

# Toward AI-Guided Smoking Cessation: Individualized Nicotine Addiction Modeling Using Gaussian Processes

Anirudh Chari<sup>1</sup>

<sup>1</sup>Illinois Mathematics and Science Academy, Aurora, IL, USA  
[achari@imsa.edu](mailto:achari@imsa.edu)

## Abstract

Cigarette smoking remains the leading cause of preventable disease and death in the United States, accounting for nearly half a million deaths annually. Given the recent rise of artificial intelligence in healthcare applications, computational assessment of smoking behaviors is a promising direction. In this study, we aim to recognize and classify addiction patterns in individual smokers' daily usage based on time series data. To this end, we leverage Gaussian process modeling to iteratively learn a function that defines a smoker's behavior as usage data is accumulated. Namely, we aim to learn weekly periodic trends in usage, and then utilize the model to predict future trends. We demonstrate that the outputted predictions resemble the actual data well, and that these informed forecasts significantly outperform those of a naive prediction model with respect to accuracy. Finally, we propose strategies for utilizing these predictions for goal-setting as part of a computer-supervised gradual cessation program.

## 1 Introduction

Smoking cessation generally comes in two forms: abrupt and gradual. Abrupt cessation has shown to be more effective in maintaining long-term abstinence than self-supervised gradual cessation [1]. However, computer-supervised gradual cessation (i.e., goal-setting algorithms) has not been widely explored, and a preliminary approach showed significantly higher long-term abstinence rates than both abrupt and self-supervised gradual cessation attempts, at least in adolescents [2]. With the recent rise and great success of artificial intelligence (AI) in healthcare applications [3], this is a promising direction.

Modeling drug use and addiction computationally has been widely explored in recent years [4]. Popular approaches include reinforcement learning [5] and dopamine-based modeling [6]. In [7], the author demonstrates that Bayesian frameworks can be effective for analyzing decision-making behaviors in drug addicts. In [8], the authors develop a computational model of nicotine addiction that classifies the severity of a user's addiction based on neurophysiological indicators.

Time series forecasting is the problem of predicting future data points or trends in a time series, given a segment of the current data. Popular time series forecasting techniques include exponential smoothing, ARIMA models, Kalman filters, long short-term memory (LSTM) models [9], and Gaussian processes (GPs) [10]. Because they operate on Bayesian inference, GPs pose the additional advantage of uncertainty quantification in outputted predictions, which can be especially helpful in evaluating their efficacy in the proposed setting. For an intuitive tutorial on the mechanics of GPs for solving regression problems, see [11].

The main challenge in forecasting is to learn the underlying patterns within potentially noisy data. This makes time series forecasting a very valuable tool in healthcare, as exemplified by its application to disease diagnosis and prognosis [12, 13]. Time series analysis and forecasting have also been applied in modeling large-scale nicotine use and cessation [14], especially in evaluating interventions [15, 16]. However, applying forecasting methods in behavior assessment of individual smokers has not been widely studied in recent years.

We propose that the nicotine use of smokers with respect to time can be modeled using GPs, and that this model can be employed for effective classification and prediction of addiction and cessation behaviors. Ultimately, we aim to leverage these insights to develop a personalized and adaptive gradual smoking cessation program.

## 2 Methods

### 2.1 Preliminary: Gaussian Processes

#### 2.1.1 Definition

A Gaussian process (GP) model describes a probability distribution over all possible functions that fit a set of points. A GP model leverages Bayesian inference to update the posterior function function, defined as the outputted means with the variances as quantified uncertainties, as new data is obtained. This posterior function can be used as a regression model to make predictions about new data.

#### 2.1.2 Kernels

The *kernel* of a GP model defines the general curve fitting behavior, i.e. our foundational beliefs on how our function should behave. More formally, a kernel  $k_{\theta}(x_1, x_2)$  defines the covariance between two function values  $x_1$  and  $x_2$  using some hyperparameters  $\theta$ . Multiple kernels can be combined through *kernel composition* in order to fit more complex functions.

### 2.1.3 Hyperparameters

Being the coefficients of our kernel function, the hyperparameters  $\theta$  determine the exact curve fitting behavior of the GP model. Estimating hyperparameters manually is difficult; a common approach is Bayesian optimization.

