ISSN: (3007-0384)
E-ISSN: (3007-0392)
مجلة وهج العلوم للعلوم الصرفة
المجلة متاحة على الرابط
https://uomosul.edu.iq/womeneducation/jwups/

Using The Cross-Validation Method to Predict Computer Model Outputs Using The Stochastic Gaussian Process Regression Model

Zahraa Duraid AL-Khdojuh[(1)], Younus Hazim Al-Taweel[(2)],

(1), (2) Department of Mathematics College of Education for Pure Science,University of Mosul

***Corresponding author e-mail:**
[(1)]zahraa.23esp76@student.uomosul.edu.iq
[(2)]younus.altaweel@uomosul.edu.iq

***Keywords*:**
Stochastic Gaussian Process Regression Models, Computer Models, Cross Validation, Output Transformer Less Circuit Function.

ARTICLE INFO
Article history:
Received 2025/2/28
Accepted 2025/4/3
Available online: 2025/6/1

Email:
journal.purescience.ge @uomosul.edu.iq

A B S T R A C T
Computer models (CMs), which are called simulators, may be computationally expensive. Gaussian process regression models (GPRMs) are statistical models that have been used as a surrogate for expensive computer models. GPRMs are built based on assumptions similar to any statistical model. Thus, we must verify the validity of GPRMs before using them. In this paper, we use GPRMs for stochastic computer models (CMs). Then, we discussed how to verify the GPRMs for stochastic CMs by using the Cross-Validation method to predict the GPRMs. The stochastic GPRM was applied to a real CM represented by OTL Circuit function as an illustrative example. The accuracy of the GPRM was verified using stochastic measures, and it was verified that the GPRM is suitable as an alternative to the complex CM.

© 2024JWUPS, College of Education for Girls, University of Mosul.

Alma' rifa Journal of Humanities
Alma' rifa Journal of Humanities

إستخدام طريقة المقاطع للتنبؤ بمخرجات الانموذج الحاسوبي باستخدام انموذج انحدار العملية الكاوسية التصادفية

زهراء دريد الخدوجة[(1)]، أ.م.د يونس حازم الطويل [(2)]
[(2)،(1)] جامعة الموصل، كلية التربية للعلوم الصرفة، قسم الرياضيات

الخلاصة:
النماذج الحاسوبية والتي تسمى محاكيات قد تكون مكلفة من الناحية الحسابية. نماذج انحدار العملية الكاوسية التصادفية تستخدم كبديل احصائي عن النماذج الحاسوبية المعقدة. يتم بناء انموذج انحدار العملية الكاوسية التصادفية بالاعتماد على بعض الفرضيات كأي انموذج احصائي. لذلك يجب التحقق من دقة انموذج انحدار العملية الكاوسية التصادفية قبل استخدامه كبديل عن الانموذج الحاسوبي التصادفي. في هذا البحث تم استخدام نماذج انحدار العملية الكاوسية التصادفية كبديل عن النماذج الحاسوبية التصادفية المعقدة. بعد ذلك ناقشنا كيفية التحقق من انموذج انحدار العملية الكاوسية

التصادفي للنماذج الحاسوبية التصادفية باستخدام طريقة المقاطع Cross–Validation (CV) للتنبؤ بمخرجات الانموذج الحاسوبي التصادفي. تم تطبيق انحدار العملية الكاوسية العشوائي على انموذج حاسوبي حقيقي متمثل بدالة الدائرة الكهربائية OTL Circuit function. وتم التحقق من دقة انموذج انحدار العملية الكاوسية التصادفي باستخدام بعض المقاييس التحقق التصادفية وتم التحقق من ان انموذج انحدار العملية الكاوسية التصادفي يصلح لان يكون بديل عن الانموذج الحاسوبي التصادفي المعقد.

الكلمات المفتاحية: نماذج انحدار العملية الكاوسية التصادفية–النماذج الحاسوبية التصادفية–طريقة المقاطع(CV)–دالة الدائرة الكهربائية

1- المقدمة:

الانموذج الحاسوبي وهو عبارة عن دالة رياضية تستخدم في العديد من مجالات الحياة كالعلوم والتكنولوجيا والطب وغيرها، كأنموذج بديل للأنظمة الواقعية. غالباً، ما تكون النماذج الحاسوبية معقدة لذا قد تحتاج الى وقت طويل للحصول على قيم مخرجاتها حيث ان الحصول على قيمة واحدة من مخرجات الانموذج الحاسوبي قد يحتاج الى ساعات او حتى ايام. لذلك تم اقتراح بناء انموذج احصائي يسمى انموذج انحدار العملية الكاوسية كبديل عن الانموذج الحاسوبي المعقد. يعتمد بناء انموذج انحدار العملية الكاوسية على مجموعتين من البيانات الاولى (نقاط التدريب) والثانية هي (نقاط التحقق). يستخدم انموذج انحدار العملية الكاوسية للتنبؤ بالمخرجات في المستقبل، ولأن هذا الانموذج احتمالي فإننا نحصل على عدم اليقين في التنبؤات. حيث اننا نكتسب السرعة على حساب عدم اليقين. لقد تم بذل الكثير من الجهد لاستخدام العملية الكاوسية لأنموذج انحدار العملية الكاوسية للنماذج الحاسوبية من قبل [3] ,[2] ,[1] . عادةً، تكون هذه النماذج حتمية اي (الانموذج الحاسوبي في كل مرة يعطي نفس قيم المخرجات عند تنفيذه لنفس قيم المدخلات). وتستخدم العملية الكاوسية والتي تعتبر واحدة من الادوات الاساسية في النمذجة الاحصائية بسبب سهولتها وتقديرها الشامل والتحليلي لعدم اليقين وايضاً توفر اطاراً قوياً لفهم الظواهر المعقدة والتنبؤ بالعمليات المستقبلية لتوليد نتائج وتنبؤات دقيقة. في السنوات الاخيرة تحول بعض الاهتمام الى انموذج انحدار العملية الكاوسية التصادفي للنماذج الحاسوبية التصادفية اي ان الانموذج الحاسوبي التصادفي في كل مرة يعطي مخرجات مختلفة عند تنفيذه لنفس قيم المدخلات [5] ,[4]. يتضمن هذا غالباً، نمذجة كيفية تغير كل من الوسط والتباين في الانموذج الحاسوبي لتغيرات قيم الادخال المختلفة.

بناء انموذج انحدار العملية الكاوسية التصادفي يتطلب بعض الافتراضات والتقديرات [6]. وهذا يلزمنا التحقق من دقة انموذج انحدار العملية الكاوسية التصادفي بمجرد انشائه، حيث يجب ان يمثل انموذج انحدار العملية الكاوسية التصادفي الأنموذج الحاسوبي التصادفي بشكل مناسب اذا كان سيتم استخدامه كأنموذج بديل. تم استخدام انموذج انحدار العملية الكاوسية في مختلف مجالات العلوم، مثلا في الاحصاء المكاني، تم استخدام انموذج العملية الكاوسية في التحليل الديناميكي للأنظمة الهندسية من قبل [7]. تم

استخدام انموذج انحدار العملية الكاوسية في علم المناخ [8] لتطوير انموذج حاسوبي للتنبؤ بحمى الضنك بناء على البيانات المتكاملة بيانات الصحة والمناخ. وفي التعلم الالي تم استخدام انموذج انحدار العملية الكاوسية التصادفي من قبل [9], [10]. تم استخدام الانموذج الحاسوبي التصادفي من قبل [11] من خلال دراسة اجراها على وباء الايبولا في عام 2014، حيث تم تحفيز الجهود المتزايدة في نمذجة الاوبئة المعدية بسبب مشكلة تحدي الايبولا في عام 2015. و في العلوم الفيزيائية استخدم [12] انموذج حاسوبي تصادفي لتصميم تركيز الاوكسجين في طبقة مائية رقيقة حوالي (2000متر) في جنوب المحيط الاطلسي. حيث وصف الانموذج الحاسوبي التصادفي من خلال معادلة انتشار الحمل الحراري. وكذلك تم استخدام انموذج العملية الكاوسية لتحليل البيانات المتسلسلة زمنياً من قبل [13]، حيث ناقش الاطار المفاهيمي للنمذجة البايزية للبيانات المتسلسلة زمنياً وقدم اسس النمذجة البايزية غير المعلمية للعملية الكاوسية. وتم استخدام انموذج انحدار العملية الكاوسية لتحليل مسار الحركة حيث قدم [14] تمثيلاً فعالاً لتحليل ومطابقة مسارات الحركة وقام بإنشاء انموذج مسار كحقل تدفق كثيف مستمر من مجموعة متفرقة من تسلسلات المنهج باستخدام انموذج انحدار العملية الكاوسية، وفي التجارب الحاسوبية تم استخدام انموذج انحدار العملية الكاوسية من قبل [15], [16]، وفي المتغيرات المتعددة تم استخدام نماذج انحدار العملية الكاوسية من قبل [17], [18]. بشكل عام توجد تفاصيل لأنموذج انحدار العملية الكاوسية التصادفي للنماذج الحاسوبية التصادفية من قبل (Baker et al., 2020a).

