Aeroengine Prognostics via Local Linear Smoothing, Filtering and Prediction

Kartik B. Ariyur and Jan Jelinek
Honeywell Labs, Minneapolis
3660 Technology Dr.
Minneapolis, MN 55418

ABSTRACT

We propose a new method for local linear smoothing, filtering and prediction of noisy data. Its novelty consists in two of its steps: a sliding window filter that uses Student's t-statistics to perform smoothing and filtering, and a trend change detection scheme that uses a convex hull construction to determine a change of slope or intercept of the local linear trend. The final linear trend detected is used for linear prediction and interval estimation. The application of the scheme to gas-turbine engine prognostics is presented.

1. MOTIVATION

Constraints to comprehensive collection of measurements in industrial systems, coupled with the need to predict, diagnose and avert failures of systems for purposes of safety, preventive maintenance, and inventory reduction predicate the need for algorithms that can make meaningful predictions from sparse and non-uniform time series. Ad hoc methods such as fuzzy logic [1] and neural networks [4], while permitting rapid deployment to products, have to be tuned carefully to the specific problem at hand. Moreover, they do not provide a quantitative estimate of the reliability of their extrapolations or predictions.

Online filtering and prediction algorithms such as simple low pass filters and Kalman filters [7], on the other hand are designed to minimize memory and computational power needed for online operation. Hence, they do not exploit all the information available in batch data to make their predictions. Moreover, the non-availability (either because of intrinsic difficulty or high cost of modeling) of fault growth models or the absence of adequate measurements in many problems makes application of model-based system identification techniques [2, 6] difficult.

To ameliorate these deficiencies, we propose algorithms based on linear statistical inference that achieve the followings aims.

1. Perform smoothing and filtering of data.
2. Determine and eliminate statistical outliers before using data for prediction.
3. Detect trend changes in data.
4. Provide a measure of reliability of the prediction.

Smoothing and filtering are performed by the sliding window filter – this involves calculating confidence intervals of mean prediction from overlapping windows in data, and choosing the smoothed estimate as the linear estimate that results in the tightest interval estimate (among the estimates computed with the windowing scheme). Changes of linear trends (slopes or intercepts) are reported when the convex hulls of upper and lower confidence bounds of mean prediction intersect. The last linear trend thus detected in data is used for linear prediction and interval estimation.

We also present some illustrative examples of application of these algorithms to the practical problem that motivated their development: minimizing false alarms in fault prognostics in predictive trend monitoring of gas turbine engines used in aircraft (main engines and auxiliary power units (APUs)). In the examples presented herein, we apply our algorithms to model-derived prediction residuals (which give a measure of parameter deviation from ideality) and to efficiency parameters.

The paper is organized as follows: Section 2 recapitulates basic linear regression and interval estimation; Section 3 presents the sliding window filter that performs smoothing and filtering; Section 4 presents trend change detection; Section 5 presents extrapolation of trends into the future; Section 6 presents illustrative examples; and Section 7 presents examples of engine parameter prognostics for aircraft APUs.

2. REVIEW OF LINEAR STATISTICAL INFERENCE

Consider the simple linear regression model with a single regressor [3]:

\[ y = \beta_0 + \beta_1 x + \varepsilon, \]  
(2.1)
where the intercept $\beta_0$, and the slope $\beta_1$, are unknown constants and $\varepsilon$ is a random error component, with zero mean and variance $\sigma^2$. If we have a set of sampled data satisfying

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n, \quad (2.2)$$

we can estimate the parameters $\beta_0$ and $\beta_1$ through the method of least squares. Least squares estimation gives the parameter estimates as

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad (2.3)$$

where

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \quad \text{and} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$
$$\quad (2.5)$$

are the averages of $y_i$ and $x_i$ respectively, and

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{and} \quad S_{xy} = \sum_{i=1}^{n} y_i (x_i - \bar{x})$$
$$\quad (2.6)$$

