Statistical Models for Unequally Spaced Time Series

Emre Erdogan * Sheng Ma † Alina Beygelzimer † Irina Rish †

January 16, 2005

Abstract

Irregularly observed time series and their analysis are fundamental for any application in which data are collected in a distributed or asynchronous manor. We propose a theoretical framework for analyzing both stationary and non-stationary irregularly spaced time series. Our models can be viewed as extensions of the well known autoregression (AR) model. We provide experiments suggesting that, in practice, the proposed approach performs well in computing the basic statistics and doing prediction. We also develop a resampling strategy that uses the proposed models to reduce irregular time series to regular time series. This enables us to take advantage of the vast number of approaches developed for analyzing regular time series.

1 Introduction

Unevenly sampled time series are common in many real-life applications when measurements are constrained by practical conditions. The irregularity of observations can have several fundamental reasons. First, any event-driven collection process (in which observations are recorded only when some event occurs) is inherently irregular. Second, in such applications as sensor networks, or any distributed monitoring infrastructure, data collection is distributed, and collection agents cannot easily synchronize with one another. In addition, their sampling intervals and policies may be different. Finally, in many applications, measurements cannot be made regularly or have to be interrupted due to some events (either foreseen or not).

Time series analysis has a long history. The vast majority of methods, however, can only handle regular time series and do not easily extend to unevenly sampled data. Continuous time series models can be directly applied for the problem (e.g., [5]), but they tend to be complicated (mostly due to the difficulty of estimating and evaluating them from discretely sampled data) and do not provide a satisfying solution in practice.

In data analysis practice, irregularity is a recognized data characteristic, and practitioners dealt with it heuristically. It is a common practice to ignore the times and treat the data as if it were regular. This can clearly introduce a significant bias leading to incorrect predictions. Consider, for example, the return of a slowly, but consistently changing stock, recorded first very frequently and then significantly less frequently. If we ignore the times, it would appear as if the stock became more rapidly changing, thus riskier, while in fact the statistical properties of the stock did not change.

Many basic questions that are well understood for regular time series, are not dealt with for unequally spaced time series. The goal of this paper is to provide such a theoretical foundation. At the very least, we would like to be able to compute the basic statistics of a given time series (e.g., its mean, variance, autocorrelation), and predict its future values.

Our contributions can be summarized as follows:

- We propose two statistical models for handling irregularly sampled time series. The first model assumes stationarity, and can be viewed as a natural extension of the classical AR(1) model for regular time seres. The second model relaxes the stationarity assumption by allowing a more general dependence on the current time, time difference, and the state of the process at a given time.
- We show how to efficiently estimate the parameters of both models using the maximum likelihood method.
- We propose and give solutions for two strategies based on the proposed models. The first strategy is to compute the basic statistics (e.g., autocorrelation function) and do prediction directly from the models. This approach does not easily extend to non-linear time series and multiple

*Columbia University, New York, NY, ee168@columbia.edu

†IBM T. J. Watson Research Center, Hawthorne, NY, {shengma,beygel,rish}@us.ibm.com

irregular time series. The second strategy avoids these problems by using the model to convert irregular time series to regular time series by resampling. The reduction reduces the problem to a problem that has already been thoroughly analyzed and for which many approaches are available. The resampling approach can be found in the full version of the paper [4].

Related work A vast amount of techniques were developed for analyzing regularly sampled time series. Unfortunately, most of these techniques do not take into account sampling times, and cannot be easily generalized to irregularly sampled time series.

As a simple heuristic, we can ignore the times and treat the values as regularly sampled. Obviously, if there is enough structure and irregularity in sampling times, we lose a lot of information about the dynamics of the system.

Many techniques have been proposed to handle time series with missing data, which in the limit can be viewed as irregularly sampled [8]. One approach is to interpolate the data to equally spaced sampling times. A survey of such interpolation techniques can be found in [1]. While this is a reasonable heuristic for dealing with missing values, the interpolation process typically results in a significant bias (e.g., smooths the data) changing the dynamics of the process, thus these models can not be applied if the data is truly unequally spaced. Another problem is that there is little understanding of which which interpolation method does best on a given dataset.