## 2.2 Modeling Smoking Habits

We observe that weekly patterns emerge in smokers' nicotine use [17]. Thus, we can formulate the problem as periodical time forecasting, for which GPs have proven to be effective [18]. At a high level, the role of the GP is to learn a function that defines the user's behavior. Our approach uses the framework provided in [18] as a foundation, and we summarize the framework as it relates to our work below.

### 2.2.1 Kernel Composition

We utilize a kernel  $K$ , which is composed by

$$K = \text{PER} + \text{LIN} + \text{RBF} + \text{SM}_1 + \text{SM}_2. \quad (1)$$

Here, the periodic kernel PER provides a weekly seasonality, the LIN kernel provides a linear trend, and the RBF,  $\text{SM}_1$  and  $\text{SM}_2$  kernels are collectively used to represent nonlinear trends. See [18] for explicit definitions of each of these kernels.

### 2.2.2 Hyperparameter Estimation

While the model aims to learn the specific parameters defining an individual's behavior, it also consists of hyperparameters  $\theta$  defining general behavior that applies to the entire function space, i.e., all users' behaviors. We begin by assigning log-normal priors to each hyperparameter, and then improve our estimations by training the model on various sample data using an iterative process called *maximum a posteriori* (MAP) estimation, as outlined in [18].

### 2.2.3 The Forecasting Problem

GP regression modeling is naturally applicable to time series forecasting problems. We can formally represent our addiction forecasting task as a regression problem

$$y = f(x) + v \quad (2)$$

where  $y$  is nicotine dose per day (measured in number of cigarettes),  $x$  is time in days,  $f \sim GP(0, k_\theta)$  is the function with the GP as a prior distribution, and  $v \sim \mathcal{N}(0, \sigma_v^2)$  is the noise with variance  $\sigma_v^2$ . Then, given  $m$  points of training data  $\mathbf{x} = (x_1, \dots, x_m)$ ,  $\mathbf{y} = (y_1, \dots, y_m)$ , and given  $n$  test inputs  $\mathbf{x}^* = (x_{m+1}^*, \dots, x_{m+n}^*)$ , the forecasting problem is to compute the function's posterior distribution  $\mathbf{f}^* = (f(x_{m+1}^*), \dots, f(x_{m+n}^*))$ .

## 3 Experiments

### 3.1 Dataset

We utilize the dataset provided in [17], which contains a sample of 62 participants who have been smoking for more than two consecutive years and who smoke more than five cigarettes a day. Participants were male and female college students between 18 and 26 years of age. Each participant recorded the number of cigarettes he or she smoked each day for up to 12 consecutive weeks. Some data was incomplete, so we consider only the participants who recorded data for the full 12-week period, which yields  $T = 50$  time series. For each time series, we form our training set using the first  $m = 56$  days and our test set using the last  $n = 28$  days. We normalize each time series to have a mean of 0 and a standard deviation of 1.

### 3.2 Baseline

We compare the performance of the GP model to that of a baseline naive model. The naive model uses the average value of the historical data as its predicted value for all future days. More formally,

$$f^*(x) = \frac{1}{m} \sum_{t=0}^m f(t), \quad \forall x \in \{m+1, \dots, m+n\}. \quad (3)$$

The naive model makes an educated guess that smoking behaviors remain constant, but it leverages no knowledge of weekly patterns in behavior.

### 3.3 Evaluation Metric

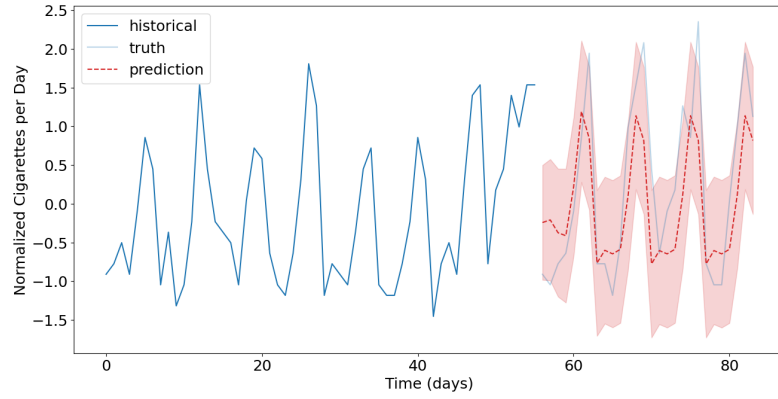
The *mean absolute error* (MAE) of the test set is given by