In terms of the parameter estimates above, the estimate of the mean response $E(y|x_0)$ at a point $x_0$ is given by

$$E(y|x_0) = \mu_{y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0 \quad (2.7)$$

Under the assumption of normal independently distributed error $\varepsilon_i$, the $100(1-\alpha)$ percent confidence interval on the mean response at $x_0$ is

$$\hat{\mu}_{y|x_0} - t_{\alpha/2, n-2} \sqrt{\frac{MSE}{n} \frac{1}{S_{xx}}}$$
$$\leq y_f \leq \hat{\mu}_{y|x_0} + t_{\alpha/2, n-2} \sqrt{\frac{MSE}{n} \frac{1}{S_{xx}}} \quad (2.8)$$

where $t_{\alpha/2, n-2}$ is inverse function of the cumulative distribution function (CDF) of Student’s $t$-distribution (Figure 1 graphs this function; as the degrees of freedom increase, it approaches the inverse of the Gaussian CDF.) with $n-2$ degrees of freedom at $\alpha/2$, and $MSE$ is an estimate of the noise variance and equal to the mean square error in the residuals:

$$MSE = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
$$\quad (2.9)$$

where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. The $100(1-\alpha)$ percent confidence interval for a future observation $y_f$ at $x_f$ is given by

$$\hat{y}_f - t_{\alpha/2, n-2} \sqrt{\frac{MSE}{n} \frac{1}{S_{xx}}} \leq y_f \leq \hat{y}_f + t_{\alpha/2, n-2} \sqrt{\frac{MSE}{n} \frac{1}{S_{xx}}} \quad (2.10)$$

where $\hat{y}_f = \hat{\beta}_0 + \hat{\beta}_1 x_f$.

### 3. SLIDING WINDOW FILTER

We present now a nonlinear filter that uses local linear statistical inference to perform smoothing of data. An
The underlying assumption is that the rate of sampling is fast enough that there are several data points within a small variation of slope and intercept of the trend. We name it the sliding window filter as it involves sliding a regression window through data to produce the smoothing. This permits usage of the following assumption for locally fitting straight lines to data:

**ASSUMPTION 3.1 – The observed data points y, follow a linear trend as given by Equation (2.2) with unknown variance \( \sigma^2 \).**

Before we go into the details of the filtering, we note that the mean square error \( \text{MSE} \) of the fit consists in both statistical error due to noise in the observations and model error due to change of slope or intercept in the linear model. This is because we are performing local linear fits to data that is not intrinsically linear (trend in data is nonlinear and possibly even discontinuous). To motivate the filtering method presented below, and the guidelines for selection of its parameters, we rewrite the model Equation (2.1) to reflect the varying slope and intercept of tangents to a general piecewise continuous curve at sample points and derive the \( \text{MSE} \) as a function of model error and statistical error:

\[
\text{MSE} = \frac{1}{n-2} \left( \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \right) = \frac{1}{n-2} \left( \sum_{i=1}^{n} (\hat{\beta}_{0,i} + \hat{\beta}_{1,i} x_i + \epsilon_i)^2 \right)
\]

In view of Equation (3.3), we would like a filter to minimize both the model error locally and get the \( \text{MSE} \) near its minimum value - variance in the data. The filter we propose attempts this in the following steps:

1. **Windowing** – This involves the partitioning of the time-series data into several overlapping windows, where a window (windows are indexed by \( k \)) denotes a contiguous sequence of ordered pairs in the data (chronological, for example). The next step of linear regression is performed over these windows. Window size should be chosen to trade off two conflicting goals – the window length \( W_L \) should be long enough to average out statistical noise, and small enough that the model parameters do not change appreciably within it. The purpose is to keep the \( \text{MSE} \) small. The window step \( \delta L \), the distance between the beginning of one window and that of its successor, is decided by computational constraints (required speed of execution on a given machine). Figure 2 depicts three possible windowing schemes for a data set: the top plot shows the simplest scheme with uniform non-overlapping time windows, the second shows a uniform windowing scheme with overlapping windows, and the third shows a scheme with non-uniform and overlapping windows. More complex windowing can be performed, and windowing can be made adaptive, e.g., larger windows in regions of little trend change and smaller windows in regions of greater trend change. Finally, it should be noted that a window should on average contain a valid statistical sample – at least three points. The minimum number of points a window must contain for regression is set as a parameter \( N_{w,\text{min}} \) in the algorithm. Typically, all values of \( N_{w,\text{min}} = 5 \) or greater produce about the same results so long as the sampling is fairly uniform.

