# $2^{nd}$ Statistical Report on Breathometer

Thomas Ng

## I. INTRODUCTION

This report is written to address and modify the calibration issues for Breathometers. Previously in the first report, we adopted the linear regression model for calibration. In this report, we examine the variation of voltage over a short interval of time after blowing alcohol into the device, which in turn motivate us to suspect that there is a chance that the device has a design flaw, i.e. certain amount of alcohol is being trapped in the device after blowing alcohol into the device and thus, leads to overestimation problems. In the light of this issue, we attempt to develop another regression model to try and overcome these issues.

## II. STATISTICAL MODEL - REGRESSION WITH DECAY

For the measurements made in $(CP_1', CP_2', CP_1, CP_2, CP_3, CP_4)(i.e., BAC = 2\%, 4\%, 2\%, 4\%, 6\%, 8\%)$, the corresponding statistical model is:

$$\begin{aligned}
\log_{10}(RLRS_1) &= \alpha + \beta \times \log_{10}(BAC_1) + \epsilon_1 \\
\log_{10}(RLRS_2) &= \alpha + \beta \times \log_{10}(BAC_2 + BAC_1 \times e^{-r(t_2 - t_1)}) + \epsilon_2 \\
\log_{10}(RLRS_3) &= \alpha + \beta \times \log_{10}(BAC_3 + BAC_2 \times e^{-r(t_3 - t_2)} + BAC_1 \times e^{-r(t_3 - t_1)}) + \epsilon_3 \\
&\vdots
\end{aligned}$$

where $\epsilon_i$ are independent noise with common mean zero and variance $\sigma^2$ and $t_i$ is the time-stamp when $i^{th}$ measurement is made. That is, for $i = 1, \cdots, 6$

$$\log_{10}(RLRS_i) = \alpha + \beta \times \log_{10}(w_i) + \epsilon_i$$

with

$$w_i = \sum_{j=1}^{i} BAC_j \times e^{-r(t_i - t_j)}$$

the noise is in the scale of $\log(RLRS)$. The parameters can be estimated by:

$$\min_{r,\alpha,\beta} \sum_{i=1}^{6} (y_i - \alpha - \beta x_i(r))^2$$

where $y_i = log_{10}(RLRS_i)$ and $x_i(r) = \log_{10}(w_i)$ which is a function of $r$. For each fix $r$, the $\alpha$ and $\beta$ can then be estimated by minimising the error sum of squares: (i.e. Regression)

$$\sum_{i=1}^{6}(y_i - \alpha - \beta x_i)^2$$

## III. COMPUTATION - NEWTON-RAPHSON METHOD

To minimise the residual sum of squares in the regression problem above, we need to use the Newton - Raphson Method, which is described below:

## IV. NEWTON RAPHSON METHOD

To minimise the function $f(x)$ below

$$f(x, y) = \sum_{i=1}^{6}(y_i - \alpha - \beta x_i(r))^2$$

, where $x_i(r) = log(w_i) = log\left(\sum_{j=1}^{i} BAC_j \times e^{-r(t_i - t_j)}\right)$.
For simplicity, we will denote $x_i(r)$ as $x_i$. The Newton Raphson method is used to estimate the parameters $\theta = (\alpha, \beta, r)^T$. The algorithm is as follow:

$$\theta^{(t+1)} = \theta^t - \left[f''(x, y|\theta^{(t)})\right]^{-1} f'(x, y|\theta^{(t)})$$

,where $f'(x, y|\theta^{(t)})$ will be the vector of score function at the $t^{th}$ iteration and $f''(x, y|\theta^{(t)})$ will be the Hessian matrix, i.e. $2^{nd}$ derivative of the $f(x, y)$ at the $t^{th}$ iteration.

Since this is a minimisation problem, we expect the Hessian matrix , i.e. $f''(x, y|\theta^{(t)})$ to be positive definite. We need to check that their eigenvalues must ALL be positive $\lambda_i > 0 \quad \forall\, i$.

