Asymptotic modelling of corn plant data using moving sums (MOSUM) technique

W Somayasa
Department of Mathematics, Hala Oleo University, Jl. Mokodompit, Kendari 93118, Indonesia
E-mail: wayan.somayasa@uho.ac.id

Abstract. Empirical model building using linear regression has been widely used in agricultural study. The validity of an assumed model is usually check using either F of likelihood ratio test under normally distributed observations. In this paper an asymptotic method based on the moving sums (MOSUM) of triangular array of ordinary least squares (OLS) residuals is proposed. Reasonable tests statistics for detecting valid model are defined as the Kolmogorov-Smirnov (KS) and Cramér-von Mises (MvM) functionals of the residual MOSUM process. The limit process is obtained for the condition under $H_0$ and under $H_1$ by applying the continuous mapping theorem to the existing limit theorem for partial sums of the residuals. The quantiles of the KS and CvM statistics under $H_0$ are approximated by simulation. The application of the method in obtaining a valid model for the speed of growth of corn plants results in the conclusion that a first-order polynomial regression model is shown to be plausible.

1. Introduction
In the present paper we propose a method of building a model empirically for the speed of growth of corn plant measured in the unit cm/day using linear regression. The response variable which is represented by the growth speed of the corn plant is observed in several points over a farm land region which is assumed to be a closed rectangle. The positions of the points on which the observations have been conducted were set according to a regular lattice of size $16 \times 21$ having the uniform distance of 75 cm between two nearby points running from east to west and from south to north. We aim at the development of a valid regression model describing the relationship between the position on the farm land and the corresponding growth speed of corn plant, so that the behavior how the speed of growth varies over the land can be studied. Moreover, by agricultural studies it is evidenced that the growth speed of corn plant can identify the level of fertility of the farm land where the corn has been planted. Hence, the fitted model can help the practitioner in accessing the variability of the fertility level of the land.

By following most common approach, we regard the data as a finite realization of a heterochedastic regression model

\[\{Z(t,s) = m(t,s) + \delta(t,s) : (t,s) \in J}\], \hspace{1cm} (1)

where $Z(t,s)$ represents the speed of growth of corn plant observed in the point $(t,s) \in J$, $\delta(t,s)$ is independent random error with $E(\delta(t,s)) = 0$ and $Var(\delta(t,s)) = \sigma(t,s)$ and $m$ is the true-unknown regression function having bounded variation on $J$. Thereby $\sigma$ is a positive function on $J$. As suggested in Härdle [1], the unknown regression function $m$ can be estimated by applying
smoothing techniques such as using either kernel or spline method. For least squares estimation in linear regression it is assumed that \( m \) lies in a linear subspace \( \mathbf{V} \) generated by known regression functions \( \{ w_1, \ldots, w_q, w_{q+1}, \ldots, w_p \} \subset L_2(\nu_0, \mathbf{J}) \). In practice the validity of an assumed model needs to be check before the fitted model is used in prediction or in predicting or inferring the condition on the experimental region on which the speed of growth attains either maximum or minimum yield, see Somayasa [2]. To this end the vector of residuals is commonly investigated by the reason the residuals can better estimate the random errors. When the observations are normally distributed the well known likelihood ratio test which coincides with \( F \) test can be applied, see Graybill [3], Seber and Lee [4] and Arnold [5]. Asymptotic test based on the limiting distribution of \( F \) statistic was studied in Arnold [6] and Pruscha [7]. Next, motivated by the pioneering works in MacNeill and et al. [8] and Xie and MacNeill [9], asymptotic model check based on set indexed cumulative sums (CUSUM) process of the matrix of ordinary least squares (OLS) residuals has been proposed in Somayasa and et al. [10] and Somayasa [11] using the geometric technique established in Bischoff and Somayasa [12].

Inspiring by an argument stated in Chu and et al. [13] that for testing model constancy, MOSUM test based on recursive residuals is more sensitive than the CUSUM test, we propose an asymptotic test by embedding the matrix of OLS residuals to a spatial random field using the MOSUM operator defined below for testing whether or not \( \mathbf{H} \) rejecting \( (\nu) \) observations. The MOSUM of the matrix of OLS residuals is defined by

\[
O_{ij}^{(n_1,n_2)} := \sum_{k=j+1}^{j+\gamma_1} \sum_{\ell=1}^{i+\gamma_2} r_{\ell k}, \quad i = 1, \ldots, n_1, \quad j = 1, \ldots, n_2, \tag{2}
\]

where the paired index \((i, j)\) corresponds to the point \((t_{n_1}, s_{n_2})\) in the experimental region \( \mathbf{J} \), such that \( F_{01}(t_{n_1}) = i/n_1 \) and \( F_{02}(s_{n_2}) = j/n_2 \). The positive integers \( \gamma_1 \) and \( \gamma_2 \) determines the size of the window which are chosen in such a way that \((i + \gamma_1)/n_1 = F_{01}(t_{n_1} + h_1) \) and \((j + \gamma_2)/n_2 = F_{02}(s_{n_2} + h_2)\), for \( h_1 \) and \( h_2 \) with \( 0 \leq h_1 \leq b_1 - a_1 \) and \( 0 \leq h_2 \leq b_2 - a_2 \). By extending the ideas of [9, 13, 14], MOSUM test for testing \( H_0 : m \in \mathbf{W} \) against \( H_1 : m \in \mathbf{V} \) can be realized by applying the KS and CvM type statistics, defined respectively by

