Guanglu Zhang
Mem. ASME
Department of Mechanical Engineering,
Texas A&M University,
College Station, TX 77840
e-mail: glzhang@tamu.edu

Douglas Allaire
Mem. ASME
Department of Mechanical Engineering,
Texas A&M University,
College Station, TX 77840
e-mail: dallaire@tamu.edu

Daniel A. McAdams[1]
Fellow ASME
Department of Mechanical Engineering,
Texas A&M University,
College Station, TX 77840
e-mail: dmcadams@tamu.edu

Venkatesh Shankar
Center for Retailing Studies,
Mays Business School,
Texas A&M University,
College Station, TX 77840
e-mail: vshankar@mays.tamu.edu

Generating Technology Evolution Prediction Intervals Using a Bootstrap Method

Technology evolution prediction is critical for designers, business managers, and entrepreneurs to make important decisions during product development planning such as R&D investment and outsourcing. In practice, designers want to supplement point forecasts with prediction intervals to assess future uncertainty and make contingency plans accordingly. However, prediction intervals generation for technology evolution has received scant attention in the literature. In this paper, we develop a generic method that uses bootstrapping to generate prediction intervals for technology evolution. The method we develop can be applied to any model that describes technology performance incremental change. We consider parameter uncertainty and data uncertainty and establish their empirical probability distributions. We determine an appropriate confidence level to generate prediction intervals through a holdout sample analysis rather than specify that the confidence level equals 0.05 as is typically done in the literature. In addition, our method provides the probability distribution of each parameter in a prediction model. The probability distribution is valuable when parameter values are associated with the impact factors of technology evolution. We validate our method to generate prediction intervals through two case studies of central processing units (CPU) and passenger airplanes. These case studies show that the prediction intervals generated by our method cover every actual data point in the holdout sample tests. We outline four steps to generate prediction intervals for technology evolution prediction in practice.
[DOI: 10.1115/1.4041860]

Keywords: prediction interval, technology evolution, uncertainty, bootstrapping, Moore's Law, Lotka-Volterra equations

1 Introduction

Technology evolution prediction, also called technological forecasting, models the performance change of a technology over time. Technology evolution prediction helps designers, business managers, and entrepreneurs make informed R&D and outsourcing decisions during product development planning [1,2]. Specifically, practitioners rely on the prediction to set reasonable R&D targets for their product in order to compete with other players in the industry. Technology evolution prediction also aids decision makers by providing estimates of when the performance of a technology will reach a desirable value for adoption.

The prediction results generated by existing models of technology evolution prediction are expressed as single numbers, which are called point forecasts. Designers often want to supplement point forecasts by computing interval forecasts to assess future uncertainty and make contingency plans accordingly [3]. For example, it is often necessary for designers to estimate the earliest and the latest time at which the technology of interest would achieve an expected performance. A comprehensive R&D plan is usually developed based on such estimation. However, prediction intervals generation for technology evolution has received scant attention in the literature. Researchers have used continuous mathematical functions or differential equations (e.g., exponential function, logistic function, and Lotka–Volterra equations) to model the performance change of a technology over time but seldom supplemented their point forecast results for future technology performance with prediction intervals [4–6].

In this paper, we introduce a broadly applicable method to generate prediction intervals for technology evolution prediction. To our knowledge, a prediction interval generation method that can be applied to any model that predicts technology performance incremental changes is a new contribution to technology evolution prediction. The method we develop is based on a bootstrapping [7–9] approach and does not rely on any parametric assumptions (e.g., assumptions of normality). We consider parameter uncertainty and data uncertainty and establish their empirical probability distributions. We determine an appropriate confidence level, α, to generate reasonable prediction intervals through a holdout sample analysis rather than set $\alpha = 0.05$ as is frequently done in the literature. In addition, our method provides the probability distribution of each parameter in a prediction model. The probability distribution is valuable for designers when parameter values are associated with the impact factors of technology evolution (e.g., performance upper limit in the logistic S-curve model or technology interaction in the Lotka–Volterra ecosystem model).

This paper begins with a short review of the related literature in Sec. 2. We analyze the uncertainty pertaining to technology evolution prediction and develop a method to generate prediction intervals based on a bootstrapping approach in Sec. 3. To ease the implementation of our method in practical projects of technology evolution prediction, we summarize four steps to implement the method in Sec. 4. In Sec. 5, we illustrate the application of our method using evolution data on central processing units (CPUs) and passenger airplanes. We conclude with a brief discussion of our contributions and outline future research directions.

2 Background and Related Work

There are three primary types of uncertainty encountered in prediction problems [10]. These uncertainties are often referred

[1]Corresponding author.

Contributed by the Design Automation Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received April 29, 2018; final manuscript received October 24, 2018; published online January 31, 2019. Assoc. Editor: Harrison M. Kim.

to as model uncertainty, parameter uncertainty, and data uncertainty. Model uncertainty comes from the structure of the model. Considerable error may occur if a prediction model is not appropriate for the problem [10,11]. Parameter uncertainty is rooted in the estimation of the model parameters. Researchers usually assume that the model parameters are constants during a data fitting process. However, the value of each parameter may change over time [3,6], and the variation of each parameter may lead to prediction error. Data uncertainty covers the observed data's random variations that are not explained by model uncertainty and parameter uncertainty. The random variation is often represented by a random error (or white noise) term, $\boldsymbol{\varepsilon}$, in prediction models [3,12–14]. Of note, researchers often categorize the sources of uncertainty as either aleatory or epistemic in engineering modeling for risk and reliability analysis [15]. Here, the model uncertainty and the parameter uncertainty belong to the epistemic category; the data uncertainty is characterized as aleatory.

Researchers have considered these three types of uncertainty and developed several methods to construct prediction intervals for time series problems [3,14,16]. Although time series models are not commonly used in technology evolution prediction, these methods associated with time series models could provide a framework for new methods in uncertainty estimation of technology evolution prediction. A typical time series prediction model has the following mathematical form

$$y_t = F_t(y_{t-1}, y_{t-2}, \ldots y_{t-p}; \boldsymbol{\theta}) + \varepsilon_t \tag{1}$$

where y_t is the variable of interest at time t, F is a function that takes arguments y at previous time $t-1$, $t-2$, …, $t-p$ and potential parameters $\boldsymbol{\theta}$, and ε is an independent random error term. Typical time series data (e.g., yearly United States' oil production) are uniformly spaced. One data point is given at each time period (e.g., monthly or yearly). The current value of the series, y_t, can be explained as a function of p consecutive past y values y_{t-1}, y_{t-2}, …, y_{t-p}, and other variables $\boldsymbol{\theta}$ (e.g., q past error ε values ε_{t-1}, ε_{t-2}, …, ε_{t-q} in a mixed autoregressive moving average model [14]).

To mitigate the model uncertainty and improve the robustness and efficiency of statistical models, researchers select an optimal prediction model from a wide range of time series models by minimizing a statistic such as Akaike's information criterion or Schwarz Bayesian information criterion [10,12]. Several procedures have been established to select an optimal model, which include, but are not limited to, the Box-Jenkins procedure for autoregressive integrated moving average (ARIMA) models and the Holt-Winters procedure for exponential smoothing models [10,12,17,18]. Model uncertainty is generally neglected once the optimal prediction model is selected via one of these procedures.

Many times, the parameter uncertainty is underemphasized in time series analysis. However, parameter uncertainty cannot be overlooked in a model with many parameters or when the number of observed data points is small [3]. Researchers often perform sensitivity analysis to assess parameter uncertainty [19]. Bayesian approach has also been widely adopted in time series problems to quantify parameter uncertainty [20].

Data uncertainty is described by the random error term, ε, in Eq. (1). Researchers often assume the random error is an independent variable that follows a parametric distribution. Assumptions of normality are commonly used for the random error term ε [3] but no complete consensus exists among researchers regarding this choice. Harvey suggested the use of a t-distribution to describe the random error for a Gaussian time series model [18]. Williams and Goodman analyzed six time series of residences and the number of business main telephones in service on the last day of the month for three Michigan cities [21]. They found that the absolute values of the prediction error approximately follow a gamma distribution [21].