A number of authors suggested to use continuous time diffusion processes for the problem. Jones [6] proposed a state-space representation,and showed that for Gaussian inputs and errors, the likelihood of data can be calculated exactly using Kalman filters. A nonlinear (non-convex) optimization can then be used to obtain maximum likelihood estimates of the parameters. Brockwell [2] improved on this model and suggested a continuous time ARMA process driven by the Lévy process. His models, however, assume stationarity, and parameter estimation is done via non-convex optimization using Kalman filtering limiting the practical use of these models.

2 Background: Basic Definitions

A *time series* $X(t)$ is an ordered sequence of observations of a variable X sampled at different points t over time. Let the sampling times be t_0, t_1, $\cdots$, t_n satisfying $0 \leq t_0 < t_1 < \ldots < t_n$. If the time points are equally spaced (i.e., $t_{i+1} - t_i = \Delta$ for all $i = 0, ..., n-1$, where $\Delta > 0$ is some constant), we call the time series *regularly sampled*. Otherwise, the sequence of pairs $\{X(t_i), t_i\}$ is called an *irregularly sampled time series.*

Definition 1. (AUTOCOVARIANCE [3]) *For a process* $\{X(t), t \in T\}$ *with* $\mathbf{var}(X(t)) < \infty$ *for each* $t \in T$, *the* auto-covariance function $\mathbf{cov}_X(X(t), X(s))$, *for* $t, s \in T$, *is given by*

$$\mathbf{E}\left[(X(t) - \mathbf{E}[X(t)])(X(s) - \mathbf{E}[X(s)])\right].$$

Definition 2. (STATIONARITY [3]) *A time series* $\{X(t), t \in T\}$ *is said to be* stationary *if*

- $\mathbf{E}[|X(t)^2|] < \infty$, $\mathbf{E}[X(t)] = c$, *for all* $t \in T$,
- $\mathbf{cov}_X(X(t), X(s)) = \mathbf{cov}_X(X(t+h), X(s+h)))$ *for all* $t, s, h \in T$.

In other words, a stationary process is a process whose statistical properties do not vary with time.

Definition 3. (AR(1) PROCESS [3]) *A regularly sampled process* $\{X(t), t = 0, 1, 2, \ldots\}$ *is said to be an* AR(1) process *if* $\{X(t)\}$ *is stationary and if for every* t,

$$X(t) = \theta X(t-1) + \sigma \epsilon_t,$$

where $\{\epsilon_t\}$ *is a series of random variables with* $\mathbf{E}(\epsilon_t) = 0$, $\mathbf{var}(\epsilon_t) = 1$, *and* $\mathbf{cov}(\epsilon_t, \epsilon_s) = 0$ *for every* $t \neq s$. Notice that by recursive substitution, we can write $X(t+h)$ for any positive integer h in terms of $X(t)$ as

$$X(t+h) = \theta^h X(t) + \sigma \sum_{j=0}^{h-1} \theta^j \epsilon_{t+1+j}.$$

The process $\{\epsilon_t\}$ is also called "white noise". We will assume that $\epsilon_t \sim N(0, 1)$ for all t.

3 Overview of our Approach

Suppose that our irregularly sampled time series $Y(t)$ can be decomposed as

$$Y(t) = a(t) + X(t), \tag{3.1}$$

where $a(t)$ is a slowly changing deterministic function called the "trend component" and $X(t)$ is the "random noise component".

In general, one can observe only the values $Y(t)$. Therefore, our first goal is to estimate the deterministic part $a(t)$, and extract the random noise component $X(t) = Y(t) - a(t)$. Our second goal is to find a satisfactory probabilistic model for the process $X(t)$, analyze its properties, and use it together with $a(t)$ to predict $Y(t)$.

Let $\{y(t_i), t_i\}$, $i = 0, 1, \ldots, n$ be a sample of $Y(t)$. We assume that $a(t)$ is a polynomial of degree p in t,

$$a_p(t) = \rho_0 + \rho_1 t + \rho_2 t^2 +, ... + \rho_p t^p, \tag{3.2}$$

where p is a nonnegative integer and $\boldsymbol{\rho} = [\rho_0; \rho_1; ...; \rho_p]$ is a vector of coefficients. A more general structure can also be used. The vector $\boldsymbol{\rho}$ can be estimated using the least squares method, by choosing the vector minimizing $\sum_{i=0}^{n}(y(t_i) - a(t_i))^2$. This is straightforward, and we will turn to developing a statistical model for $X(t)$.