$$\frac{1}{n} \sum_{t=m}^{m+n} |f(t) - f^*(t)|. \quad (4)$$

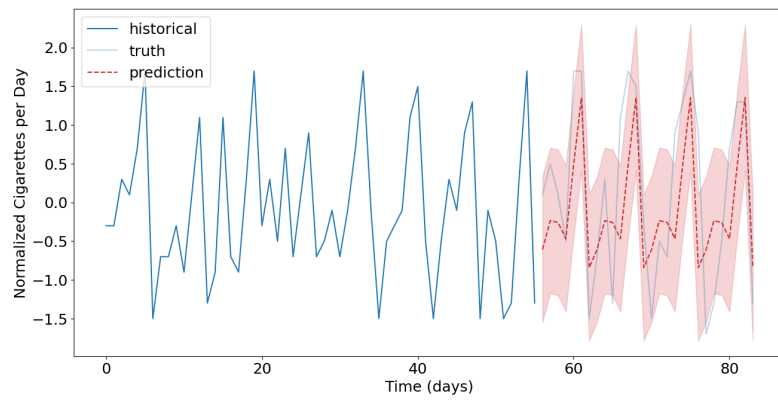
Lower values of MAE indicate better model performance. MAE is a common metric for evaluating regression models on normalized data.

### 3.4 Results

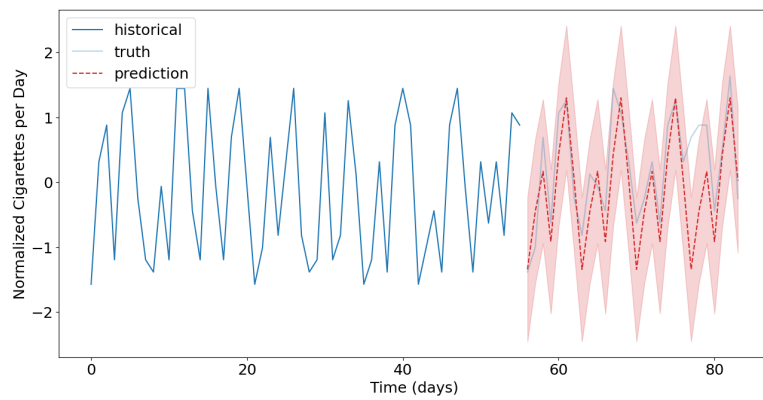
The MAE values across the time series were significantly lower for the GP model ( $M = 0.823$ ,  $SD = 0.168$ ) than for the naive model ( $M = 0.908$ ,  $SD = 0.150$ ),  $t(49) = 4.78$ ,  $p < .001$ . The GP model successfully discerns a wide variety of periodicities (Fig. 1), and the predicted values and associated error bounds resemble the actual data well,  $\chi^2(26, N = 50) = 767.65$ ,  $p < .001$ .



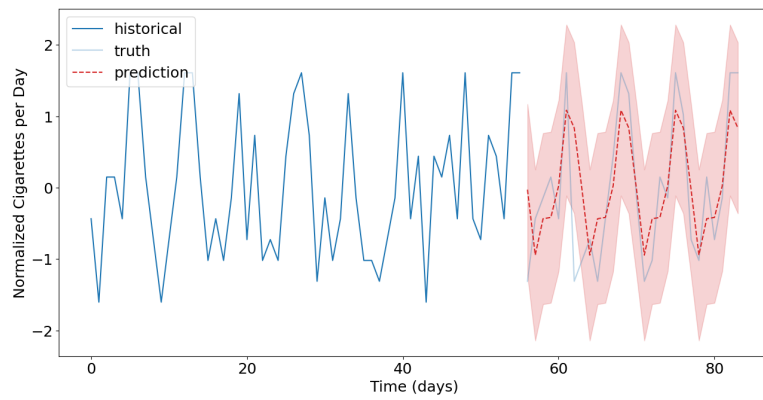
(a)