### A. Score function

$$f'(x,y|\theta) = \begin{bmatrix} \frac{\partial f}{\partial \alpha} \\ \frac{\partial f}{\partial \beta} \\ \frac{\partial f}{\partial r} \end{bmatrix} = \begin{bmatrix} -2\sum_i (y_i - \alpha - \beta x_i) \\ -2\sum_i (y_i - \alpha - \beta x_i)\, x_i \\ -2\beta \sum_i (y_i - \alpha - \beta x_i) \frac{dx_i}{dr} \end{bmatrix}$$

, where

$$\frac{dx_i}{dr} = \frac{1}{ln(10)} \frac{1}{w_i} \frac{dw_i}{dr} \tag{IV.1}$$

$$= \left( -\frac{1}{ln(10)} \cdot \frac{1}{w_i} \cdot \sum_{j=1}^{i} BAC_j e^{-r(t_i - t_j)} \cdot (t_i - t_j) \right) \tag{IV.2}$$

### B. $2^{nd}$ derivative matrix

The $2^{nd}$ derivative matrix is:

$$\begin{bmatrix} \frac{\partial^2 f}{\partial \alpha^2} & \frac{\partial^2 f}{\partial \alpha \partial \beta} & \frac{\partial^2 f}{\partial \alpha \partial r} \\ \frac{\partial^2 f}{\partial \beta \partial \alpha} & \frac{\partial^2 f}{\partial \beta^2} & \frac{\partial^2 f}{\partial \beta \partial r} \\ \frac{\partial^2 f}{\partial \alpha \partial r} & \frac{\partial^2 f}{\partial r \partial \beta} & \frac{\partial^2 f}{\partial r^2} \end{bmatrix}$$

Individual elements are computed as follow:

$$\frac{\partial^2 f}{\partial \alpha^2} = 2 \times 6 = 12$$

$$\frac{\partial^2 f}{\partial \beta^2} = 2 \cdot \sum_i x_i^2$$

$$\frac{\partial^2 f}{\partial r^2} = (-2\beta) \left[ \sum_i (y_i - \alpha - \beta x_i) \frac{d^2 x_i}{dr^2} + \sum_i (-\beta) \cdot \frac{dx_i}{dr} \cdot \frac{dx_i}{dr} \right]$$

$$\frac{\partial^2 f}{\partial\alpha\partial\beta} = \frac{\partial^2 f}{\partial\beta\partial\alpha} = 2\sum_i x_i$$

$$\frac{\partial^2 f}{\partial\alpha\partial r} = \frac{\partial^2 f}{\partial r\partial\alpha} = 2\beta\sum_i \frac{dx_i}{dr}$$

$$\frac{\partial^2 f}{\partial\beta\partial r} = \frac{\partial^2 f}{\partial r\partial\beta} = (-2)\cdot\sum_i (y_i - \alpha - \beta x_i)\frac{dx_i}{dr} + (-2)(-\beta)\sum_i \frac{dx_i}{dr}x_i$$

, where

$$\begin{aligned}\frac{\partial^2 x_i}{\partial r^2} &= \left(-\frac{1}{ln(10)}\cdot\frac{1}{w_i^2}\cdot\frac{dw_i}{dr}\cdot\frac{dw_i}{dr} + \frac{1}{ln(10)}\cdot\frac{1}{w_i}\cdot\frac{d^2w_i}{dr^2}\right) \\ &= -\frac{1}{ln(10)}\cdot\frac{1}{w_i^2}\cdot\frac{dw_i}{dr}\cdot\underbrace{\sum_{j=1}^{i} BAC_j\cdot -(t_i - t_j)\cdot e^{-r(t_i-t_j)}}_{\frac{dw_i}{dr}} + \frac{1}{ln(10)}\cdot\frac{1}{w_i}\underbrace{\sum_{j=1}^{i} BAC_j(t_i-t_j)^2 e^{-r(t_i-t_j)}}_{\frac{d^2w_i}{dr^2}}\end{aligned}$$

The initial values of the parameters $\beta$ and $\alpha$ for newton method are chosen to be the OLS parameter estimate. So, I will just plug the OLS estimate into the newton algorithm.

## V. FINGDINGS

Figure V.1 shows that regression with decay outperforms the standard line approach adopted by the engineers, as the median of the estimates from regression with decay method are closer to the true $BAC$ levels than that of estimates from standard line method. This is also illustrated in Figure V.2, where the *absolute out-sample prediction error difference between standard line approach and regression with decay* are positive.