We propose two parametric statistical models for analyzing $X(t)$. The first model, described in the next section, is a direct extension of the classical AR(1) model given in Definition 3 and assumes that $X(t)$ is stationary. The second model, presented in Section 5, relaxes the stationarity assumption and allows a more general dependence of $X(t + \Delta)$ on t, Δ, and $X(t)$.

4 A Statistical Model for Stationary $X(t)$

Suppose that $X(t)$ obtained from $Y(t)$ after removing the trend component $a_p(t)$ is a stationary process. We define an irregularly sampled stationary AR(1) process as follows.

Definition 4. (IS-AR(1) PROCESS) *A time series $X(t)$ is an irregularly sampled stationary AR(1) process if it is stationary and if for every t and $\Delta \geq 0$,*

$$X(t + \Delta) = \theta^{\Delta} X(t) + \sigma_{\Delta}\epsilon_{t+\Delta}, \tag{4.3}$$

where $\epsilon_t \sim N(0, 1)$ and $\mathbf{cov}(\epsilon_t, \epsilon_s) = 0$ for every $t \neq s$ and $\sigma_{\Delta}^2 = \sigma^2(\frac{1-\theta^{2\Delta}}{1-\theta^2})$ for some $\sigma > 0$.

If the times are regularly spaced, IS-AR(1) can be reduced to the original AR(1) process by observing that $\sigma \sum_{j=0}^{h-1} \theta^j \epsilon_{t+1+j} \sim N(0, \sigma^2(\frac{1-\theta^{2h}}{1-\theta^2}))$ and comparing with Equation 4.3.

4.1 Parameter estimation

In this section, we show how to estimate parameters θ and σ, given a set of observations $\{x(t_0), x(t_1), \ldots, x(t_n)\}$ of $X(t)$.

Define $\Delta_i = t_{i+1} - t_i$ for all $i = 0, \ldots, n-1$. We can assume that all $\Delta_i \geq 1$; otherwise, we can rescale each Δ_i by $\min_i\{\Delta_i\}$.

Since $\mathbf{E}[\epsilon_t] = 0$ and $\mathbf{cov}(\epsilon_t, \epsilon_s) = 0$ for all $t \neq s$, we can estimate θ by the least squares method. We need to find $\widehat{\theta} \in (-1, 1)$ minimizing $\sum_{i=0}^{n-1}(X(t_{i+1}) - \theta^{\Delta_i}X(t_i))^2$. Since $\Delta_i \geq 1$ for all i, this sum is a convex function of θ that can be efficiently minimized using convex optimization.

To estimate σ, we set $z_i = x(t_{i+1}) - (\widehat{\theta})^{\Delta_i}x(t_i)$. By Definition 4, we have $z_i \sim N(0, \sigma_{\Delta_i}^2)$ where $\sigma_{\Delta_i}^2 = \sigma^2(\frac{1-(\widehat{\theta})^{2\Delta_i}}{1-(\widehat{\theta})^2})$. We can thus estimate σ by maximizing the Gaussian likelihood of the residuals $z(t_0), \ldots, z(t_{n-1})$ at times $t_0, t_1, \ldots, t_{n-1}$. The maximum likelihood estimator of σ is given by

$$\widehat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}\frac{(x(t_{i+1}) - (\widehat{\theta})^{\Delta_i}x(t_i))^2}{\rho_i}}, \tag{4.4}$$

where $\rho_i = (\frac{1-\widehat{\theta}^{2\Delta}}{1-\widehat{\theta}^2})$ for all i. The derivation is omitted due to page limit (see [4]).

4.2 Prediction using the IS-AR(1) model

We first establish conditions for $X(t_0)$ under which $X(t)$ is a stationary process. Then assuming the stationarity of $X(t)$, we derive equations for one-step prediction and the auto-covariance function. We can assume without loss of generality that $t_0 = 0$.

