Quantifying the Trendiness of Trends

Andreas Kryger Jensen and Claus Thorn Ekstrøm
Biostatistics, Institute of Public Health, University of Copenhagen
aeje@sund.ku.dk, ekstrom@sund.ku.dk

30 May, 2020

Text formatted like this denotes a revision. Comments to the revisions can be found in the associated reply letter.

Abstract

A statement often seen in the news concerning some public health outcome is that some trend has changed or been broken. Such statements are often based on longitudinal data from surveys, and the change in trend is typically claimed by the news media to have occurred at the time of the latest data collection - if no change had occurred the story is much less likely to be printed. These types of statistical assessments are very important as they may potentially influence public health decisions on a national level.

We propose two measures for quantifying the trendiness of trends. Under the assumption that reality evolves in continuous time we define what constitutes a trend and a change in a trend, and we introduce a probabilistic Trend Direction Index. This index has the interpretation of the probability that a latent characteristic has changed monotonicity at any given time conditional on observed data. We also define a global index of Expected Trend Instability quantifying the expected number of times that a trend has changed on an interval.

Using a latent Gaussian process model we show how the Trend Direction Index and the Expected Trend Instability can be estimated from data in a Bayesian framework and use the proposed method for two applications: the development of the proportion of smokers in Denmark during the last 20 years and the development of new cases that tested positive for COVID-19 in Italy from February 24th and 90 days on.

Keywords: Functional Data Analysis, Gaussian Processes, Trends, Bayesian Statistics

1 Introduction

Trend detection has received increased attention in many fields, and while many important applications have their roots in the fields of economics (stock development) and environmental change (global temperature), trend identification has important ramifications in industry (process monitoring), medicine (disease development) and public health (changes in society).

This manuscript is concerned with the fundamental problem of estimating an underlying trend based on random variables observed repeatedly over time. In addition to this problem we also wish to assess points in time where it is possible that such a trend is changing. Our motivation comes from two recent examples: in the first, the news media in Denmark stated that the trend in the proportion of smokers in Denmark has changed at the end of the year 2018 such that the proportion is now increasing whereas it had been decreasing for the previous 20 years. This statement was based on survey data collected yearly since 1998 and reported by the Danish Health Authority (The Danish Health Authority 2019), and it is critical for the Danish Health Authorities to be able to evaluate and react if there is an actual change in trend. The second example relates to the recent

outbreak of COVID-19 in Italy where it is of tremendous importance to determine if the disease spread is increasing or slowing down by considering the trend in number of new cases (see Figure 1).

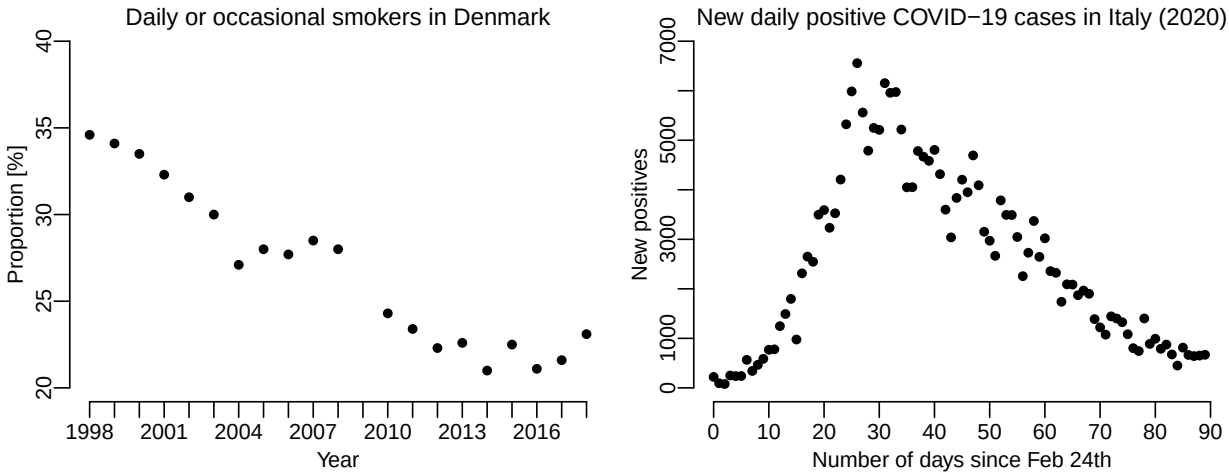

Figure 1: Left panel: The proportion of daily or occasional smokers in Denmark during the last 20 years estimated from survey data and reported by the Danish Health Authority. The 2009 measurement is missing due to problems with representativeness. Right panel: The number of daily new cases tested positive for COVID-19 in Italy.

The concept of "trend" is not itself well-defined and it is often up to each researcher to define exactly what is meant by "trend" (Esterby 1993). Consequently, it is not obvious what constitutes a *change* in trend, and that makes it difficult to compare statistical methods to detect trend changes, since they attempt to address slightly different problems.

Change-point analysis is an often-used approach for detecting if a change (of a prespecified type) has taken place (Basseville and Nikiforov 1993; Barry and Hartigan 1993). However, change-point analysis is marred by the fact that change-points that happen close to the boundary of the observed time frame are notoriously difficult to detect, and that they assume that change-points are abrupt changes that occur at specific time point.

Another common approach to analyse trends in time series is to apply a low-pass smoothing filter to the observed data in order to remove noise and extract the underlying latent trend (Chandler and Scott 2011). But in addition to the smoothing filter, it is also necessary to circumvent the problem of what constitutes a change in trend so it is necessary to specify a decision criterion to determine if any filtered changes are relevant.

Gottlieb and Müller (2012) define a stickiness coefficient for longitudinal data and use the stickiness coefficient to summarize the extent to which deviations from the mean trajectory tend to co-vary over time. While the stickiness coefficient determines changes from the expected trend it is a measure that is not easily interpreted.

Kim et al. (2009) proposed an ℓ_1 Trend Filtering approach based on an idea by Hodrick and Prescott (1997) that imposes sparsity on the 1st order differences of the conditional mean. This produces trend estimates that are piece-wise linear with the inherent assumption that changes to the time series are abrupt. The Trend Filtering approach has been further extended to sparsity of kth order differences and to a Bayesian framework, where the flexibility of the difference of the conditional mean is controlled through the prior distributions (Ramdas and Tibshirani 2016; Kowal, Matteson,

and Ruppert 2019)

We propose a new method for evaluating changes in the latent trend, f, which starts by a clear specification of the problem and relevant measures of the trend: The Trend Direction Index gives a local probability of the monotonicity of f and provides an answer to the research question "What is the probability that the latent trend is increasing?". The Expected Trend Instability index gives the expected number of changes in the monotonicity of f over a time-interval. The Expected Trend Instability index provides information about the volatility of f and can be used to evaluate research questions about the rate of change in direction of f. In particular it can be used to gauge how surprising such statements as „It is the first time in 20 years that f has stopped decreasing and started to increase" is. Besides providing exact answers to specific and relevant research questions, our proposed method estimates the actual probability that the trend is changing.

The manuscript is structured as follows: In Section 2 we present our statistical model based on a latent Gaussian process formulation giving rise to explicit expressions for the Trend Direction Index and the Expected Trend Instability index conditional on observed data. Section 3 is concerned with estimating the models parameters. In Section 4 we undertake a simulation study to show the performance of the proposed method and to compare it with trend filtering, and in Section 5 we provide extended applications to our two cases: the development of the proportion of smokers in Denmark during the last 20 years and the development of new COVID-19 cases in Italy. We conclude with a discussion.

Proofs of propositions are given in the Supplementary Material, and reproducible code and Stan implementations are available at the first author's GitHub repository (Kryger Jensen 2019).

2 Methods

In the following we assume that reality evolves in continuous time $t \in \mathcal{T} \subset \mathbb{R}$ and that there exists a random, latent function $f = \{f(t) : t \in \mathcal{T}\}$ sufficiently smooth on a compact subset of the real line, $\mathcal{T}$, that governs the underlying evolution of some observable characteristic in a population. We can observe this latent characteristic with noise by sampling f at discrete time points according to the additive model $Y_i = f(t_i) + \varepsilon_i$ where ε_i is a zero mean random variable independent of $f(t_i)$ with hyper-parameters Θ. Given observations of the form $(Y_i, t_i)_{i=1}^n$ we are interested in modeling the dynamical properties of f.

The trend of f is defined as its instantaneous slope given by the function $df(t) = \left(\frac{\mathrm{d}f(s)}{\mathrm{d}s}\right)(t)$, and f is increasing and has a positive trend at t if $df(t) > 0$, and f is decreasing with a negative trend at t if $df(t) < 0$. A change in trend is defined as a change in the sign of df, i.e., when f goes from increasing to decreasing or vice versa.

As f is a smooth random function there are no points-in-time where the underlying process abruptly changes sign. There is instead a gradual and continuous change in the monotonicity of f, and an assessment of a change in trend is defined by the probability of the sign of df. This stands in contrast to traditional change-point models which assume that there are one or more exact time points where a sudden change in function or its parameterization occurs (Carlstein, Müller, and Siegmund 1994).

The probability of a positive trend for f at time $t + \delta$ is quantified by the Trend Direction Index

$$\mathrm{TDI}(t,\delta) = P(df(t+\delta) > 0 \mid \mathcal{F}_t), \quad t \in \mathcal{T} \tag{1}$$

where $\mathcal{F}_t$ is a σ-algebra of available information observed up until time t. The value of $\mathrm{TDI}(t,\delta)$ is a local probabilistic index, and it is equal to the probability that f is increasing at time $t+\delta$ given everything known about the data generating process up until and including time t. A similar definition can be given for a negative trend but that is equal to $1-\mathrm{TDI}(t,\delta)$ and therefore redundant. The sign of δ determines whether the Trend Direction Index estimates the past ($\delta \leq 0$) or forecasts the future ($\delta > 0$). Most of the examples seen in the news concerning public health outcomes are concerned with t being equal to the current calendar time and $\delta = 0$. This excludes the usage of both change-point and segmented regression models (Quandt 1958) as there are no observations available beyond the stipulated change-point. A useful reparameterization of the Trend Direction Index is $\mathrm{TDI}(\max \mathcal{T}, t - \max \mathcal{T})$ with $t \leq \max \mathcal{T}$. This parameterization conditions on the full observation period and looks back in time whereas setting $t = \max \mathcal{T}$ corresponds to the current Trend Direction Index at the end of the observation period.

In addition to the Trend Direction Index we define a global measure of trend instability. Informally, we say that a random function f is *trend stable* on an interval $\mathcal{I}$ if its sample paths maintain their monotonicity so that the trends do not change sign on the interval. To quantify the trend *in*stability we propose to use the expected number of zero-crossings by df on $\mathcal{I}$. We define the Expected Trend Instability as

$$\mathrm{ETI}(\mathcal{I}) = \mathrm{E}\left[\#\left\{t \in \mathcal{I} : df(t) = 0\right\} \mid \mathcal{F}\right] \tag{2}$$

equal to the expected value of the size of the random set of zero-crossings by df on $\mathcal{I}$ when conditioning on a suitable σ-algebra $\mathcal{F}$. A common case is when $\mathcal{F}$ is generated by all observed data on $\mathcal{T}$ and $\mathcal{I} \subseteq \mathcal{T}$. The lower $\mathrm{ETI}(\mathcal{I})$ is, the more stable is the trend of f on $\mathcal{I}$ and vice versa.