\[
\mathcal{K}S_{n_1 \times n_2} := \max_{1 \leq i \leq n_1, 1 \leq j \leq n_2} \left| \frac{1}{\sqrt{n_1 n_2}} O_{ij}^{(n_1,n_2)} \right| \tag{3}
\]

\[
\mathcal{C}M_{n_1 \times n_2} := \frac{1}{n_1 n_2} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} \left( O_{ij}^{(n_1,n_2)} \right)^2,
\]

rejecting \( H_0 \) for large values of \( \mathcal{K}S_{n_1 \times n_2} \) or \( \mathcal{C}M_{n_1 \times n_2} \).

The asymptotic size \( \alpha \) rejection regions of the \( \mathcal{K}S \) and \( \mathcal{C}M \) MOSUM tests can be obtained by deriving the limiting distributions of \( \mathcal{K}S_{n_1 \times n_2} \) and \( \mathcal{C}M_{n_1 \times n_2} \) under \( H_0 \), whereas the limiting distributions under \( H_1 \) are important for investigating the powers of the tests. This subject will be discussed in Section 2. The application of the method to the speed of growth of corn plant will be demonstrated in Section 3. The paper will be closed with some conclusions and remarks for future researches in Section 4.
2. Asymptotic distributions of the KS and CvM type statistics

Before proceed we define some preliminary notations which are important for our result. For fixed $h_1$ and $h_2$, with $0 \leq h_1 \leq b_1 - a_1$ and $0 \leq h_2 \leq b_2 - a_2$, we define a difference operator $\Delta^{(h_1,h_2)}$ on $C(J_{h_1h_2})$, such that for every $f \in C(J_{h_1h_2})$,

$$\Delta^{(h_1,h_2)}f(t,s) := f(t + h_1, s + h_2) - f(t, s + h_2) - f(t + h_1, s) + f(t, s).$$

**Proposition 2.1** For any fixed $h_1$ and $h_2$, with $0 \leq h_1 \leq b_1 - a_1$ and $0 \leq h_2 \leq b_2 - a_2$, the operator $\Delta^{(h_1,h_2)}$ is linear and continuous on $C(J_{h_1h_2})$.

**Proof:** Suppose that $f_1$ and $f_2$ are any two functions in $C(J_{h_1h_2})$ and $\alpha_1, \alpha_2$ are any two constants. We first show that $\Delta^{(h_1,h_2)}[\alpha_1 f_1 + \alpha_2 f_2] = \alpha_1 \Delta^{(h_1,h_2)} f_1 + \alpha_2 \Delta^{(h_1,h_2)} f_2$. By the definition of $\Delta^{(h_1,h_2)}$, for every $(t,s) \in J_{h_1h_2}$ we get

$$\Delta^{(h_1,h_2)}[\alpha_1 f_1 + \alpha_2 f_2](t,s) = [\alpha_1 f_1 + \alpha_2 f_2](t + h_1, s + h_2) - [\alpha_1 f_1 + \alpha_2 f_2](t, s + h_2)
- [\alpha_1 f_1 + \alpha_2 f_2](t + h_1, s) + [\alpha_1 f_1 + \alpha_2 f_2](t, s)
= \alpha_1 [f_1(t + h_1, s + h_2) - f_1(t, s + h_2) - f_1(t + h_1, s) + f_1(t, s)]
+ \alpha_2 [f_2(t + h_1, s + h_2) - f_2(t, s + h_2) - f_2(t + h_1, s) + f_2(t, s)]
= [\alpha_1 \Delta^{(h_1,h_2)} f_1] + [\alpha_2 \Delta^{(h_1,h_2)} f_2](t,s),$$

showing the linearity of $\Delta^{(h_1,h_2)}$. To prove the continuity, suppose $f_1$ and $f_2$ are any two functions in $C(J_{h_1h_2})$, then by recalling the definition of $\Delta^{(h_1,h_2)}$, for any $(t,s) \in J_{h_1h_2}$, it holds

$$\left| (\Delta^{(h_1,h_2)} f_1)(t,s) - (\Delta^{(h_1,h_2)} f_2)(t,s) \right| \leq |f_1(t + h_1, s + h_2) - f_2(t + h_1, s + h_2)| + |f_1(t, s + h_2) - f_2(t, s + h_2)|
+ |f_1(t + h_1, s) - f_2(t + h_1, s)| + |f_1(t, s) - f_2(t, s)| \leq 4 \| f_1 - f_2 \|_\infty .$$

The last inequality implies that $\| \Delta^{(h_1,h_2)} f_1 - \Delta^{(h_1,h_2)} f_2 \|_\infty \leq 4 \| f_1 - f_2 \|_\infty$. This means for every $\varepsilon > 0$, there exists $\delta := \varepsilon/4$, such that $\| \Delta^{(h_1,h_2)} f_1 - \Delta^{(h_1,h_2)} f_2 \|_\infty \leq \varepsilon$, whenever $\| f_1 - f_2 \|_\infty \leq \delta$, finishing the proof.