2. **Regression** – The second step is to fit lines to the data points within the windows: estimate the slope as \( \hat{\beta}_{1,k} \) and intercept as \( \hat{\beta}_{0,k} \) as in Equations (2.4) and (2.3) from window \( k \), estimate intervals of mean prediction and future prediction as in Equations (2.8) and (2.10) respectively. Because the windows overlap, this step results in multiple interval estimates for every point. The confidence level \( \alpha \) for the interval estimates is chosen to trade off between precision and reliability of the interval estimates: small \( \alpha \) will result in a wide interval in Equations...
(2.8) and (2.10) with $1 - \alpha$ (high) probability of the mean or predicted response lying between the limits; conversely, larger $\alpha$ will result in tighter confidence intervals and lower probability of the prediction lying between the bounds.

3. **Outlier Elimination** — Where needed, statistical outliers can be eliminated from the data used to calculate the regression parameters and interval estimates. Figure 3 is a schematic of the operation of outlier elimination, which uses the specific windowing scheme shown. The left sample window, test window, and right sample window slide through data. The size of the test window for outlier elimination $L_t$, is chosen to ensure at least one sample inside of every test window. This can be made larger for faster execution. At every step, a least squares line is fit to data in the left sample window and another least squares line fit to data in the right sample window. The predictions of these regression lines at a specified level of confidence at a data location in the test window are then compared with the data in the test window. If the data does not fall in the confidence cone of prediction of either the left or the right sampling window, it is marked as an outlier. For greater accuracy, left and right prediction are performed by several overlapping left and right sample windows, and outlier marking is done through voting by the windows. The voting can be either weighted by the reciprocal of the interval estimates produced by the window, or it can be unweighted (simple majority).

4. **Smoothing** — The interval estimates of mean prediction at each point $(x_i, y_i)$ from different regression estimates (available from overlapping windows that perform the regression several times for the same point) are compared, and the minimum width interval is determined:

$$i_{l,k}^{(t)} = \min_k \left\{ 2t_{\alpha/2,n_k-2} \sqrt{\text{MSE}_k \left( \frac{1}{n_k} + \frac{(x_i - \bar{x}_k)^2}{S_{xx,k}} \right)} \right\},$$

where $i_{l,k}^{(t)}$ is the minimum width interval corresponding to $k = k^*$, $n_k$ is the number of data points inside window $k$, and the subscript $k$ refers to calculations made from linear regression in window $k$. The smoothed estimate of the trend is taken as the estimate

$$\hat{y}_{l,k^*} = \beta_{l,k^*} x_i + \beta_{0,k^*},$$

corresponding to $i_{l,k}^{(t)}$.

Notes — The method outlined above is suboptimal, in that we do not attempt to arrive at the minimum possible width of the confidence interval widths (or equivalently, minimum MSE). This is because we do not perform the minimization over all possible (statistically meaningful) windows in data, but only over a subset of them as permitted by our capability to compute. The prime advantage of this suboptimal method, over a Kalman filter is the ability to deal with unknown covariances of measurement noise and process noise. Moreover, in the absence of a model of the underlying physical process, we are able to avoid the imposition of an arbitrary model structure upon the data to perform smoothing and prediction (as would be needed to implement a Kalman filter). Besides, there is also the ability to naturally account for lapses in data collection and non-uniform sampling. These simply raise the widths of the confidence intervals of our predictions.