We note, thanks to a comment by an anonymous reviewer, that the Expected Trend Instability represents an upper bound for the probability of observing at least one zero-crossing event, $df(\cdot) = 0$, over $\mathcal{I}$ since

$$\begin{aligned}
\mathrm{ETI}(\mathcal{I}) &= \sum_{k=0}^{\infty} P(\#\{t \in \mathcal{I} : df(t) = 0\} = k) \cdot k \\
&\geq \sum_{k=1}^{\infty} P(\#\{t \in \mathcal{I} : df(t) = 0\} = k) \\
&= P(\#\{t \in \mathcal{I} : df(t) \geq 0\} \geq 1)
\end{aligned}$$

where the inequality becomes sharp if $\sum_{k\geq 2} P(\#\{t \in \mathcal{I} : df(t) = 0\} = k)$ is small. The Expected Trend Instability can therefore be used over smaller intervals $\mathcal{I}$ to provide statements about the expected probability that a change in trend will happen.

These two general definitions of trendiness will be evaluated in the light of a particular statistical model in the next sections leading to expressions of their estimates.

2.1 Latent Gaussian Process Model

The definitions in the previous section impose restrictions on the latent function f. We shall assume that f is a Gaussian process on $\mathcal{T}$. From a Bayesian perspective this is equivalent to imposing an infinite dimensional prior distribution on the latent characteristic that governs the observed outcomes. Statistical models with Gaussian process priors are a flexible approach for non-parametric regression (Radford 1999). Using a latent Gaussian process also provides an analytically tractable way for performing statistical inference on its derivatives. The general idea of our model is to apply the properties of the Gaussian process prior on the latent characteristic to update the finite dimensional distributions by conditioning on the observed data. This results in a posterior Gaussian process and its derivatives from which estimates of the trendiness indices will be derived.

A random function f is a Gaussian process if and only if the vector $(f(t_1), \ldots, f(t_n))$ has a multivariate normal distribution for every finite set of evaluation points $(t_1, \ldots, t_n)$, and we write $f \sim \mathcal{GP}(\mu(\cdot), C(\cdot,\cdot))$ where μ is the mean function and C is a symmetric, positive definite covariance function (Cramer and Leadbetter 1967). We observed dependent data in terms of outcomes and their associated sampling times, $(Y_i, t_i)_{i=1}^n$, and we assume that the data are generated by the hierarchical model

$$\begin{aligned} f \mid \beta, \theta &\sim \mathcal{GP}(\mu_\beta(\cdot), C_\theta(\cdot,\cdot)) \\ Y_i \mid t_i, f(t_i), \Theta &\overset{iid}{\sim} N(f(t_i), \sigma^2), \quad \Theta = (\beta, \theta, \sigma) \end{aligned} \tag{3}$$

where β is a vector of parameters for the mean function of f, θ is a vector of parameters governing the covariance of f, and σ is the conditional standard deviation of the observations.

Assumption 1. *We assume the following regularity conditions.*

A1: f is a separable Gaussian process.

A2: $\mathrm{E}[f(t) \mid \beta, \theta] = \mu_\beta(t)$ *is a twice continuously differentiable function.*

A3: $\mathrm{Cov}[f(s), f(t) \mid \beta, \theta] = C_\theta(s,t)$ *has mixed third-order partial derivatives continuous at the diagonal.*

A4: The joint distribution of $(df(t), d^2f(t) \mid \beta, \theta)$ *is non-degenerate at any t i.e.,* $\mathrm{Cov}[df(t), df(t) \mid \beta, \theta] > 0$ *and* $-1 < \mathrm{Cor}[df(t), d^2f(t) \mid \beta, \theta] < 1$.

Implications of the assumptions: A1 is a strictly technical condition required to ensure that functionals of f defined on an uncountable index set can form random variables. All continuous Gaussian processes are separable. A2 is required in order to make $\mathrm{E}[df \mid \beta, \theta]$ and $\mathrm{E}[d^2f \mid \beta, \theta]$ well-defined in Equation (4). This is a modeling choice and can be satisfied by choosing μ_β accordingly. A3 is similarly required in order to make the prior covariance matrix in Equation (4) well-defined. We discuss this issue further in Section 2.4. A4 implies that one cannot choose a covariance function where i) the associated covariance of df becomes degenerate with no variability or ii) a covariance function where the first and second order derivatives are perfectly correlated. This is in order to ensure that the Trend Direction Index and the Expected Trend Instability are well-defined quantities. This is again a modeling choice and can always be verified in practice for a particular choice of C_θ.

Under the above assumptions, an important property of a Gaussian process is that it together with its first and second derivatives is a multivariate Gaussian process with explicit expressions for the

joint mean, covariance and cross-covariance functions. Specifically, the joint distribution of the latent function, f, and its first and second derivatives, df and d^2f, is the multivariate Gaussian process

$$\begin{bmatrix} f(s) \\ df(t) \\ d^2f(u) \end{bmatrix} \mid \beta, \theta \sim \mathcal{GP}\left(\begin{bmatrix} \mu_\beta(s) \\ d\mu_\beta(t) \\ d^2\mu_\beta(u) \end{bmatrix}, \begin{bmatrix} C_\theta(s,s') & \partial_2 C_\theta(s,t) & \partial_2^2 C_\theta(s,u) \\ \partial_1 C_\theta(t,s) & \partial_1\partial_2 C_\theta(t,t') & \partial_1\partial_2^2 C_\theta(t,u) \\ \partial_1^2 C_\theta(u,s) & \partial_1^2\partial_2 C_\theta(u,t) & \partial_1^2\partial_2^2 C_\theta(u,u') \end{bmatrix} \right) \tag{4}$$

where $d^k\mu_\beta$ is the k'th derivative of μ_β and ∂_j^k denotes the k'th order partial derivative with respect to the j'th variable (Cramer and Leadbetter 1967). Proposition 1 states the joint posterior distribution of (f, df, d^2f) conditional on the observed data.

Proposition 1. *Let the data generating model be defined as in Equation (3) and* $\mathbf{Y} = (Y_1, \ldots, Y_n)$ *the vector of observed outcomes together with its sampling times* $\mathbf{t} = (t_1, \ldots, t_n)$*. Then by the conditions in Assumption 1 the joint distribution of* (f, df, d^2f) *conditional on* $\mathbf{Y}$ *and* $\mathbf{t}$ *evaluated at the vector* $\mathbf{t}^*$ *of* p *time points is*

$$\begin{bmatrix} f(\mathbf{t}^*) \\ df(\mathbf{t}^*) \\ d^2f(\mathbf{t}^*) \end{bmatrix} \mid \mathbf{Y}, \mathbf{t}, \Theta \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\mu} \in \mathbb{R}^{3p}$ *is the column vector of posterior expectations and* $\boldsymbol{\Sigma} \in \mathbb{R}^{3p \times 3p}$ *is the joint posterior covariance matrix. Partitioning these as*

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_f(\mathbf{t}^* \mid \Theta) \\ \mu_{df}(\mathbf{t}^* \mid \Theta) \\ \mu_{d^2f}(\mathbf{t}^* \mid \Theta) \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_f(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) & \Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) & \Sigma_{f,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) \\ \Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta)^T & \Sigma_{df}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) & \Sigma_{df,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) \\ \Sigma_{f,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta)^T & \Sigma_{df,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta)^T & \Sigma_{d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) \end{bmatrix}$$

the individual components are given by

$$\begin{aligned}
\mu_f(\mathbf{t}^* \mid \Theta) &= \mu_\beta(\mathbf{t}^*) + C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\
\mu_{df}(\mathbf{t}^* \mid \Theta) &= d\mu_\beta(\mathbf{t}^*) + \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\
\mu_{d^2f}(\mathbf{t}^* \mid \Theta) &= d^2\mu_\beta(\mathbf{t}^*) + \partial_1^2 C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \\
\Sigma_f(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) &= C_\theta(\mathbf{t}^*, \mathbf{t}^*) - C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} C_\theta(\mathbf{t}, \mathbf{t}^*) \\
\Sigma_{df}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) &= \partial_1\partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} \partial_2 C_\theta(\mathbf{t}, \mathbf{t}^*) \\
\Sigma_{d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) &= \partial_1^2\partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - \partial_1^2 C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} \partial_2^2 C_\theta(\mathbf{t}, \mathbf{t}^*) \\
\Sigma_{f,df}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) &= \partial_2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} \partial_2 C_\theta(\mathbf{t}, \mathbf{t}^*) \\
\Sigma_{f,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) &= \partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} \partial_2^2 C_\theta(\mathbf{t}, \mathbf{t}^*) \\
\Sigma_{df,d^2f}(\mathbf{t}^*, \mathbf{t}^* \mid \Theta) &= \partial_1\partial_2^2 C_\theta(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_\theta(\mathbf{t}^*, \mathbf{t}) \left(C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right)^{-1} \partial_2^2 C_\theta(\mathbf{t}, \mathbf{t}^*)
\end{aligned}$$

In addition to the closed-form expressions of the conditional distributions of the latent function and its derivatives given in Proposition 1 it can also be useful to consider the predictive distribution of the model. That is: the conditional distribution of a new observations given the observed data. This is given by

$$Y^*(t^*) \mid t^*, \mathbf{Y}, \mathbf{t}, \Theta \sim N\left(\mu_f(t^* \mid \Theta), \Sigma_f(t^*, t^* \mid \Theta) + \sigma^2\right) \tag{5}$$

where Y^* is a new random variable predicted at some arbitrary time point t^*.

Proposition 1 grants the foundation for the rest of the methodological development. In the following two subsections we show how the Trend Direction Index and the Expected Trend Instability can be expressed under the data generating model in Equation (3) using the results of Proposition 1.

2.2 The Trend Direction Index

The Trend Direction Index was defined generally in Equation (1). Conditioning on the σ-algebra of all observed data we may express the Trend Direction Index under the model in Equation (3) through the posterior distribution of df. The following proposition states this result.

Proposition 2. *Let $\mathcal{F}_t$ be the σ-algebra generated by $(\mathbf{Y}, \mathbf{t})$ where $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is the vector of observed outcomes and $\mathbf{t} = (t_1, \ldots, t_n)$ the associated sampling times. Furthermore, assume that assumptions A1-A3 above are fulfilled. The Trend Direction Index defined in Equation (1) can then be written in terms of the posterior distribution of df as*

$$\begin{aligned}\mathrm{TDI}(t, \delta \mid \Theta) &= P(df(t+\delta) > 0 \mid \mathbf{Y}, \mathbf{t}, \Theta) \\ &= \frac{1}{2} + \frac{1}{2}\,\mathrm{Erf}\left(\frac{\mu_{df}(t+\delta \mid \Theta)}{2^{1/2}\Sigma_{df}(t+\delta, t+\delta \mid \Theta)^{1/2}}\right)\end{aligned}$$

where $\mathrm{Erf}\colon x \mapsto 2\pi^{-1/2}\int_0^x \exp(-u^2)\mathrm{d}u$ *is the error function and μ_{df} and Σ_{df} are the posterior mean and covariance of the trend defined in Proposition 1.*

Proposition 2 shows that the Trend Direction Index is equal to 0.5 when $\mu_{df}(t+\delta \mid \Theta) = 0$ corresponding to when the expected value of the posterior of f is locally constant at time $t+\delta$. A decision rule based on $\mathrm{TDI}(t, \delta \mid \Theta) \lessgtr 50\%$ is therefore a natural choice for assessing the local trendiness of f. However, different thresholds based on external loss or utility functions can be used depending on the application. Note that the requirements of assumptions A2 and A3 can be reduced to first order differentiable (A2) and mixed second derivatives (A3), respectively, if only the TDI is to be used.