Using Equation 4.3, independence of ϵ_t and $X(0)$, and the fact that $\mathbf{E}[\epsilon_t] = 0$ for all t, we can express $\mathbf{E}[X(t)], \mathbf{var}[X(t)]$, and $\mathbf{cov}[X(t), X(t+\Delta)]$ in terms of $X(0)$ as follows.

$$\mathbf{E}[X(t)] = \theta^t \mathbf{E}[X(0)] \tag{4.5}$$

$$\mathbf{var}[X(t)] = \theta^{2t}\,\mathbf{var}[X(0)] + \sigma^2\frac{1-\theta^{2t}}{1-\theta^2} \tag{4.6}$$

$$\mathbf{cov}[X(t), X(t+\Delta)] = \theta^{\Delta}\,\mathbf{var}[X(0)] \tag{4.7}$$

PROPOSITION 4.1. *Assume that $\mathbf{E}[X(t)^2] < \infty$ and $\mathbf{E}[X(0)] = 0$. Then $X(t)$ in Definition 4 is a stationary process if $\mathbf{var}[X(0)] = \frac{\sigma^2}{1-\theta^2}$ and $\mathbf{cov}_X(\Delta) = \mathbf{cov}[X(t), X(t+\Delta)] = \theta^{\Delta}\frac{\sigma^2}{1-\theta^2}$.*

Proof. For $\mathbf{var}[X(0)] = \frac{\sigma^2}{1-\theta^2}$, Equation 4.6 gives

$$\mathbf{var}[X(t)] = \theta^{2t}\frac{\sigma^2}{1-\theta^2} + \sigma^2\frac{1-\theta^{2t}}{1-\theta^2} = \frac{\sigma^2}{1-\theta^2},$$

yielding

$$\mathbf{cov}[X(t), X(t+\Delta)] = \theta^{\Delta}\,\mathbf{var}[X(t)] = \theta^{\Delta}\frac{\sigma^2}{1-\theta^2}.$$

Since $\mathbf{cov}[X(t), X(t+\Delta)]$ does not depend on t, $X(t)$ is stationary. ∎

A one-step predictor of $X(t+\Delta)$ given $X(t)$ for any $\Delta > 0$ is given by the conditional expectation of $X(t+\Delta)$ (using Equation 4.5):

$$\widehat{X}(t+\Delta) = \mathbf{E}[X(t+\Delta)|X(t)] = \theta^{\Delta}X(t).$$

4.3 Analyzing $Y(t)$ with a Stationary Component $X(t)$:

The following procedure can be used for estimating the auto-covariance function $\mathbf{cov}_Y(\Delta)$ of irregularly sampled time series $Y(t)$ and for predicting $Y(t+\Delta)$ given $Y(t)$. First, we fit a polynomial $a_p(t)$ to $Y(t)$ as described before, and set $X(t) = Y(t) - a_p(t)$. We then estimate θ as $\widehat{\theta} =$

$\arg\min_\theta \sum_{i=0}^{n-1}(X(t_{i+1}) - \theta^{\Delta_i} X(t_i))^2$, and σ using $\widehat{\sigma}$ in Equation (4.4). Since $a_p(t)$ is deterministic, set $\mathbf{cov}_Y(\Delta) = \mathbf{cov}_X(\Delta) = \theta^\Delta \frac{\sigma^2}{1-\theta^2}$. Finally, prediction is given by

$$\begin{aligned}\widehat{Y}(t+\Delta) &= a_p(t+\Delta) + \widehat{X}(t+\Delta)\\ &= a_p(t+\Delta) + \mathbf{E}[X(t+\Delta)|X(t)]\\ &= a_p(t+\Delta) + \theta^\Delta X(t).\end{aligned}$$

5 A model for non-stationary $X(t)$

The model introduced in the previous section assumes that $X(t)$ is stationary. Mean and variance of a stationary process are time independent, and the covariance between any two observations depends only on their time difference. This allowed us to derive a simple expression for the auto-covariance function. In practice, however, one may not have stationarity – statistical properties of $X(t)$ may vary with time. Thus it is natural to focus on estimating these properties in the near future instead of trying to obtain some global, time-independent values. To achieve this goal, we model $X(t+\Delta)$ as a function of t, Δ, and $X(t)$, plus a random noise whose variance also depends on t, Δ, and $X(t)$. As before, after fitting a polynomial of degree p to $Y(t)$ we get $X(t) = Y(t) - a_p(t)$.