في هذا البحث تم توضيح كيفية بناء انموذج انحدار العملية الكاوسية التصادفي للنماذج الحاسوبية التصادفية وكيفية استخدام بعض المقاييس للتحقق من دقة هذا الأنموذج باستخدام طريقة المقاطع (CV).

2- انموذج انحدار العملية الكاوسية التصادفي للنماذج الحاسوبية التصادفية:

انموذج انحدار العملية الكاوسية هو عملية عشوائية لها مدخلات ومخرجات فاذا كان لدينا المتغير $\mathbf{x}$ المخرج له $y = \mathcal{Y}(X) + \varepsilon$ حيث ان $X = (\mathrm{x}_1, \dots, \mathrm{x}_p) \in \chi_1 \times \dots \times \chi_P = \chi \subset \Re^p$ عبارة عن متجه يحتوي p من المتغيرات و ε الخطأ العشوائي. اذا كان لدينا n من المدخلات $\mathbf{X} = (X_1, \dots, X_n)$ فالمخرجات لها تكتب على شكل انحدار خطي

$$Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

حيث ان $Y = \left(y_1 = \mathcal{Y}(X_1), \dots, y_n = \mathcal{Y}(X_n)\right)$. المخرجات تتوزع توزيع طبيعي بوسط محدد ودالة تغاير مضاف لها التباين الاضافي (الجوهري) σ^2 [20]. بشكل عام، تمثل دالة الوسط $\mathrm{m}(X)$ الاتجاه العام وتستخدم لإعطاء معلومات سابقة حول شكل العلاقة، اما دالة التغاير $\mathrm{K}(X, X')$ فتحدد النقاط المحلية مما يؤدي الى اعطاء شكل سهل ومفهوم. غالباً، ما يتم اضافة حد الى القطر الرئيسي لمصفوفة التغاير يسمى التباين الاضافي σ^2 الموزع بشكل طبيعي للأنموذج الحاسوبي التصادفي وهي تعني الضوضاء بالنسبة لأنموذج انحدار العملية الكاوسية التصادفي للنماذج الحاسوبية التصادفية. وتعد معلمة التباين الاضافي ضرورية في المشكلات التصادفية.

2-1 بناء انموذج انحدار العملية الكاوسية التصادفي للنماذج الحاسوبية التصادفية

بداية يتم تمثيل التوزيع الاولي حول $\mathcal{Y}(X)$ بواسطة العملية الكاوسية بدالة وسط $\mathrm{m}(X)$ ودالة تغاير $\mathrm{K}(X,X')$ مضاف لها التباين الاضافي باستخدام صيغة هرمية:

$$\mathcal{Y}(X)|\boldsymbol{\beta},\alpha^2,\boldsymbol{\varphi},\sigma^2 \sim GP(\mathrm{m}(X),\mathrm{K}(X,X')+\sigma^2) \qquad (1)$$

إذ ان دالة الوسط $\mathrm{m}(X)$ تعطى بالشكل التالي

$$\mathrm{m}(X)=\mathrm{h}(X)^T\boldsymbol{\beta} \qquad (2)$$

حيث ان $\mathrm{h}(X):\chi\subset\Re^p\longmapsto\Re^q$ هي متجه لدوال مدخلات معلومة، حيث يمكن ان يكون بعد q مختلف عن بعد p وان q تمثل عدد المعلمات مضاف لها الواحد $q=p+1$ فاذا كان لدينا متغير واحد عدد المتغيرات زائد واحد يساوي اثنان يمثل عدد المعلمات وان $\boldsymbol{\beta}$ هي متجه غير معلوم للمعاملات $\boldsymbol{\beta}=(\beta_1,\dots,\beta_p)$. ويتم اختيار الدالة $\mathrm{h}(X)$ حسب شكل الدالة $\mathcal{Y}(X)$. حيث تم اقتراح الصيغة الخطية لدالة المتوسط من قبل [21]، وقام بوصف الاستدلال تحليليا بالنسبة الى $\boldsymbol{\beta}$. بالرغم، من انه يجب اختيار $\mathrm{h}(X)$ حسب شكل الانموذج الحاسوبي $\mathcal{Y}(X)$ الا انه يعد امرا اختياريا، هناك عدة اشكال تمثل $\mathrm{h}(X)$ سنوضح بعض منها، لكل $\mathbf{x}$ $\mathrm{h}(X)=1$، $q=1$، تعتبر ابسط حالة من حالات دوال الانحدار حيث تكون دالة المتوسط في هذه الحالة مساوية الى المعلمة $\boldsymbol{\beta}$ هذا يدل على وجود توقع مسبق لعدم امكانية وجود اتجاه لكيفية استجابة مخرجات التباين للمدخلات، يوجد شكل اخر للدالة $\mathrm{h}(X)$ اخرى وهي ان تكون $\mathrm{h}(X)^T=(1,X^T)$ تكون في هذه الحالة عدد المعلمات $q=p+1$ حيث ان دالة المتوسط تكون في هذه الحالة $\mathrm{m}(X)=\beta_\circ+\beta_1\mathrm{x}_1+\cdots+\beta_p\mathrm{x}_p$ وهذا يدل على وجود توقع مسبق بأن ناتج الانموذج الحاسوبي يمثل اتجاه خطي استجابة لكل متغير من متغيرات الادخال. دالة التغاير تكون بالشكل $\mathrm{K}(X,X')=\alpha^2C(X,X';\boldsymbol{\varphi})$ حيث α^2 معلمة التباين العام غير المعلومة و $C(X,X';\boldsymbol{\varphi})$ هي دالة ارتباط معلومة مع معلمة ارتباط $\boldsymbol{\varphi}$ غير معلومة. في هذا البحث تم اختيار دالة الارتباط الاسية التربيعية والتي يتم تمثيلها بالصيغة التالية:

$$C(X,X';\boldsymbol{\varphi})=exp\left\{\sum_i^p\left(\frac{\mathrm{x}_i-\mathrm{x}_i'}{\varphi_i}\right)^2\right\} \qquad (3)$$

نفرض ان $Y=(y_1=\mathcal{Y}(X_1),\dots,y_n=\mathcal{Y}(X_n))$ مخرجات الانموذج الحاسوبي عند نقاط التصميم $X_1,\dots,X_n$ في مجال الادخال χ، وهذه البيانات تمثل مجموعة بيانات التدريب. طبقا للمعادلة (1) يكون توزيع المخرجات توزيع طبيعي متعدد المتغيرات:

$$Y|\boldsymbol{\beta},\alpha^2,\varphi,\sigma^2 \sim \mathrm{N}_n(\mathrm{H}\boldsymbol{\beta},\alpha^2\mathrm{A}+\sigma^2) \qquad (4)$$

حيث $\mathrm{H}=[\mathrm{h}(X_1),\dots,\mathrm{h}(X_n)]^T$ ، وكذلك A هي مصفوفة عناصرها $\mathrm{A}_{(i,j)}$ حيث:

$$\mathrm{A}_{(i,j)} = \begin{pmatrix} 1 & \mathrm{K}(X_1,X_2) & \cdots & \mathrm{K}(X_1,X_n) \\ \mathrm{K}(X_2,X_1) & 1 & \cdots & \mathrm{K}(X_2,X_n) \\ \vdots & & & \vdots \\ \mathrm{K}(X_n,X_1) & \cdots & \cdots & 1 \end{pmatrix} + \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 1 \end{pmatrix}$$

$(i = 1,2, \dots \quad , j = 1,2, \dots)$ بشرط $i \neq j$. باستخدام التوزيع الطبيعي الشرطي متعدد المتغيرات نحصل على:

$$\mathcal{Y}(X^*)|\boldsymbol{\beta}, \alpha^2, \varphi, \sigma^2, Y \sim \mathrm{N}(\mathcal{M}(X^*), \mathcal{V}(X^*) + \sigma^2) \qquad (5)$$

حيث

$$\mathcal{M}(X^*) = \mathrm{m}(X^*) + \mathrm{K}(X^*,X)(\mathrm{K}(X,X) + \sigma^2\mathrm{I})^{-1}\big(Y - \mathrm{m}(X)\big)$$
$$\mathcal{V}(X^*) = \mathrm{K}(X^*,X^*) + \sigma^2\mathrm{I} - \mathrm{K}(X^*,X)(\mathrm{K}(X,X) + \sigma^2\mathrm{I})^{-1}\mathrm{K}(X,X^*)$$

حيث ان $\mathrm{K}(X^*,X)$ تمثل مصفوفة تغاير بين القيم المتنبأ بها مع قيم التدريب، $\mathrm{K}(X,X)$ تمثل مصفوفة تغاير بين قيم التدريب مع نفسها، $\mathrm{K}(X^*,X^*)$ مصفوفة تغاير بين قيم التحقيق مع نفسها، $\mathrm{m}(X)$ هو توقع الدالة من X ، $\mathrm{m}(X^*)$ هو توقع الدالة من X^* ، I هي المصفوفة المحايدة.