If we define current time as t and want to predict the value of variable y at time $t+k$ as y_{t+k}, the general form of a $100(1-\alpha)$ % prediction interval for y_{t+k} in a time series analysis is [3]

$$F_{t+k} \pm z_{\alpha/2}\sqrt{\mathrm{var}(\varepsilon_{t+k})}, \tag{2}$$

where F_{t+k} is the value of the function F at time $t+k$, $z_{\alpha/2}$ is the $\alpha/2$ percentage point of the parametric distribution of the random error ε, var (ε_{t+k}) is the variance of random error ε at time $t+k$. In practice, researchers often set confidence level $\alpha = 0.05$ and compute the 95% prediction intervals for their problem. Equation (2) also assumes the model prediction, F_{t+k}, is unbiased with expected mean squared prediction error. The analytic form of var (ε_{t+k}) could be derived for several time series modes. For example, the random walk with drift model has the form [14]

$$y_{t+1} = \delta + y_t + \varepsilon_{t+1} \tag{3}$$

where the constant δ is called drift. The variance of the random error, ε, at time $t+k$ has the form

$$\mathrm{var}(\varepsilon_{t+k}) = k\sigma_\varepsilon^2 \tag{4}$$

where σ_ε^2 equals var (ε_{t+1}), which is the variance of the one step ahead prediction error [3,22]. For the time series models for which the analytic form of var (ε_{t+k}) is not available, researchers have to use approximate formulas [22] or numerical approaches (e.g., Monte Carlo simulation or bootstrapping) to estimate the variance [3,23]. Of note, typical time series models (e.g., ARIMA and exponential smoothing [12]) represented by Eq. (1) require uniformly spaced data. Advanced time series models (e.g., ACD-GARCH model [24]) exist for time series data with nonuniform spacing.

Time series models and the associated prediction interval generation methods have been successfully applied in several areas such as econometrics, demography, marketing, and medical science [12,14,16]. However, time series models are not commonly used in technology evolution prediction. Researchers often believe that an underlying law governs the incremental change in technology evolution that involves a dominant design [25–27]. Such law is usually described by continuous mathematical functions or differential equations involving time rather than time series models. Moreover, a typical technology evolution problem involves fewer than 30 data points. Statistical theorems, such as the central limit theorem, are not applicable with such a small sample size. It is also hard to validate a parametric distribution assumption for the error term ε because a parametric distribution test (e.g., normality test) has less power to reject the null hypothesis due to the small sample size [28].

Popular technology evolution prediction models include, but are not limited to, Moore's Law, the logistic S-curve model, and the Lotka-Volterra ecosystem model [4,6,29]. Researchers have fitted these models to nonuniformly spaced data and predict technology performance through mathematical extrapolation. However, researchers seldom supplemented their point forecast results with prediction intervals. The limited publications that consider uncertainty in technology evolution prediction include the works from Farmer and Lafond [30], Arendt et al. [31], Naim and Lewis [32], and Nagy et al. [33]. Farmer and Lafond modified Moore's Law as a correlated geometric random walk with drift model and constructed the prediction intervals through an approximate approach [30]. However, the model and the approach developed by Farmer and Lafond only work for uniformly spaced time series data. Arendt et al. fitted technology evolution data by an Erto-Lanzotti S-curve model and considered the parameter uncertainty through a Monte Carlo simulation [31]. The method provided by Arendt et al. is valuable for decision making in design, but the method cannot be applied to the technology evolution that does not follow the Erto-Lanzotti S-curve model. Naim and Lewis introduced an n-dimensional growth model for engineering system performance evolution [32]. They considered the uncertainty of one parameter (performance upper limit) through a Monte Carlo simulation. The data uncertainty and the uncertainty pertaining to other parameters in their model were neglected. Nagy et al. tested six technology evolution models in their paper [33]. The technology performance Nagy et al. discussed was production or price rather than technical performance metrics (e.g., speed or capacity) that are of primary interest to designers.

3 Prediction Intervals Generation Using a Bootstrap Method

Based on the capability void discussed in Sec. 2, we introduce a generic method to generate prediction intervals for technology evolution prediction in this section. The method we introduce can be applied to any technology evolution prediction model that describes the incremental change in technology performance. The discussion of more fundamental or radical technological changes (e.g., changes in system configuration or functionality, and disruptive innovations) is beyond the scope of this paper. We discuss the three types of uncertainty in technology evolution prediction and establish the empirical probability distributions of parameter uncertainty and data uncertainty through bootstrapping. We also present a holdout sample analysis to determine the confidence level, α, for prediction intervals generation. In addition, we construct the probability distribution of each parameter in a prediction model. Of note, researchers also could transform the nonuniformly spaced technology evolution data to uniformly spaced data through an interpolation approach. Typical time series models (e.g., ARIMA or exponential smoothing [12]) could then be used for technology evolution prediction. However, the interpolation approach introduces artificial data points that bring fresh uncertainty in technology evolution prediction, and this new uncertainty is hard to address in prediction intervals generation.

As a generic problem of technology evolution prediction, there are n technology performance data points from start time, T_1, to current time, T_2. Designers want to predict the technology performance from current time, T_2, to future time, $T_2+\tau$, with prediction intervals. The n technology performance data at time $t_1, t_2, \ldots, t_n$ are denoted by $y_1, y_2, \ldots, y_n$, respectively. A technology evolution prediction model $M(t;\ \boldsymbol{\varphi})$ is chosen to fit the data, where t is time and $\boldsymbol{\varphi} = (A, B, C, \ldots)$ represents the constant parameters in model M. Each technology performance data point is expressed as

$$y_i = M(t_i; \boldsymbol{\varphi}_0) + \varepsilon_i \tag{5}$$

where $i \in \{1, 2, \ldots, n\}$, $\boldsymbol{\varphi_0} = (A_0, B_0, C_0, \ldots)$ denotes estimated parameters derived from a data fitting process, and ε_i is the deviation from the model at each data point.

The point forecast of technology performance at time t_e, where $T_2 < t_e < T_2+\tau$ is obtained by mathematical extrapolation as $M(t_e;\ \boldsymbol{\varphi_0})$. An optimal prediction model should be selected to minimize the model uncertainty. It is a challenging task to select an optimal model from a wide range of technology evolution prediction models. Such discussion is beyond the scope of this paper. We assume that $M\ (t;\ \boldsymbol{\varphi})$ is the optimal model and do not consider the model uncertainty. To incorporate the model uncertainty, designers could review the work of Chatfield [11], Meade and Islam [29], Young [34], and Draper [35].

