Evaluating Gaussian processes for sparse irregular
spatio-temporal data

Mehmet Süzen * and Abed Ajraou

Abstract
A practical approach to evaluate performance of a Gaussian process regression models
(GPR) for irregularly sampled sparse time-series is introduced. The approach entails con-
struction of a secondary autoregressive model using the fine scale predictions to forecast a
future observation used in GPR. We build different GPR models for Ornstein-Uhlenbeck
and Fractional processes for simulated toy data with different sparsity levels to assess the
utility of the approach.

Keywords: Gaussian Processes, autoregressive models.

1. Introduction

Time-series analysis is very common subject that manifest itself in sciences and in industry
Hamilton (1994)]. Temporal data rarely available in regular intervals and with sufficient
sample size, i.e., sparse, irregularly occurred and noisy [Richards et al. (2011)]. In these
circumstances, standard modelling techniques would not be appropriate. Gaussian proce-
ses provide a powerful alternative [Williams and Rasmussen (2006)], where a prior knowl-
edge can be used without any restrictions on regularity or sparsity on the temporal data
[Roberts et al. (2013)]. But a measure of goodness of fit would be an issue in this setting.
A usual approach is to measure goodness of fit by comparing the results based on a gold
standard result. Here we propose an approach to measure performance of GP without re-
sorting to a gold standard result in sparse time-series via building a secondary model based
on the resulting.

2. Gaussian Processes

A short sketch of the machinery of the Gaussian Processes [Williams and Rasmussen (2006)]
is presented as follows. The starting point for Gaussian process is to define an arbitrary
function that explains the outcome with a variance \( \sigma_y \) and noise \( \epsilon_t \),

\[
y(t) = f(t) + \sigma_y \epsilon_t
\]

The primary approach in GP is that training time-series \((y, x)\) and \((y^*, x^*)\) the time-
series that is to be learned can be expressed in a joint distribution, which is a multivariate
Gaussian distribution,

* Correspondence E-mail: mehmet.suzen@physics.org
\begin{align*}
p(y, y^*) &= \mathcal{N}\left( \begin{bmatrix} x \\ x^* \end{bmatrix}, \begin{bmatrix} K_{xx} & K_{xx^*} \\ K_{xx^*} & K_{xx^*} \\ \end{bmatrix} \right) \end{align*}

where \( x \in \mathbb{R}^{nb} \) and \( x^* \in \mathbb{R}^{md} \), where \( n \) and \( m \) are number of rows and \( d \) is the number of predictors, the Kernel matrices can be computed as follows, in multivariate setting,

\[ K_{i,j}(x_i, x_j) = -\sigma^2 \exp\left( -\beta \sum_{k=1}^{d} \frac{(x_{ik} - x_{jk})^2}{l_d^2} \right)^{\alpha_d} \]

\( i = 1, \ldots, n \) and \( j = 1, \ldots, m \), where \( d \) is the number of features or variates at each time point. Kernel choice here is not unique but this is a generalised form of the square exponential kernel. Hyperparameters \( (\sigma^2, \beta, l, \alpha) \) can be interpreted as signal variance, scaling factor, length scale and roughness on the time-series respectively. Hence, GP as a function approximation can be obtained in a closed form, the mean function and the covariance matrix,

\[ L = (K_{xx} + \sigma^2 I)^{-1} \]

\[ y^* = K_{x^*x}L^{-1}y \]

\[ y^*_{cov} = K_{x^*x^*} - K_{x^*x}L^{-1}K_{xx^*} \]

Note that \( y^* \) is the observations that is to be learned from the training data.

### 3. Evaluation Technique

A variable of interest \( y \) appears regularly over time. A temporal evolution of \( y \) can be expressed with ordered set \( \Omega^* = (t_i^*, y_i^*) \), where \( i = 1, \ldots, n \). Only a subset of \( \Omega^* \) may be observed in irregularly spaced intervals and not too frequently, i.e., sparse. These sparse observations of \( y \) appear in the subset \( \Omega = (t_j, y_k) \), where \( \{k \in \{1, \ldots, n\}\} \), as an ordered sequence but irregular, i.e., irregular time-series. Building a Gaussian Process model to construct the original time series \( \Omega^* \) using the partial information \( \Omega \) is one of the most promising approach for sparse irregular temporal data [Richards et al. (2011)]. The resulting series can be denoted by \( \Omega^*_{gp} = (t_i^*, Y_i^*) \).

Performance of the resulting reconstruction \( \Omega^*_{gp} \) usually measured against a gold standard method [Roberts et al. (2013)]. We propose using a secondary Autoregressive Model (AR) based on the resulting set \( \Omega^*_{gp} \) is proposed. Construction of \( n - 1 \) different autoregressive models to predict observation points \( y_j \). Procedure is as follows.

1. Fixed a horizon \( h \), for Autoregressive model prediction.
2. Select the \( m - 1 \) subsets of \( \Omega^*_{gp} \) each up to a point \( y_j, j_k - h \), denoting each subset, \((\Omega_{gp}^*)_k \), where \( k = 2, \ldots, m \).
3. Build an AR model for each \((\Omega_{gp}^*)_k \) and predict next \( y_k \) as \( y_k^{ar} \).
4. Mean Absolute Percent Error (MAPE-AR) can be computed, \( M = \frac{1}{m-1} \sum_{k=2}^{m}(y_k - \frac{y_k^{ar}}{y_k})/y_k \)

MAPE-AR value will quantify the goodness-of-fit for GP regression without resorting to a gold standard method.
3.1 Simulated Data

A pair of toy data is generated, using generalised form of the square exponential kernel, Ornstein-Uhlenbeck and Fractional process with Kernel hyperparameters \((\alpha, \beta, l, \sigma)\), \((1.0,1.0,2.0,1.0)\) and \((1.3,1.0,2.0,1.0)\) respectively. Simulated data contains 351 observations points with regular time spacing of 0.02. New subset of observations generated by randomly selecting 3, 5 and 7 percent of the simulated data, at least 5 time-steps apart. This constitutes different sparsity levels.

3.2 Experiments

We fit Gaussian Process on these sparse data sets using square exponential kernel. Results for 3 percent sparse data sets are shown on the Figure 1.

Using seasonal ARIMA\((1,1,1)(1,1,1)\) as a secondary autoregressive model, MAPE-AR measure is summarized in Figure 2, for demonstration purposes.
4. Summary

A technique to measure goodness-of-fit in Gaussian processes for sparse temporal data is proposed based on building secondary autoregressive model to construct the regularly spaced data. In our empirical investigation we have demonstrated the utility of the approach using simulated data.

References

James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

Joseph W Richards, Dan L Starr, Nathaniel R Butler, Joshua S Bloom, John M Brewer, Arien Crellein-Quick, Justin Higgins, Rachel Kennedy, and Maxime Rischard. On machine-learned classification of variable stars with sparse and noisy time-series data. *The Astrophysical Journal*, 733(1):10, 2011.

Stephen Roberts, M Osborne, M Ebden, Steven Reece, N Gibson, and S Aigrain. Gaussian processes for time-series modelling. *Phil. Trans. R. Soc. A*, 371(1984):20110550, 2013.

Christopher KI Williams and Carl Edward Rasmussen. Gaussian processes for machine learning. *the MIT Press*, 2(3):4, 2006.