كذلك يتم توزيع ونمذجة التباين الاضافي بواسطة العملية الكاوسية فيتم نمذجة لوغارتم التباين الاضافي كعملية كاوس بدلاً من التباين الاضافي نفسه. هذا ما قام به الكثيرون [22], [23], [24], [25] حيث تم نمذجة لوغارتم التباين الاضافي بدالة وسط $\mathrm{m}_\sigma(X)$، ودالة تغاير $\mathrm{K}_\sigma(X,X')$ باستخدام الصيغة التالية:

$$\log\big(\sigma^2(X)\big)\big|\boldsymbol{\beta}, \alpha^2, \varphi \sim GP\big(\mathrm{m}_\sigma(X), \mathrm{K}_\sigma(X,X')\big) \qquad (6)$$

إذ ان دالة الوسط للتباين الاضافي تعطى بالصيغة التالية $\mathrm{m}_\sigma(X) = \mathrm{h}_\sigma(X)^T\boldsymbol{\beta}_\sigma$ ، كما ان دالة التغاير تعطى بالشكل $\mathrm{K}_\sigma(X,X') = \alpha_\sigma^2 C_\sigma(X,X';\boldsymbol{\varphi})$، نفرض ان

$$\log\big(\sigma^2(X)\big) = \Big(\log\big(\sigma^2(\mathrm{x})\big)_1 = \log\big(\sigma^2(X_1)\big), \dots, \log\big(\sigma^2(\mathrm{x})\big)_n = \log\big(\sigma^2(X_n)\big)\Big)$$

نفرض ان مخرجات الانموذج الحاسوبي بالنسبة الى لوغارتم التباين الاضافي عند نقاط التصميم $X_1, \dots, X_n$ في مجال الادخال χ. طبقا للمعادلة (6) يكون توزيع المخرجات توزيع طبيعي متعدد المتغيرات

$$\log\big(\sigma^2(X)\big)\big|\boldsymbol{\beta}, \alpha^2, \boldsymbol{\varphi} \sim GP(\mathrm{H}_\sigma\boldsymbol{\beta}_\sigma, {\alpha^2}_\sigma\mathrm{A}_\sigma) \qquad (7)$$

$\mathrm{H}_\sigma = [\mathrm{h}_\sigma(X_1), \dots, \mathrm{h}_\sigma(X_n)]^T$، وكذلك A_σ هي مصفوفة عناصرها $\mathrm{A}_{\sigma(i,j)}$ حيث $\mathrm{A}_{\sigma(i,j)} = C_\sigma\big(X_i, X_j; \boldsymbol{\varphi}\big)$، $(i = 1,2, \dots \quad , j = 1,2, \dots)$ بشرط $i \neq j$

$$\mathrm{A}_{\sigma(i,j)} = \begin{pmatrix} 1 & \mathrm{K}_\sigma(X_1,X_2) & \cdots & \mathrm{K}_\sigma(X_1,X_n) \\ \mathrm{K}_\sigma(X_2,X_1) & 1 & \cdots & \mathrm{K}_\sigma(X_2,X_n) \\ \vdots & & & \vdots \\ \mathrm{K}_\sigma(X_n,X_1) & \cdots & \cdots & 1 \end{pmatrix}$$

باستخدام التوزيع الطبيعي الشرطي متعدد المتغيرات نحصل على

$$\log\big(\sigma^2(X^*)\big)\big|\boldsymbol{\beta}, \alpha^2, \varphi, \log\big(\sigma^2(X)\big) \sim \mathrm{N}\big(\mathcal{M}_\sigma(X^*), \mathcal{V}_\sigma(X^*)\big) \qquad (8)$$
$$\mathcal{M}_\sigma(X^*) = \mathrm{m}_\sigma(X^*) + \mathrm{K}_\sigma(X^*,X)\mathrm{K}_\sigma(X,X)^{-1}\Big(\log\big(\sigma^2(X)\big) - \mathrm{m}_\sigma(X)\Big),$$

$$\mathcal{V}_\sigma(X^*) = \mathrm{K}_\sigma(X^*, X^*) - \mathrm{K}_\sigma(X^*, X)\mathrm{K}_\sigma(X, X)^{-1}\mathrm{K}_\sigma(X, X^*).$$

بالنسبة للمعلمات $\boldsymbol{\beta}, \alpha, \boldsymbol{\varphi}$ يتم استخدام طرق تقدير مثل طريقة الامكان الاعظم (ML) وللمزيد من المعلومات [22], [23], [24], [25] وايضا [5]

2-2 طريقة المقاطع

طريقة المقاطع Cross-Validation (CV) Method، هي: طريقة تستخدم؛ للحصول على تنبؤات للقيم الحقيقية للأنموذج الحاسوبي العشوائي. ففي البداية يتم بناء انموذج انحدار العملية الكاوسية العشوائي، بعد ذلك يتم حذف المشاهدة y_i، ثم الحصول على التوزيع اللاحق للقيمة y_i بالاعتماد على بقية القيم Y_{-i}، وإنّ Y_{-i} تمثل نقاط مخرجات الانموذج الحاسوبي العشوائي التي تم تنفيذها على نقاط التدريب ما عدا المشاهدة التي ترتيبها i. يتم تنفيذ طريقة المقاطع CV مع مدخلات التدريب لاشتقاق التوزيع اللاحق لكل قيمة للمخرجات $y_i = \mathcal{Y}(X_i) + \epsilon_i$ بالاعتماد على نقاط التدريب الاخرى Y_{-i}.

3- توليد نقاط التصميم

نستخدم في هذا البحث طريقة شائعة الاستخدام في الوقت الحاضر وهي طريقة المكعب اللاتيني (Maximin Latin Hypercube Design (Maximin LHD)) وتعني تعظيم القيمة الصغرى للمسافة بين النقاط وهي احد تصاميم ملىء الفراغ (Space Filling)، تعمل على تجزئة النقاط وتوزيعها بشكل متساوي حتى تضمن تغطية مساحة العمل بالكامل وايضاً تعمل على تقليل التباين، حيث يتم اعطاء Maximin LHD بالصيغة التالية:

$$\mathbf{X} = \max_{\mathbf{X}\subset\chi} \min_{\{X,X'\}\in\mathbf{X}} dis(X_i, X_j) \qquad (9)$$

لذلك يضمن استخدام طريقة (Maximin LHD) ان نقاط التصميم منتشرة بشكل متساوي وتغطي مجال الادخال بالكامل [26].

4- مقاييس التحقق من دقة انموذج انحدار العملية الكاوسية التصادفي

في هذا القسم سنوضح بعض طرق التحقق والمقاييس للتحقق من دقة انموذج انحدار العملية الكاوسية التصادفي للمقارنة بين مخرجات الانموذج الحاسوبي التصادفية ومخرجات انموذج انحدار العملية الكاوسية التصادفي، وتنقسم المقاييس الى مقاييس التحقق الحتمية و المقاييس التصادفية. المقاييس الحتمية تم توضيحها من قبل [6] و [27] سنركز في هذا البحث على المقاييس التصادفية.

4-1 المقاييس التصادفية

في هذا القسم استخدمنا بعض المقاييس للتحقق من دقة الوسط والتباين والتي تم توضيحها من قبل [28].

4-2 تقييم المتوسط

في هذا القسم تم توضيح كيفية التحقق من الوسط. ولكن بالنسبة لعينة المتوسطات حيث تم حساب قيم عدم التوقع (unexpectedness) U [28] على النحو التالي:

$$U = 2\left(0.5 - p\left(Z \le \hat{Z}\right)\right) \qquad (10)$$

حيث ان Z كمية عشوائية و$\hat{Z}$ هي مشاهدة لتلك الكمية العشوائية. حيث ان $p(Z \leq \hat{Z})$ يجب ان تكون بين $[0,1]$ و يجب ان تكون قيم U محصورة بين الفترة $[-1,1]$. القيمة الموجبة الكبيرة لـ U تتوافق مع قيمة صغيرة بشكل غير متوقع للمشاهدة. وبالتالي من المحتمل ان يبالغ انموذج انحدار العملية الكاوسية التصادفي في تقدير Z والقيمة السالبة تعني قيمة كبيرة بشكل غير متوقع لـ Z ومن المحتمل ان يقلل انموذج انحدار العملية الكاوسية التصادفي من Z.

4-3 تقييم التباين

استخدمنا النوع الثاني من عمليات التحقق من انموذج انحدار العملية الكاوسية التصادفي للنماذج الحاسوبية العشوائية التصادفية وهو تقييم التباين وكما هو الحال في تقييم المتوسط يمكن حساب Variance Unexpectedness من المعادلة (10) حيث يجب ان تتراوح قيمة Variance Unexpectedness بين $[-1,1]$ كما في حالة تقييم الوسط.