Let $W := \{ W(t,s) : (t,s) \in J \}$ be a centered Gaussian random field with the control measure $\nu_0$ and the covariance function given by

$$C_W(t,s; t', s') = \int_{[a_1,t] \times [a_2,s] \cap [a_1,t'] \times [a_2,s']} \sigma(x,y) \nu_0(dx,dy), \quad (t,s), (t',s') \in J,$$

see also MacNeill and et al. [8] and Somayasa [11]. When $\sigma$ is constant and $\nu_0$ is the uniform probability measure, then $W$ reduces to the well-known Brownian sheet which is also sometimes called Wiener-Chentsov field, cf. Lifshits [18], p. 11. The application of the operator $\Delta^{(h_1,h_2)}$ to $W$ leads us to the process $S_{h_1,h_2;\nu_0} := \Delta^{(h_1,h_2)} W$, which is defined as

$$S_{h_1,h_2;\nu_0}(t,s) = W(t + h_1, s + h_2) - W(t, s + h_2) - W(t + h_1, s) + W(t, s),$$

for every $(t,s) \in J_{h_1h_2}$. It can be easily shown that $S_{h_1,h_2;\nu_0}$ is a centered Gaussian process with the covariance function given by

$$C_{S_{h_1,h_2;\nu_0}}(t,s; t', s') = \int_{[t,t+h_1] \times [s,s+h_2] \cap [t',t'+h_1] \times [s',s'+h_2]} \sigma(x,y) \nu_0(dx,dy).$$
which reduces to the well-known Slepian random field denoted by $S_{J_{h_1 h_2}}$ when $\sigma$ and $\nu_0$ are set respectively to a constant and the uniform probability measure, with the covariance function

$$C_{S_{J_{h_1 h_2}}}(t_1, s_1; t_2, s_2) = (h_1 - |t_1 - t_2|)(h_2 - |s_1 - s_2|),$$

where $x^+ := \max\{x, 0\}$. The definition and some results regarding the small ball probability of $S_{J_{h_1 h_2}}$ have been well documented in Gao and Li [19]. The Cameron-Martin density of the Slepian process on interval can be studied in Bischoff and Grgg [20].

The MOSUM of the OLS residuals (2) can be written as a random function as follows. If the paired index $(i, j)$ is such that there exists a point $(t, s) \in J_{h_1 h_2} := [a_1, b_1 - h_1] \times [a_2, b_2 - h_2]$, with $i/n_1 = F_{01}(t)$ and $j/n_2 = F_{02}(s)$ and the constants $h_1$ and $h_2$ with $0 \leq h_1 \leq b_1 - a_1$ and $0 \leq h_2 \leq b_2 - a_2$, such that $(i + \gamma_1)/n_1 = F_{01}(t + h_1)$ and $(j + \gamma_2)/n_2 = F_{02}(s + h_2)$, then the following equation holds true

$$O^{(n_1, n_2)}_{(ij)}(t, s) := \sum_{k=[n_2 F_{02}(s)]}^{[n_2 F_{02}(s)+h_2]} \sum_{\ell=[n_1 F_{01}(t)]}^{[n_1 F_{01}(t)+h_1]} r_{\ell k}, \quad (4)$$

for $(t, s) \in J_{h_1 h_2}$, where $[x] := \max\{z \in \mathbb{Z} \geq 0 \mid z \leq x\}$. Throughout this paper the sum on the right-hand side of (4) will be denoted by $O^{(n_1, n_2)}_{(h_1 h_2; h_2)}(t, s)$, for $(t, s) \in J_{h_1 h_2}$. As a special case, when $\nu_0$ is given by the uniform probability measure $\lambda$ and $J$ is the unit rectangle $I = [0, 1] \times [0, 1]$, having the probability distribution function $F_0(t, s) = ts$, $(t, s) \in I$, the design of experiment will be an $n_1 \times n_2$ regular lattice, and we have

$$O^{(n_1, n_2)}_{(h_1 h_2; h_2)}(t, s) := \sum_{k=[n_2 s]}^{[n_2 s]+[n_2 h_2]} \sum_{\ell=[n_1 t]}^{[n_1 t]+[n_1 h_1]} r_{\ell k},$$

Throughout this paper we call the random processes (2) the residual MOSUM process. The sample paths of the process are in the Skorohod space $C(J_{h_1 h_2})$ of right continuous functions with left limit on $J_{h_1 h_2}$. The space is furnished with the Skorohod topology induced by the supremum norm $\| \cdot \|_\infty$, cf. Billingsley [17].

Now we are ready to state the limiting process of $O^{(n_1, n_2)}_{(h_1 h_2; h_0)}$ under $H_0$ and $H_1$.