The Trend Direction Index is illustrated in Figure 2 in the noise free case, $\sigma = 0$, with known values of β and θ, and where the prior expected value of f is set equal to $\mu_\beta(t) = 0$. The first and second rows of the plot show 150 random sample paths from the posterior distribution of (f, df) with the posterior expectations in bold lines, and the third row shows the Trend Direction Index. Since the parameters, Θ, of f are known in this example, the Trend Direction Index is a deterministic function of time. The three columns in the plot show how the posterior of f, df, and the TDI are updated after one, two and four observations both forwards and backwards in time. The figure shows how the inclusion of additional observations results in changes of the posterior distribution of the trend and the Trend Direction Index. The variation of the forecasts remains unchanged, whereas the posterior distribution back in time restricts the variation of the curves to accommodate

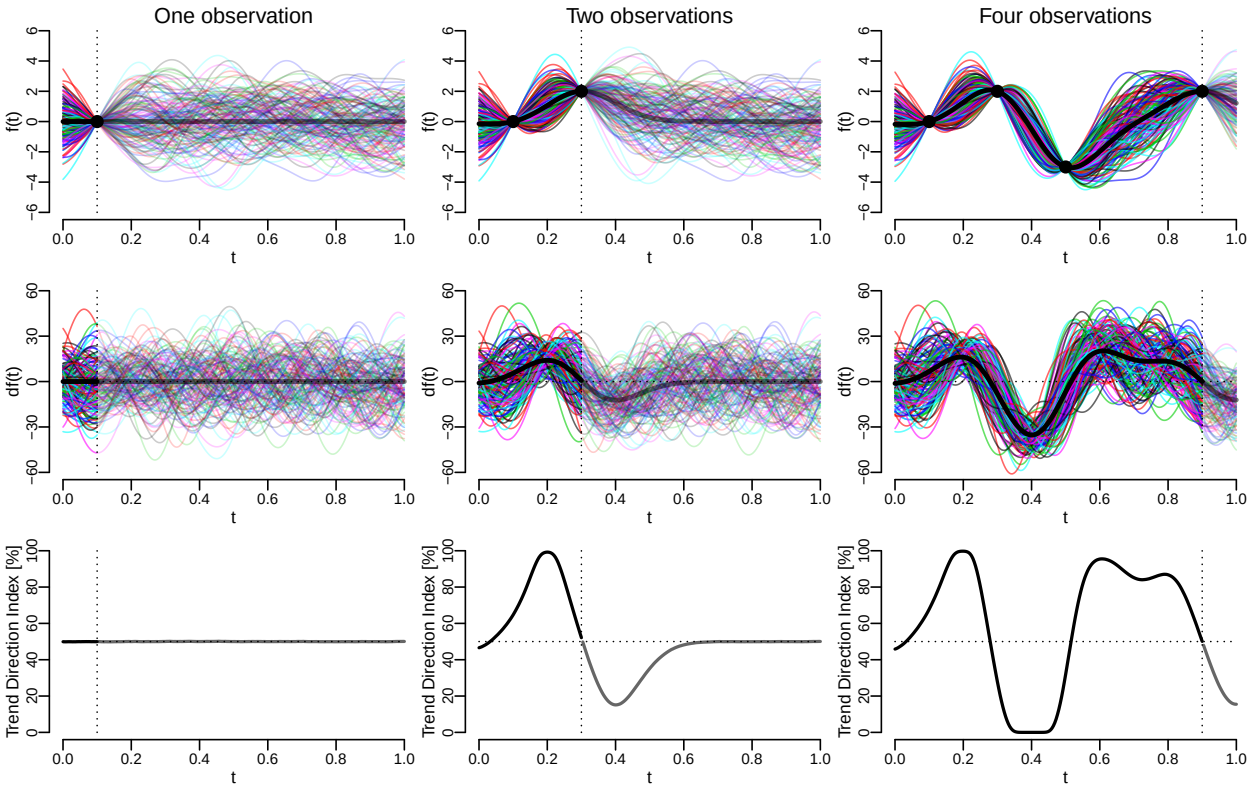

Figure 2: 150 realizations from the posterior distribution of f (top row), df (middle row) with expected values in bold and the Trend Direction Index (bottom row) conditional on one, two and four noise free observations. Dotted vertical lines show the points in time after which forecasting takes place.

the observations. The vertical dotted lines denote the point in time after which forecasting occurs. When forecasting beyond the last observation, the posterior of f becomes more and more dominated by its prior implying that df becomes symmetric around zero so that the Trend Direction Index stabilizes around 50%.

2.3 The Expected Trend Instability

The Expected Trend Instability was defined in Equation (2) as the expected number of roots of the trend on an interval conditional on observed data. Now we make this concept more precise and frame it in the context of the latent Gaussian process model. Let $\mathcal{I}$ be a compact subset of the real line and consider the random càdlàg function

$$N_{\mathcal{I}}(t) = \#\left\{s \leq t : df(s) = 0, s \in \mathcal{I}\right\}$$

counting the cumulative number of points in $\mathcal{I}$ up to time t where the trend is equal to zero. The Expected Trend Instability on $\mathcal{I}$ is equal to

$$\mathrm{ETI}(\mathcal{I} \mid \Theta) = \mathrm{E}\left[N_{\mathcal{I}}\left(\max \mathcal{I}\right) \mid \Theta, \mathcal{F}\right]$$

giving the expected total number of zero-crossings by df on $\mathcal{I}$. The following proposition gives the expression for the Expected Trend Instability under the data generating model in Equation (3).

Proposition 3. *Let $\mathcal{F}$ be the σ-algebra generated by the observed data $(\mathbf{Y}, \mathbf{t})$ and μ_{df}, μ_{d^2f}, Σ_{df}, Σ_{d^2f} and Σ_{df,d^2f} the moments of the joint posterior distribution of (df, d^2f) as stated in Proposition 1, and assume that all assumptions A1-A4 are fulfilled. The Expected Trend Instability is then*

$$\mathrm{ETI}(\mathcal{I} \mid \Theta) = \int_{\mathcal{I}} d\mathrm{ETI}(t \mid \Theta)\mathrm{d}t$$

where dETI is the local Expected Trend Instability given by

$$d\mathrm{ETI}(t, \mathcal{T} \mid \Theta) = \lambda(t \mid \Theta)\phi\left(\frac{\mu_{df}(t \mid \Theta)}{\Sigma_{df}(t, t \mid \Theta)^{1/2}}\right)\left(2\phi(\zeta(t \mid \Theta)) + \zeta(t \mid \Theta)\,\mathrm{Erf}\left(\frac{\zeta(t \mid \Theta)}{2^{1/2}}\right)\right)$$

and $\phi\colon x \mapsto 2^{-1/2}\pi^{-1/2}\exp(-\frac{1}{2}x^2)$ is the standard normal density function, $\mathrm{Erf}\colon x \mapsto 2\pi^{-1/2}\int_0^x \exp(-u^2)\mathrm{d}u$ *is the error function, and λ, ω and ζ are functions defined as*

$$\begin{aligned}
\lambda(t \mid \Theta) &= \frac{\Sigma_{d^2f}(t, t \mid \Theta)^{1/2}}{\Sigma_{df}(t, t \mid \Theta)^{1/2}}\left(1 - \omega(t \mid \Theta)^2\right)^{1/2} \\
\omega(t \mid \Theta) &= \frac{\Sigma_{df,d^2f}(t, t \mid \Theta)}{\Sigma_{df}(t, t \mid \Theta)^{1/2}\Sigma_{d^2f}(t, t \mid \Theta)^{1/2}} \\
\zeta(t \mid \Theta) &= \frac{\mu_{df}(t \mid \Theta)\Sigma_{d^2f}(t, t \mid \Theta)^{1/2}\omega(t)\Sigma_{df}(t, t \mid \Theta)^{-1/2} - \mu_{d^2f}(t \mid \Theta)}{\Sigma_{d^2f}(t, t \mid \Theta)^{1/2}\left(1 - \omega(t \mid \Theta)^2\right)^{1/2}}
\end{aligned}$$

To illustrate the Expected Trend Instability, Figure 3 shows 25 random Gaussian processes on the unit interval simulated under three different values of ETI. The sample paths are paired so that each function in the first row has an associated trend function in the second row. The different values of ETI are set to 0.25, 0.5 and 1 corresponding to the expected number of times that the functions change monotonicity or equivalently that the trend crosses zero on that interval. Sample paths that are *trend stable*, i.e., always increasing/decreasing, are shown by solid blue lines, and sample paths that are *trend unstable*, i.e., the derivatives crosses zero, are shown by dashed gold colored lines. It is seen that for low values of ETI most of the sample paths preserve their monotonicity on the interval and their associated derivatives are correspondingly either always positive or negative. For larger values of ETI, more of the trends start crossing zero implying less stability in the monotonicity of f. We note that even though we are only modeling a single curve, the Expected Trend Instability is defined in terms of a posterior distribution of random curves which is what the figure illustrates.

2.4 Modeling the prior

To complete the model specification in Equation (3) we must decide on the functional forms of the mean and covariance functions of the Gaussian process prior, but there is an inherent ambiguity in how to choose these. To explain this we note that a square-integrable Gaussian process f on a

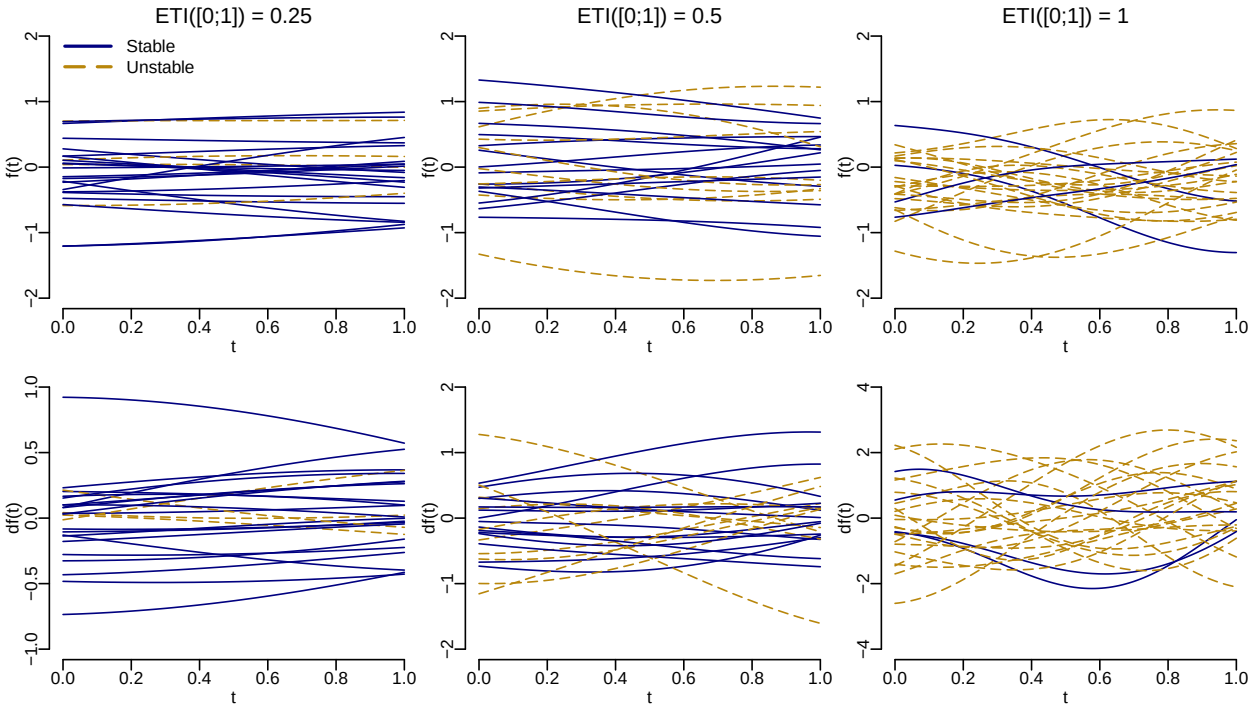

Figure 3: 25 random pairs sampled from the joint distribution of a Gaussian Process (f) and its derivative (df) with different values of Expected Trend Instability (ETI). The first row shows samples from f, and the second row shows samples from df. The columns define different values of ETI. Sample paths that are trend stable are shown by solid blue lines, and unstable sample paths are shown by dashed gold colored lines.