5- مثال دالة الدائرة الكهربائية

في هذا القسم، سنأخذ مثال حقيقي على الأنموذج الحاسوبي التصادفي متمثل بدالة الدائرة الكهربائية Output Transformer Less Circuit Function حيث Output Transformer Less (OTL) ذات الابعاد الستة. حيث تعمل دالة الدائرة الكهربائية على تصميم انموذج دائرة سحب ودفع دون محول اخراج. تم استخدام دالة الدائرة الكهربائية OTL Circuit Function من قبل [29] كمثال لنمذجة البيانات في التجارب الحاسوبية حيث تم عمل مقارنة تجريبية بين انموذج Kriging مع انحدار متتابعة الاسقاط ، كما استخدمها [30] لتصميم وتحليل التجارب الحاسوبية لفحص متغيرات الادخال وايضا تم استخدامها من قبل [31] لفحص المجموعة المستندة الى الحساسية على مرحلتين في التجارب الحاسوبية.

وتعطى دالة الدائرة الكهربائية بالشكل التالي:

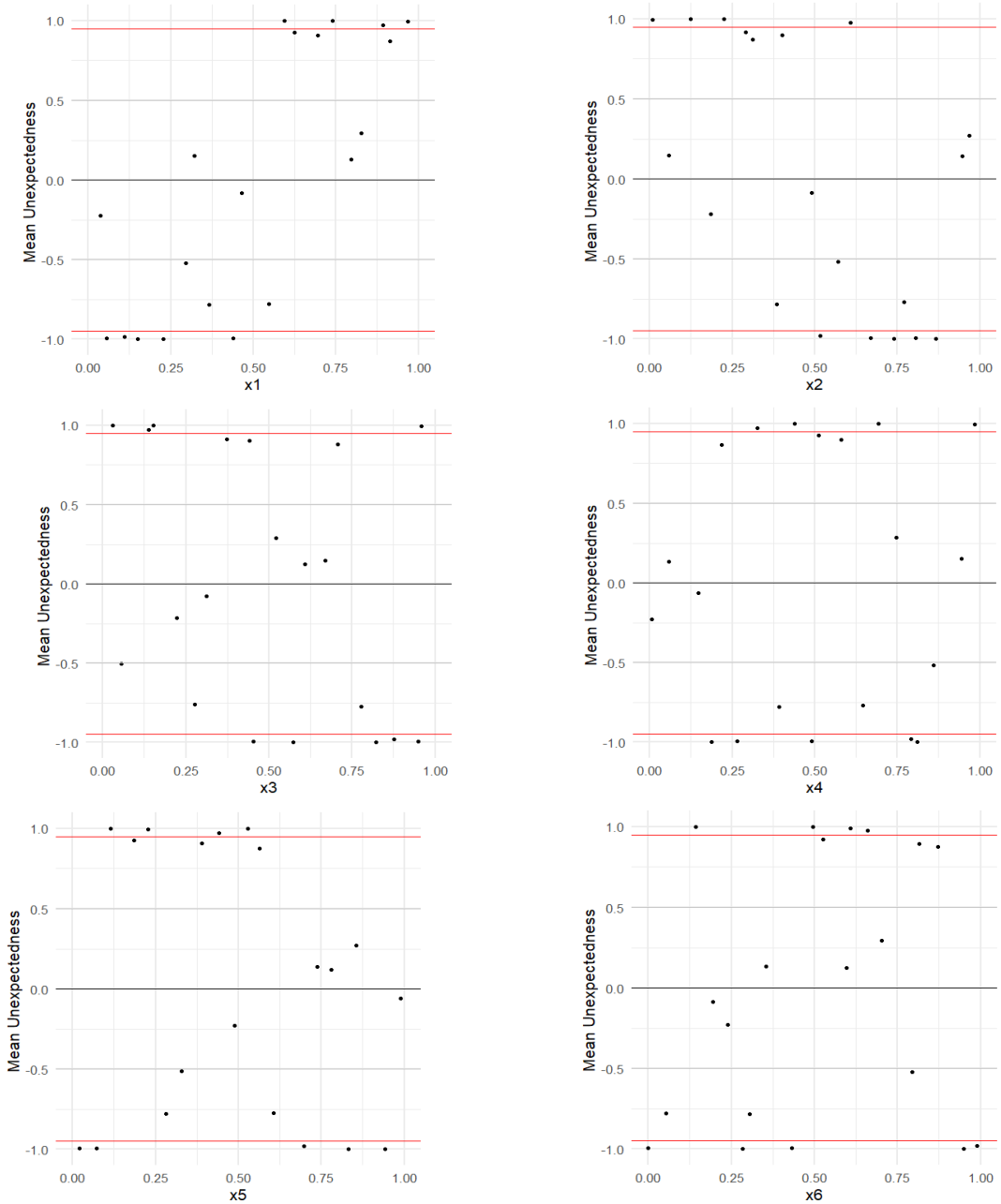
$$V_m(X) = \frac{(V_{b1} + 0.74)\gamma(R_{c2} + 9)}{\gamma(R_{c2} + 9) + R_f} + \frac{11035R_f}{\gamma(R_{c2} + 9)} + \frac{0.74R_f\gamma(R_{c2} + 9)}{\big(\gamma(R_{c2} + 9) + R_f\big)R_{c1}},$$

حيث $V_{b1} = \frac{12R_{b2}}{R_{b1}+R_{b2}}$ ، تحتوي دالة الدائرة على ست متغيرات ادخال وهي $X = \big(R_{b1}, R_{b2}, R_f, R_{c1}, R_{c2}, \gamma\big)$، متغير الاستجابة V_m يمثل جهد نقطة المنتصف، متغيرات الادخال ومجال الادخال لها تعطى كالاتي: $R_{b1} \in [50,150]$ يمثل المقاومة $b1$ ، $R_{b2} \in [25,70]$ يمثل المقاومة $b2$، $R_f \in [0.5,3]$ يمثل المقاومة f ، $R_{c1} \in [1.2,2.5]$ يمثل المقاومة $c1$، $R_{c2} \in [0.25,1.2]$ يمثل المقاومة $c2$، $\gamma \in [50,300]$ يمثل مكسب التيار، وتقاس كل متغيرات الدالة بوحدة (K-Ohms) ما عدا γ يقاس بوحدة (Amperes) .

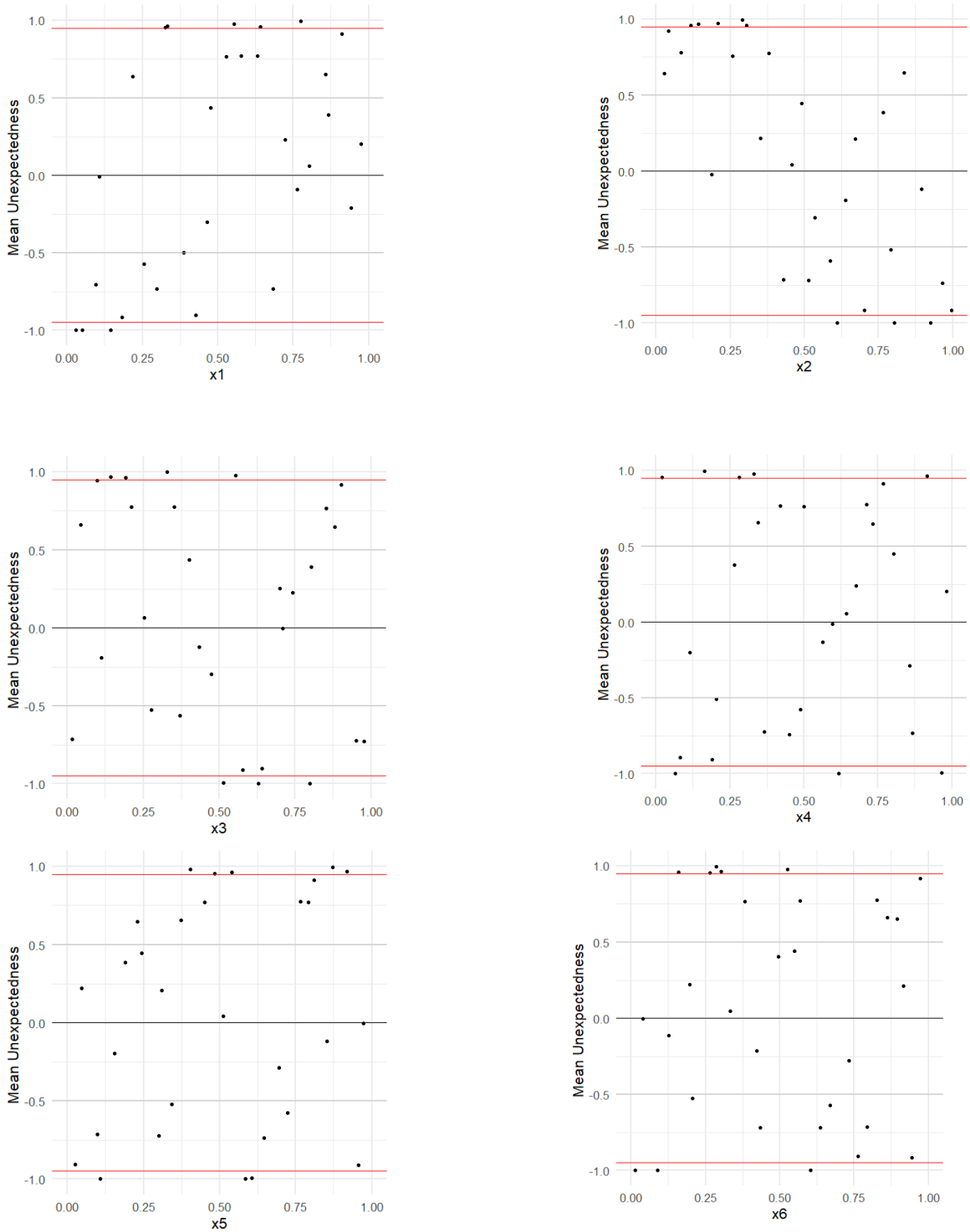
الأن، تم بناء انموذج انحدار العملية الكاوسية التصادفي حيث تم استخدام دالة الوسط الخطية (2) واستخدام دالة الارتباط الاسية التربيعية (3) حيث ان دالة التغاير كانت $\mathrm{K}(X, X') = \alpha^2 C(X, X'; \boldsymbol{\varphi})$ وتم تقدير المعلمات باستخدام دالة الامكان الاعظم (ML) واستخدمنا طريقة المقاطع (CV) للتنبؤ بمخرجات