**Theorem 2.2** Suppose the regression functions $\{w_1, \ldots, w_q\}$ are orthonormal in $L_2 (\nu_0, J)$ and have bounded variation on $J$. For fixed $h_1$ and $h_2$ with $0 \leq h_1 \leq b_1 - a_1$ and $0 \leq h_2 \leq b_2 - a_2$, if $H_0$ is true, then it holds

$$\frac{1}{\sqrt{n_1 n_2}} O^{(n_1, n_2)}_{(h_1 h_2; h_0)} \Rightarrow W_{h_1 h_2; h_0}, \quad \text{as } n_1, n_2 \to \infty,$$

where

$$W_{h_1 h_2; h_0} := S_{h_1 h_2; h_0} - \frac{1}{q} \sum_{i=1}^{q} \left( \int_{J} w_i(x, y) dW(x, y) \right) \Delta^{(h_1 h_2)}S_{w_i}.$$  

**Proof**: Let $h_1$ and $h_2$ be fixed with $0 \leq h_1 \leq b_1 - a_1$ and $0 \leq h_2 \leq b_2 - a_2$. Then by recalling (4), for every $(t, s) \in J_{h_1 h_2}$, we have

$$\sum_{k=[n_2 F_{02}(s)]+1}^{[n_2 F_{02}(s)+h_2]} \sum_{\ell=[n_1 F_{01}(t)]+1}^{[n_1 F_{01}(t)+h_1]} r_{\ell k} = \sum_{k=[n_2 F_{02}(a_2)]}^{[n_2 F_{02}(a_2)+h_2]} \sum_{\ell=[n_1 F_{01}(a_1)]}^{[n_1 F_{01}(a_1)+h_1]} r_{\ell k}.$$
where $1_A$ is the indicator of any set $A$. Thus, we get the following equation

$$
\frac{1}{\sqrt{n_1 n_2}} \mathcal{G}^{(n_1 n_2)}_{h_1 h_2; \nu_0} = \Delta^{(h_1 h_2)} \frac{1}{\sqrt{n_1 n_2}} Z_n (R_{n_1 \times n_2}),
$$

where $Z_n (R_{n_1 \times n_2})$ is the partial sums (CUSUM) process of the OLS residuals, defined and investigated in [8, 9, 10, 11], where it was shown therein that under $H_0$,

$$
\frac{1}{\sqrt{n_1 n_2}} Z_n (R_{n_1 \times n_2}) \Rightarrow \mathcal{W} - \sum_{i=1}^{q} \left( \int_{\mathbb{J}} w_i(x, y) d\mathcal{W}(x, y) \right) S_{w_i},
$$

where $S_{w_i}(t, s) = \int_{[a_1, t] \times [a_2, s]} w_i(x, y) \nu_0(dx, dy)$. Furthermore, since by Proposition 2.1 the operator $\Delta^{(h_1 h_2)}$ is continuous and linear, then by applying the well-known continuous mapping theorem, cf. Billingsley [17], we get

$$
\frac{1}{\sqrt{n_1 n_2}} \mathcal{G}^{(n_1 n_2)}_{h_1 h_2; \nu_0} \Rightarrow \mathcal{W}_{h_1, h_2, \tilde{w}, \nu_0} := \Delta^{(h_1 h_2)} \mathcal{W} - \Delta^{(h_1 h_2)} \sum_{i=1}^{q} \left( \int_{\mathbb{J}} w_i(x, y) d\mathcal{W}(x, y) S_{w_i} \right) \Delta^{(h_1 h_2)} S_{w_i},
$$

establishing the proof of the theorem.

It is worth noting that the limit process is denoted by $\mathcal{W}_{h_1, h_2, \tilde{w}, \nu_0}$ by the dependence of the probability distribution on the window size $h_1$ and $h_2$, the vector of regression functions $\tilde{w} = (w_1, \ldots, w_q)^T$ and the design $\nu_0$. This means that the limit process must be computed case by case. For example, when under $H_0$ we consider a constant model with $\mathbf{W} = [w_1]$, where $w_1 \equiv 1$, then the limit process is given by

$$
\mathcal{W}_{h_1, h_2, \tilde{w}, \nu_0}(t, s) = S_{h_1, h_2; \nu_0}(t, s) - \mathcal{W}(h_1, h_2) \nu_0([a_1, t] \times [a_2, s]), (t, s) \in \mathbf{J}_{h_1, h_2},
$$

which is a centered Gaussian process with the covariance function

$$
\mathcal{C}_{\mathcal{W}}_{h_1, h_2, \tilde{w}, \nu_0}(t; s; t', s') = \nu_0 \left( [t, t + h_1] \times [s, s + h_2] \cap [t', t' + h_1] \times [s', s' + h_2] \right) - \nu_0 \left( [t, t + h_1] \times [s, s + h_2] \cap [t', t' + h_1] \times [s', s' + h_2] \right).
$$

The limiting distribution of the KS and CvM statistics can now be straightforwardly obtained by utilizing the result in Theorem 2.2 and the continuous mapping theorem.
Table 1. Simulated $\tilde{t}_{1-\alpha}$ associated with constant and first-order model for several chosen values of $\alpha$. The samples are generated using 100 $\times$ 100 regular lattice on the unit rectangle. The error terms are generated from $N(0, \sigma(t, s))$, with $\sigma(t, s) = ts$ under 10000 runs.