compact domain can by written as $f(t) = \mu_\beta(t) + \sum_{p=1}^{\infty} Z_p \phi_p(t)$ where Z_p are independent, zero mean and unit variance normally distributed random variables, and ϕ_p are pair-wise orthonormal functions. This is known as the Karhunen–Loève representation. Furthermore, Mercer's theorem states that any continuous covariance function C admits the representation $C(s,t) = \sum_{p=1}^{\infty} \lambda_p \phi_p(s) \phi_p(t)$ uniformly on $\mathcal{T} \times \mathcal{T}$ where ϕ_p are the eigenfunctions of C and form an orthonormal basis of L^2. This shows that f can be written as a sum of its mean and a possibly infinite weighted sum of functions defined implicitly through the spectral decomposition of the covariance function. So if e.g., the linear function is an eigenfunction of the covariance function it will be superfluous in the mean. This results in a trade-off situation where a rich specification of the mean, μ_β, gives little space for flexibility added by the covariance function, C_θ. On the opposite end, a simple mean structure requires a flexible covariance structure. The latter approach is the most prevalent in applied Gaussian Process regression modeling where a very simple mean structure (often just a constant or linear term) is used in the prior but in combination with a flexible covariance function. We refer to Kaufman et al. (2011) for further discussions of these issues. We note that this issue arises because our approach focuses on a single realization of a random function. In the case of multiple observations of f the issue would be different.

Regarding the choice of covariance function, the Squared Exponential (SE), the Rational Quadratic (RQ), the Matern 3/2 (M3/2), and the Matern 5/2 (M5/2) are commonly used covariance functions for Gaussian process regression (C. E. Rasmussen and Williams 2006). These covariance functions

are given by

$$
\begin{aligned}
C_\theta^{\mathrm{SE}}(s,t) &= \alpha^2 \exp\left(-\frac{(s-t)^2}{2\rho^2}\right), \quad \theta = (\alpha,\rho) > 0 \\
C_\theta^{\mathrm{RQ}}(s,t) &= \alpha^2 \left(1+\frac{(s-t)^2}{2\rho^2\nu}\right)^{-\nu}, \quad \theta = (\alpha,\rho,\nu) > 0 \\
C_\theta^{\mathrm{M3/2}}(s,t) &= \alpha^2 \left(1+\frac{\sqrt{3}\sqrt{(s-t)^2}}{\rho}\right) \exp\left(-\frac{\sqrt{3}\sqrt{(s-t)^2}}{\rho}\right), \quad \theta = (\alpha,\rho) > 0 \\
C_\theta^{\mathrm{M5/2}}(s,t) &= \alpha^2 \left(1+\frac{\sqrt{5}\sqrt{(s-t)^2}}{\rho}+\frac{5(s-t)^2}{3\rho^2}\right) \exp\left(-\frac{\sqrt{5}\sqrt{(s-t)^2}}{\rho}\right), \quad \theta = (\alpha,\rho) > 0
\end{aligned}
\tag{6}
$$

The SE covariance function is characterized by it giving rise to very smooth estimates due to it being infinitely differentiable. This can be disadvantageous in some applications as that will make it difficult for the posterior to adapt to localized changes in smoothness. The other covariance functions listed try to remedy this issue. Specifically, the RQ covariance function can be derived as an infinite scale mixture of SE covariance functions in terms of ρ^{-2}. This enables the resulting estimates to operate on different time-scales simultaneously. Smaller values of ν will give rise to more wiggly posteriors, and as it is one of the model's hyper-parameters its value and thus the resulting adaptivity is data-driven. For $\nu \to \infty$ the RQ covariance function converges on $\mathcal{T} \times \mathcal{T}$ to the SE covariance function. The Matern covariance functions shown here explicitly for a generally continuous parameter $\nu = 3/2$ and $\nu = 5/2$ belong to a larger class of covariance functions. The general expression is involved and includes a modified Bessel function for which derivatives are difficult to calculate. However, in this class ν directly controls the differentiability which again can be chosen in a data-driven manner.

If non-stationary components are suspected to be presented in the data generating process, these can be included in the covariance structure as part of defining the prior. One example is the non-stationary covariance function for periodic components discussed by MacKay (1998). Another interesting example is the neural network covariance structure of C. K. Williams (1998). In fact, Neal (2012) showed that a Bayesian neural network with one hidden layer converges to a Gaussian Process when the number of hidden neurons goes to infinity. The actual form of covariance function implied by the neural network depends on the prior for the network weights and the activation functions.

It should be noted that sums and products of covariance functions gives valid covariance functions. This enables to formulate a wide range of prior covariance structures based on simple building blocks.

We conclude this section by noting two types of covariance functions that cannot be used for our application. This is because they are both in violation with our Assumption A3 which can be verified by a straight-forward calculation. The first case in the M3/2 covariance function in Equation (6). While it can be used for estimation of TDI, it cannot be used if an estimate of ETI is also required. This is because it does not fulfill the requirement of mixed third-order partial derivatives of the assumption. Another example is $C_\theta^{\mathrm{OU}}(s,t) = \alpha^2 \exp(-|s-t|/\rho)$. This is the covariance function of the Ornstein-Uhlenbeck process which is mean-squared continuous but not mean-squared differentiable. This also fails to satisfy the required assumption.

2.5 The Gaussian assumptions and non-normal outcomes

The closed-form expressions derived in the sections above depend on the assumptions of Gaussianity on the latent scale and for the conditional distribution of the observed data. For outcomes that are conditionally non-normal the posterior distributions are in general analytically intractable. In some cases one can see our assumptions as an approximation facilitated by the central limit theorem e.g., in the case where the outcome is a proportion and the number of experiments is large or if the outcome is count data and the rate is not too low. The conditional variance, σ^2 in Equation (3), could therefore be changed to reflect the variance of this limiting distribution.

One possibility for altering the model is to retain the model structure but apply a transformation of the observed data. This amounts to altering the second part of Equation (3) to

$$g(Y_i) \mid t_i, f(t_i), \Theta \overset{iid}{\sim} N(f(t_i), \sigma^2)$$

where g is a known, monotone function. Our model could directly be fitted to the transformed outcomes, but the posterior estimates will also be on the transformed scale and may therefore be difficult to interpret.

Another possibility is to consider both the outcome and the latent function on the same transformed scale. This alters Equation (3) to

$$\begin{aligned} g(f) \mid \beta, \theta &\sim \mathcal{GP}(\mu_\beta(\cdot), C_\theta(\cdot, \cdot)) \\ g(Y_i) \mid t_i, g(f(t_i)), \Theta &\overset{iid}{\sim} N(g(f(t_i)), \sigma^2) \end{aligned}$$

Our model is again directly applicable under this alternative data generating model. It follows that the trend on this scale is $g(f(t))' = f'(t)g'(f(t))$. The Trend Direction Index will therefore be equal to $\text{TDI}(t, \delta \mid \Theta) = P(f'(t+\delta)g'(f(t+\delta)) > 0 \mid g(\mathbf{Y}), \mathbf{t}, \Theta)$. This can, however, be transformed back to the original scale of *df* by normalization since we can sample from the posterior distribution of $g(f(t))$ by the Gaussian assumptions on the transformed scale. Therefore we can also obtain samples from $g'(f(t))$ since g^{-1} and g' are known. The Trend Direction Index on the original scale can thus be obtained by a Monte Carlo approximation using these samples.

3 Estimation

The trendiness indices depend on the parameters Θ of the latent Gaussian process. These parameters must be estimated from the observed data, and we consider two different estimation procedures: maximum likelihood estimation and a fully Bayesian estimator.

Remark: The difference between the maximum likelihood and the Bayesian method is that they give rise to two different versions of the Trend Direction Index. In the former setting the index is a deterministic function, while in the latter it is a random function governed by the posterior distribution of Θ. The maximum likelihood estimator is also called an empirical Bayes approach since the prior distributions are estimated from data using the marginal likelihood of the model. This can be seen as an approximation to the Bayesian model where the hyper-parameters are fixed at their most likely values instead of being integrated out.

The maximum likelihood estimator consists of finding the values of the parameters that maximize the marginal likelihood of the observed data and plugging these into the expressions of the posterior

distributions and the trendiness indices. The marginal distribution of $\mathbf{Y}$ is found by integrating out the distribution of the latent Gaussian process in the conditional observation model $\mathbf{Y} \mid f(\mathbf{t}), \mathbf{t}, \Theta$ in Equation (3). Since the observation model consists of normal distributed random variables conditional on the latent Gaussian process, the marginal distribution is multivariate normal with expectation $\mu_\beta(\mathbf{t})$ and $n \times n$ covariance matrix $C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I$. The marginal log likelihood function is therefore

$$\log L(\Theta \mid \mathbf{Y}, \mathbf{t}) \propto -\frac{1}{2} \log |C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I| - \frac{1}{2} (\mathbf{Y} - \mu_\beta(\mathbf{t}))^T \left[C_\theta(\mathbf{t}, \mathbf{t}) + \sigma^2 I \right]^{-1} (\mathbf{Y} - \mu_\beta(\mathbf{t})) \tag{7}$$

and the maximum likelihood estimate $\widehat{\Theta^{\mathrm{ML}}} = \arg\sup_{\Theta=(\beta,\theta,\sigma)} \log L(\Theta \mid \mathbf{Y}, \mathbf{t})$ can be obtained by numerical optimization or found as the roots to the score equations $\nabla_\theta \log L(\Theta \mid \mathbf{Y}, \mathbf{t}) = 0$. This estimate can then be plugged in to the expressions for the posterior distributions of $(f, df, d^2 f)$ in Proposition 1 enabling simulation of the posterior distributions at any vector of time points. Point estimates of the Trend Direction Index and the Expected Trend Instability are then $\mathrm{TDI}(t, \delta \mid \widehat{\Theta^{\mathrm{ML}}})$ and $\mathrm{ETI}(\mathcal{I} \mid \widehat{\Theta^{\mathrm{ML}}})$ according to Propositions 2 and 3 respectively. Similarly, the predictive distribution of a new observation is equal to Equation (5) with the plug-in estimate $\Theta = \widehat{\Theta^{\mathrm{ML}}}$.