We estimate the parameter uncertainty using a bootstrap method. The bootstrap method, also called bootstrapping or the resample method, was introduced by Efron [7]. The method has been used to construct prediction intervals for regression, time series, and growth curve models [8,9,23,36,37]. Here, an original sample is comprised of the deviation terms in Eq. (5) as $\boldsymbol{E_0} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$. We create a resample $\boldsymbol{E_1}$ of size n by drawing elements from the original sample, $\boldsymbol{E_0}$, with replacement. Each element in sample $\boldsymbol{E_0}$ has an equal probability of $1/n$ to be drawn. The resampling process is repeated R times to generate R resamples as

$$\boldsymbol{E_j} = (\varepsilon_1^j, \varepsilon_2^j, \ldots, \varepsilon_n^j) \tag{6}$$

where $j \in \{1, 2, \ldots, R\}$. Typically, R is at least 1000 for prediction intervals generation [9]. Each resample is made up of the elements in the original sample ($\varepsilon_i^j \in \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$). We then derive the R bootstrapped technology performance data sets as $\boldsymbol{Y_j} = (y_1^j, y_2^j, \ldots, y_n^j)$ where

$$y_i^j = M(t_i; \boldsymbol{\varphi}_0) + \varepsilon_i^j \tag{7}$$

The model parameters can be estimated using a specified data fitting method (e.g., ordinary least squares) for the bootstrapped technology performance data sets $\boldsymbol{Y_j}$ as $\boldsymbol{\varphi_j} = (A_j, B_j, C_j, \ldots)$. An empirical probability distribution $\boldsymbol{M_e}(t_e;\ \boldsymbol{\varphi})$ at time t_e is built by setting the same probability, $1/R$, at each point $M(t_e;\ \boldsymbol{\varphi_1})$, $M(t_e;\ \boldsymbol{\varphi_2}), \ldots, M(t_e;\ \boldsymbol{\varphi_R})$ [38]. The empirical probability distribution $\boldsymbol{M_e}(t_e;\ \boldsymbol{\varphi})$ describes the parameter uncertainty in technology evolution prediction. Here, we also can build the empirical probability distribution of each parameter in the prediction model M from $\boldsymbol{\varphi_j}$. For example, we can build an empirical probability distribution $\boldsymbol{A_e}$ for parameter A by setting the same probability, $1/R$, at each point $A_1, A_2, \ldots, A_R$. If the parameters have a clear interpretation in technology evolution context (e.g., technology performance upper limit in the logistic S-curve model or technology interaction in the Lotka-Volterra ecosystem model), the probability distributions of the parameters built here could be a key reference for designers to make R&D and outsourcing decisions, which is illustrated by the passenger airplane case study in Sec. 5.2.

The data uncertainty is represented by the deviation term, ε. We assume the deviation is an independent random variable that is applied to any technology performance data from T_1 to $T_2+\tau$. It is hard to assume any parametric distribution (e.g., normal distribution) for the random variable $\boldsymbol{\varepsilon}$ because a typical technology evolution problem has less than 30 data points. We construct an empirical probability distribution $\boldsymbol{E_e}\ (\varepsilon)$ from the sample $\boldsymbol{E_0} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ by setting the same probability of $1/n$ to each element in the sample $\boldsymbol{E_0}$ [38]. The empirical probability distribution, $\boldsymbol{E_e}(\varepsilon)$, describes the data uncertainty in technology evolution prediction.

Above all, we could generate the prediction intervals using the empirical probability distributions $\boldsymbol{M_e}(t_e;\ \boldsymbol{\varphi})$ and $\boldsymbol{E_e}(\varepsilon)$. The percentile method is used because of its convenience [8]. A $100(1-\alpha)$ % prediction interval for technology performance y at time t_e is given by

$$[M_e^{\alpha/2} + \varepsilon_e^{\alpha/2}, M_e^{1-(\alpha/2)} + \varepsilon_e^{1-(\alpha/2)}] \tag{8}$$

where $M_e^{\alpha/2}$ and $M_e^{1-(\alpha/2)}$ are the $\alpha/2$ and $1-\alpha/2$ percentiles of the $\boldsymbol{M_e}\ (t_e;\ \boldsymbol{\varphi})$ distribution, $\varepsilon_e^{\alpha/2}$ and $\varepsilon_e^{1-(\alpha/2)}$ are the $\alpha/2$ and $1-\alpha/2$ percentiles of the $\boldsymbol{E_e}(\varepsilon)$ distribution. For example, a 95% prediction interval ($\alpha = 0.05$) for technology performance at time t_e is

$$\left[M_e^{2.5\%} + \varepsilon_e^{2.5\%}, M_e^{97.5\%} + \varepsilon_e^{97.5\%}\right] \tag{9}$$

Of note, a $100(1-\alpha)$ % prediction interval does not cover $100(1-\alpha)$ % possible technology performance in a future time period [3,21,22]. The value of α is considered a designer's choice. It is of significance for designers to determine an appropriate value for the confidence level α to generate reasonable prediction intervals. Such reasonable prediction intervals should capture the uncertainty of future technology performance in a modest range rather than an exaggerated wide range. In practice, researchers often follow the convention in statistical hypothesis testing and take $\alpha = 0.05$ to generate 95% prediction intervals. However, the 95% prediction intervals sometimes are too wide in technology evolution prediction (e.g., Figs. 1 and 2 in Sec. 5), and such wide intervals are not useful for designers to make R&D decisions. We suggest using holdout sample analysis to determine an appropriate

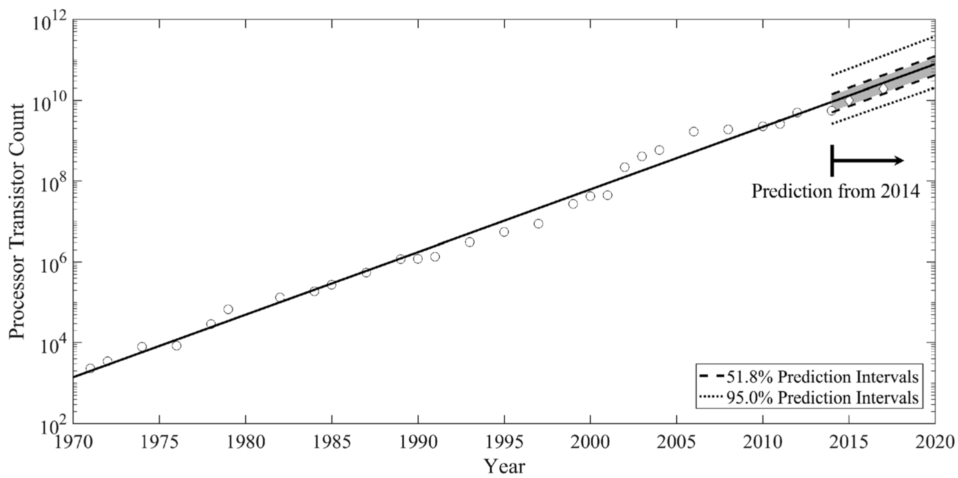

Fig. 1 CPU transistor count evolution prediction intervals from 2014 to 2018

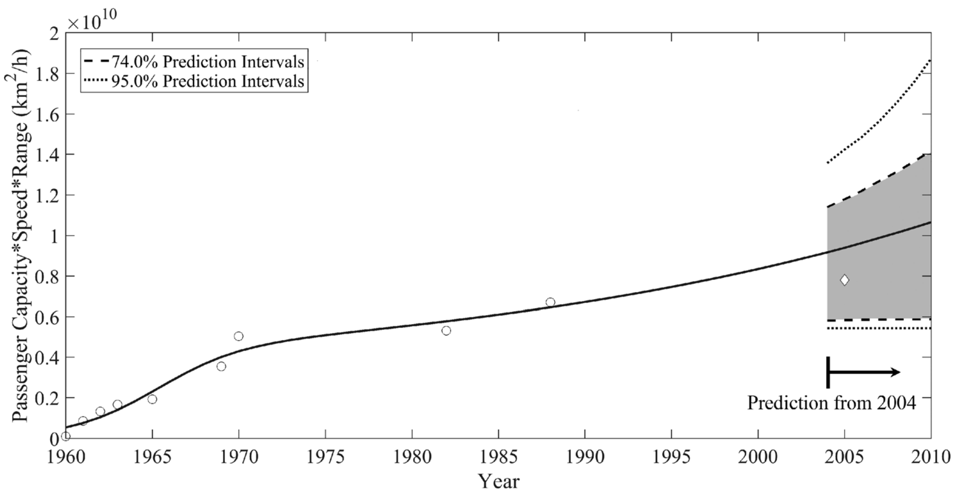

Fig. 2 Passenger airplane performance evolution prediction from 2004 to 2008

value for the confidence level α and generate reasonable prediction intervals accordingly. The suggested holdout sample analysis gives an appropriate value of α based on prior technology performance evolution data rather than a constant value of α. This strategy considers the difference between diverse technologies on evolution trend and the associated uncertainty. First, the same model M is used to fit the data points from start time T_1 to time $T_2-\tau$. The empirical probability distributions $\boldsymbol{M_e}(t_e;\,\boldsymbol{\varphi})$ and $\boldsymbol{E_e}(\varepsilon)$ at time t_e, where $T_2-\tau<t_e<T_2$, can be derived following the preceding procedure. There are k known data points from time $T_2-\tau$ to time T_2, from which the narrowest prediction intervals that just cover the k data points can be identified. The confidence level α_e associated with the narrowest prediction intervals is used as the appropriate value of confidence level. We then use the appropriate confidence level α_e to generate the prediction intervals from current time, T_2, to future time, $T_2+\tau$. The underlying assumption of this method is that the prediction uncertainty from time T_2 to time $T_2+\tau$ is smaller than or equal to the uncertainty from time $T_2-\tau$ to time T_2. This assumption is validated through two case studies of CPUs and passenger airplanes evolution predictions in Sec. 5. It is shown that the prediction intervals generated by the confidence level α_e suffice to cover every actual data point in the holdout sample tests. Meanwhile, the 95% prediction intervals are much wider than the prediction intervals generated by the confidence level α_e.