| Model      | $h_1$ | $h_2$ | 0.01000 | 0.02500 | 0.05000 | 0.10000 | 0.15000 | 0.20000 |
|------------|-------|-------|---------|---------|---------|---------|---------|---------|
| Constant   | 0.10  | 0.10  | 0.03649 | 0.03446 | 0.03271 | 0.03124 | 0.03030 | 0.02950 |
|            | 0.15  | 0.20  | 0.03695 | 0.03461 | 0.03313 | 0.03136 | 0.03030 | 0.02956 |
|            | 0.20  | 0.30  | 0.03630 | 0.03438 | 0.03282 | 0.03132 | 0.03032 | 0.02953 |
| First-order| 0.40  | 0.30  | 0.03650 | 0.03456 | 0.03309 | 0.03139 | 0.03035 | 0.02957 |
|            | 0.35  | 0.20  | 0.036807 | 0.03447 | 0.03302 | 0.03128 | 0.03026 | 0.02945 |
|            | 0.20  | 0.30  | 0.03647 | 0.03428 | 0.03288 | 0.03123 | 0.03027 | 0.02954 |

Corollary 2.3 Under the condition of Theorem 2.2, we have

$$\begin{align*}
\mathcal{KS}_{n_1 \times n_2} &\Rightarrow \sup_{(t,s)\in J_{h_1,h_2}} \left| W_{h_1,h_2,\tilde{\nu}_t}(t,s) \right| \\
\mathcal{CM}_{n_1 \times n_2} &\Rightarrow \int_{J_{h_1,h_2}} \left( W_{h_1,h_2,\tilde{\nu}_t}(t,s) \right)^2 \nu_0(dt,ds).
\end{align*}$$

Proof: By the property of the MOSUM operator, the KS statistic can be represented by

$$\begin{align*}
\mathcal{KS}_{n_1 \times n_2} &= \max_{1 \leq i \leq n_1 \atop 1 \leq j \leq n_2} \frac{1}{\sqrt{n_1 n_2}} \left| O_{ij}^{(n_1,n_2)} \right| = \sup_{(t,s)\in J_{h_1,h_2}} \left| O_{h_1 h_2,\nu_0}(t,s) \right| = \| O_{h_1 h_2,\nu_0} \|_\infty.
\end{align*}$$

Next, let $\nu_{n_1 n_2}$ be a discrete uniform probability measure on $(\mathbf{J}, \mathcal{B}(\mathbf{J}))$, defined by

$$\nu_{n_1 n_2}(B) := \frac{1}{n_1 n_2} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} 1_A(t_{n_1 i}, s_{n_2 j}),$$

where $(t_{n_1 i}, s_{n_2 j})$ is a design point that satisfies $F_0(t_{n_1 i}, s_{n_2 j}) = \frac{i j}{n_1 n_2}$, cf. [16]. Then $\mathcal{CM}$ can be written as

$$\mathcal{CM}_{n_1 \times n_2} = \int_{J_{h_1,h_2}} \left( O_{h_1 h_2,\nu_0}^{(n_1,n_2)}(t,s) \right)^2 \nu_{n_1 n_2}(dt,ds).$$

Since, supremum norm is continuous and $\nu_{n_1 n_2}$ converge in distribution to $\nu_0$, the results follow. Next, let $\hat{t}_{1-\alpha}$ be a constant, such that $\mathbf{P} \left\{ \sup_{(t,s)\in J_{h_1,h_2}} | W_{m,h_1,h_2,\nu_0}(t,s) | \geq \hat{t}_{1-\alpha} \right\} = \alpha$. If KS test is used, then $H_0$ will be rejected asymptotically at level $\alpha$, if and only if $\mathcal{KS}_{n_1 \times n_2} \geq \hat{t}_{1-\alpha}$. Similarly, let $\hat{q}_{1-\alpha}$ be a constant, such that $\mathbf{P} \left\{ \int_{J_{h_1,h_2}} ( W_{m,h_1,h_2,\nu_0}(t,s) )^2 \nu_0(dt,ds) \geq \hat{q}_{1-\alpha} \right\} = \alpha$. If CVM test is used, then $H_0$ will be rejected at level $\alpha$, if and only if $\mathcal{CM}_{n_1 \times n_2} \geq \hat{q}_{1-\alpha}$. However in the practice the decision is usually drawn based on the corresponding $p$-values of the tests denoted respectively by $\hat{\alpha}_{KS}$ and $\hat{\alpha}_{CM}$. Suppose $\mathcal{KS}_{n_1 \times n_2}$ and $\mathcal{CM}_{n_1 \times n_2}$ be the observed values of $\mathcal{KS}_{n_1 \times n_2}$ and $\mathcal{CM}_{n_1 \times n_2}$, respectively. Then both $\hat{\alpha}_{KS}$ and $\hat{\alpha}_{CM}$ are computed by the following
The 3rd International Conference On Science

Table 1 and 2 present respectively the approximation values of

For a given observation, $H_0$ will be rejected for all values of $\alpha$ larger than $\hat{\alpha}_{KS}$ or $\hat{\alpha}_{CM}$.