5.1 General IN-AR(1) Process Let $\boldsymbol{\theta} \in \mathbf{R}^m$ be an m-dimensional drift parameter vector and $\sigma \in \mathbf{R}$ be a scalar variance parameter. Let $\boldsymbol{\alpha}(\Delta, t, X(t)) : [0,\infty)\times[0,\infty)\times\mathbf{R} \to \mathbf{R}^m$ and $\beta(\Delta, t, X(t)) : [0,\infty)\times[0,\infty)\times\mathbf{R} \to \mathbf{R}$ be functions of Δ, t, and $X(t)$.

Definition 5. (IN-AR(1) PROCESS) *An irregularly sampled non-stationary IN-AR(1) process is defined as*

$$\begin{aligned}X(t+\Delta) = X(t) &+ \boldsymbol{\theta}^T\boldsymbol{\alpha}(\Delta, t, X(t))\\ &+ \sigma\beta(\Delta, t, X(t))\epsilon_{t+\Delta},\end{aligned}$$

where $\epsilon_{t+\Delta} \sim N(0,1)$, $\mathbf{cov}(\epsilon_t, \epsilon_s) = 0$ for all $t \neq s$, $\boldsymbol{\theta}$ is the vector of drift parameters, $\boldsymbol{\alpha}(\Delta, t, X(t))$ is the drift function, σ is the variance parameter, and $\beta(\Delta, t, X(t))$ is the variance function. In addition, if $\Delta = 0$ then the functions $\boldsymbol{\alpha}$ and β satisfy

$$\boldsymbol{\alpha}(0, t, X(t)) = \mathbf{0} \text{ and } \beta(0, t, X(t)) = 0.$$

Since the above condition is the only assumption on the structure of $\boldsymbol{\alpha}$ and β, the model covers a wide range of irregularly sampled time series.

5.2 Parameter Estimation Since ϵ_t and ϵ_s are independent for all $t \neq s$, the distribution of $X(t_{i+1})$ given $X(t_i)$ is normal with mean $X(t_i) + \boldsymbol{\theta}^T\boldsymbol{\alpha}(\Delta_i, t_i, X(t_i))$ and variance $\beta^2(\Delta_i, t_i, X(t_i))\sigma^2$. Therefore, $\boldsymbol{\theta}$ and σ can be estimated by maximizing the Gaussian likelihood of observations $x(t_1), \ldots, x(t_n)$ at times $t_1, ..., t_n$.

PROPOSITION 5.1. *The maximum likelihood estimators of $\boldsymbol{\theta}$ and σ are given by*

$$\begin{aligned}\widehat{\boldsymbol{\theta}} &= \left(\sum_{i=1}^n \frac{\boldsymbol{\alpha}(\Delta_i, t_i, x(t_i))\boldsymbol{\alpha}^T(\Delta_i, t_i, x(t_i))}{\beta^2(\Delta_i, t_i, x(t_i))}\right)^{-1}\\ &\qquad\cdot \sum_{i=1}^n \frac{(x(t_{i+1}) - x(t_i))\boldsymbol{\alpha}(\Delta_i, t_i, x(t_i))}{\beta^2(\Delta_i, t_i, x(t_i))},\\ \widehat{\sigma} &= \sqrt{\frac{1}{n}\sum_{i=1}^n \frac{(x(t_{i+1}) - x(t_i) - \widehat{\boldsymbol{\theta}}^T\boldsymbol{\alpha}(\Delta_i, t_i, x(t_i)))^2}{\beta^2(\Delta_i, t_i, x(t_i))}}.\end{aligned}$$

Proof. See the full version of the paper [4].

The functions $\boldsymbol{\alpha}(\Delta, t, X(t))$ and $\beta(\Delta, t, X(t))$ can be chosen by first selecting a set of candidate functions (which can be based on data generation process, some preliminary analysis, or interaction with domain experts). Proposition 5.1 can then be used to estimate the parameters for each pair of candidate functions, choosing the pair that gives the best fit to the data. One can also use various greedy search-based methods in more general families of candidate functions.