The maximum likelihood estimator is easy to implement and fast to compute, but it is difficult to propagate the uncertainties of the parameter estimates through to the posterior distributions and the trendiness indices. This is disadvantageous since in order to conduct a qualified assessment of trendiness we are not only interested in point estimates but also the associated uncertainties. A Bayesian estimator, while slower to compute, facilitates a straightforward way to encompass all uncertainties in the final estimates. To specify a Bayesian estimator we must augment the data generating model in Equation (3) with another hierarchical level specifying a prior distribution of the model parameters. We therefore add the following level

$$(\beta, \theta, \sigma) \sim G(\Theta \mid \Psi, \mathbf{t})$$

to the specification where G is some family of distribution functions indexed by a vector of hyper-parameters Ψ that are prior parameters for the parameters Θ, that govern the mean and covariance of f. The joint distribution of the model can be factorized as

$$P(\mathbf{Y}, f(\mathbf{t}), \Theta \mid \Psi, \mathbf{t}) = P(\mathbf{Y} \mid f(\mathbf{t}), \Theta, \Psi, \mathbf{t}) P(f(\mathbf{t}) \mid \Theta, \Psi, \mathbf{t}) G(\Theta \mid \Psi, \mathbf{t})$$

and each conditional probability is specified by a sub-model in the augmented hierarchy. We always condition on Ψ and $\mathbf{t}$ as they are considered fixed data. The posterior distribution of Θ given the observed outcomes is then

$$\begin{aligned} P(\Theta \mid \mathbf{Y}, \Psi, \mathbf{t}) &= \frac{G(\Theta \mid \Psi, \mathbf{t}) P(\mathbf{Y} \mid \Theta, \Psi, \mathbf{t})}{P(\mathbf{Y} \mid \Psi, \mathbf{t})} \\ &= \frac{G(\Theta \mid \Psi, \mathbf{t}) \int P(\mathbf{Y} \mid f(\mathbf{t}), \Theta, \Psi, \mathbf{t}) dP(f(\mathbf{t}) \mid \Theta, \Psi, \mathbf{t})}{\int\int P(\mathbf{Y} \mid f(\mathbf{t}), \Theta, \Psi, \mathbf{t}) dP(f(\mathbf{t}) \mid \Theta, \Psi, \mathbf{t}) dG(\Theta \mid \Psi, \mathbf{t})} \end{aligned}$$

Let $\widetilde{\Theta} \sim P(\Theta \mid \mathbf{Y}, \Psi, \mathbf{t})$. The posterior distribution of Θ induces corresponding distributions over the trendiness indices according to $\text{TDI}(t, \delta \mid \widetilde{\Theta})$, $d\text{EDI}(t, \mathcal{T}, \mid \widetilde{\Theta})$ and $\text{EDI}(\mathcal{I} \mid \widetilde{\Theta})$. For example, the Trend Direction Index in the Bayesian formulation is a surface in (t, δ) where each value is a distribution over probability values. We suggest to summarize the trendiness indices by their posterior quantiles. For example, for the Trend Direction Index we summarize its posterior distribution by functions Q_τ such that

$$P\left(\text{TDI}\left(t, \delta \mid \widetilde{\Theta}\right) \leq \tau\right) = Q_\tau(t, \delta)$$

with for example $\tau \in \{0.025, 0.5, 0.975\}$ to obtain 95% credible intervals. In the Baysian model the predictive distribution in Equation (5) should be averaged across the posterior distribution of the hyper-parameters. This leads to the posterior predictive distribution

$$Y^*(t^*) \mid t^*, \mathbf{Y}, \mathbf{t}, \Psi = \int_\Theta P(Y^*(t^*) \mid t^*, \mathbf{Y}, \mathbf{t}, \Theta, \Psi) dP(\Theta \mid \mathbf{Y}, \Psi, \mathbf{t})$$

where the integral is in practice approximated through the MCMC samples.

We have implemented both the maximum likelihood and the Bayesian estimator in Stan (Carpenter et al. 2017) and R (R Core Team 2018) in combination with the `rstan` package (Stan Development Team 2018). Stan is a probabilistic programming language enabling full Bayesian inference using Markov chain Monte Carlo sampling. The Stan implementation of the maximum likelihood estimator requires the marginal maximum likelihood estimates of the parameters supplied as data, and from these it will simulate random functions from the posterior distribution of (f, df, d^2f) on a user-supplied grid of time points as well as return point estimates of TDI and dETI. The latter can then be integrated numerically to obtain the Expected Trend Instability on an interval. The Bayesian estimator requires values of the hyper-parameters Ψ supplied as data and from these it will generate random samples $(\widetilde{\Theta}_1, \ldots, \widetilde{\Theta}_K)$ from $P(\Theta \mid \mathbf{Y}, \Psi, \mathbf{t})$ by Markov chain Monte Carlo. These samples are then used to approximate the posterior distribution of the Trend Direction Index by $(\text{TDI}(t, \delta \mid \widetilde{\Theta}_1), \ldots, \text{TDI}(t, \delta \mid \widetilde{\Theta}_K))$ and similarly for dETI.

3.1 Model selection

In connection to Section 2.4 on choosing the parameters for the Gaussian process prior we propose to specify a set of candidate models and perform a selection of the best fitting model according to a cross-validation procedure. Different types of cross-validation can be performed depending on the purpose of the analysis. In some cases one stands at the end of the data collection and wants to look at what has happened. In this cases it would be natural to condition on all the observed data and perform a leave-one-out cross-validation. In other cases one is interested in forecasting the trend and the associated indices, and here it would be natural to take the direction of time in to account. This could be done by e.g., a one-step-ahead cross-validation procedure in which the observed data are compared to model forecasts one step ahead in time based on successive partitioning of the time series. For more information on such procedures see e.g., Bergmeir and Benítez (2012).

To perform the leave-one-out cross-validation we consider a set of candidate models indexed by $\mathcal{M}$. This set would typically include different parameterization of the mean and covariance function of the latent process. Turn by turn a single data pair (Y_i, t_i) is excluded and we let the leave-one-out data

be denoted $(\mathbf{Y}_{-i}, \mathbf{t}_{-i})$. Based on the leave-one-out data sets the hyper-parameters $\Theta_{-i}^{\mathcal{M}} = (\beta, \theta, \sigma^2)$ can be estimated for each model by maximizing the marginal log-likelihood. These estimates are given by

$$\widehat{\Theta}_{-i}^{\mathcal{M}} = \arg\sup_{\Theta} \log L(\Theta \mid \mathbf{Y}_{-i}, \mathbf{t}_{-i})$$

where the log likelihood function is given in Equation (7). These estimates are then plugged into the expression for the posterior expectation of f in Proposition 1 to obtain the leave-one-out predictions at times t_i. The mean squared error of prediction (or another loss function) can then be calculated by comparing the leave-one-out predictions and the observed values averaged across all data points

$$\mathrm{MSPE}_{\mathrm{LOO}}^{\mathcal{M}} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \mathrm{E}[f(t_i) \mid \mathbf{Y}_{-i}, \mathbf{t}_{-i}, \widehat{\Theta}_{-i}^{\mathcal{M}}]\right)^2$$

The selected model among the candidate set is then $\mathcal{M}_{\mathrm{opt}} = \arg\min_{\mathcal{M}} \mathrm{MSPE}_{\mathrm{LOO}}^{\mathcal{M}}$. Different cross-validation schemes can be performed similarly by a modification of how the leave-one-out data is constructed.

We note that with our model being implemented in Stan, efficient approximate leave-one-out cross-validation and model comparison using the LOOIC criterion can directly be performed by using the `loo` package (Vehtari et al. 2019).

4 Simulation study

To assess the performance of our method we performed a simulation study. We generated $10,000$ random Gaussian processes on the unit interval with zero mean and the Squared Exponential (SE) covariance function, see Equation (6), with parameters $\alpha = 1$ and $\rho = \frac{\sqrt{3}}{2\pi}$ in 15 different scenarios in which we varied the number of observation points $(n = 25, 50, 100)$ and the measurement noise $(\sigma = 0.025, 0.05, 0.1, 0.15, 0.2)$. The Supplementary Material shows 50 random sample paths for each scenario.

In each $r = 1, \ldots, 10,000$ simulation we know the true latent functions (f_r, df_r) and by fitting our model we obtain estimates $\widehat{f_r^{\mathrm{GP}}}$ and $\widehat{df_r^{\mathrm{GP}}}$ corresponding to the posterior means μ_f and μ_{df} given in Proposition 1 and also TDI_r and $d\mathrm{ETI}_r$. We compare the estimates to the truths using two different measures: an integrated residual and the squared L^2 norm. The integrated residuals are defined as $\int_0^1 (f_r(t) - \widehat{f_r^{\mathrm{GP}}}(t))\mathrm{d}t$ and similarly for $\widehat{df_r^{\mathrm{GP}}}$. For TDI the cumulative residual is defined as $\int_0^1 (1(df_r(t) > 0) - \mathrm{TDI}_r(t))\mathrm{d}t$ where 1 denotes the indicator function. For ETI the cumulative residual is defined as $\int_0^1 (N_r(t) - d\mathrm{ETI}_r(t))\mathrm{d}t$ where $N_r(t)$ is the càdlàg counting process that jumps with 1 every time df_r has a root on the interval. If our estimates are unbiased we expect these integrated residuals to have zero mean across the simulations. The squared L^2 norms are defined in a similar manner for all the quantities as e.g., $\int_0^1 (f_r(t) - \widehat{f_r^{\mathrm{GP}}}(t))^2\mathrm{d}t$ and reflect the variability of the estimates.

For comparison we employed the Trend Filtering method as implemented in the R package `genlasso` (Arnold and Tibshirani 2019) on the same simulated data and reported similar measures for its estimated mean, $\widehat{f^{\mathrm{TF}}}$, and derivative, $\widehat{df^{\mathrm{TF}}}$ using a 10-fold cross-validation of the penalty parameter. Note that we only compare with the estimates of the latent mean and its derivative since Trend

Table 1: Summary statistics from the simulation study. Each value is the average across 10,000 simulations except for ETI where the median is reported. Superscript GP denotes our proposed method and TF denotes Trend Filtering. Numbers have been rounded to three decimal places.