4 Four Steps to Generate Prediction Intervals for Technology Evolution Prediction

In Sec. 3, we presented a method to generate prediction intervals for a generic problem of technology evolution prediction. In this section, we summarize that procedure to implement the method in four steps. Designers can follow these guidelines to supplement point forecasts with prediction intervals in technology evolution prediction.

Step 1—Data collection and pretreatment. Designers first collect the technology performance data for a time interval of interest. The time interval usually begins with a past time, T_1, and ends in current time, T_2. There may be more than one data point in a specific time period (e.g., several CPUs with different transistor counts are introduced in the same year). Designers retain only the greatest performance value in each time period because that data point represents the best available technology performance at the time. Designers also delete performance values that are lower than those during previous time periods. The remaining data points are used for technology evolution prediction. Designers can transform (e.g., a log transformation) or normalize (e.g., dimensionless treatment) the data if necessary.

Step 2—Model selection. Designers select a technology evolution prediction model for their specific problem. Descriptive models (e.g., S-curve models or Moore's Law) are simple

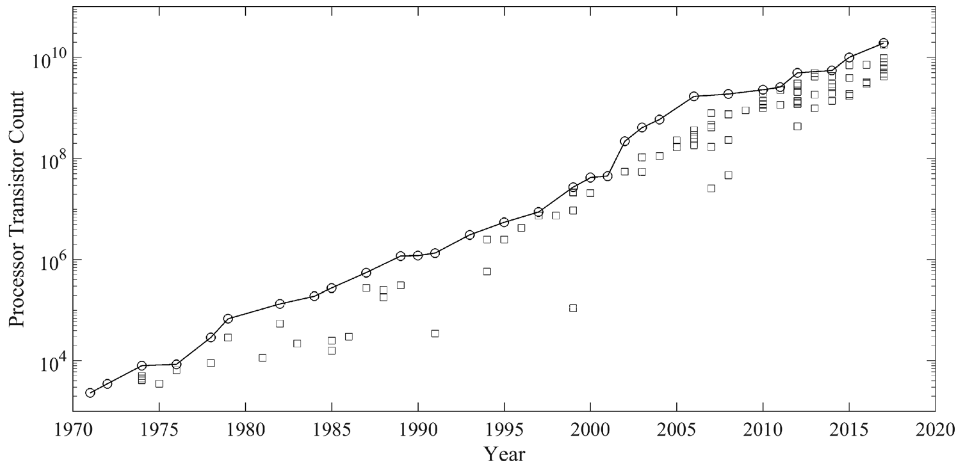

Fig. 3 CPU transistor count data during 1970–2018

mathematical functions that are applicable for fundamental technology evolution prediction in which technology interaction is ignored. If technology interaction needs to be considered, a system model (e.g., the Lotka-Volterra ecosystem model) is preferable [6,39]. Of note, the best fit model usually is not the best prediction model [11]. The model selection should be based on the results from several holdout sample tests as well as the advice from domain or subject matter experts.

Step 3—Confidence level determination. Designers conduct a holdout sample analysis using the technology evolution data from time T_1 to time $T_2-\tau$. The technology evolution prediction model selected in step 2 is estimated using the holdout data set, and the empirical probability distributions for model prediction $\boldsymbol{M_e}(t_e;\,\boldsymbol{\varphi})$ and deviation $\boldsymbol{E_e}(\varepsilon)$ at time t_e, where $T_2-\tau<t_e<T_2$, are derived as described in Sec. 3. Designers calculate a confidence level α associated with each data point during time interval $(T_2-\tau,\,T_2)$. The smallest value of confidence level α corresponds to the narrowest prediction intervals that can cover every data point during time interval $(T_2-\tau,\,T_2)$. The smallest confidence level is then used as the appropriate value α_e in the next step.

Step 4—Prediction intervals generation. Designers predict technology performance and generate prediction intervals in the future time interval $(T_2,\,T_2+\tau)$ using data from start time, T_1, to current time, T_2. The technology evolution prediction model selected in step 2 is estimated using the entire data set. Of note, the parameter estimates derived from the data fitting process are not the same as the parameter estimates from the holdout sample analysis in step 3. Designers obtain the empirical probability distributions for model prediction $\boldsymbol{M_e}(t_e;\,\boldsymbol{\varphi})$ and deviation $\boldsymbol{E_e}(\varepsilon)$ in the future time interval $(T_2,\,T_2+\tau)$ through the method developed in Sec. 3. The probability distribution of each parameter in the model also could be built if necessary. The upper and lower limits of the prediction intervals are generated by Eq. (8) using confidence level α_e derived in step 3. Designers should substitute the latest technology performance (y_n in Sec. 3) for the lower limit(s) of the prediction intervals generated by Eq. (8) if the lower limit(s) is smaller than the latest technology performance because technology performance monotonically increases over time.

5 Case Studies of Central Processing Units and Passenger Airplanes

In this section, we use CPUs and passenger airplanes as two case studies to illustrate the four steps of the process outlined in Sec. 4. In the holdout sample tests of the two case studies, the prediction intervals generated by our method cover every actual data point. These results validate our method and the associated assumptions.

The holdout sample analysis also shows that the 95% prediction intervals are much wider than the $100(1-\alpha_e)$ % prediction intervals generated by our method. In practice, we suggest that designers use $100(1-\alpha_e)$ % prediction intervals for R&D planning and decision making. Designers also could estimate the earliest and the latest time at which the technology of interest would achieve the expected performance level using the $100(1-\alpha_e)$ % prediction intervals. The 95% prediction intervals can be used as a reference to make contingency plans if necessary.

Our method also provides the probability distribution of each parameter in a prediction model. In the passenger airplane case study, the probability distributions of the parameters C_{01} and C_{10} in the Lotka-Volterra ecosystem model indicate the interaction between the system (passenger airplane) and component (turbofan aero-engine) technologies. The probability distributions help designers to make more informed R&D and outsourcing decisions.

5.1 Central Processing Units Transistor Count Evolution Prediction. The prediction of CPU transistor count is important in the semi-conductor industry. The prediction results are critical for high technology companies in R&D planning [4]. To validate our method, we use the CPU transistor count data from 1970 to 2014 and generate prediction intervals during 2014–2018.

There are 118 data points in total as shown in Fig. 3. We choose the top performance data point of each year and then remove the data points that are lower than any previous data point. Only 30 data points (dots on Fig. 3) are selected to represent the CPU performance evolution from 1970 to 2018. As is common in the semi-conductor industry, we take the natural logarithm of the 30 performance values because of fast improvements in CPU performance.