4. **TREND CHANGE DETECTION**

We propose a novel method of determining a trend change in the data, i.e., determining when the slope and intercept of a line change from their previous values. The idea behind the method is fundamental: *that the convex hulls of upper and lower confidence bounds of mean prediction intersect when the deviation of the trend from linearity is large enough to be perceptible above the noise extant in data*. Thus, we are using the fact that the intersection of the convex hulls of upper and lower bounds contradicts linearity of the trend, to detect deviations from linearity. Figure 4 shows three graphical examples of convex hulls of mean prediction confidence bounds (as calculated by Equation (2.8)) for data with additive noise: the top plot for a continuous straight line;

![Linear trend with noise](image1)

![Piecewise linear trend with noise](image2)

![Exponential trend with noise](image3)

Figure 4. Convex Hulls of Confidence Bounds

1 A subset $C$ of $\mathbb{R}^n$ is said to be convex if $(1 - \lambda)x + \lambda y \in C$ whenever $x \in C$, $y \in C$, and $0 < \lambda < 1$. [5]
the middle plot for a piecewise continuous line, with a jump in its intercept at \( x = 250 \); the bottom plot depicts an exponential curve. The equations for the different plots are respectively (top-middle-bottom):

\[
y = 0.01x + 5\varepsilon \\
y = \begin{cases} 
5 + 0.01x + 3\varepsilon & \text{if } x \in [5, 245] \\
10 + 0.01x + 3\varepsilon & \text{if } x \in [250, 500] 
\end{cases} \\
y = e^{0.005x} + 5\varepsilon ,
\]

where \( \varepsilon \) is a normal random variable with mean zero and standard deviation unity. In the top plot, the upper and lower convex hulls (upper and lower bows) do not intersect, because, by definition, for a line (Equation (4.1)), the upper confidence limits are above the line and the lower confidence limits are below it. However, in the middle plot, since there are two lines (Equation (4.2)), the second above the other, the convex hulls (upper and lower parallelograms) intersect. Similarly, the convex hulls of the confidence bounds of the exponential (Equation (4.3)) intersect as the slope of the curve has changed significantly within the sample considered. A point to note is that if the noise level was higher, or if we set the confidence levels higher, the convex hulls may not intersect.

All of the plots in Figure 4 use the actual value of the underlying function (\( y \) – noise in Equations, (4.1) – (4.3)) rather than its estimate \( \hat{y} \) to calculate the confidence intervals. If there is sufficient sampling and the window step for the sliding window filter is sufficiently small (leading to correspondingly larger overlap between windows), the estimates \( \hat{y} \) will approach their ideal values. Thus, in the ideal case of identifying the perfect smoothed curve, the intersection of upper and lower convex hulls will detect deviations from linearity of the unknown curve. This leads to two mathematical questions:

1. Given that the convex hulls of upper and lower bounds at confidence level \( 100(1 - \alpha)\% \) intersect, what is the probability of a change of slope or intercept of the local line?

2. Conversely, given that the trend has deviated from linearity (a certain change of slope or intercept), what is the probability that the convex hulls of upper and lower confidence bounds at confidence level \( 100(1 - \alpha)\% \) intersect?

We attempt to give a partial answer to the questions by clarifying them. The intersection of the convex hulls implies one of the following two statements:

1. A point on the line segment joining two of the upper confidence bounds lies below one of the lower confidence bounds. Mathematically,

\[
\lambda(\hat{y}_i + \hat{e}_i) + (1 - \lambda)(\hat{y}_j + \hat{e}_j) \leq \hat{y}_k + \hat{e}_k, \quad (4.4)
\]

for some \( i, j, k \), where the positive quantities

\[
\hat{e}_i = t_{\alpha/2, n_i - 2} \sqrt{\frac{1}{n_i} (\frac{x_i - \bar{x}_i)^2}{S_{xx,i}}}, \quad (4.5)
\]

are the minimum confidence interval semi-widths estimated at \( x_i \) using the corresponding window lengths \( n_i \) on which they are based (these give the \( S_{xx,i} \) and \( \bar{x}_i \)). Moving the positive semi-width terms to one side of the inequality, and using the identity \((1 - \lambda) + \lambda = 1\), we rewrite Equation (4.4) as

\[
\lambda(\hat{y}_k - \hat{y}_i) + (1 - \lambda)(\hat{y}_k - \hat{y}_j) \leq \lambda(\hat{e}_k + \hat{e}_i) + (1 - \lambda)(\hat{e}_k + \hat{e}_j) \quad (4.6)
\]