		Integrated residual						Squared L2 norm					
n	σ	$\widehat{f^{\text{GP}}}$	$\widehat{df^{\text{GP}}}$	$\widehat{f^{\text{TF}}}$	$\widehat{df^{\text{TF}}}$	TDI	ETI	$\widehat{f^{\text{GP}}}$	$\widehat{df^{\text{GP}}}$	$\widehat{f^{\text{TF}}}$	$\widehat{df^{\text{TF}}}$	TDI	ETI
25	0.025	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.041	0.504	0.011	0.008
25	0.050	0	0	0.000	0.000	0.000	-0.002	0.001	0.001	0.130	1.188	0.021	0.018
25	0.100	0	0	0.002	0.003	0.001	-0.008	0.003	0.004	0.408	2.988	0.037	0.040
25	0.150	0	0	-0.001	-0.002	-0.001	-0.020	0.005	0.008	0.818	5.248	0.051	0.064
25	0.200	0	0	-0.003	-0.003	-0.001	-0.034	0.009	0.014	1.343	8.248	0.063	0.094
50	0.025	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.026	0.535	0.009	0.006
50	0.050	0	0	0.000	0.000	0.000	-0.002	0.000	0.001	0.081	1.287	0.016	0.012
50	0.100	0	0	0.000	0.000	0.000	-0.006	0.001	0.002	0.244	3.348	0.028	0.028
50	0.150	0	0	0.000	0.001	0.000	-0.012	0.003	0.005	0.469	6.353	0.038	0.045
50	0.200	0	0	0.000	0.003	0.001	-0.020	0.005	0.008	0.759	9.832	0.050	0.068
100	0.025	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.399	0.007	0.004
100	0.050	0	0	0.000	0.001	0.000	-0.001	0.000	0.000	0.050	1.020	0.012	0.009
100	0.100	0	0	0.000	0.000	0.000	-0.003	0.001	0.001	0.150	2.187	0.022	0.019
100	0.150	0	0	0.000	-0.003	0.000	-0.007	0.002	0.002	0.283	4.454	0.030	0.031
100	0.200	0	0	0.002	0.002	0.001	-0.012	0.003	0.004	0.450	6.140	0.038	0.045

Filtering does not provide a probability distribution for the derivative which is required by the definition of TDI.

Summary statistics from the simulations are shown in Table 1. It is seen that both our proposed model and the Trend Filtering model provide unbiased estimates in all scenarios. Looking at the squared L^2 norm our estimates of f and df show very small variability across scenarios, while the estimates from Trend Filtering showed an increase in variability for increasing measurement noise and a decrease for increasing number of observations. We note that this is expected as we simulate processes with continuous sample paths, and Trend Filtering estimates a piece-wise linear approximation. While this leads to unbiased results, as the table shows, the variability of these estimates will necessarily be larger, and this is more pronounced for smaller number of observations.

The variability of TDI and ETI as measured by the squared L^2 norm were low in all scenarios but increased with the amount of measurement noise as expected. In a few cases the estimated ETI was far away from its true value. This was especially pronounced in scenarios with a low number of observations and a high observation noise. This was caused by $\widehat{f^{\text{GP}}}$ degenerating to either a constant function or to a perfect interpolation of the observations. This reason for this behavior is that the hyper-parameters become weekly identified under such circumstances. These cases are still included in the reported summary statistics in Table 1. In such cases, the model can be regularized through the priors of the hyper-parameters, but this must be determined on a case by case basis.

5 Applications

Trend of proportion of Danish smokers

A report published by The Danish Health Authority in January 2019 updated the estimated proportion of daily or occasional smokers in Denmark with new data from 2018 (The Danish Health

Authority 2019). The data was based on an online survey including 5017 participants. The report also included data on the proportion of smokers in Denmark during the last 20 years which was shown in Figure 1. The report was picked up by several news papers under headlines stating that the proportion of smokers in Denmark had significantly increased for the first time in two decades (Navne, Schmidt, and Rasmussen 2019). The report published no statistical analyses for this statement but wrote, that because the study population is so large, then more or less all differences become statistically significant at the 5% level (this was written as a 95% significance level in the report).

This data set provides an instrumental way of exemplifying our two proposed measures of trendiness. In this application we wish to assess the statistical properties of the following questions:

Q1: Is the proportion of smokers increasing in the year 2018 conditional on data from the last 20 years?

Q2: If the proportion of smokers is currently increasing, when did this increase probably start?

Q3: Is it the first time during the last 20 years that the trend in the proportion of smokers has changed?

A naive approach for trying to answer questions Q1 and Q2 is to apply sequential χ^2-tests in a 2×2 table. Table 2 shows the p-values for the χ^2-test of independence between the proportion of smokers in 2018 and each of the five previous years. Using a significance level of 5% the conclusion is ambiguous. Compared to the previous year, there was no significant change in the proportion in 2018. Three out of these five comparisons fail to show a significant change in proportions. It is therefore evident that such point-wise testing is not sufficiently perceptive to catch the underlying continuous development.

Table 2: p-values obtained from χ^2-tests of independence between the proportion of smokers in 2018 and the five previous years. Numbers in bold are statistically significant differences at the 5% level.

2018	2017	2016	2015	2014	2013
p-value	0.074	**0.020**	0.495	**0.012**	0.576

Similarly, a simple approach for trying to answer question Q3 would be to look at the cumulative number of times that the difference in proportion between consecutive years changes sign. In the data set there were nine changes in the sign of the difference between the proportion in each year and the proportion in the previous year giving this very crude estimate of the number of times that the trend has changed. This approach suffers from the facts that it is based on a finite difference approximation at the sampling points to the continuous derivative, and that it uses the noisy measurements instead the latent function. Consequently, the changes in trend are quite unstable.

We now present an analysis of the data set using our method. As a specification of the latent function we considered three different mean functions, a constant mean $\mu_\beta(t) = \beta_0$, a linear mean $\mu_\beta(t) = \beta_0 + \beta_1 t$, and a quadratic mean $\mu_\beta(t) = \beta_0 + \beta_1 t + \beta_2 t^2$ and the four different covariance functions given in Equation (6). This gives a total of 12 different candidate models to compare. Since we condition on all the observed data and are not interested in forecasting in this application we performed the model comparison by leave-one-out cross-validation as discussed in Section 3.1.

Table 3 shows the mean squared error of prediction for each candidate model. For the models with a Rational Quadratic covariance function and a linear and a quadratic mean function the

parameter ν diverged numerically implying convergence to the Squared Exponential covariance function. Comparing the leave-one-out mean squared error of prediction, the prior distribution of f in the optimal model has a constant mean function and a Rational Quadratic covariance function. The marginal maximum likelihood estimates of the parameters in the optimal model were

$$\widehat{\beta_0^{\mathrm{ML}}} = 28.001, \quad \widehat{\alpha^{\mathrm{ML}}} = 4.543, \quad \widehat{\rho^{\mathrm{ML}}} = 4.438, \quad \widehat{\nu^{\mathrm{ML}}} = 1.020, \quad \widehat{\sigma^{\mathrm{ML}}} = 0.622 \tag{8}$$

Table 3: Leave-one-out cross-validated mean squared error of prediction for each of the 12 candidate models. $\Leftarrow$ indicates numerical convergence to the SE covariance function.

	SE	RQ	Matern 3/2	Matern 5/2
Constant	0.682	**0.651**	0.687	0.660
Linear	0.806	$\Leftarrow$	0.896	0.865
Quadratic	0.736	$\Leftarrow$	0.800	0.785

Figure 4 shows the fit of the model by the maximum likelihood method. The plots were obtained by plugging the maximum likelihood estimates into the expressions for the posterior distributions of f and df defined in Proposition 1, the Trend Direction Index in Proposition 2, and the Expected Trend Instability in Proposition 3. The predictions were performed on an equidistant grid of 500 time points spanning the 20 years. The plot of the posterior trend (top right) shows two regions in time where the posterior mean of the derivative is positive, one around $[2004; 2008]$ and one shortly after 2015 and until the end of the observation period. The Trend Direction Index (bottom left) quantifies this positive trendiness as a probability standing in 2018 and looking back in time while also taking the uncertainty into account. The bottom right panel shows the local Expected Trend Instability and its integral is to the expected number of times that the trend has changed sign. Table 4 summarizes the maximum likelihood estimates of the Trend Direction Index at the end of the observation period and the previous five years as well as the Expected Trend Instability during the full observation period and during only the last ten years. Crosspoint is the first point in time during the last ten years where the Trend Direction Index became greater than 50% i.e.,

$$\text{Crosspoint} = \underset{t\in[-10;0]}{\arg\min} \{2018 + t : \text{TDI}(2018, 2018 - t) \geq 50\%\}$$

Based on the results from the maximum likelihood analysis we may answer the questions by stating that the expected proportion of smokers in Denmark is currently increasing with a probability of 95.24%. This is, however, not a recent development as the probability of an increasing proportion has been greater than 50% since the middle of 2015. This can be compared to the sequential χ^2-tests in Table 2, which gave a cruder and less consistent result. The estimated values of ETI in Table 4 show that there has been an average of 3.68 changes in the monotonicity of the proportion during the last 20 years. This value does support the statement by the news outlets that it is the first time in 20 years that the trend has changed. The value, however, reduces to 1.39 when only looking 10 years back, which is slightly less than half the ETI for the longer period.

We also applied the Bayesian estimator to the data using the same prior mean and covariance structure. The Bayesian estimator requires a prior distribution for the model parameters. We used independent priors of the form

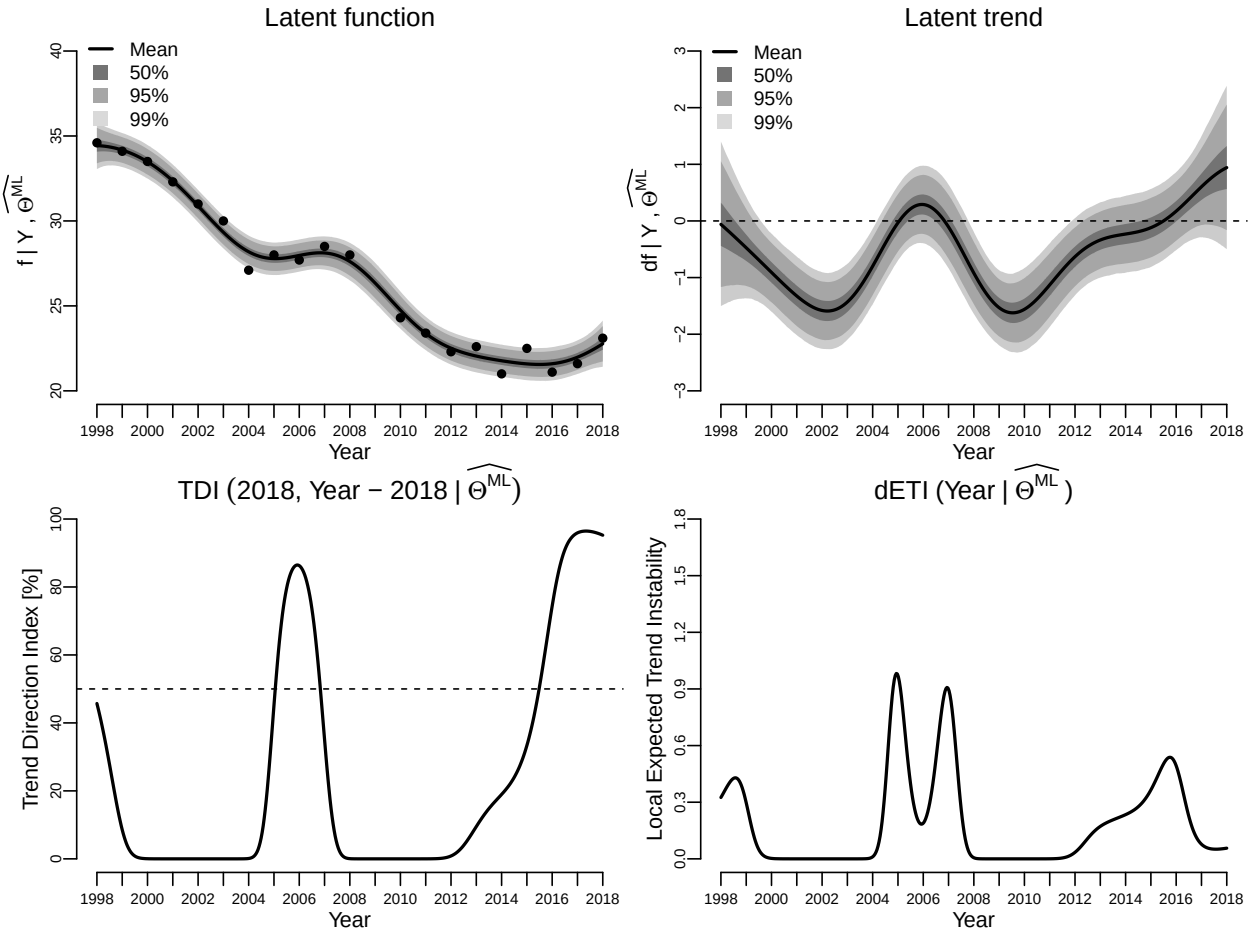

Figure 4: Results from fitting the latent Gaussian process model by maximum likelihood. The first row shows the posterior distributions of f (left) and df (right) with the posterior means in bold and gray areas showing point-wise probability intervals for the posterior distribution. The second row shows the estimated Trend Direction Index (left) and the local Expected Trend Instability (right).