الانموذج الحاسوبي التصادفي حيث يتم تنفيذ طريقة المقاطع (CV) مع مدخلات التدريب للحصول على التوزيع اللاحق لكل قيمة من المخرجات $y_i = \mathcal{Y}(X_i) + \varepsilon_i$ بالاعتماد على النقاط الاخرى y_{-i}. ثم بعد ذلك نقوم بالتحقق من دقة اداء انموذج انحدار العملية الكاوسية التصادفي للنماذج الحاسوبية التصادفية باستخدام بعض المقاييس التصادفية. وباستخدام طريقة (Maximin LHD) تم توليد نقاط التصميم.

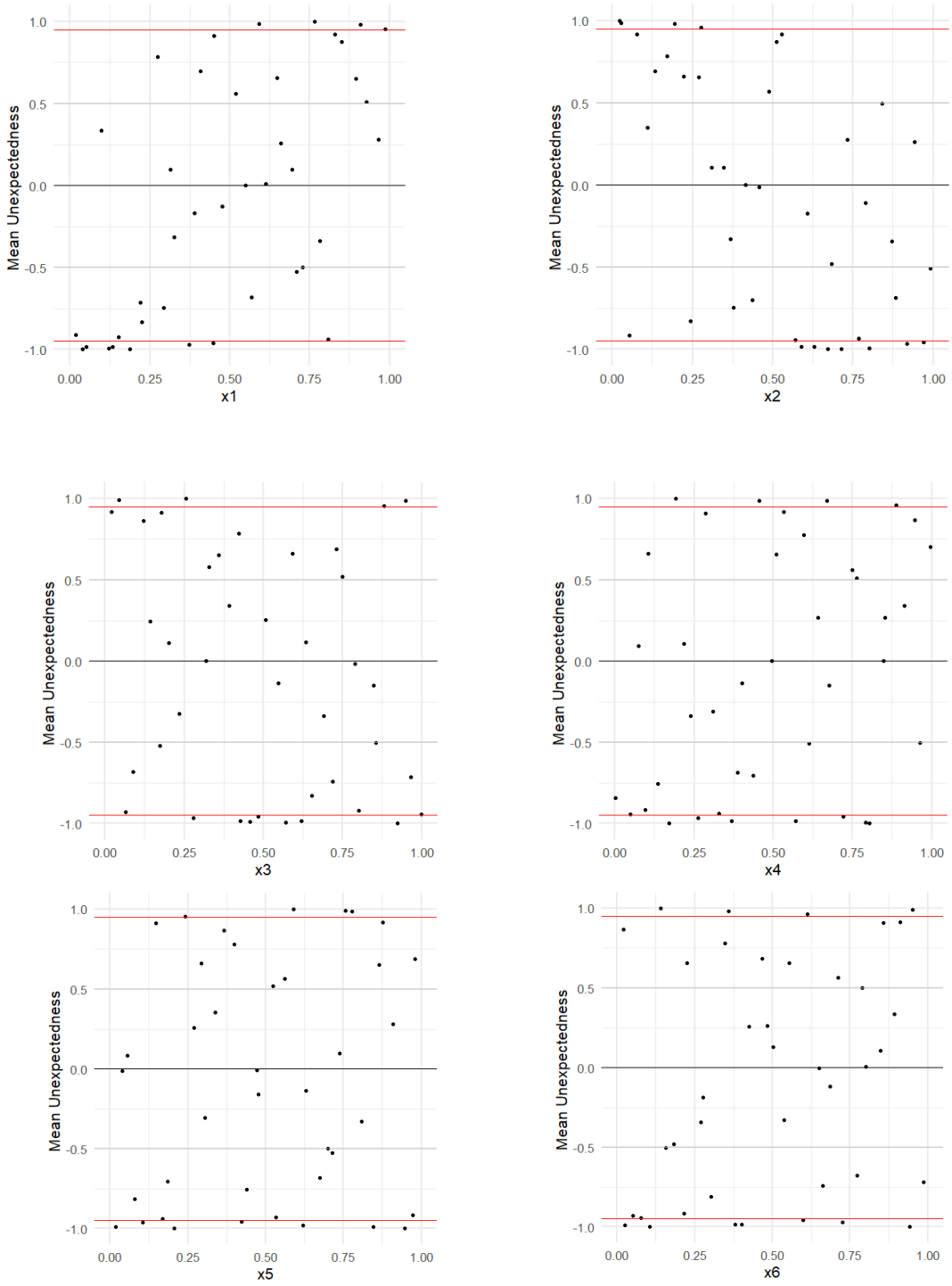
بداية نقوم بتنفيذ الانموذج الحاســـوبي مع المدخلات 60,40,30,20 نقطة من نقاط التدريب التي تم توليدها بواسـطة Maximin LHD. الان نقوم بتطبيق المقاييس التصــادفية للتحقق من انموذج انحدار العملية الكاوسـية التصــادفي للنماذج التصــادفية حيث توضــح الاشـكال من 1−4 عدم التوقع للوسـط Mean Unexpectedness لأنموذج انحدار العملية الكاوسية التصادفي.



