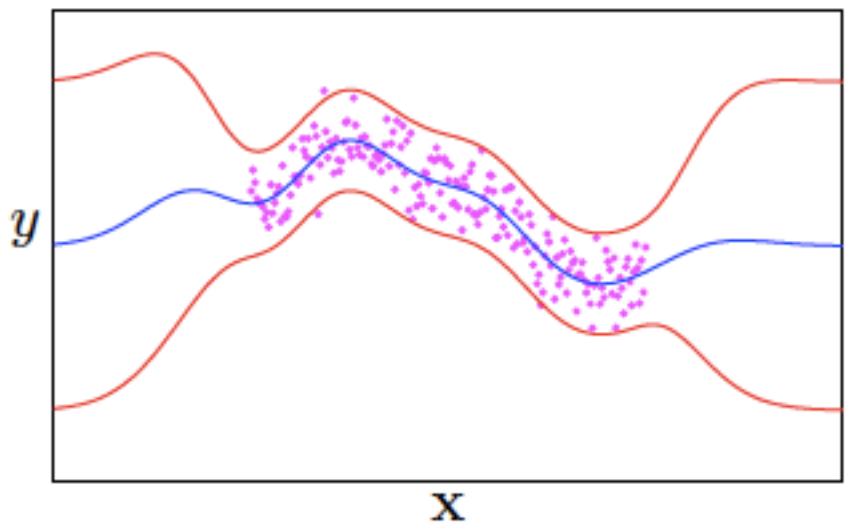
  The posterior probability of H_α is given by

$$P(H_\alpha \mid D) \propto P(H_\alpha) P(D \mid H_\alpha)$$

Prediction: $p(y \mid D) = \sum_\alpha P(y \mid H_\alpha) P(H_\alpha \mid D)$

GP Regression

- It is a nonlinear regression when you need to learn a function *f* with uncertainties from data $D = \{X, y\}$



Ref: Eurandom 2010, Z. Ghahramani

GP Regression (Cont.)

- A Gaussian process defines a distribution over functions p(f) which can be used for Bayesian regression

$$p(f|D) = p(f)p(D|f)/p(D)$$

- Gaussian processes (GPs) are parameterized by a mean function, $\mu(x)$, and a covariance function, or kernel, $K(x, x')$.
- The covariance matrix K is between all the pair of points x and x' and specifies a distribution on functions

$$p(f(x), f(x')) = \mathrm{N}(\mu, \Sigma)$$

$$\mu = \begin{bmatrix} \mu(x) \\ \mu(x') \end{bmatrix} \quad \Sigma = \begin{bmatrix} K(x,x) & K(x,x') \\ K(x',x) & K(x',x') \end{bmatrix}$$

and similarly for $p(f(x_1), \ldots, f(x_n))$ where now μ is an n x 1 vector and Σ is an n x n matrix

Imagine observing a data set $\mathcal{D} = \{(\mathbf{x}_i, y_i)_{i=1}^n\} = (\mathbf{X}, \mathbf{y})$.

Model:

$$\begin{aligned} y_i &= f(\mathbf{x}_i) + \epsilon_i \\ f &\sim \mathrm{GP}(\cdot|0, K) \\ \epsilon_i &\sim \mathrm{N}(\cdot|0, \sigma^2) \end{aligned}$$

Prior on f is a GP, likelihood is Gaussian, therefore posterior on f is also a GP.

We can use this to make predictions

$$p(y_*|\mathbf{x}_*, \mathcal{D}) = \int p(y_*|\mathbf{x}_*, f, \mathcal{D})\, p(f|\mathcal{D})\, df$$

We can also compute the marginal likelihood (evidence) and use this to compare or tune covariance functions

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|f, \mathbf{X})\, p(f)\, df$$

GP Regression (Cont.)