2. A point on the line segment joining two of the lower confidence bounds lies above one of the upper confidence bounds. Mathematically,

\[
\lambda(\hat{y}_i - \hat{e}_i) + (1 - \lambda)(\hat{y}_j - \hat{e}_j) \geq \hat{y}_k + \hat{e}_k, \quad (4.7)
\]

for some \( i, j, k \). Moving the positive semi-width terms to one side of the inequality as before, and using the identity \((1 - \lambda) + \lambda = 1\), we rewrite Equation (4.7) as

\[
\lambda(\hat{y}_i - \hat{y}_k) + (1 - \lambda)(\hat{y}_j - \hat{y}_k) \geq \lambda(\hat{e}_k + \hat{e}_i) + (1 - \lambda)(\hat{e}_k + \hat{e}_j) \quad (4.8)
\]

Hence, we can write that the condition for convex hull intersection as

\[
\lambda(\hat{y}_i - \hat{y}_k) + (1 - \lambda)(\hat{y}_j - \hat{y}_k) \geq \lambda(\hat{e}_k + \hat{e}_i) + (1 - \lambda)(\hat{e}_k + \hat{e}_j) \quad (4.9)
\]

We see that qualitatively, this needs either a large/small change in slope (and/or intercept) over a small time/large period. We note two issues here: first, that usage of higher confidence levels results in small trend deviations being undetected; second, that changes of slope and intercept can compensate for each other and make detection less likely.

5. PROGNOSTICS

In this section, we supply a method for extrapolation from time series data building upon the results of the sliding window filter and trend change detection. The idea is straightforward; find the largest segment of data from the end into the past for which the convex hulls of upper and lower confidence bounds do not intersect; fit a straight
line to this last segment of data, and calculate future predictions along with interval estimates. This can be achieved in several ways, and we present the outline of one such algorithm (implemented on the APU 331-350):

**ALGORITHM 5.1 (PROGNOSTIC USING CONVEX HULLS OF CONFIDENCE BOUNDS)**

1. Calculate convex hulls of the upper and lower confidence bounds of mean prediction.
2. If they intersect, repeat the previous step using only the confidence interval bounds after the first intersection.
3. If they do not intersect, fit a least squares line through all data points from the end for which the convex hulls do not intersect, and calculate needed future predictions along with confidence interval estimates (of mean prediction or future prediction as may be needed).

**Discussion** – The overall algorithm for arriving at the predictions is an application of linear statistical inference to batch data. It attempts to maximize the precision in identifying trend borders in data by filtering out outliers in successive iterations and using convex hulls of mean prediction confidence bounds. The algorithm has two kinds of properties: those that arise purely as a result of the statistical nature of data, and those that arise from assuming linear underlying trends. Under the first category, we have the following: we cannot detect trend changes smaller than the variance when our sample sizes in the sample windows are small; we cannot precisely distinguish the points where trends change, but only in some region around them. Under the second category, slow continuous trend changes that appear cannot be distinguished as they are locally linear and change cumulatively; we also may end up breaking up a continuous trend into several linear pieces and thereby reduce the extent to which we can extrapolate from the data. However, the method supplies clear information as to when greater frequency of data collection is needed (infrequent collection of data leads to wide interval estimates for future prediction), and as to when we cannot reasonably extrapolate from data available up to the current time. This happens when a trend change is detected very near the end of available data, and there are very few points from which to extrapolate in the future.