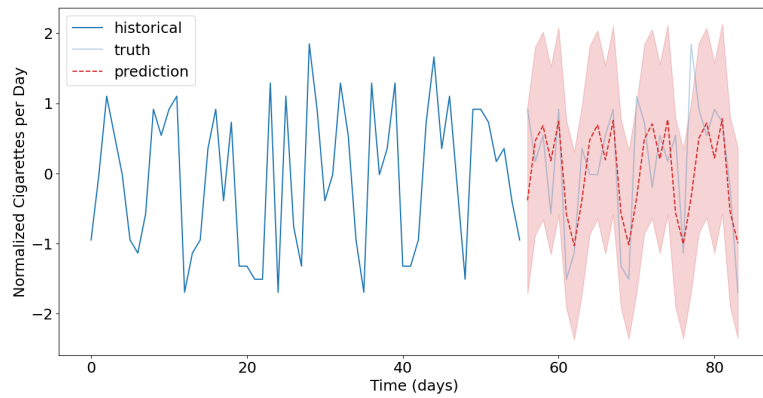
(b)



(c)



(d)



(e)

Figure 1: Example outputs from GP forecasting.

## 4 Guided Cessation

We have shown that GPs are an effective tool for modeling nicotine addiction behaviors. We expect that the model will be useful in designing a computer-supervised gradual smoking cessation program. However, this requires an additional goal-setting procedure that actually leverages the model’s insights. Here, we outline some suggested procedures.

### 4.1 Naive Bounding Procedure

At time  $t$ , let  $\mathbf{f}^* = (f(x_1^*), \dots, f(x_m^*))$  denote the posterior distribution of  $f \sim GP(0, k_{\theta})$ , i.e. the regression output up to  $x = x_m$  where  $t < x_m$ . Then a simple goal-setting function is

$$g(t) = \max(0, \lfloor k f^*(t) \rfloor) \quad (5)$$

for some constant  $k \in (0, 1)$  to enforce a downward trend.

### 4.2 Bayesian Inference Procedure

More generally, we can express our goals-setting function as

$$g(t) = \max(0, \lfloor \lambda f^*(t) \rfloor) \quad (6)$$

where  $\lambda$  is the user’s *learning rate* – the rate at which he or she can successfully reduce nicotine intake. The learning rate can be somewhat interpreted as the strength of the user’s addiction. If  $\lambda$  is too low, the goals are unachievable and the user is prone to relapse. On the other hand, if  $\lambda$  is too high, progress is minimal the cessation process is unnecessarily prolonged. In some sense, our learning rate here is analogous to the learning rate in gradient descent (usually denoted  $\alpha$ ), which determines the step size in each iteration. Each user has a unique learning rate that might even change throughout the cessation attempt, and thus we are tasked with learning the optimal  $\lambda$  over time.

We claim that  $\lambda$  can be learned using Bayesian inference. On day  $t = 0$  we begin with some prior distribution, say  $\lambda \sim \mathcal{N}(0.9, 0.2)$ . Then for all days  $t > 0$ , we update our distribution with an iteration of Bayes’ theorem using the new evidence

$$E_t = \frac{f(t-1)}{g(t-1)} \quad (7)$$

and we sample our new  $\lambda$  from the posterior distribution.

## 5 Conclusion

In this paper, we showed that Gaussian processes are an effective method for individualized modeling of nicotine addiction behaviors. In particular, GPs perform well when tasked with finding weekly patterns in a user’s smoking

habits, and the outputted predictions for future usage are significantly more accurate than those of a naive forecasting model. We also proposed methods for leveraging the GP forecast outputs for goal-setting as part of a guided gradual cessation program. In future work, we intend to conduct a user study that validates the effectiveness of GPs in computer-guided smoking cessation.