$$G(\Theta \mid \Psi, \mathbf{t}) = G(\beta_0 \mid \Psi_{\beta_0}, \mathbf{t})G(\alpha \mid \Psi_\alpha, \mathbf{t})G(\rho \mid \Psi_\rho, \mathbf{t})G(\nu \mid \Psi_\nu, \mathbf{t})G(\sigma \mid \Psi_\sigma, \mathbf{t})$$

where each prior was a heavy-tailed distribution with a moderate variance centered at the maximum likelihood estimates. We used the following distributions

$$\begin{aligned} \beta_0 &\sim T\left(\widehat{\beta_0^{\mathrm{ML}}}, 3, 3\right), \qquad \alpha \sim \text{Half-}T\left(\widehat{\alpha^{\mathrm{ML}}}, 3, 3\right), \quad \rho \sim \text{Half-}N\left(\widehat{\rho^{\mathrm{ML}}}, 1\right) \\ \nu &\sim \text{Half-}T\left(\widehat{\nu^{\mathrm{ML}}}, 3, 3\right), \quad \sigma \sim \text{Half-}T\left(\widehat{\sigma^{\mathrm{ML}}}, 3, 3\right) \end{aligned}$$

where the maximum likelihood values are given in Equation (8) and Half-$T(\cdot, \cdot, \mathrm{df})$ and Half-$N(\cdot, \cdot)$ denotes the location-scale half T- and normal distribution functions with df degrees of freedom due to the requirement of positivity. We ran four independent Markov chains for 25,000 iterations each

with half of the iterations used for warm-up and discarded. Convergence was assessed by trace plots of the MCMC draws and the potential scale reduction factor, $\widehat{R}$, of Gelman and Rubin (1992). The trace plots are included in the Supplementary Material.

Figure 5 shows results from the Bayesian estimator. In this model both trendiness indices are time-varying posterior distributions and the top row of the shows the posterior distributions of the Trend Direction Index (left) and the Local Expected Trend Instability (right) summarized by time-dependent quantiles. The bottom row shows posterior density estimates of the Expected Trend Instability during the last twenty years (left) and during the last ten years (right). The same summary statistics as for the maximum likelihood analysis are given in Table 4 but here stated in terms of posterior medians and 2.5% and 97.5% posterior quantiles.

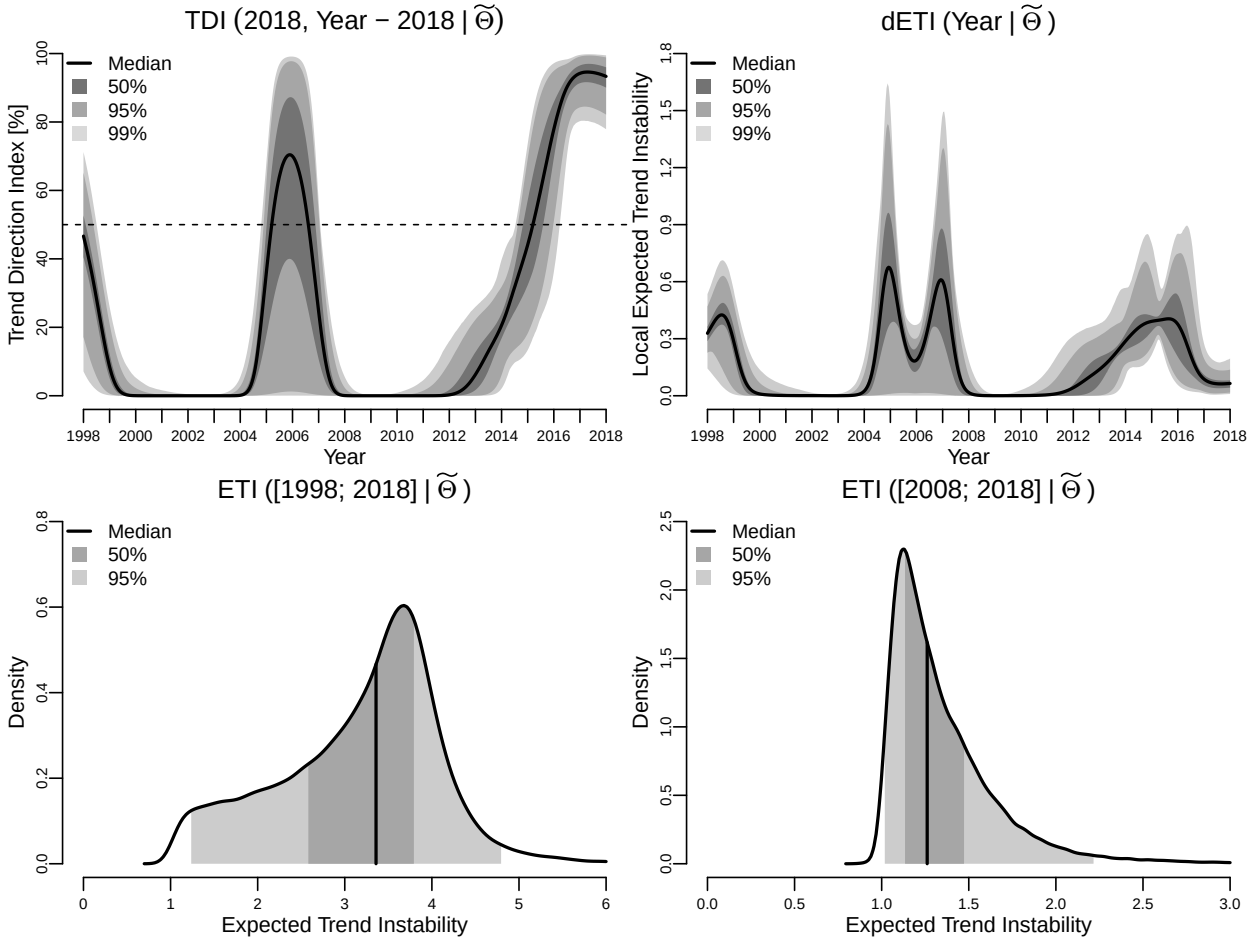

Figure 5: Results from fitting the latent Gaussian process model by Bayesian analysis. The first row shows the posterior distributions of TDI (left) and local ETI (right) with the posterior means in bold and gray areas showing point-wise 95% and 99% probability intervals for the posterior distribution. The second row shows densities and probability intervals for the expected trend instability for the 20-year period back-in-time from 2018 (left) and 10 year back-in-time (right).

The results from the two analyses generally agree, but there are two differences that we wish to address. Both analyses showed a local peak in trendiness around 2006. In the maximum likelihood analysis this occurred at 2005.94 with a Trend Direction Index of 86.47%. In the Bayesian analysis the peak occurs at 2005.87 with a median Trend Direction Index of 79.43% and a 95% posterior

Table 4: Summary measures from the Maximum Likelihood and Bayesian analyses. The rows show the estimated Trend Direction Index for 2013 to 2018 and the Expected Trend Instability for the last 10 and 20 years all conditional on data from 1998 to 2018. For the Bayesian analysis posterior medians and 95% posterior probability intervals are given. Crosspoint is the time during the last 10 years where TDI first exceeded 50%.

	Maximum Likelihood	Bayesian Posterior	
TDI$(2018, 0)$	95.24%	93.32%	[82.15%; 98.86%]
TDI$(2018, -1)$	95.92%	94.21%	[84.28%; 99.11%]
TDI$(2018, -2)$	74.41%	77.87%	[51.02%; 94.94%]
TDI$(2018, -3)$	33.36%	44.11%	[18.23%; 69.19%]
TDI$(2018, -4)$	18.96%	20.60%	[6.05%; 31.82%]
TDI$(2018, -5)$	9.50%	6.21%	[0.03%; 22.21%]
Crosspoint	2015.48	2015.19	[2014.62; 2015.96]
ETI$([1998, 2018])$	3.68	3.36	[1.24; 4.79]
ETI$([2008, 2018])$	1.39	1.25	[1.02; 2.22]

probability interval of [1.23%; 97.73%]. The added uncertainty estimates facilitated by the Bayesian estimator shows this trendiness is so variable that there is no reason to believe in a substantial increase in proportions at that point in time. This insight could not have been obtained from the maximum likelihood analysis.

The second difference is that the Bayesian model seems to generally induce more sluggish estimates due to mixing over the underlying parameters. This can be seen from the plot of the median local Expected Trend Instability in Figure 5 which is generally lower than its corresponding maximum likelihood point estimates in Figure 4. This is similarly reflected in the median ETI estimates in Table 4 which are lower than their values under maximum likelihood. Looking at the posterior distributions of the covariance parameters θ (not shown), we see that this is mainly a result of not restricting the parameter ν to its maximum likelihood value. The 95% probability interval of the posterior distribution of ν was $[0.328; 10.743]$ which is highly right-skewed compared to the maximum likelihood estimate of $\widehat{\nu^{\mathrm{ML}}} = 1.020$.

To understand the effect of ν on the smoothness of the fitted models we can compare the local expected number of crossings by a Gaussian process and its derivative at their mean values in the simple case of a zero-mean process with either the Rational Quadratic or the Squared Exponential covariance function. In this case the formula in Proposition 3 simplifies immensely, and as shown in the Supplementary Material the local expected number of mean-crossings by f is equal to $\pi\rho^{-1}$ for both covariance functions. However, for df the local expected number of mean-crossings is equal to $3^{1/2}\pi^{-1}\rho^{-1}$ for the Squared Exponential covariance function and $3^{1/2}\pi^{-1}\rho^{-1}(1+\nu^{-1})^{1/2}$ for the Rational Quadratic covariance function. We note that $1 < (1+v^{-1})^{1/2} < \infty$ for $0 < \nu < \infty$ and monotonically decreasing for $\nu \to \infty$ with a limit of one. Therefore, the value of ν has no effect on the crossing intensity of the process itself, but its derivative is always larger under a Rational Quadratic covariance function compared to the Squared Exponential covariance function with equality in the limit. A right-skewed posterior distribution of ν therefore favors fewer crossings of the trend leading to a more stable trend and a smaller value of the Expected Trend Instability.