5.3 Prediction using the general IN-AR(1) model Since $a_p(t)$ is deterministic, we only need to predict $X(t+\Delta)$. We have

$$\begin{aligned}\widehat{Y}(t+\Delta) &= \mathbf{E}[Y(t+\Delta)|Y(t)]\\ &= a_p(t+\Delta) + \mathbf{E}[X(t+\Delta)|X(t)],\\ \mathbf{var}[Y(t+\Delta)|Y(t)] &= \mathbf{var}[X(t+\Delta)|X(t)],\\ \mathbf{cov}[Y(t+\Delta_1+\Delta_2), &Y(t+\Delta_1)|Y(t)]\\ &= \mathbf{cov}[X(t+\Delta_1+\Delta_2), X(t+\Delta_1)|X(t)].\end{aligned}$$

Since we did not assume that $X(t)$ is stationary, we might not have a time independent expression for the mean, variance, and the auto-covariance function.

Using Definition 5, the independence of $\epsilon_{t+\Delta}$ and $X(t)$, and the assumption that $\mathbf{E}[\epsilon_t] = 0$ for all t, we can write the conditional expectation of $X(t+\Delta)$ given $X(t)$ and conditional variance as

$$\begin{aligned}&\mathbf{E}[X(t+\Delta)|X(t)] = X(t) + \boldsymbol{\theta}^T\boldsymbol{\alpha}(\Delta, t, X(t)) +\\ &\quad \sigma\beta(\Delta, t, X(t))\mathbf{E}[\epsilon_{t+\Delta}] = X(t) + \widehat{\boldsymbol{\theta}}^T\boldsymbol{\alpha}(\Delta, t, X(t))\\ &\mathbf{var}[X(t+\Delta)|X(t)] = \mathbf{E}\Big[(X(t+\Delta)-\\ &\quad - \mathbf{E}[X(t+\Delta)|X(t)])^2|X(t)\Big] = \sigma^2\beta^2(\Delta, t, X(t))\end{aligned}$$

Let $\Delta_1, \Delta_2 > 0$. The conditional covariance between $X(t+\Delta_1+\Delta_2)$ and $X(t+\Delta_1)$ given $X(t)$ is

$$\begin{aligned}
&\mathbf{cov}[X(t+\Delta_1+\Delta_2), X(t+\Delta_1)|X(t)] = \sigma^2\beta^2(\Delta, t, X(t)) \\
&\quad + \mathbf{E}[\boldsymbol{\theta}^T\boldsymbol{\alpha}(\Delta_2, t+\Delta_1, X(t+\Delta_1))X(t+\Delta_1)|X(t)] \\
&\quad - \left(X(t) + \widehat{\boldsymbol{\theta}}^T\boldsymbol{\alpha}(\Delta, t, X(t))\right)\cdot \\
&\qquad \cdot \mathbf{E}[\boldsymbol{\theta}^T\boldsymbol{\alpha}(\Delta_2, t+\Delta_1, X(t+\Delta_1))|X(t)]
\end{aligned}$$

It can be further simplified using the structure of $\boldsymbol{\alpha}$. The expressions inside the expectation operators are functions of $\epsilon_{t+\Delta_1}$, and thus are independent of $X(t)$, so the operators can be removed. Finally, $\mathbf{cov}[X(t+\Delta_1+\Delta_2), X(t+\Delta_1)|X(t)]$ is a function of $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma$ and $X(t)$. Since $X(t)$ is given, if we replace $\boldsymbol{\theta}$ and σ by their estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\sigma}$, we can estimate $\mathbf{cov}[X(t+\Delta_1+\Delta_2), X(t+\Delta_1)|X(t)]$.

A one-step predictor of $X(t+\Delta)$ given $X(t)$ for any $\Delta > 0$ is given by:

$$\widehat{X}(t+\Delta) = \mathbf{E}[X(t+\Delta)|X(t)] = X(t) + \widehat{\boldsymbol{\theta}}^T\boldsymbol{\alpha}(\Delta, t, X(t)).$$

5.4 Analyzing $Y(t)$ with a non-stationary $X(t)$: To predict $Y(t+\delta)$ given $Y(t)$, we first fit a polynomial $a_p(t)$ to $Y(t)$ as in Section 4.3, and estimate $\widehat{\boldsymbol{\theta}}$ and $\widehat{\sigma}$ using Proposition 5.1. Then a predictor $\widehat{Y}(t+\delta)$ is given by