We use Moore's Law as the technology evolution prediction model for CPU performance evolution. Moore's Law is the most influential model used widely in the semiconductor industry to predict the performances of CPU and dynamic random-access memory [4]. The mathematical model is given by

$$y = e^{A+Bt} \tag{10}$$

where y is technology performance, t is time, and A and B are constant parameters. We make a natural logarithm transformation to Eq. (10) and obtain a simple regression model as

$$z = A + Bt \tag{11}$$

where $z=\ln(y)$. Equation (11) is a linear model that is used to fit the natural logarithm transformed data set.

To validate our prediction intervals generation method, we conduct a holdout sample test. In this test, we check whether our prediction intervals cover the actual data points in the following years. Imagine we are in 2014 and want to predict the CPU transistor count evolution in the subsequent four years ($\tau = 4$) using the data from 1970 to 2014 ($T_1 = 0$, $T_2 = 44$).

To determine an appropriate value for confidence level α_e, we perform a holdout sample analysis using the CPU transistor count data from 1970 to 2010 ($T_1 = 0$, $T_2 - \tau = 40$). We fit the 25 data points ($n = 25$) to minimize the sum of squared errors. The fitted model is

$$z_i = 7.183 + 0.3607 t_i + \varepsilon_i \quad (12)$$

where $i \in \{1, 2, \ldots, 25\}$, z_i is the natural logarithm transformed transistor count, t_i is time, where $T_1 \le t_i \le T_2 - \tau$, and ε_i is the deviation at each data point. We create an original sample from the deviation term of Eq. (12) as $\boldsymbol{E_0} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{25})$. Resamples with size $n = 25$ are generated by drawing elements from sample $\boldsymbol{E_0}$ with replacement. Each element in sample $\boldsymbol{E_0}$ has a probability of 1/25 to be drawn. We create 1000 resamples ($R = 1000$) as

$$\boldsymbol{E_j} = \left(\varepsilon_1^j, \varepsilon_2^j, \ldots, \varepsilon_n^j\right) \quad (13)$$

where $j \in \{1, 2, \ldots, 1000\}$. We then derive 1000 bootstrapped CPU transistor count data sets as $\boldsymbol{Z_j} = (z_1^j, z_2^j, \ldots, z_{25}^j)$ where

$$z_i^j = 7.183 + 0.3607 t_i + \varepsilon_i^j \quad (14)$$

We estimate bootstrapped model parameters A_j and B_j for each bootstrapped CPU transistor count data set $\boldsymbol{Z_j}$ as $\boldsymbol{\varphi_j} = (A_j, B_j)$. We build an empirical probability distribution $\boldsymbol{M_e}(t_e; \boldsymbol{\varphi})$ at time t_e ($40 < t_e < 44$) by setting probability 1/1000 at each point $M(t_e; \boldsymbol{\varphi_1})$, $M(t_e; \boldsymbol{\varphi_2})$, …, $M(t_e; \boldsymbol{\varphi_{1000}})$ where

$$M(t_e; \boldsymbol{\varphi_j}) = A_j + B_j t_e \quad (15)$$

The empirical probability distribution $\boldsymbol{M_e}(t_e; \boldsymbol{\varphi})$ describes the parameter uncertainty in CPU transistor count prediction.

To estimate data uncertainty in CPU transistor count prediction, we construct an empirical probability distribution $\boldsymbol{E_e}(\varepsilon)$ from the sample $\boldsymbol{E_0} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{25})$ by setting a probability of 1/25 to each element in the sample $\boldsymbol{E_0}$.

Above all, we derive a $100(1-\alpha)$ % prediction intervals for natural logarithm transformed CPU transistor count z at time t_e ($40 < t_e < 44$) by

$$[M_e^{\alpha/2} + \varepsilon_e^{\alpha/2}, M_e^{1-(\alpha/2)} + \varepsilon_e^{1-(\alpha/2)}] \quad (16)$$

where $M_e^{\alpha/2}$ and $M_e^{1-(\alpha/2)}$ are the $\alpha/2$ and $1-\alpha/2$ percentiles of the $\boldsymbol{M_e}(t_e; \boldsymbol{\varphi})$ distribution, $\varepsilon_e^{\alpha/2}$ and $\varepsilon_e^{1-(\alpha/2)}$ are the $\alpha/2$ and $1-\alpha/2$ percentiles of the $\boldsymbol{E_e}(\varepsilon)$ distribution. We have the actual data points in 2011, 2012, and 2014 ($t_e = 41$, 42, and 44). We could determine the distribution percentile P% that overlap these actual data points at time $t_e = 41$, 42, and 44. The three values of confidence level α associated with each prediction interval can be calculated by $1-|1-P/50|$. A $100(1-\alpha)$ % prediction interval corresponds to each α value. The results of the distribution percentile, the confidence level, and the prediction interval at each actual data point are listed in Table 1.

Table 1 shows that the 51.8% prediction intervals cover the three data points during 2010–2014, resulting in corresponding confidence level of $\alpha_e = 0.482$ which is used to generate prediction intervals in the next step.

As the final step, we fit the 28 data points of CPU transistor count ($n = 28$) from 1970 to 2014 ($T_1 = 0$, $T_2 = 44$) with Eq. (11). The ordinary least-squares estimation produces the fitted model as follows:

Table 1 Distribution percentile, confidence level, and prediction interval at 2011, 2012, and 2014 of CPU transistor count evolution

Year	Distribution percentile (%)	Confidence level α	Prediction interval
2011	36.0	0.720	28.0%
2012	56.0	0.880	12.0%
2014	24.1	0.482	51.8%

$$z_i = 7.242 + 0.3568 t_i + \varepsilon_i \quad (17)$$

Following the same procedure as stated earlier in this section, we generate 51.8% prediction intervals for CPU transistor count from 2014 to 2018. The results are shown in Fig. 1. The 51.8% prediction intervals create a gray area that covers the two actual data points (rhombuses in Fig. 1) at 2015 and 2017. We also show 95% prediction intervals ($\alpha = 0.05$) in Fig. 1. Figure 1 shows that 95% prediction intervals are much wider than the 51.8% prediction intervals on the log coordinates.

5.2 Passenger Airplane Performance Evolution Prediction.

The evolution of the passenger airplane has led to airliners with higher passenger capacity, faster speed, and longer range over time. The aircraft engine (hereafter referred to as the aero-engine) is a major component of the passenger airplane. The interaction between the passenger airplane and the aero-engine should be considered in the passenger airplane performance evolution prediction [6]. Thus, we use the Lotka–Volterra equations [6] to model the technology interaction and predict the passenger airplane performance evolution. Here, the passenger airplane is the system technology. We take passenger capacity·speed·range (km^2/h) as the performance metric of the system technology. The turbofan aero-engine is the component technology. Take-off thrust (kN) is used as the performance metric of the component technology. The point forecasts of the passenger airplane and the turbofan aero-engine performance evolutions are found in our previous publication [6]. We focus on prediction intervals generation in this section. To test our method, we use the performance evolution data of the passenger airplane and the turbofan aero-engine from 1960 to 2004 and generate prediction intervals during 2004–2008.

We collect the performance evolution data of the passenger airplane and the turbofan aero-engine during 1960–2008 [40–42]. We choose the top performance data point of each year and then remove the data points that are lower than any previous data point. We select ten data points that represent the passenger airplane performance evolution and 15 data points that represent the turbofan aero-engine performance evolution from 1960 to 2008. To create dimensionless metrics, we divide these data values by the corresponding characteristic values [6]. The latest performance in a specified time interval is used as the characteristic value for each technology

In this case study, the simplified Lotka–Volterra equations are

$$\frac{dy_0^*}{dt} = a_0 y_0^* - b_0 y_0^{*2} + C_{01} y_0^* y_1^* \quad (18)$$

$$\frac{dy_1^*}{dt} = a_1 y_1^* - b_1 y_1^{*2} + C_{10} y_1^* y_0^* \quad (19)$$

where y_0^* is the dimensionless passenger airplane performance, y_1^* is the dimensionless turbofan aero-engine performance, and a_0, b_0, C_{01}, a_1, b_1, and C_{10} are constant parameters.