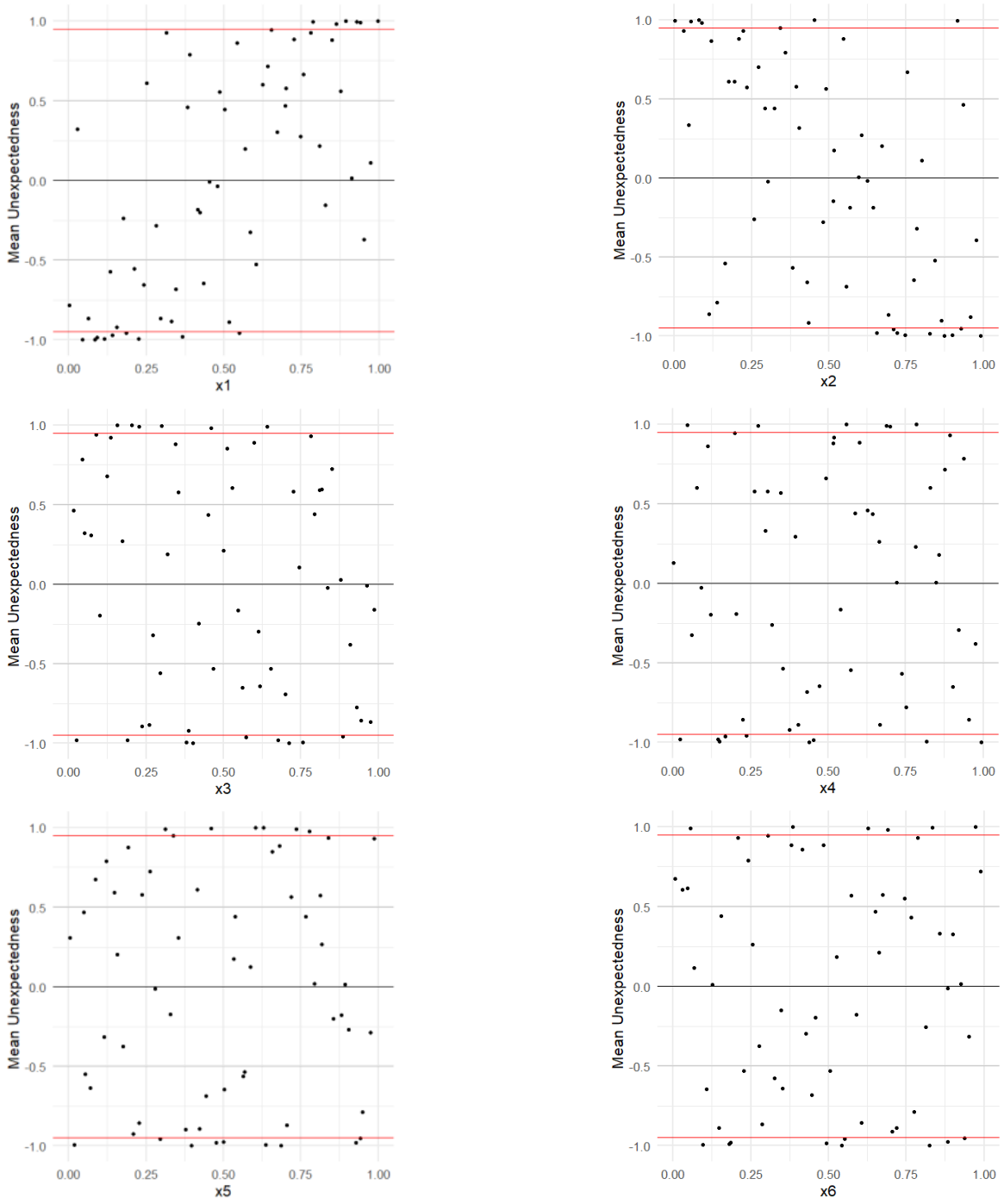
الشكل(1) يوضح عدم التوقع للوسط Mean Unexpectedness عند نقاط التصميم 20



الشكل (2) يوضح عدم التوقع للوسط Mean Unexpectedness عند نقاط التصميم 30



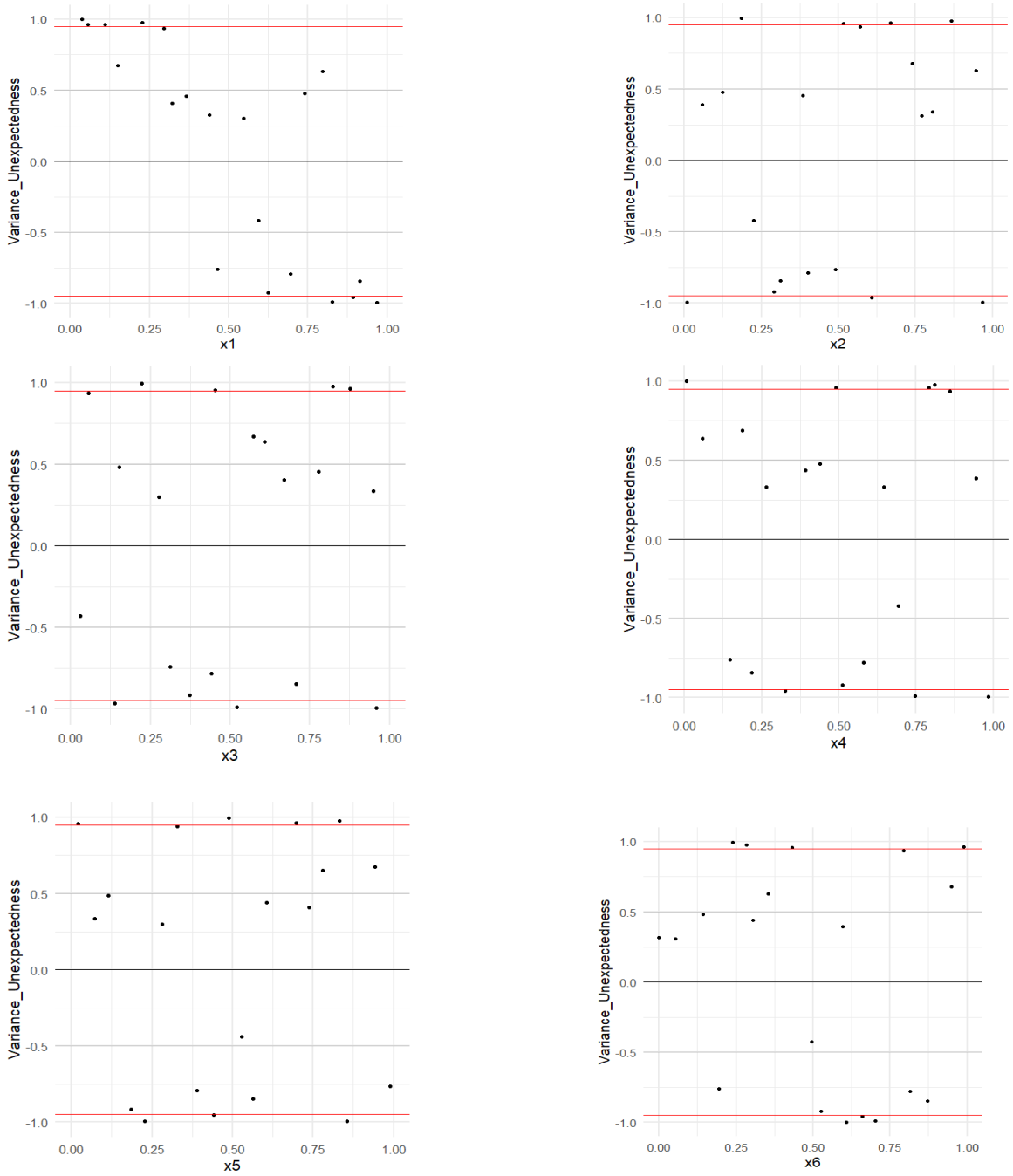
الشكل (3) يوضح عدم التوقع للوسط Mean Unexpectedness عند نقاط التصميم 40



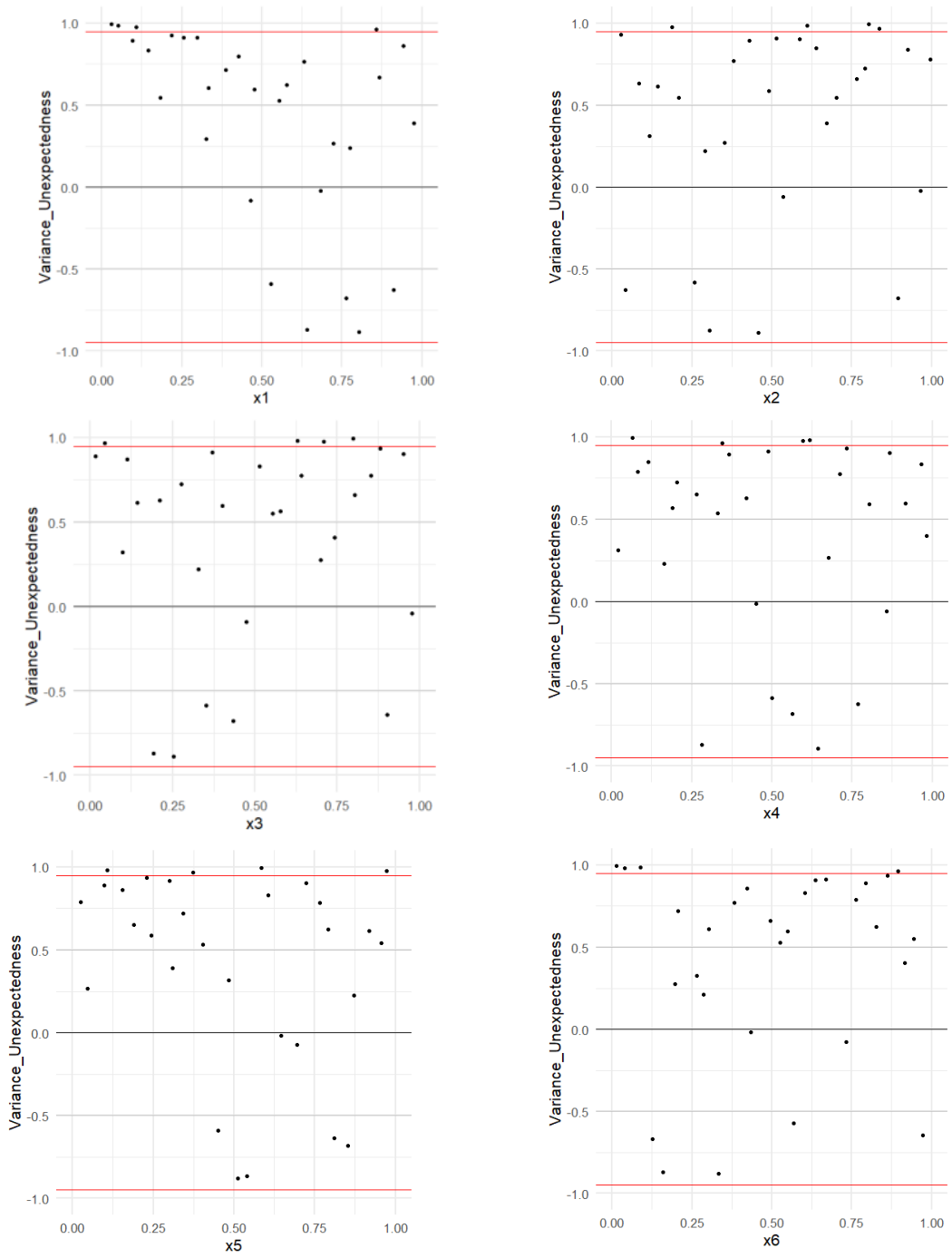
الشكل (4) يوضح عدم التوقع للوسط Mean Unexpectedness عند نقاط التصميم 60