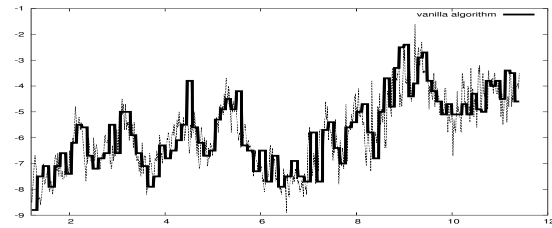
$$\begin{aligned}
\widehat{Y}(t+\delta) &= a_p(t+\delta) + \widehat{X}(t+\delta) \\
&= a_p(t+\delta) + X(t) + \widehat{\boldsymbol{\theta}}^T\boldsymbol{\alpha}(\delta, t, X(t)).
\end{aligned}$$

Similarly, we can estimate the variance and autocovariance functions of $Y(t)$ for given δ_1 and δ_2:

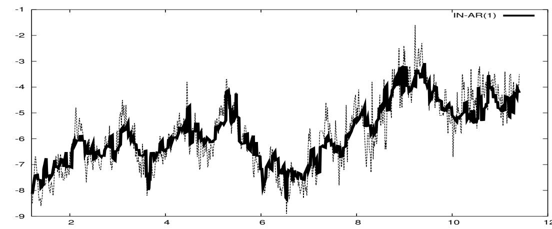
$$\begin{aligned}
&\mathbf{var}[Y(t+\delta)|Y(t)] = \widehat{\sigma}^2\beta^2(\delta, t, X(t)), \\
&\mathbf{cov}[Y(t+\delta_1+\delta_2), Y(t+\delta_1)|Y(t)] = \\
&\qquad \mathbf{cov}[X(t+\delta_1+\delta_2), X(t+\delta_1)|X(t)].
\end{aligned}$$

6 Computational Experiments

We tested prediction abilities of our IN-AR(1) model on several real datasets. Figure 1 shows a dataset containing a historical isotopic temperature record from the Vostok ice core (about 1K irregularly sampled points), due to Petit et al. [7]. Figure 1(a) overlays the dataset with a 10-point prediction given by the model trained on 100 points. For comparison, we did a similar prediction using a vanilla algorithm that always predicts the last value it sees (Figure 1(b)). The vanilla algorithm produces a step function with a good overall fit to the data, but with no attempt to give accurate short-term predictions. The curve produced by the IN-AR(1) model provides a smoother, much more accurate fit to the data. See the full version of the paper for more experiments [4].



(a)



(b)

Figure 1: 10-step prediction (a) vanilla, (b) IS-AR(1)

7 Conclusion

We proposed two AR(1)-type models for analyzing irregularly sampled time series. The first model is based on the assumption of stationarity, the second model relaxes this assumption. Both models are extremely simple and can be efficiently fit to the data. The presented approach can be extended to higher order auto-regression processes AR(p), moving average MA(q) and autoregressive moving average processes ARMA(p,q). An interesting research question is to develop a model for analyzing irregularly sampled time series with a non-Gaussian noise.

References

[1] H. M. Adorf. Interpolation of irregularly sampled data series. In R. Shaw, H. Payne, and J. Hayes, editors, *Astronomical Data Analysis Software and Systems IV*, *APS Series* 77, 460–463, 1995.

[2] P. J. Brockwell. Lévy driven carma processes. *Ann.Inst.Statist.Math.*, 53(1):113–124, 2001.

[3] P. J. Brockwell and R. A. Davis. *Time series: Theory and Methods*. Springer Verlag, 1991.

[4] E. Erdogan, S. Ma, A. Beygelzimer, and I. Rish. Technical Report, IBM T. J. Watson Center, 2005.

[5] R. H. Jones and P. V. Tryon. Continuous time series models for unequally spaced data applied to modeling atomic clocks. *SIAM J. Sci. Stat. Comput.*, 4(1):71–81, January 1987.

[6] R. H. Jones. *Time series analysis with unequally spaced data*. In E. Hannan, P. Krishnaiah, and M. Rao, editors, *Handbook of statistics*, 5, 157–178, North Holland, Amsterdam, 1985.

[7] J. R. Petit, et al. Climate and atmospheric history of the past 420,000 years from the vostok ice core, antarctica. *Nature*, 399:429–436, 1999.

[8] D. B. Rubin. *Statistical Analysis With Missing Data*. Wiley, New York, 2002.