## References

- [1] N. Lindson-Hawley, M. Banting, R. West, S. Michie, B. Shinkins, and P. Aveyard, "Gradual versus abrupt smoking cessation: A randomized, controlled noninferiority trial," *Ann. Intern. Med.*, vol. 164, no. 9, p. 585, 2016.
- [2] W. Riley, A. Jerome, A. Behar, and S. Zack, "Feasibility of computerized scheduled gradual reduction for adolescent smoking cessation," *Subst. Use Misuse*, vol. 37, no. 2, pp. 255–263, 2002.
- [3] A. Bohr and K. Memarzadeh, "The rise of artificial intelligence in healthcare applications," in *Artificial Intelligence in Healthcare*, Elsevier, 2020, pp. 25–60.
- [4] J. A. Mollick and H. Kober, "Computational models of drug use and addiction: A review," *J. Abnorm. Psychol.*, vol. 129, no. 6, pp. 544–555, 2020.
- [5] K. C. Berridge, "From prediction error to incentive salience: mesolimbic computation of reward motivation," *Eur. J. Neurosci.*, vol. 35, no. 7, pp. 1124–1143, 2012.
- [6] W. Schultz, P. Dayan, and P. R. Montague, "A neural substrate of prediction and reward," *Science*, vol. 275, no. 5306, pp. 1593–1599, 1997.
- [7] K. Friston, "A theory of cortical responses," *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, vol. 360, no. 1456, pp. 815–836, 2005.
- [8] S. Metin and N. S. Sengor, "From occasional choices to inevitable musts: A computational model of nicotine addiction," *Comput. Intell. Neurosci.*, vol. 2012, pp. 1–14, 2012.
- [9] R. Williams, S. Hochreiter, and J. Schmidhuber, "Long Short-Term Memory," *Cmu.edu*, 1997. [Online]. Available: <https://deeplearning.cs.cmu.edu/S23/document/readings/LSTM.pdf>. [Accessed: 02-Oct-2023].
- [10] S. Roberts, M. Osborne, M. Ebdon, S. Reece, N. Gibson, and S. Aigrain, "Gaussian processes for time-series modelling," *Philos. Trans. A Math. Phys. Eng. Sci.*, vol. 371, no. 1984, p. 20110550, 2013.
- [11] J. Wang, "An intuitive tutorial to Gaussian processes regression," *Arxiv.org*. [Online]. Available: <http://arxiv.org/abs/2009.10862>. [Accessed: 02-Oct-2023].



- [12] C. Bui, N. Pham, A. Vo, A. Tran, A. Nguyen, and T. Le, “Time series forecasting for healthcare diagnosis and prognostics with the focus on cardiovascular diseases,” in 6th International Conference on the Development of Biomedical Engineering in Vietnam (BME6), Singapore: Springer Singapore, 2018, pp. 809–818.
- [13] A. S. Billis and P. D. Bamidis, “Employing time-series forecasting to historical medical data: an application towards early prognosis within elderly health monitoring environments,” in Proceedings of the 3rd International Conference on Artificial Intelligence and Assistive Medicine - Volume 1213, 2014, pp. 31–35.
- [14] L. Dierker et al., “Tobacco, alcohol, and marijuana use among first-year U.s. college students: A time series analysis,” *Subst. Use Misuse*, vol. 43, no. 5, pp. 680–699, 2008.
- [15] S. J. Hoffman, M. J. P. Poirier, S. Rogers Van Katwyk, P. Baral, and L. Sritharan, “Impact of the WHO Framework Convention on Tobacco Control on global cigarette consumption: quasi-experimental evaluations using interrupted time series analysis and in-sample forecast event modelling,” *BMJ*, vol. 365, p. l2287, 2019.
- [16] J. M. Lightwood, S. Anderson, and S. A. Glantz, “Smoking and healthcare expenditure reductions associated with the California Tobacco Control Program, 1989 to 2019: A predictive validation,” *PLoS One*, vol. 18, no. 3, p. e0263579, 2023.
- [17] J. F. Rosel et al., “Pooled time series modeling reveals smoking habit memory pattern,” *Front. Psychiatry*, vol. 11, 2020.
- [18] G. Corani, A. Benavoli, and M. Zaffalon, “Time series forecasting with Gaussian Processes needs priors,” *Arxiv.org*. [Online]. Available: <http://arxiv.org/abs/2009.08102>. [Accessed: 02-Oct-2023].