Regression with decay seems to perform better than standard linear regression, based on Figure V.1. However, if we look at the absolute out-sample prediction error difference between standard linear regression and regression with decay in Figure V.3, it seems that both standard linear regression and regression with decay seem to perform equally well for $CP1, CP2, CP3$ and standard linear regression performs better than regression with decay for $CP4$.

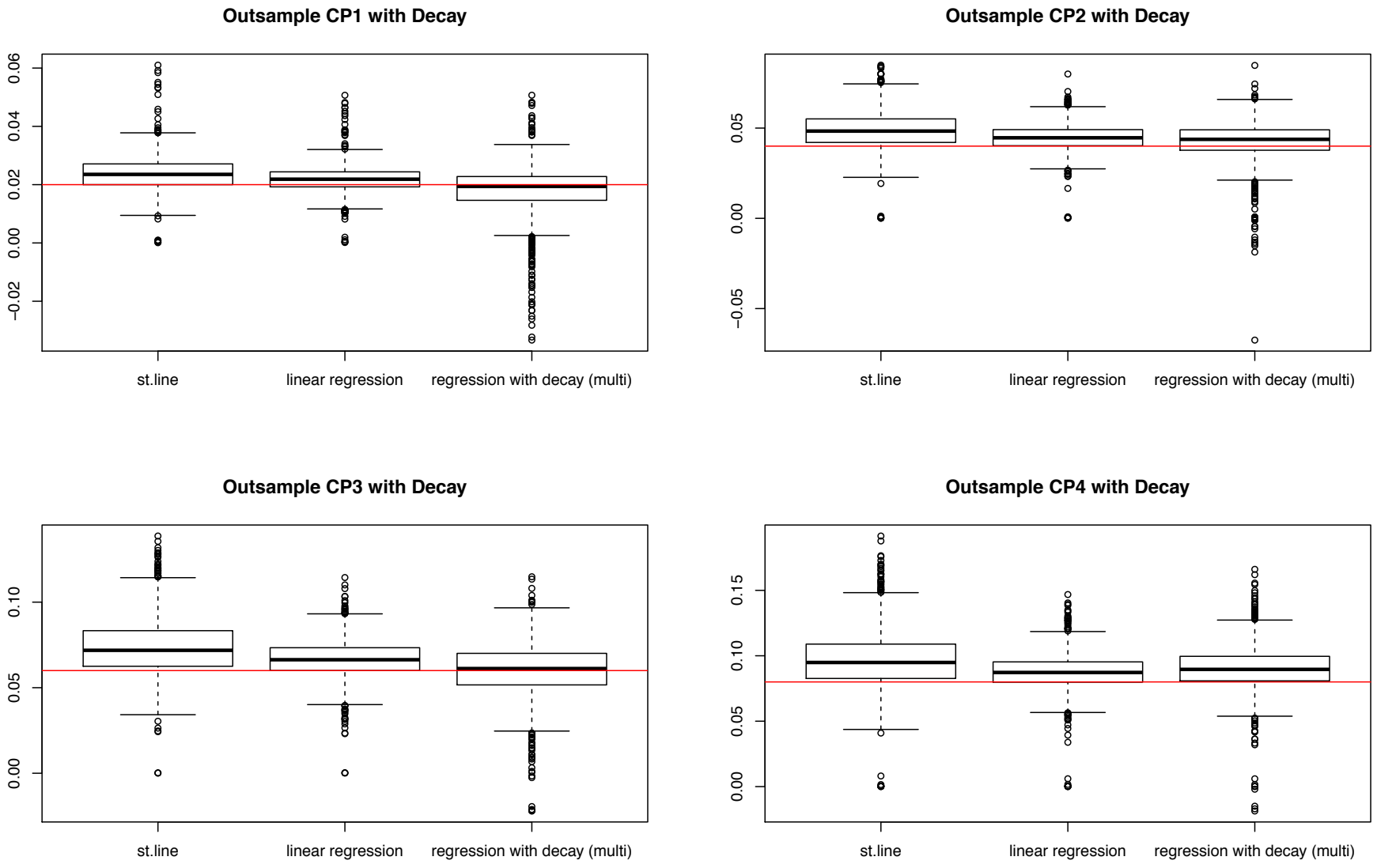


Figure V.1: Out-Sample Predictions

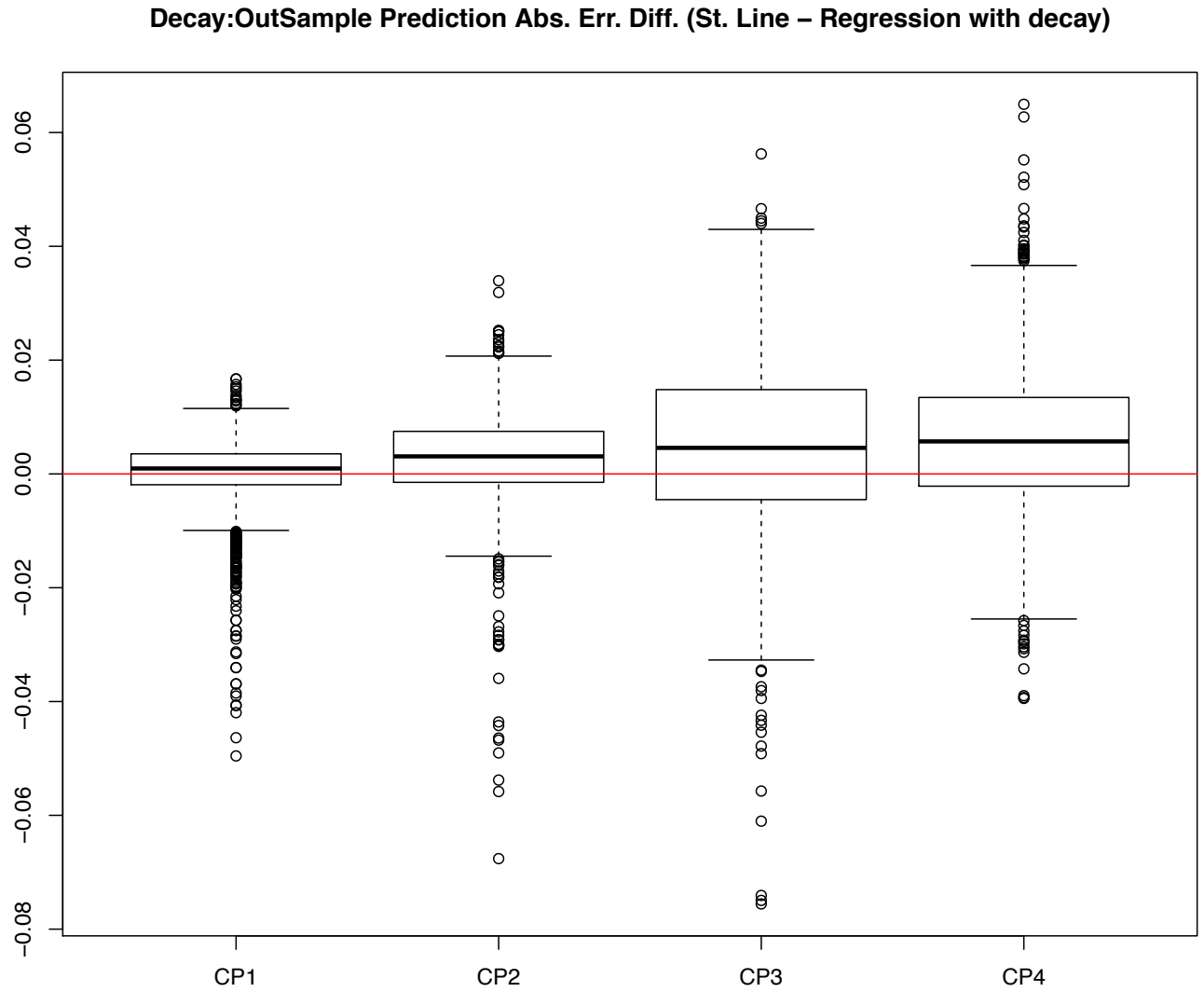


Figure V.2: Out-Sample Predictions Absolute Error Difference (st.line - regression with decay)