- Set of random variables, any finite collection of which follows joint Gaussian distribution
- Gaussian distribution specified by mean, $\mu(x)$, and kernel function, $k(x_1,x_2)$
- May also be defined using standard linear model form

$$y = f(x) + \varepsilon$$

Kernel Function

- Kernel function = covariance function
- Kernel function determines correlation between data points
  - Can model trends in data such as periodicity
- Popular example is squared-exponential kernel function

$$k(x_1, x_2) = \sigma^2 \exp\left[\frac{\|x_1 - x_2\|^2}{2l^2}\right]$$

- l is the characteristic length-scale of the process (showing, "how far apart" two points have to be for X to change significantly
- Used to develop overall covariance matrix *K* for vector of data

Kernel Function Hyperparameters

- Kernel functions contain unknown hyperparameters
- Hyperparameters can be chosen by maximizing log-marginal likelihood given by

$$\log p(y / X) = -\frac{1}{2} y^T K^{-1} y - \frac{1}{2} \log K - \frac{n}{2} \log 2\pi$$

- Achieved using conjugate gradient optimization technique
  - Built-in option in available software packages
- Chi-square goodness-of-fit can also be applied to test data to assess model

Log Gaussian Processes

- Observed data Y are strictly positive
- Assume log(Y) is normally distributed

$$\begin{bmatrix} \log(Y) \\ \log(f^*) \end{bmatrix} \sim N\left(\begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} K(X,X) & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix} \right)$$

- Can use conditional probability to determine prediction for f^*

$$\log(f^*)/\log(Y) \sim N(\mu,\Sigma)$$

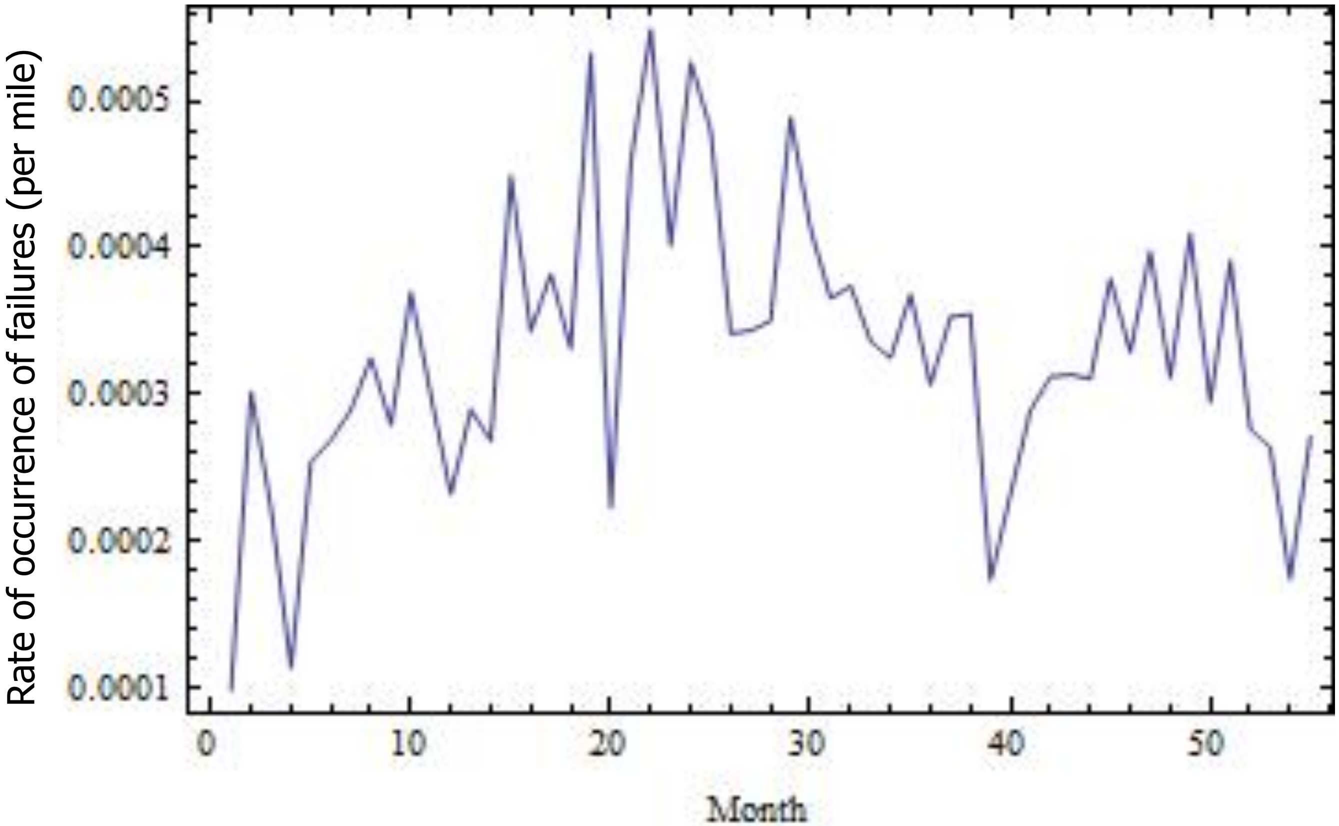
$$\mu = b + K(X^*,X)K(X,X)^{-1}(\log(Y)-b)$$

$$\Sigma = K(X^*,X^*) - K(X^*,X)^{-1}K(X,X^*)K(X,X^*)$$

- Inverse transform can be used to proper domain

Application to Fleet of Vehicles

- Use Log GPR to model rate-of-occurrence of failures per month for fleet of vehicles



