WTLS solutions (derived in (a)):

$$\hat{a}_{WTLS} = \frac{S_{yy} - \rho S_{uu} + \sqrt{(S_{yy} - \rho S_{uu})^2 + 4\rho S_{yu}^2}}{2 S_{yu}}$$

$$\hat{b}_{WTLS} = \bar{y} - \hat{a}_{WTLS} \bar{u}$$

$$\hat{u}_{i,WTLS} = \frac{\hat{a}_{WTLS}(y_i - \hat{b}_{WTLS}) + \rho u_i}{(\alpha^2 + \rho)}$$

$$\hat{y}_{i,WTLS} = \hat{a}_{WTLS} \hat{x}_{WTLS} + \hat{b}_{WTLS}$$

(2) a) Here, I consider CO_2 concentration as input variable and temperature deviation as output variable

This is because CO_2 concentration measurements are likely to have more errors. And, from an engineering viewpoint, CO_2 emission causes temperature deviation

N = 31 samples.

~~Perform~~ Computing sample statistics using MATLAB we get,

$$\bar{y} = \frac{\sum_i y_i}{N} = 0.8512$$

$$\bar{u} = \frac{\sum_i u_i}{N} = \cancel{0.0875} \quad 369.4197$$

$$S_{yy} = \frac{1}{N} \sum_i (y_i - \bar{y})^2 = 0.0875$$

$$S_{yu} = \frac{1}{N} \sum_i (y_i - \bar{y})(u_i - \bar{u}) = 4.192$$

$$S_{uu} = \frac{1}{N} \sum_i (u_i - \bar{u})^2 = 255.74$$

$$\hat{\alpha}_{OLS} = \frac{S_{yu}}{S_{uu}} = 0.01639$$

$$\hat{\beta}_{OLS} = \bar{y} - \hat{\alpha}\,\bar{u} = -5.2044$$

OLS estimate of CO_2 concentration reqd. for 3.6°F raise $= \dfrac{3.6 - \hat{\beta}_{OLS}}{\hat{\alpha}_{OLS}}$

$$= \boxed{537.1075 \text{ ppm}}$$

$$\hat{\alpha}_{TLS} = \frac{S_{yy} - S_{uu} + \sqrt{(S_{yy} - S_{uu})^2 + 4S_{yu}^2}}{2\,S_{yu}} = 0.016393$$

$$\hat{\beta}_{TLS} = \bar{y} - \hat{\alpha}_{TLS}\,\bar{u} = -5.20482$$

TLS estimate of CO_2 concentration reqd. for 3.6°F raise $= \dfrac{3.6 - \hat{\beta}_{TLS}}{\hat{\alpha}_{TLS}}$

$$= \boxed{537.0952 \text{ ppm}}$$

b) Now, we keep time (year) as input variable u

and CO_2 as output variable y.

We know that there is no error in the input, i.e., year. So we choose to go with an OLS estimator

Computing sample statistics in MATLAB,

$\bar{y} = 369.4197$ ppm $\quad \bar{u} = 1999$

$s_{yy} = 255.7407 \quad s_{yu} = 142.5342 \quad s_{uu} = 80$

$$\hat{\alpha}_{OLS} = \frac{s_{yu}}{s_{uu}} = 1.7817$$

$$\hat{\beta}_{OLS} = \bar{y} - \alpha\bar{u} = -3192.153.$$

i) Finding time taken to reach the OLS estimate of concentration

$$t = \frac{537.1025 - \hat{\beta}}{\hat{\alpha}} = \boxed{2093.1179}$$

ii) Finding time taken to reach the TLS estimate of concentration

$$t = \frac{537.0952 - \hat{\beta}}{\hat{\alpha}} = \boxed{2093.111}$$

Both the estimates are similar. Conservatively, we can say by year $\boxed{2093}$ & (72 more years) we will reach the maximum level of CO_2 to cause 3.6 °F rise

(3) a) Since we are performing OLS as well as ROLS:
choice of input-output doesnt matter for this analysis

Set input, u as ~~EF~~ EP method
output, y as CF method.