نلاحظ من الاشكال من 1-4 ان جميع نقاط التدريب تقع داخل الفترة [1,1−] وان جميع النقاط منتشرة في جميع الانحاء الا ان هناك بعض النقاط التي تقع قريبة من حدود الفترة [1,1−] نقول عنها متطرفة حيث عندما استخدمنا النقاط 20 كانت النقاط عند الاطراف اكثر وعندما قمنا بزيادة النقاط الى 30، 40، 60 تحسن اداء انموذج انحدار العملية الكاوسية التصادفي بالرغم من وجود بعض النقاط القريبة من حدود الفترة الا ان اداء الانموذج كان جيداً.

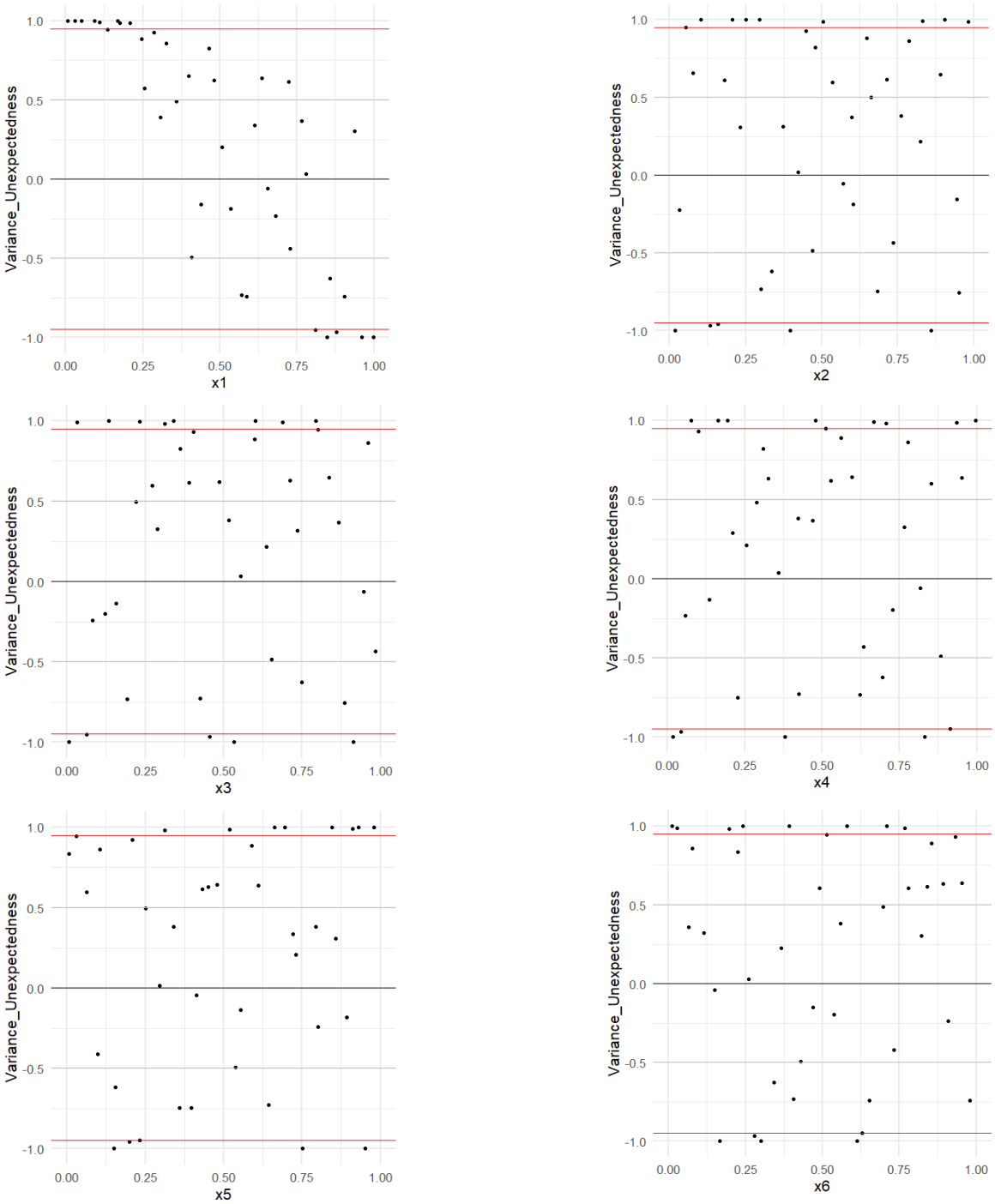
الان نقوم بالتحقق من دقة التباين حيث توضح الاشكال من 5-8 عدم التوقع للتباين Variance Unexpectedness عند نقاط التدريب 60,40,30,20.



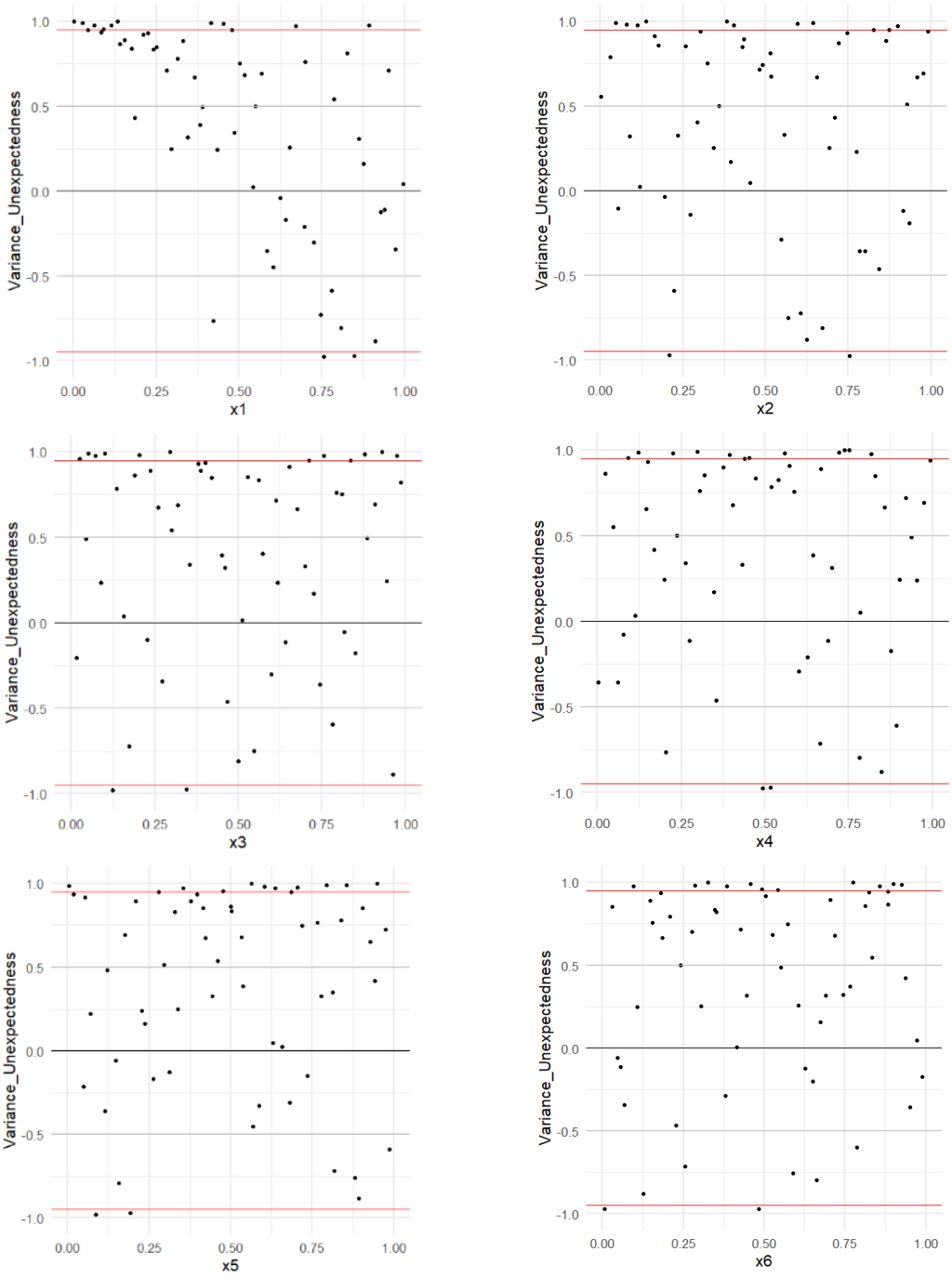
الشكل (5) يوضح عدم التوقع للتباين Variance Unexpectedness عند نقاط التصميم 20