Covariance Function Choices

- Data appear to have periodic behavior along with noise
- Examine kernel function alternatives to describe correlation within data
  - Combinations of kernel functions are also kernel functions
  - Options include Squared Exponential (SE), Noise, Periodic, Polynomial, etc.
- Can use negative log-likelihood to discriminate between possible kernel functions
  - Smaller values indicate higher likelihood
  - Better description of data

Kernel Function Likelihood

Kernel Function Alternative	Description	Negative Log Likelihood
1	SE, Periodic, Noise	7.84
2	SE, Noise	8.54
3	Noise	15.83
4	SE, Periodic	8.13
5	SE	8.54
6	SE, Polynomial, Noise	8.54
7	Polynomial	13.33
8	Rational quadratic, Noise	8.51
9	Rational quadratic	8.51

Sum of squared exponential, periodic, and noise kernels yields highest likelihood and best description of data

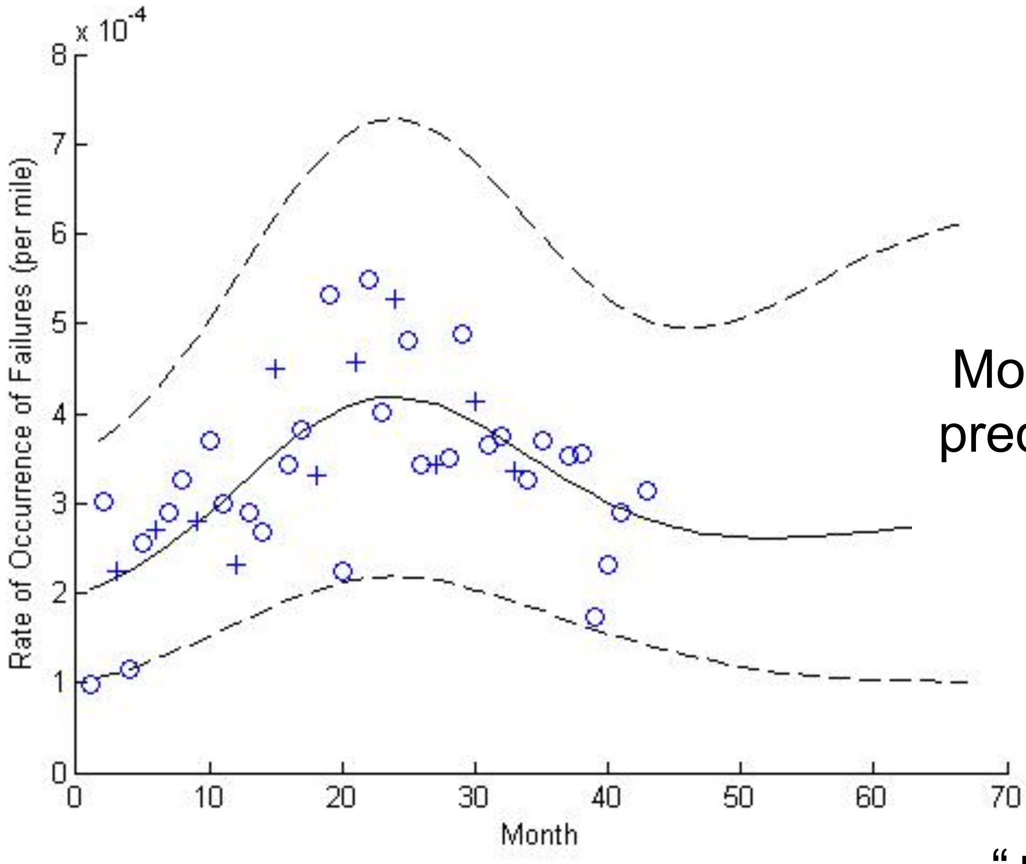
Covariance Function and Hyperparameters

$$k(x_1, x_2) = \theta_1^{\ 2} \exp\left[-\frac{(x_1 - x_2)^2}{2\theta_2^{\ 2}}\right] + \theta_3^{\ 2} \exp\left[-\frac{2\sin^2\left(\pi \frac{(x_1 - x_2)}{\theta_4}\right)}{\theta_5^{\ 2}}\right] + \sigma_n \partial(x_1, x_2)$$

$$\partial(x_1, x_2) = \begin{matrix} 1, \ x_1 = x_2 \\ 0, \quad o.w. \end{matrix}$$