Sample statistics. [All computations done in MATLAB]

$S_{yy} = 0.9232$ $S_{yu} = 0.9241$ $S_{uu} = 0.9311$

$\bar{y} = 1.9505$ $\bar{u} = 2.0155$

Our aim to see whether the 95% confidence interval of $\hat{\alpha}$ contains 1 (so that $y = u$ ~~line~~ model comes) and if $\hat{\beta}$ is close to zero

95% CI : $\hat{\alpha} \pm$ ~~ds~~ $2.16\,\sigma_\varepsilon^2$

$\because \sigma_\varepsilon^2$ is unknown, $\hat{\sigma}_\varepsilon^2 = \sum \frac{(y_i - \hat{\alpha} u_i - \hat{\beta})^2}{N-2}$

OLS (Assumption: EP method is free of error, CF method has error)

$\hat{\alpha}_{OLS} = \frac{S_{yu}}{S_{uu}} = \boxed{0.9924}$ $\hat{\beta}_{OLS} = \boxed{-0.0497}$

95% confidence interval of $\hat{\alpha}$: $[0.9776, 1.0073]$

C.I. contains 1! $\hat{\beta}$ close to 0!

IOLS

Assumption : EP has error, CF method has no error.

$$\hat{\alpha}_{IOLS} = \frac{S_{yy}}{S_{yu}} = \boxed{0.9991}$$

$$\hat{\beta}_{IOLS} = \bar{y} - \hat{\alpha}_{IOLS}\bar{u} = \boxed{-0.0564}$$

95% confidence interval for $\hat{\alpha}_{IOLS}$: ~~0.9809~~ $[0.9842, 1.0074]$

C.I. contains 1! for $\hat{\alpha}$ $\hat{\beta}$ close to 0!

TLS

Assumption : both EP & CF method has error.

$$\hat{\alpha}_{TLS} = \frac{S_{yy} - S_{uu} + \sqrt{(S_{yy} - S_{uu})^2 + 4S^2_{yu}}}{2 S_{yu}} = \boxed{0.9957}$$

~~0.216893~~

$$\hat{\beta}_{TLS} = \bar{y} - \hat{\alpha}_{TLS}\bar{u} = \boxed{-0.0564}$$

~~8.2048~~

95% confidence interval for $\hat{\alpha}_{TLS}$ = $[0.9809, 1.0106]$

C.I. contains $\hat{\alpha} = 1$! $\hat{\beta}$ close to 0

All 3 approaches have $\hat{\alpha}$ close to 1 and $\hat{\beta}$ close to 0

So we can conclude that the new method (CF) is a good ~~sg~~ substitute for the established method (EP)

b) OLS : Input is error free → EP measurement is error free

~~⇒ Anth = u_i =~~

⇒ Estimate = u_i = $\boxed{2.31 \text{ mg/L}}$

IOLS

Output is error free → CF is error free

⇒ Estimate = y_i = $\boxed{2.20 \text{ mg/L}}$

TLS   Approximating $\alpha = 1$, $\beta = 0$,

~~Assuming~~ $\alpha = 1$,

$$\text{estimate } \hat{y}_i = \hat{u}_i = \frac{\alpha(y_i - \beta) + u_i}{\alpha^2 + 1}$$

$$= \frac{y_i + u_i}{2} = \frac{2.31 + 2.2}{2}$$

⇒ estimate of phytic acid = $\boxed{2.255 \text{ mg/L}}$

(4) The objective is to find if the defect is aligned across vertical or horizontal axis.

~~&~~ So by performing Linear regression, ~~if~~

→ If we get equation as $y = \beta$ it is parallel to x-axis

→ if we get equation as $x = \beta$ it is parallel to y-axis

Essentially we can do 2 regression for each data:

$$y = \alpha_1 x + \beta_1 \& \quad x = \alpha_2 y + \beta_2$$

and check if α_1 or α_2 is small enough / close to zero

Since swapping x & y is eqnt to IOLS we need ~~to~~ not do that separately. Just need to do OLS ~~and~~ to cover IOLS also.