الشكل (6) يوضح عدم التوقع للتباين Variance Unexpectedness عند نقاط التصميم 30



الشكل (7) يوضح عدم التوقع للتباين Variance Unexpectedness عند نقاط التصميم 40



الشكل (8) يوضح عدم التوقع للتباينVariance Unexpectedness عند نقاط التصميم 60

لاحظنا من خلال الاشكال من 5-8 ايضا ان النقاط جميعها تقع داخل الفترة [−1,1] وهناك بعض النقاط الواقعة قريبة من حدود الفترة [−1,1]. وايضا عندما قمنا باستخدام النقاط 20 كانت النقاط عند الاطراف اكثر وبزيادة عدد النقاط الى 30، 40، 60 تحسن اداء انموذج انحدار العملية الكاوسية التصادفي نستنتج من ذلك ان طريقة المقاطع (CV) للتنبؤ بمخرجات الانموذج الحاسوبي التصادفي جيدة بالرغم من وجود بعض النقاط المتطرفة الا انها تصلح للتنبؤ خاصة اذا كان لدينا عدد قليل من نقاط التدريب.

6- الاستنتاجات

في هذا البحث تم استخدام انموذج انحدار العملية الكاوسية التصادفي كبديل عن الانموذج الحاسوبي التصادفي المعقد وتم التحقق من دقة انموذج انحدار العملية الكاوسية التصادفي باستخدام مثال حقيقي متمثل بدالة الدائرة الكهربائية حيث قمنا باختبار اداء الانموذج الحاسوبي من خلال استخدام طريقة المقاطع (CV) والتحقق باستخدام المقاييس التصادفية وهي التحقق من الوسط والتباين واستنتجنا بان طريقة المقاطع (CV) كانت جيدة للتنبؤ بمخرجات الانموذج الحاسوبي التصادفي واستنتجنا ايضاً بان طريقة المقاطع (CV) تصلح لا تكون بديلة عن الطرق الاخرى خاصة عندما تكون النقاط قليلة. ومن خلال هذا تم التوصل الى ان انموذج انحدار العملية الكاوسية التصادفي يصلح لان يكون بديل عن الانموذج الحاسوبي التصادفي المعقد.

7- شكر وتقدير

الباحثون يتقدمون بالشكر والامتنان الى قسم الرياضيات في كلية التربية للعلوم الصرفة/جامعة الموصل. على الجهود المبذولة والتسهيلات المقدمة والتي ساعدت على تحسين واكمال جودة هذا البحث

المصادر

[1] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.

[2] J. Oakley and A. O'hagan, "Bayesian inference for the uncertainty distribution of computer model outputs," *Biometrika*, vol. 89, no. 4, pp. 769–784, 2002.

[3] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical science*, vol. 4, no. 4, pp. 409–423, 1989.

[4] B. Ankenman, B. L. Nelson, and J. Staum, "Stochastic Kriging for Simulation Metamodeling," *Operations Research*, vol. 58, no. 2, pp. 371–382, Apr. 2010, doi: 10.1287/opre.1090.0754.

[5] A. Boukouvalas, "Emulation of random output simulators," PhD Thesis, Aston University, 2010. Accessed: Nov. 04, 2024. [Online]. Available: http://publications.aston.ac.uk/id/eprint/39333/

[6] L. S. Bastos and A. O'Hagan, "Diagnostics for Gaussian process emulators," *Technometrics*, vol. 51, no. 4, pp. 425–438, 2009.

[7] F. DiazDelaO and S. Adhikari, "Structural dynamic analysis using Gaussian process emulators," *Engineering Computations*, vol. 27, no. 5, pp. 580–605, 2010.

[8] G. Manogaran and D. Lopez, "A Gaussian process based big data processing framework in cluster computing environment," *Cluster Computing*, vol. 21, pp. 189–204, 2018.

[9] M. Seeger, "Gaussian processes for machine learning," *International journal of neural systems*, vol. 14, no. 02, pp. 69–106, 2004.
[10] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.
[11] A. Fadikar, D. Higdon, J. Chen, B. Lewis, S. Venkatramanan, and M. Marathe, "Calibrating a stochastic, agent-based model using quantile-based emulation," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 6, no. 4, pp. 1685–1706, 2018.
[12] R. Herbei and L. M. Berliner, "Estimating ocean circulation: an MCMC approach with approximated likelihoods via the Bernoulli factory," *Journal of the American Statistical Association*, vol. 109, no. 507, pp. 944–954, 2014.
[13] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain, "Gaussian processes for time-series modelling," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, p. 20110550, 2013.
[14] K. Kim, D. Lee, and I. Essa, "Gaussian process regression flow for analysis of motion trajectories," in *2011 International Conference on Computer Vision*, IEEE, 2011, pp. 1164–1171.
[15] R. W. Al-Naser and Y. Al-Taweel, "Using Kriging models for approximating computer models and quantifying their uncertainty," in *AIP Conference Proceedings*, AIP Publishing, 2023.
[16] Y. H. Al-Taweel and N. Sadeek, "A comparison of different methods for building Bayesian kriging models," *Pakistan Journal of Statistics and Operation Research*, pp. 73–82, 2020.
[17] Y. Al-Taweel, "Uncertainty quantification of multivariate Gaussian process regression for approximating multivariate computer codes," *TWMS Journal of Applied and Engineering Mathematics*, 2024.
[18] J. P. Kleijnen and E. Mehdad, "Multivariate versus univariate Kriging metamodels for multi-response simulation models," *European Journal of Operational Research*, vol. 236, no. 2, pp. 573–582, 2014.
[19] S. R. Baker, N. Bloom, S. J. Davis, K. Kost, M. Sammon, and T. Viratyosin, "The unprecedented stock market reaction to COVID-19," *The review of asset pricing studies*, vol. 10, no. 4, pp. 742–758, 2020.
[20] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.
[21] A. O'Hagan, "Curve fitting and optimal design for prediction," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 40, no. 1, pp. 1–24, 1978.
[22] A. Binois, A. Bridault, G. Pion, and T. Ducrocq, "Dental development pathology in wild artiodactyls: Two prehistoric case studies from France," *International Journal of Paleopathology*, vol. 4, pp. 53–58, 2014.

[23] A. Boukouvalas, D. Cornford, and M. Stehlík, "Optimal design for correlated processes with input-dependent noise," *Computational Statistics & Data Analysis*, vol. 71, pp. 1088–1102, 2014.
[24] D. P. Goldberg *et al.*, "The validity of two versions of the GHQ in the WHO study of mental illness in general health care," *Psychological medicine*, vol. 27, no. 1, pp. 191–197, 1997.
[25] A. Kersting *et al.*, "Complicated grief after traumatic loss: A 14-month follow up study," *Eur Arch Psychiatry Clin Neurosc*, vol. 257, no. 8, pp. 437–443, Dec. 2007, doi: 10.1007/s00406-007-0743-1.
[26] M. D. Morris and T. J. Mitchell, "Exploratory designs for computational experiments," *Journal of statistical planning and inference*, vol. 43, no. 3, pp. 381–402, 1995.
[27] Y. Al-Taweel, "Diagnostics and Simulation-Based Methods for Validating Gaussian Process Emulators," PhD Thesis, University of Sheffield, 2018.
[28] E. Baker, P. Challenor, and M. Eames, "Diagnostics for stochastic Gaussian process emulators," *arXiv preprint arXiv:1902.01289*, 2019.
[29] E. N. Ben-Ari and D. M. Steinberg, "Modeling data from computer experiments: an empirical comparison of kriging with MARS and projection pursuit regression," *Quality Engineering*, vol. 19, no. 4, pp. 327–338, 2007.
[30] H. Moon, "Design and analysis of computer experiments for screening input variables," PhD Thesis, The Ohio State University, 2010.
[31] H. Moon, A. M. Dean, and T. J. Santner, "Two-stage sensitivity-based group screening in computer experiments," *Technometrics*, vol. 54, no. 4, pp. 376–387, 2012.