**6. ILLUSTRATIVE EXAMPLES**

To enable clear understanding of the working of the prognostic tool, and the conditions under which it produces valid results, we provide illustrative examples in this section. We generate time series data with \( t \) ranging from 5 to 500 in steps of 5 using the same functions used to illustrate convex hull intersections:

\[
y = 0.01x + 5\varepsilon \quad \text{(6.1)}
\]

\[
y = \begin{cases} 
5 + 0.01x + 2\varepsilon & \text{if } x \in [5, 245] \\
10 + 0.01x + 2\varepsilon & \text{if } x \in [250, 500] 
\end{cases} \quad \text{(6.2)}
\]

\[
y = e^{0.005x} + 3\varepsilon \quad \text{(6.3)}
\]

The smoothing/prediction (with interval estimation for mean prediction) results shown in Figure 5. The smoothed curves are continuous; the data points are in dots with line segments joining them. The confidence intervals of mean prediction into the future are shown as light lines extending beyond an abscissa of 500. These examples use the sliding window filter without outlier elimination. The parameters used in the algorithm are (window length, window step, minimum number of samples in a window, and confidence level respectively):

\[
L_w = 40, \delta L_w = 10, N_w, \min = 3, \alpha = 0.01
\]

We can make the following inferences from the plot in Figure 5:

1. The sliding window algorithm can identify different smoothed trends within the data depending upon the parameters values used, the variance in data, and the size of the sample.
2. Trends can get distorted by the presence of outliers. This is seen from the linear trend in the top plot being detected as a more complicated curve.
3. The location of trend change at \( x = 250 \) in the middle plot is not detected exactly at \( x=250 \), but close by. This is seen in the smoothed data.
4. The prediction for nonlinear curves usually results in wider confidence bounds – this is seen in the prognostic from the exponential curve in the bottom plot. This is because the change in trend necessitates use of only a small part of data for prediction – the part for which the convex hulls of upper and lower confidence bounds do not intersect.

7. ENGINE PROGOSTICS

In this section, we present some examples of application of the algorithms on integrated APU residual from APU331-350 data. The plots show raw data points and asterixes for outliers (which are not considered in generating final predictions). Upper and lower confidence interval bounds are plotted as squares, final prognostics upper and lower bounds are given by six-pointed stars joined by lines.

The algorithm parameters used in the examples were

\[ \delta L_w = 50\text{hrs}, \quad L_w = 200\text{hrs}, \quad L_t = 50\text{hrs}, \]
\[ N_{w,\text{min}} = 5, \quad \alpha = 0.05 \]

where \( L_t \) is the length of the test window in outlier elimination (see Figure 3). The predictions are either through calculation of mean prediction confidence intervals from the last trend in data (if the last detected trend contains more than \( 2N_{min} = 10 \) samples), or through linear regression on the last 500 hours of data (the convex confidence interval bounds methods is not used due to a lack of valid data points).

The plots in Figures 6 and 7 are described below:

1. **Figure 6**: This figure shows the deterioration of the power section. The prognostic is made possible by the detection of the last trend in spite of the sparse data collection during recent engine.

2. **Figure 7**: In this figure, we arise in lube oil temperature, and again, a region of possible future movement of the trend is captured between the prognostic lines. In spite of irregular data collection and noisy data, we are able to make useful predictions.

A prognostic module based on the algorithms reported in this paper is being used to predict the health of Honeywell engines with several airline customers. Results thus far have been encouraging: several faults have been anticipated.

CONCLUSIONS

We have developed a new framework for smoothing, filtering and prediction based on statistical interval estimation. We have found the approach well suited to provide prognoses from time series data with non-uniform sampling and irregular collection. The algorithms are usable in many places where traditional filtering techniques are used on scalar time series. Examples are fusion of data from multiple sensors, online filtering for slow processes (since it is a batch method that is computationally intensive compared to recursive filters), and noise model identification.

ACKNOWLEDGEMENTS
This work was supported by the Honeywell Engines, Systems and Services group. We also acknowledge useful discussions with Dinkar Mylaraswamy.

REFERENCES


CONTACT

Please direct all inquiries to Kartik B. Ariyur at kartik.ariyur@honeywell.com.