In addition, we can check using TLS. Also, value of β doesn't matter in terms of deciding defect orientation

Defect - 1

~~$y = \alpha_1 x + \beta_1$ for~~.

$\bar{y} = 824.044 \quad \bar{x} = 3774.6$

$S_{yy} = 267.0202 \quad S_{yx} = 171.9279 \quad S_{xx} = 1876.5$

Model 1: $y = \alpha_1 x + \beta_1$

$$\alpha_{1,OLS} = \frac{S_{yy}}{S_{yu}} = \underline{0.0092}$$

$$\alpha_{1,TLS} = \frac{S_{yy} - S_{uu} + \sqrt{(S_{yy} - S_{uu})^2 + 4S_{yu}^2}}{2 S_{yu}}$$

$$= \underline{0.0093}$$

Both OLS & TLS suggest, the line is of the form

$$\underline{y = \beta}$$

~~Model 2: $x = \alpha_2 y + \beta_2$~~ (~~Note that actual slope is inverse of~~ this

So we have successfully identified the defect along x-axis

$$\therefore \boxed{\text{Defect 1 is due to corrosion of steel reinforcement bars}}$$

Defect 2

$\bar{u} = 2143.9$ $\bar{y} = 3190.2$

$S_{yy} = 1904.0$ $S_{yu} = 1708.0$

$S_{uu} = 2153.9$; Model ①: $y = \hat{\alpha} x + \hat{\beta}$

$$\hat{\alpha}_{1,OLS} = \frac{S_{yu}}{S_{yy}} = \cancel{0.0092}\ \underline{0.793}$$

$$\hat{\alpha}_{1,TLS} = \frac{S_{yy} - S_{uu} + \sqrt{(S_{yy} - S_{uu})^2 + 4S_{yu}^2}}{2S_{yu}} = \underline{0.9295}$$

Both $\hat{\alpha}$ estimates are close to 1. Trying with model 2.

Model ② : $x = \hat{\alpha} y + \hat{\beta}$

$$\hat{\alpha}_{2,OLS} = \frac{S_{yu}}{S_{yy}} = \underline{0.8971}$$

(using u, y nomenclature as defined earlier)

$$\hat{\alpha}_{2,TLS} = \frac{S_{yu} - S_{yy} + \sqrt{(S_{uu} - S_{yy})^2 + 4S_{uy}^2}}{2S_{uy}} = \underline{1.0258}$$

Once again both are close to 1. and not the equation is of the form $y = x$ along vertical or horizontal axis (instead oriented at 45° - along $y = x$)

$\Rightarrow$ Defect 2 is not due to corrosion of steel reinforcement bars

Defect 3

$\bar{y} = 3522.3$ $\quad \bar{x} = 2832.7$ $\quad S_{yy} = 3099.5$

$S_{yx} = 910.12$ $\quad S_{xx} = 5152.5$

Model ①: $y = \hat{\alpha} x + \hat{\beta}$

$\hat{\alpha}_{1,OLS} = \cancel{28.4298}\ 0.1766$ $\quad \hat{\alpha}_{1,TLS} = \cancel{0.1766}\ 28.4298$

Model ②: $x = \hat{\alpha} y + \hat{\beta}$

$\hat{\alpha}_{2,OLS} = 0.0294$ $\quad \hat{\alpha}_{2,TLS} = 0.0352$

We see that $\hat{\alpha}_{2,OLS}$ & TLS are close to zero.

$\Rightarrow$ the line is of $x = \hat{\beta}$. the defect is along the y-axis

$\therefore$ Defect 3 is due to corrosion of steel reinforcement bars.

conclusion: Defects 1 & 2 are due to corrosion of and steel reinforcement bars