Parameter	Value
θ_1 (SE 1)	4.38
θ_2 (SE 2)	0.31
θ_3 (Periodic 1)	0.13
θ_4 (Periodic 2)	1.03
θ_5 (Periodic 3)	0.26
σ_n (Noise)	0.12

Predicted Result

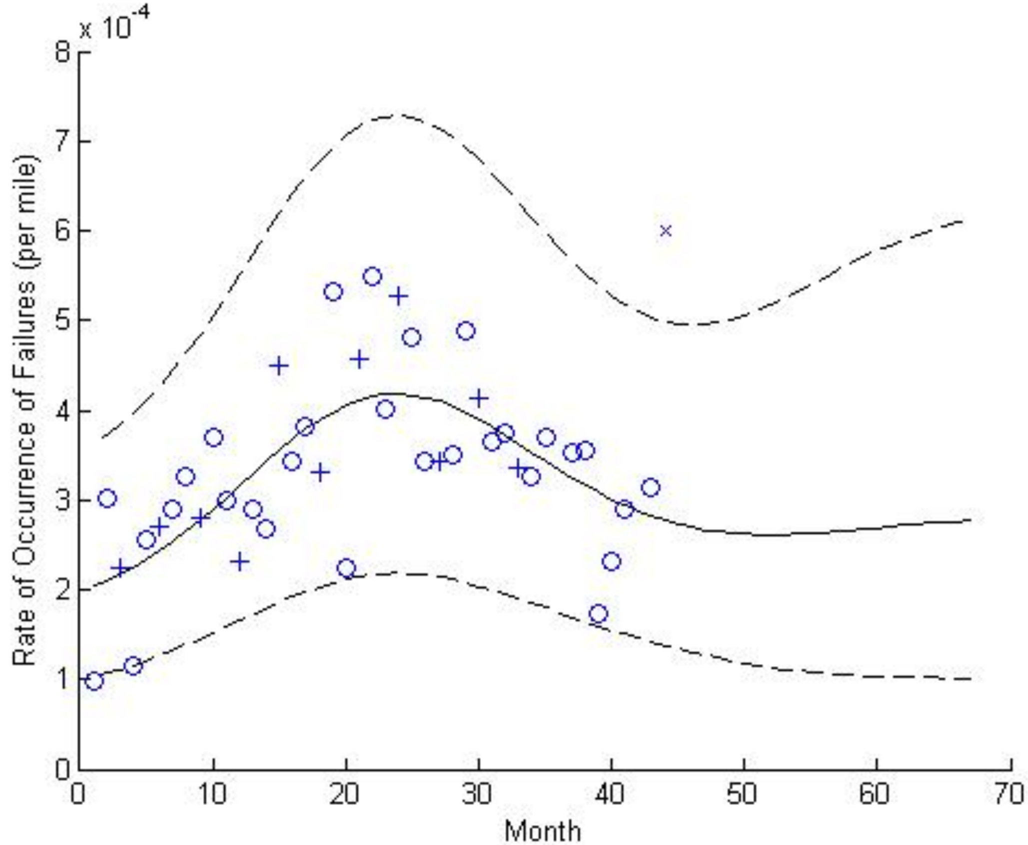


Model provides probabilistic prediction of failures in future months

"+" indicates test data used in goodness-of-fit test

**Dashed lines indicate 95% probability interval

Anomaly Detection Example



**Dashed lines indicate 95% probability interval

Conclusions and Possible Extensions

- LOG GP Regression is powerful technique for modeling complex nonlinear behavior
  - Provides probabilistic indication of reliability problems vs. typical trends within data
- Can be extended to model more complex nonlinear relationships
  - Only require appropriate kernel function
- Can also handle multidimensional inputs
  - Systems in different locations, different ages, etc.