Unfortunately, the exact values of the quantiles of the limiting distribution of the KS and CM tests cannot be computed analytically. In this paper they are approximated by simulation. Table 1 and 2 present respectively the approximation values of $\hat{r}_{1-\alpha}$ and $\hat{q}_{1-\alpha}$ for several different choices of $\alpha$, $h_1$ and $h_2$ associated with constant and first-order models.

The power function of the tests can be obtained by deriving the limiting distribution of the residual MOSUM process under $H_1$. For this purpose, we consider the local alternative by observing the localized version of Model 1, defined by

$$Z(t, s) = \frac{1}{\sqrt{N_1N_2}} m(t, s) + \delta(t, s), \quad (t, s) \in J.$$  

By applying the geometric approach due to [10, 11, 12], it was shown for the observation of Model 5, that

$$\frac{1}{\sqrt{N_1N_2}} \mathbf{Z}_n(R_{n_1 \times n_2}) \Rightarrow \varphi_m + \mathcal{W} = \sum_{i=1}^{q} \left( \int_{J_i}^{(R)} w_i(x, y) d\mathcal{W}(x, y) \right) S_{w_i},$$

where

$$\varphi_m := S_m - \sum_{i=1}^{q} \left( \int_{J_i}^{(R)} w_i(x, y) dS_m(x, y) \right) S_{w_i}.$$  

Hence, by the continuity of the operator $\Delta^{(h_1, h_2)}$, it holds

$$\frac{1}{\sqrt{N_1N_2}} \mathcal{O}^{(n_1, n_2)}_{h_1 h_2 \nu_0} \Rightarrow \Delta^{(h_1, h_2)} \varphi_m + \mathcal{W}_{m; h_1 h_2 \nu_0},$$

where for every $(t, s) \in J_{h_1 h_2}$,

$$\Delta^{(h_1, h_2)} \varphi_m (t, s) = \int_{[t, t+h_1] \times [s, s+h_2]} m(x, y) \nu_0(dx, dy)$$

$$- \sum_{i=1}^{q} \left( \int_{J_i}^{(R)} w_i(x, y) dS_m(x, y) \right) \int_{[t, t+h_1] \times [s, s+h_2]} w_i(x, y) \nu_0(dx, dy).$$

When $H_0$ is true, there exist real constants $\beta_1, \ldots, \beta_q$, such that $m = \sum_{i=1}^{q} \beta_i w_i$. By substituting $m$ under the integral sign with the last equation, we get for arbitrary fixed $i \in \{1, \ldots, q\}$,

$$\int_{J_i}^{(R)} w_i(x, y) dS_m(x, y) = \int_{J_i}^{(R)} w_i(x, y) m(x, y) \nu_0(dx, dy) = \beta_i.$$  

Hence by applying the linear property of the integral, we further have

$$\sum_{i=1}^{q} \left( \int_{J_i}^{(R)} w_i(x, y) dS_m(x, y) \right) \int_{[t, t+h_1] \times [s, s+h_2]} w_i(x, y) \nu_0(dx, dy)$$

$$= \int_{[t, t+h_1] \times [s, s+h_2]} \sum_{i=1}^{q} \beta_i w_i(x, y) \nu_0(dx, dy) = \int_{[t, t+h_1] \times [s, s+h_2]} m(x, y) \nu_0(dx, dy).$$
which implies $\Delta^{(h_1,h_2)}\varphi_m = 0$. This means that by considering the localized model, the asymptotic size $\alpha$ rejection region of both tests are not altered. In other word, if the samples are generated by the localized model, the size of the tests are not altered in one hand, but on the other hand the behavior of the test can be investigated when $H_1$ is true. The limiting power functions of the size $\alpha$ KS and CvM MOSUM tests are given respectively by

$$
k_{KS}(\hat{t}_{1-\alpha},\varphi_m) := \mathbb{P}\left\{ \sup_{(t,s) \in J_{h_1,h_2}} \left| (\Delta^{(h_1,h_2)}\varphi_m)(t,s) + W_{m;h_1,h_2;\nu_0}(t,s) \right| \geq \hat{t}_{1-\alpha} \right\},
$$

$$
k_{CM}(\hat{q}_{1-\alpha},\varphi_m) := \mathbb{P}\left\{ \int_{J_{h_1,h_2}} \left( (\Delta^{(h_1,h_2)}\varphi_m)(t,s) + W_{m;h_1,h_2;\nu_0}(t,s) \right)^2 \nu_0(dt,ds) \geq \hat{q}_{1-\alpha} \right\},
$$

where $\hat{t}_{1-\alpha}$ and $\hat{q}_{1-\alpha}$ are the constants that satisfy $k_{KS}(\hat{t}_{1-\alpha},0) = \alpha$ and $k_{CM}(\hat{q}_{1-\alpha},0) = \alpha$. This condition are fulfilled when $m$ belongs to $W$, otherwise they will move away from $\alpha$. Unfortunately, it is impossible to conduct analytic computation to $k_{KS}(\hat{t}_{1-\alpha},\varphi_m)$ as well as $k_{CM}(\hat{q}_{1-\alpha},\varphi_m)$. One way of studying the finite sample size behaviors of the tests are by approximating the powers using Monte Carlo simulation.