Again, we conduct a holdout sample test to validate our prediction intervals generation method. In the holdout sample test, we check whether our prediction intervals cover the actual data point(s) in the following years. We assume that we are in 2004 and want to predict the passenger airplane performance evolution for the subsequent four years ($\tau = 4$) using the data from 1960 to 2004 ($T_1 = 0$, $T_2 = 44$).

To determine an appropriate value for confidence level α_e, we perform a holdout sample analysis using the performance evolution data of the passenger airplane and the turbofan aero-engine from 1960 to 2000 ($T_1 = 0$, $T_2 - \tau = 40$). We fit the 23 data points ($n = 23$) to minimize the sum of squared errors using the trust region reflective algorithm [43]. In the data fitting process, the simplified Lotka–Volterra equations (18) and (19) are solved using high-order Runge–Kutta method [44,45]. The fitted model is

$$\frac{dy_0^*}{dt} = 0.280y_0^* - 0.616y_0^{*2} + 0.481y_0^*y_1^* \tag{20}$$

$$\frac{dy_1^*}{dt} = 0.0509y_1^* - 0.0319y_1^{*2} + 2.33 \cdot 10^{-14}y_1^*y_0^* \tag{21}$$

$$y_0^*(t = 0) = 0.0796 \tag{22}$$

$$y_1^*(t = 0) = 0.241, \tag{23}$$

where $t = 0$ represents the start year of 1960. Here, we treat the technology performances at the start year as unknown parameters in the data fitting process. We derive the deviation term, ε_i, by subtracting the corresponding solution of Eqs. (20)–(23) from the actual data at each data point, where $i \in \{1, 2, \ldots, 23\}$. Here, we do not treat the deviation terms of two technologies (the passenger airplane and the turbofan aero-engine) differently in the following bootstrapping process because every data point is normalized in the same range (0, 1] by the dimensionless treatment. The cluster bootstrapping approach in technology evolution prediction using a system model may be explored in future research. Thus, we create an original sample from the deviation term as $\boldsymbol{E_0} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{23})$. Resamples with size $n = 23$ are generated by drawing elements from the sample $\boldsymbol{E_0}$ with replacement. Each element in the sample $\boldsymbol{E_0}$ has a probability of 1/23 to be drawn. We create 1000 resamples ($R = 1000$) as

$$\boldsymbol{E_j} = \left(\varepsilon_1^j, \varepsilon_2^j, \ldots, \varepsilon_n^j\right) \tag{24}$$

where $j \in \{1, 2, \ldots, 1000\}$. We then derive 1,000 bootstrapped passenger airplane and turbofan aero-engine data sets as $\boldsymbol{Y_j} = (y_1^j, y_2^j, \ldots, y_{23}^j)$ by adding the bootstrapped deviation term ε_i^j to the corresponding solutions of Eqs. (20)–(23) for each data point.

We estimate bootstrapped model parameters $\boldsymbol{\varphi_j} = (y_{t0}^j, a_0^j, b_0^j, C_{01}^j, y_{t1}^j, a_1^j, b_1^j, C_{10}^j)$ for each bootstrapped technology performance data set $\boldsymbol{Y_j}$, where $y_{t0} = y_0^*(t = 0)$ and $y_{t1} = y_1^*(t = 0)$. We build an empirical probability distribution $\mathbf{MP}_e(t_e; \boldsymbol{\varphi})$ at time t_e ($40 < t_e < 44$) by setting a probability of 1/1000 at each point $MP(t_e; \boldsymbol{\varphi_1})$, $MP(t_e; \boldsymbol{\varphi_2})$,…, $MP(t_e; \boldsymbol{\varphi_{1000}})$, where $MP(t_e; \boldsymbol{\varphi_j})$ is the solution of Eqs. (18) and (19) for the passenger airplane dimensionless performance y_0^* at time t_e with eight parameter values given by $\boldsymbol{\varphi_j}$. The empirical probability distribution $\boldsymbol{MP_e}(t_e; \boldsymbol{\varphi})$ describes the parameter uncertainty in the passenger airplane performance prediction. The empirical probability distribution $\boldsymbol{MA_e}(t_e; \boldsymbol{\varphi})$ for the dimensionless turbofan aero-engine performance could be derived in the same manner.

To estimate data uncertainty, we construct an empirical probability distribution $\boldsymbol{E_e}(\varepsilon)$ from the sample $\boldsymbol{E_0} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{23})$ by setting a probability of 1/23 to each element in the sample $\boldsymbol{E_0}$.

Above all, we derive a $100(1 - \alpha)$ % prediction intervals for the dimensionless passenger airplane performance y_0^* at time t_e ($40 < t_e < 44$) by

$$[MP_e^{\alpha/2} + \varepsilon_e^{\alpha/2}, MP_e^{1-(\alpha/2)} + \varepsilon_e^{1-(\alpha/2)}] \tag{25}$$

where $MP_e^{\alpha/2}$ and $MP_e^{1-(\alpha/2)}$ are the $\alpha/2$ and $1 - \alpha/2$ percentiles of the $\boldsymbol{MP_e}(t_e; \boldsymbol{\varphi})$ distribution, $\varepsilon_e^{\alpha/2}$ and $\varepsilon_e^{1-(\alpha/2)}$ are the $\alpha/2$ and $1 - \alpha/2$ percentiles of the $\boldsymbol{E_e}(\varepsilon)$ distribution. Similarly, the $100(1 - \alpha)$ % prediction intervals for the dimensionless turbofan aero-engine performance y_1^* at time t_e ($40 < t_e < 44$) could be derived as

$$[MA_e^{\alpha/2} + \varepsilon_e^{\alpha/2}, \quad MA_e^{1-(\alpha/2)} + \varepsilon_e^{1-(\alpha/2)}] \tag{26}$$

We have one turbofan aero-engine data point at 2002 ($t_e = 42$). The distribution percentile 87.0% overlaps that actual data point at 2002. The 87.0% distribution percentile corresponds to 74.0% prediction interval. The confidence level α associated with the 74.0% prediction interval equals 0.260. Thus, we use the confidence level α_e=0.260 to generate prediction intervals in the next step.

As the final step, we fit the 24 data points of the passenger airplane and the turbofan aero-engine performance evolutions ($n = 24$) from 1960 to 2004 ($T_1 = 0$, $T_2 = 44$) with Eqs. (18) and (19). The ordinary least-squares estimation produces the fitted model as follows:

$$\frac{dy_0^*}{dt} = 0.267y_0^* - 0.612y_0^{*2} + 0.580y_0^*y_1^* \tag{27}$$

$$\frac{dy_1^*}{dt} = 0.0346y_1^* - 1.71 \cdot 10^{-8}y_1^{*2} + 4.32 \cdot 10^{-10}y_1^*y_0^* \tag{28}$$

$$y_0^*(t = 0) = 0.0797 \tag{29}$$

$$y_1^*(t = 0) = 0.224 \tag{30}$$

Following the procedure described earlier in this section, we generate 74.0% prediction intervals ($\alpha_e = 0.260$) for the passenger airplane performance evolution from 2004 to 2008. The results are shown in Fig. 2. The 74.0% prediction intervals create a gray area that covers the actual data point (rhombuses on Fig. 2) in 2005. We also show the 95% prediction intervals ($\alpha = 0.05$) in Fig. 2. Again, Fig. 2 shows that the 95% prediction intervals are much wider than the 74.0% prediction intervals generated by our method.

Of note, the lower limits of 74.0% prediction intervals are smaller than the latest passenger airplane performance (the passenger airplane performance at 1988) during 1960–2004. In this case, designers should substitute the latest system technology performance for the lower limits of the prediction intervals generated by our method because technology performance monotonically increases over time. Although some technology performances in future periods may not exceed the top technology performance in the past, designers usually ignore these performances and focus on the better performance that a technology may achieve in a future period.