As a final remark on this application we note that the observed data are proportions and therefore by nature not normally distributed. Commonly used transformations for proportions towards normality are the isometric log ratio or the arcsine-square-root or logit functions. In accordance

with the comments in Section 2.5 we also performed the trend analysis on the logit transformed outcomes. These results are included in the Supplementary Material and did not give rise to different interpretations. This is perhaps not surprising as the proportions observed are far from the edges of the parameter space where the normality approximation is more likely to hold.

Number of new COVID-19 positive cases in Italy

As a second application we look at the development of the number of new COVID-19 positives in Italy since February 24th 2020. Data was updated each day and made available at the GitHub repository of the Italian Civil Protection Department (Consiglio dei Ministri - Dipartimento della Protezione Civile 2020). 90 days had passed when this analysis was performed.

The data set provides a direct way to assess the COVID-19 disease progression and to monitor the impact of political initiatives to reduce disease spread. In this application we wish to assess the statistical properties of the following questions:

Q1: Is the disease spread currently under control or does the number of new positive cases seems to be on the rise?

Q2: How did the Italian government's decision to lockdown most of Italy on March 9th 2020 reflect in the spread of the virus?

For this application we fitted the Gaussian Process model using the maximum likelihood method and as hyper-parameters we used a constant mean function and the Rational Quadratic covariance function.

Figure 6 shows the result of the analysis. Between day five and six the Trend Direction Index crossed 95% probability and continued to climb to 100%. After 29 days (March 24th, 2020) the index proceeded to sharply decrease where it remained for some time. The Italian government imposed a quarantine on most of Italy from March 9th, and the sharp drop coincides nicely with the (as of present) expected incubation period of 2–14 days which roughly reflects the incubation and testing period for the individuals who contracted the virus before the March 9th lockdown. After March 24th, the trend was clearly not positive for a long time.

However, at day 88 (May 22nd, 2020) the Trend Direction Index had again increased and crossed 50%, and at the time of this analysis the index was equal to 54%. It is therefore currently unclear whether the number of new positives is increasing or decreasing but a slight probability favors the former. It should be stressed, that the Trend Direction Index monitors the *sign* and not the *size* of the trend and this recent increase of TDI to 54% may reflect that the number of new cases is merely starting to level off ($df(t) \approx 0$). The Trend Direction Index can be used as a monitoring tool to determine if the Italian authorities need to take extra actions.

6 Discussion

In this article, we have proposed two new measures - the Trend Direction Index and the Expected Trend Instability - in order to quantify the behavior of an underlying trend. Using a Gaussian process model for the latent structure we show how the TDI and the ETI can be estimated from data in both a maximum likelihood and a Bayesian framework and provide probabilistic statements about the monotonicity of the latent development of an observed outcome over time.

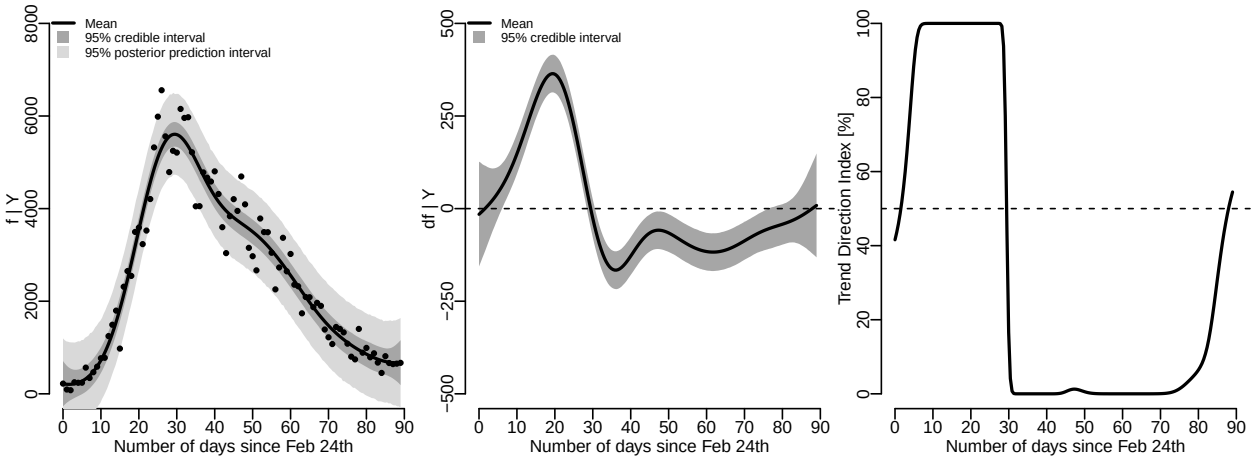

Figure 6: Results from the trend analysis of the Italian COVID-19 data. The left panel shows the number of new positives since February 24th along with the posterior mean of f and 95% credible and posterior prediction intervals. The middle panel shows the posterior distribution of the trend (df), and the right panel shows the Trend Direction Index.

Both indices have an intuitive interpretation that directly refers to properties of the latent trend and the proposed approach exploits the underlying assumed continuity to circumvent discretization and allows us to compute a different (and in many cases more relevant) probability than what can be obtained from e.g., pairwise comparisons in discrete time.

The comparison with Trend Filtering showed that our proposed method provides unbiased estimates of the underlying function f and its derivative, df but that we are able to relax the assumptions of piece-wise linearity of the underlying process, while in addition providing a probability distribution for the derivative which is used directly in the Trend Direction Index and which directly provides the answers to the research questions often posed.

It is worth noting that the indices are scale-free and thus they do not tell anything about the magnitude of a trend. Consequently, the two indices should therefore always be accompanied by plots of the posterior of df. If a prespecified magnitude, u, of a trend is desired then the threshold in the definition of the TDI is easily modified to accommodate this as $\text{TDI}_u(t, \delta) = P(df(t + \delta) > u \mid \mathcal{F}_t)$.

In conclusion, we have introduced a method for evaluation of trends that specifically address questions such as "Has the trend changed?". Our approach is based on two intuitive measures that are easily interpreted, provide well-defined measures of trend behavior, and which can be applied to a large number of situations. The flexibility of the model is further improved by the minimum of assumptions necessary to provide about the underlying latent trend.

Acknowledgements

The authors would like to thank two anonymous reviewers and the associate editor for providing thoughtful comments that substantially improved the manuscript.

7 Bibliography

Arnold, Taylor B., and Ryan J. Tibshirani. 2019. *genlasso: Path Algorithm for Generalized Lasso Problems*. https://CRAN.R-project.org/package=genlasso.

Barry, Daniel, and J. A. Hartigan. 1993. "A Bayesian Analysis for Change Point Problems." *Journal of the American Statistical Association* 88 (421). [American Statistical Association, Taylor & Francis, Ltd.]: 309–19. http://www.jstor.org/stable/2290726.

Basseville, Michele, and Igor V. Nikiforov. 1993. *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall.

Bergmeir, Christoph, and José M Benítez. 2012. "On the Use of Cross-Validation for Time Series Predictor Evaluation." *Information Sciences* 191. Elsevier: 192–213.

Carlstein, Edward, Hans-Georg Müller, and David Siegmund, eds. 1994. *Change-Point Problems*. Vol. 23. Lecture Notes – Monograph Series. Institute of Mathematical Statistics.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76 (1).

Chandler, R., and M. Scott. 2011. *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. Statistics in Practice. Wiley.

Consiglio dei Ministri - Dipartimento della Protezione Civile, Presidenza del. 2020. "COVID-19 Italia - Monitoraggio Situazione." https://github.com/pcm-dpc/COVID-19.

Cramer, Harald, and M. R. Leadbetter. 1967. *Stationary and Related Stochastic Processes – Sample Function Properties and Their Applications.* John Wiley & Sons, Inc.

Esterby, S. R. 1993. "Trend Analysis Methods for Environmental Data." *Environmetrics* 4 (4): 459–81. doi:10.1002/env.3170040407.

Gelman, Andrew, and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7 (4): 457–72.

Gottlieb, Andrea, and Hans-Georg Müller. 2012. "A Stickiness Coefficient for Longitudinal Data." *Computational Statistics & Data Analysis* 56 (12): 4000–4010.

Hodrick, Robert J., and Edward C. Prescott. 1997. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking* 29 (1). [Wiley, Ohio State University Press]: 1–16. http://www.jstor.org/stable/2953682.

Kaufman, Cari G, Derek Bingham, Salman Habib, Katrin Heitmann, Joshua A Frieman, and others. 2011. "Efficient Emulators of Computer Experiments Using Compactly Supported Correlation Functions, with an Application to Cosmology." *The Annals of Applied Statistics* 5 (4). Institute of Mathematical Statistics: 2470–92.

Kim, Seung-Jean, Kwangmoo Koh, Stephen Boyd, and Dimitry Gorinevsky. 2009. "ℓ_1 Trend Filtering." *SIAM Review* 51: 339–60.

Kowal, Daniel R., David S. Matteson, and David Ruppert. 2019. "Dynamic Shrinkage Processes."

JRSS-B 81: 781–804.

Kryger Jensen, Andreas. 2019. "GitHub Repository for the Trendiness of Trends." https://github.com/aejensen/TrendinessOfTrends.

MacKay, D. J. C. 1998. "Introduction to Gaussian Process." *Neural Networks and Machine Learning*. Springer.

Navne, Helene, Anders Legarth Schmidt, and Lars Igum Rasmussen. 2019. "Første Gang I 20 år: Flere Danskere Ryger." https://politiken.dk/forbrugogliv/sundhedogmotion/art6938627/Flere-danskere-ryger.

Neal, R.M. 2012. *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics. Springer New York.

Quandt, Richard E. 1958. "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes." *Journal of the American Statistical Association* 53 (284): 873–80.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Radford, Neal M. 1999. "Regression and Classification Using Gaussian Process Priors (with Discussion)." In *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting*, edited by A. P. Dawid José M. Bernardo James O. Berger and Adrian F. M. Smith, 475–501.

Ramdas, Aaditya, and Ryan J. Tibshirani. 2016. "Fast and Flexible Admm Algorithms for Trend Filtering." *Journal of Computational and Graphical Statistics* 25 (3). Taylor & Francis: 839–58. doi:10.1080/10618600.2015.1054033.

Rasmussen, C. E., and C. K. I. Williams. 2006. *Gaussian Processes in Machine Learning*. MIT Press.

Stan Development Team. 2018. "RStan: The R Interface to Stan." http://mc-stan.org/.

The Danish Health Authority. 2019. "Danskernes Rygevaner 2018." www.sst.dk/da/udgivelser/2019/danskernes-rygevaner-2018.

Vehtari, Aki, Jonah Gabry, Mans Magnusson, Yuling Yao, and Andrew Gelman. 2019. "loo: Efficient Leave-One-Out Cross-Validation and Waic for Bayesian Models." https://mc-stan.org/loo.

Williams, Christopher KI. 1998. "Computation with Infinite Neural Networks." *Neural Computation* 10 (5). MIT Press: 1203–16.