### 3. Application to corn plant data

In this section we discuss the application of the developed test method to the corn plant data introduced in the beginning of Section 1. In contrast to the classical $F$ or likelihood ratio test, for our approach we do not need to check the probability distribution of the speed of growth of the corn plants, since the test will be conducted asymptotically without assuming any distribution model. The drop line scatter plot as well as the corresponding perspective plot of the observations with respect to their positions are presented respectively in Figure 1 and Figure 2, from which we can infer that a first-order polynomial model $Z(t,s) = \beta_0 + \beta_1 t + \beta_2 s + \delta(t,s)$ is fit to the data, having positive slope along both $t$ and $s$ directions, where the error terms are assumed to be heterochedastic with the variance function $\sigma(t,s)$, for $(t,s) \in J$, see also Somayasa [21] for the description of the data.

The drop line scatter plot and the associated perspective plot of the residuals under the first-order assumption are exhibited respectively in Figure 3 and Figure 4. The residuals of the observations fluctuate around the point $Z = 0$ showing a descriptive fact that the assumed model is a valid model. However we aim at the drawing of decision using asymptotic statistical inference based on the MOSUM process of the OLS residuals. The first step of the method suggests to

| Model   | $h_1$ | $h_2$ | 0.01000 | 0.02500 | 0.05000 | 0.10000 | 0.15000 | 0.20000 |
|---------|-------|-------|---------|---------|---------|---------|---------|---------|
| Constant | 0.15  | 0.35  | 1.16e-05| 1.15e-05| 1.14e-05| 1.13e-05| 1.12e-05| 1.12e-05|
|         | 0.30  | 0.40  | 1.11e-05| 1.10e-05| 1.09e-05| 1.08e-05| 1.07e-05| 1.06e-05|
| First-order | 0.15 | 0.35  | 1.16e-05| 1.15e-05| 1.14e-05| 1.13e-05| 1.12e-05| 1.11e-05|
|         | 0.30  | 0.40  | 1.11e-05| 1.10e-05| 1.09e-05| 1.08e-05| 1.07e-05| 1.06e-05|
|         | 0.40  | 0.25  | 1.12e-05| 1.11e-05| 1.10e-05| 1.09e-05| 1.08e-05| 1.08e-05|
|         | 0.40  | 0.25  | 1.12e-05| 1.11e-05| 1.10e-05| 1.09e-05| 1.08e-05| 1.07e-05|

Table 2. Simulated $\tilde{q}_{1-\alpha}$ associated with constant and first-order model for several chosen values of $\alpha$. The samples are generated using $100 \times 100$ regular lattice on the unit rectangle. The error terms are generated from $N(0,\sigma(t,s))$, with $\sigma(t,s) = ts$ under 10000 runs.
estimate $\sigma(t, s)$ under the assumed model. By the motivation that $\mathbb{E}(r_{lk}^2) = \text{Var}(r_{lk}) + [\mathbb{E}(r_{lk})]^2$, with $\mathbb{E}(r_{lk}) = 0$, for $1 \leq \ell \leq 16$ and $1 \leq k \leq 21$, and $r_{lk}$ estimates $\delta(t_{n1}, s_{n2k})$, we estimate $\sigma(t, s)$ by regressing $\text{Var}(r_{lk})$ on the observed values of $r_{lk}^2$, cf. [8]. The drop line scatter and perspective plot of the $r_{lk}^2$ are presented in Figure 5 and Figure 6, respectively. By ignoring the outliers, we fit a first-order model $r^2(t, s) = \gamma_0 + \gamma_1 t + \gamma_2 s + \varepsilon(t, s)$, with $\mathbb{E}(\varepsilon(t, s)) = 0$ using least squares method. The resulting fitted model of the variance is given by

$$\sigma(t, s) = 0.002882 + 0.011217t + 0.003842s, \ (t, s) \in J$$

Suppose we assume that a constant model is plausible to the speed of growth of the corn plants. By using (3), we get the corresponding values of the KS and CM statistics for the speed of growth data, that are $KS_{16 \times 21}^* = 0.0158444$ and $CM_{16 \times 21}^* = 0.00001741$ which result in the approximated p-values $\hat{\alpha}_{KS} = 0.0013230$ and $\tilde{\alpha}_{CM} = 0.0000014$. The chosen window size are $h_1 = 0.30$ and $h_2 = 0.40$. Both p-values strongly suggest that constant model is not significant for describing the variability of the speed of growth over the experimental region.

Next, we assume under $H_0$ that the first-order model holds true. Similarly, the corresponding values of the KS and CM statistics for $h_1 = 0.30$ and $h_2 = 0.40$ are $KS_{16 \times 21} = 0.0145473$ and $CM_{16 \times 21} = 0.0000148$. The simulated p-values of the KS and the CvM tests are $\hat{\alpha}_{KS} = 0.083230$ and $\hat{\alpha}_{CM} = 0.025300$. By these values, $H_0$ is not rejected for all values of $\alpha$ smaller than 0.083230 when we use the KS test. Similarly, under the CvM test, $H_0$ may not be rejected for all values smaller than 0.025300. Thus, there exist possibilities in which the first-order model is plausible for the variable speed of growth of corn plant. The least squares fitted model is given by

$$\hat{Z} = 0.244245 + 0.159307t + 0.079332s, \ (t, s) \in J,$$

with positive slope along both sampling directions from the west to the east and from the south to the north. This means that the speed of growth of the corn plant gets higher as the position of the points go away from the origin which is located in the south-west corner of the region. This conclusion coincides also with the data presented in Figure 1 and Figure 2. The corns planted in the north-east corner seem to have the largest speed of growth, whereas those observed in the south-west corner appear to have the smallest speed of growth. Since in general speed of growth of plant reflects the fertility level of the land, this means that the north-east region is the region with the largest fertility level. The fitted model can be used in studying such characteristic of the farm land without knowing the probability distribution model of the sample.
4. Discussion