The Lotka–Volterra ecosystem model allows designers to predict the performance of system and component technologies with improved accuracy. The ecosystem model also quantifies the interaction between the technologies [6]. For example, the parameter C_{01} in Eq. (18) captures the impact of the turbofan aero-engine on the passenger airplane performance evolution; the parameter C_{10} in Eq. (19) reflects the impact of the passenger airplane on the turbofan aero-engine performance evolution. Importantly, the method presented in this paper provides the probability distribution of each parameter in a prediction model. For system models, designers could evaluate the interaction between the technologies from the probability distribution and make more informed R&D and outsourcing decisions. For example, we can construct empirical probability distributions for the eight parameters in Eqs. (20)–(23) from $\boldsymbol{\varphi_j}$. The histograms of C_{01}/a_0 and C_{10}/a_1 are shown on Figs. 4 and 5. The results indicate that the development of turbofan aero-engine has considerable impact on the passenger airplane evolution because the value of C_{01}/a_0 has a median of 1.44 and a probability of 80.8% that $C_{01}/a_0 > 0.1$. Meanwhile, the advancement of passenger airplane performance

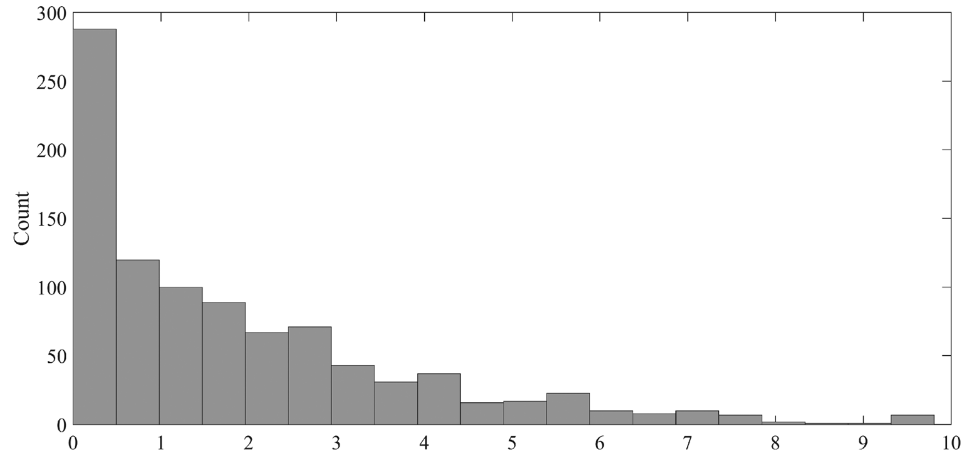

Fig. 4 Histogram of C_{01}/a_0 distribution in Eq. (20)

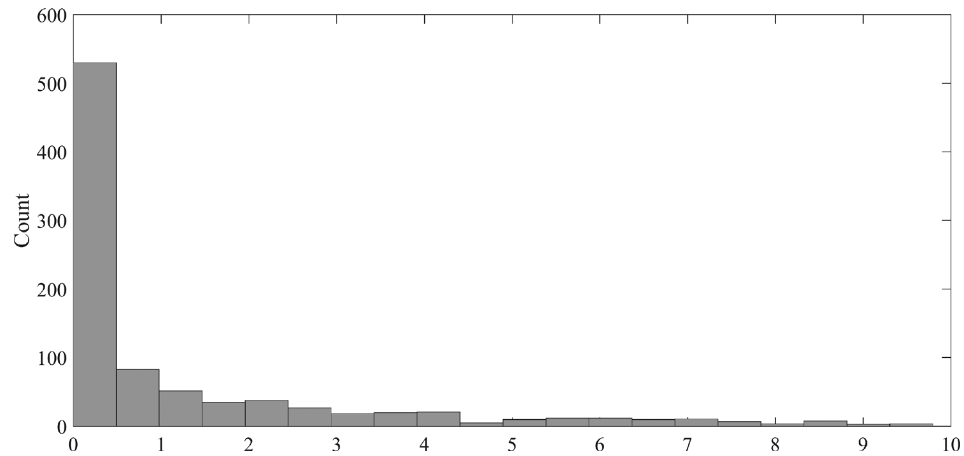

Fig. 5 Histogram of C_{10}/a_1 distribution in Eq. (21)

has limited impact on the turbofan aero-engine evolution. The value of C_{10}/a_1 has a median of 0.337 and a probability of 54.7% that $C_{01}/a_0 > 0.1$. Thus, in order to improve the passenger airplane performance, the first priority is to invest in the R&D of the turbofan aero-engine. However, the opposite is not true. Investment in the R&D of the passenger airplane is unlikely to be an effective strategy to improve the performance of the turbofan aero-engine.

In this passenger airplane case study, we model only the interaction between one system technology and one component technology. In the case of one system technology interacting with multiple component technologies, designers could use a system model associated with our method to identify the key component technologies that have significant impact on the system technology evolution (e.g., through the probability distributions of parameters C in the Lotka–Volterra ecosystem model [6]). Designers would consider developing and manufacturing these key component technologies in-house, or establish a partner relationship (e.g., through joint R&D, cross-shareholding, or exclusive supply) with the suppliers of these component technologies. Meanwhile, other component technologies that have limited impact on the system technology evolution may be considered for outsourcing [6].

6 Conclusions and Future Work

We introduced a general method that uses bootstrapping to generate prediction intervals for technology evolution prediction. The novelty of our work is in the application of bootstrapping to estimate parameter and data uncertainty for technology evolution prediction and in determining confidence level α from a holdout sample analysis. The method can be applied to any technology evolution prediction model based on mathematical functions or differential equations involving time that predicts the incremental change in technology performance. We consider parameter uncertainty and data uncertainty and establish their empirical probability distributions. We determine the appropriate confidence level α required to generate prediction intervals using a holdout sample analysis rather than setting $\alpha = 0.05$ as is frequently done in previous research. In addition, our method provides the probability distribution of each parameter in a prediction model. We outline four steps for designers to generate prediction intervals in technology evolution prediction in practice. We use CPUs and passenger airplanes prediction intervals generation as two case studies to illustrate these steps and validate our method.

The use of bootstrapping and holdout sample analysis can provide practically useful prediction intervals for the trend that technology evolution is expected to follow. These prediction intervals can be used by designers and decision makers to estimate earliest-latest technology availability, set reasonable R&D targets, and make contingency plans. The probability distribution of each parameter is valuable for designers when the parameter value is associated with the impact factors of technology evolution (e.g., performance upper limit in the logistic S-curve model or technology interaction in the Lotka–Volterra ecosystem model). In system models, the probability distribution of a corresponding parameter indicates the interaction between the system and the component technologies. Designers could evaluate the interaction between the technologies from the probability distribution and make more informed R&D and outsourcing decisions.

This paper has some limitations that offer opportunities for future research. We use CPUs and passenger airplanes as two case studies to validate our method. The method could be applied to more case studies from other industries. Future case studies could also explore time periods when the prediction intervals become sufficiently wide and lead to a discontinuity in technology evolution that signifies potential technology substitution. The proposed method, in particular, the strategy of using a holdout sample analysis to determine an appropriate confidence level α, is also applicable to diverse domains (e.g., econometrics, demography, and marketing) beyond technology evolution prediction. Moreover, our method does not consider model uncertainty in prediction intervals generation. Researchers could consider developing mathematical procedures for selecting optimal prediction models or estimating the model uncertainty in technology evolution prediction through a Bayesian approach. In addition, there are alternative methods to estimate parameter uncertainty and data uncertainty in technology evolution prediction. For example, researchers could use sensitivity analysis or Bayesian approach to estimate parameter uncertainty in the future. If there are more than 30 technology evolution data points or if researchers have practical reasons to make a parametric assumption for the random error term, Monte Carlo simulation could be used to estimate data uncertainty in technology evolution prediction. The results of these methods could be compared with those of our approach in the future. Finally, researchers may use the proposed method to explore the interaction between technology evolution and marketing models in the future. The proposed method provides the probability distribution of each parameter in a technology evolution prediction model. The probability distribution of a parameter could be associated with a marketing model, the empirical formula for which may be developed by future research. For example, the value of parameter a_0 in Eq. (18) is related to R&D expenditures on system technology. Greater R&D investments on system design and optimization increase the value of a_0 in Eq. (18). Practitioners could establish an empirical formula between the amount of R&D investment and the value of parameter a_0 in Eq. (18) to predict the impact of their investment on system technology performance.