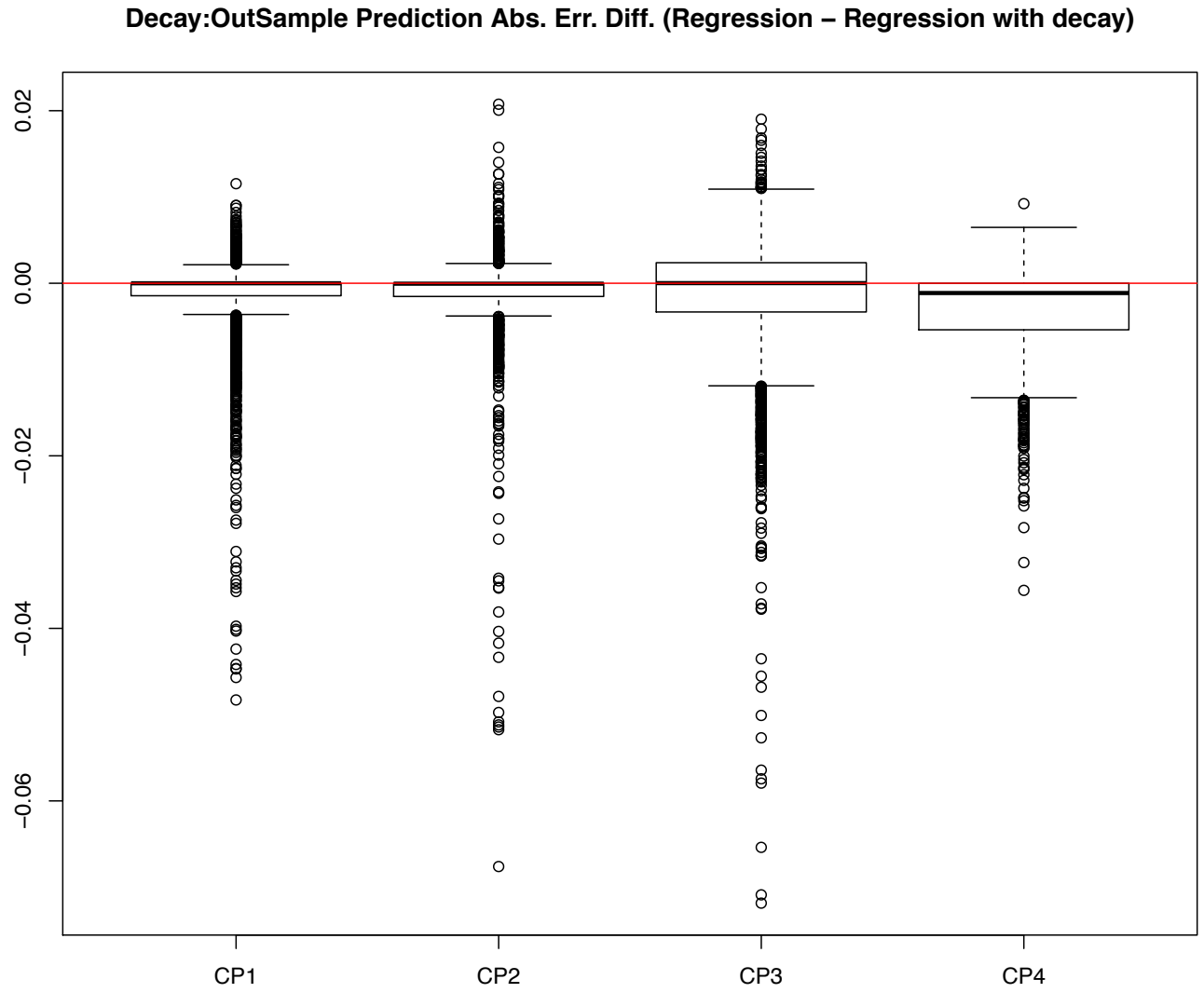


Figure V.3: Out-Sample Predictions Absolute Error Difference (st.line - regression with decay)