The corn plant data considered in this paper has been investigated in the work of Somayasa [21] using the CUSUM process of the so called recursive residual of the observations instead of the OLS residuals. The resulting plausible model therein is also first-order polynomial model with positive slopes along both $x$ and $y$-axis, see [21] for more detailed information. The advantages of applying the last method is that the limiting distribution of the CUSUM process is readily obtained by the fact the recursive residuals are uncorrelated in contrast to the OLS residuals which are correlated each other. However intuitively, our current method based on the MOSUM process of the residuals is more sensitive in the ability of detecting the true-unknown model. Thus the two different approach successfully estimate the same model for the speed of growth of corn plant.

Other methods such as the classical $F$ test may be applied. However they need to check the normality of the observations. The application of the asymptotic $F$ test need to check that the
design matrix satisfies Huber’s condition. By these reasons, we conduct no analysis for the corn plant data using such mentioned methods.

5. Concluding remarks
We establish the limit theorem for the MOSUM process of the OLS residuals obtained from heteroskedastic linear regression model with closed bounded experimental region. The asymptotic spatial process is expressed in a formula which depends on the $(h_1 h_2)$—Slepian field. The Kolmogorov-Smirnov and Cramér-von Mises functionals of the $(h_1 h_2)$—Slepian field can be used in detecting the validity of the assumed regression model. The test can be implemented by approximating the quantiles of the limiting distributions of the KS and CM statistics by simulation. The application of the method in real data can help the practitioners in agriculture area in developing a map of the fertility level of a land farm.

In the forthcoming work we extend the scope of the study to the case of spatial linear regression model with second order stationary observations. For this end we need a functional central limit theorem for spatial stationary process and a kind of mixing process, see e.g. El Machkouri and Wu [22]. More effort is clearly needed.

Other interested subject in modelling of corn plant data is the optimal prediction (kriging) of the growth speed of corn plants in unobserved point. This interest will be investigated in the next research.

Acknowledgments
A special thank is dedicated to RISTEK-DIKTI for supporting the work through the basic research award 2019. The author is grateful to all lecturers and students of the Department of Mathematics of Halu Oleo University for hospitality.

References
[1] Härde W 2013 *Applied Nonparametric Regression* (London: Cambridge University Press)
[2] Somayasa W 2016 *J. Phys.: Conf. Ser.* **725** 012002
[3] Graybill F A 1994 *Regression Analysis: Concepts and Applications* (New York: Duxbury Press)
[4] Seber G A F and Lee A J 2003 *Linear Regression Analysis (2nd ed.)* (New York: John Wiley & Sons)
[5] Arnold S F 1981 *The Theory of Linear Models and Multivariate Analysis* (New York: John Wiley & Sons)
[6] Arnold S F 1984 *J. Multivariate Analysis* **15** 325
[7] Pruscha H 2000 *Vorlesungen über Mathematische Statistik* (Stuttgart: B.G. Teubner)
[8] Macneill I B, Mao Y and Xie L 1994 *The Canadian Journal of Statistics* **22** (4) 529
[9] Xie L and MacNeill I B 2006 *South African Statist. J.* **4** 33.
[10] Somayasa W, Ruslan, Cahyono E and Engkoimani L O 2015 *Fareast J. Math. Sci.* **96**(8) 933
[11] Somayasa W 2016 *IOP Conf. Proceedings* **1707** 080018
[12] Bischoff W and Somayasa W 2009 *J. Multivariate Analysis* **100** 2167–77
[13] Chu C S J, Hornik K and Kuan C M 1995 *Biometrika* **82** 603
[14] Kramér W, Ploberger W and Alt R 1988 *Econometrica* **56** 1355.
[15] Sen P K (1982) *Ann. Stat.* **10**(1) 307
[16] Somayasa W and Budiman H 2018 *Statistics and Its Interface* **11** 61
[17] Billingsley P 1999 *Convergence of Probability Measures (2nd ed.)* (New York: John Wiley & Sons)
[18] Lifshits M 2012 *Lecture on Gaussian Processes* (Berlin: Springer)
[19] Gao F and Li W V 2007 *Trans. Amer. Math. Soc.* **359** (5) 1339
[20] Bischoff W and Gegg A 2016 *J. Theoretical Probability* **29** (2) 707
[21] Somayasa W 2019 *WSEAS Trans. Math.* **18** 62
[22] El Machkouri M and Wu W B 2013 *Stochastic Process and their Applications* **123**(1) 1