Funding Data

- This material is based in part upon work supported by the National Science Foundation under Award Number CMMI-1550002. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Otto, K. N., and Wood, K. L., 2001, *Product Design: Techniques in Reverse Engineering and New Product Development*, Prentice Hall,Upper Saddle River, NJ.

[2] Betz, F., 1993, *Strategic Technology Management*, McGraw-Hill, New York.

[3] Chatfield, C., 1993, "Calculating Interval Forecasts," J. Bus. Econ. Stat., **11**(2), pp. 121–135.

[4] Schaller, R. R., 1997, "Moore's Law: Past, Present and Future," IEEE Spectrum, **34**(6), pp. 52–59.

[5] Modis, T., 2014, *An S-Shaped Adventure: Predictions—20 Years Later*, Growth Dynamics, Lugano, Switzerland.

[6] Zhang, G., McAdams, D. A., Shankar, V., and Mohammadi Darani, M., 2018, "Technology Evolution Prediction Using Lotka–Volterra Equations," ASME J. Mech. Des., **140**(6), p. 061101.

[7] Efron, B., 1979, "Bootstrap Methods: Another Look at the Jackknife," Ann. Stat., **7**(1), pp. 1–26.

[8] Mooney, C., and Duval, R., 1993, *Bootstrapping: A Nonparametric Approach to Statistical Inference*, SAGE Publications, Newbury Park, CA.

[9] Efron, B., and Tibshirani, R., 1986, "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," Stat. Sci., **1**(1), pp. 54–75.

[10] Chatfield, C., 1996, "Model Uncertainty and Forecast Accuracy," J. Forecasting, **15**(7), pp. 495–508.

[11] Chatfield, C., 1988, "What is the 'Best' Method of Forecasting?," J. Appl. Stat., **15**(1), pp. 19–38.

[12] Hyndman, R. J., and Athanasopoulos, G., 2014, *Forecasting: Principles and Practice*, OTexts, Columbia, SC.

[13] Montgomery, D. C., Peck, E. A., and Vining, G. G., 2006, *Introduction to Linear Regression Analysis*, Wiley-Interscience, Hoboken, NJ.

[14] Shumway, R. H., and Stoffer, D. S., 2011, *Time Series Analysis and Its Applications*, Springer, New York.

[15] Kiureghian, A. D., and Ditlevsen, O., 2009, "Aleatory or Epistemic? Does It Matter?," Struct. Saf., **31**(2), pp. 105–112.

[16] Chatfield, C., 2004, *The Analysis of Time Series*, Chapman & Hall/CRC, Boca Raton, FL.

[17] de Gooijer, J. G., Abraham, B., Gould, A., and Robinson, L., 1985, "Methods for Determining the Order of an Autoregressive-Moving Average Process: A Survey," Int. Stat. Rev./Revue Internationale de Statistique, **53**(3), pp. 301–329.

[18] Harvey, A. C., 1990, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, UK.

[19] Saltelli, A., Chan, K., and Scott, E. M., 2009, *Sensitivity Analysis*, Wiley, New York.

[20] Congdon, P., 2014, *Applied Bayesian Modelling*, Wiley, Chichester, UK.

[21] Williams, W. H., and Goodman, M. L., 1971, "A Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts," J. Am. Stat. Assoc., **66**(336), pp. 752–754.

[22] Makridakis, S., Hibon, M., Lusk, E., and Belhadjali, M., 1987, "Confidence Intervals: An Empirical Investigation of the Series in the M-Competition," Int. J. Forecasting, **3**(3–4), pp. 489–508.

[23] Thombs, L. A., and Schucany, W. R., 1990, "Bootstrap Prediction Intervals for Autoregression," J. Am. Stat. Assoc., **85**(410), pp. 486–492.

[24] Ghysels, E., and Jasiak, J., 1997, *GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model*, CIRANO, Montreal, QC, Canada.

[25] Anderson, P., and Tushman, M. L., 1990, "Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change," Adm. Sci. Q., (4), pp. 604–633.

[26] Sahal, D., 1985, "Technological Guideposts and Innovation Avenues," Res. Policy, **14**(2), pp. 61–82.

[27] Christensen, C., 2013, *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Harvard Business Review Press, Cambridge, MA.

[28] Ghasemi, A., and Zahediasl, S., 2012, "Normality Tests for Statistical Analysis: A Guide for Non-Statisticians," Int. J. Endocrinol. Metabolism, **10**(2), pp. 486–489.

[29] Meade, N., and Islam, T., 1998, "Technological Forecasting—Model Selection, Model Stability, and Combining Models," Manage. Sci., **44**(8), pp. 1115–1130.

[30] Farmer, J. D., and Lafond, F., 2016, "How Predictable is Technological Progress?," Res. Policy, **45**(3), pp. 647–665.

[31] Arendt, J. L., McAdams, D. A., and Malak, R. J., 2012, "Uncertain Technology Evolution and Decision Making in Design," ASME J. Mech. Des., **134**(10), p. 100904.

[32] Naim, A. M., and Lewis, K., 2017, "Modeling the Dynamics of Innovation in Engineered Systems," ASME Paper No. DETC2017-68180.

[33] Nagy, B., Farmer, J. D., Bui, Q. M., and Trancik, J. E., 2013, "Statistical Basis for Predicting Technological Progress," PLoS ONE, **8**(2), p. e52669.

[34] Young, P., 1993, "Technological Growth Curves: A Competition of Forecasting Models," Technol. Forecasting Soc. Change, **44**(4), pp. 375–389.

[35] Draper, D., 1995, "Assessment and Propagation of Model Uncertainty," J. R. Stat. Soc. Ser. B (Methodol.), **57**(1), pp. 45–97.

[36] Freedman, D. A., 1981, "Bootstrapping Regression Models," Ann. Stat., **9**(6), pp. 1218–1228.

[37] Meade, N., and Islam, T., 1995, "Prediction Intervals for Growth Curve Forecasts," J. Forecasting, **14**(5), pp. 413–430.

[38] Rohatgi, V. K., 1984, *Statistical Inference*, Wiley, New York.

[39] Zhang, G., McAdams, D. A., Shankar, V., and Darani, M. M., 2017, "Modeling the Evolution of System Technology Performance When Component and System Technology Performances Interact: Commensalism and Amensalism," Technol. Forecasting Soc. Change, **125**, pp. 116–124.

[40] Iatrou, K., 2014, *100 Years of Commercial Aviation*, HERMES Air Transport Club, Montreal, QC, Canada.

[41] Daly, M., and Gunston, B., 2008, *Jane's Aero-Engines*, Jane's Information Group Limited, Surry, UK.

[42] Wilkinson, P. H., 1970, *Aircraft Engines of the World 1970*, Paul H. Wilkinson, Washington, DC.

[43] Moré, J., and Sorensen, D., 1983, "Computing a Trust Region Step," SIAM J. Sci. Stat. Comput., **4**(3), pp. 553–572.

[44] Dormand, J. R., and Prince, P. J., 1980, "A Family of Embedded Runge-Kutta Formulae," J. Comput. Appl. Math., **6**(1), pp. 19–26.

[45] Lapidus, L., and Seinfeld, J. H., 1971, *Numerical Solution of Ordinary Differential Equations*, Academic Press, New York.