

电子商务视角下基于局部多项式回归的信息物理系统可靠性模型

周文倩^{1,2}, 杨剑锋^{2,3}, 黄嘉悦^{1,2}

¹贵州大学数学与统计学院, 贵州 贵阳

²贵大·贵安科创超级计算算力算法应用实验室, 贵州 贵阳

³南宁师范大学数学与统计学院, 广西 南宁

收稿日期: 2024年8月30日; 录用日期: 2024年9月24日; 发布日期: 2024年11月28日

摘要

随着互联网的高度发展,信息物理系统的可靠性对于电子商务至关重要,因为这些系统支撑着在线交易、数据管理和用户体验。在电子商务视角下,本文提出了一种基于局部多项式回归的信息物理系统可靠性模型,并将其与两种传统的信息物理系统可靠性模型进行比较。此外,对模型中的参数估计采用了最小二乘估计法。最后,基于超级计算机BlueGene/L的系统的真实失效数据,利用jupyter notebook对这3类信息物理系统可靠性模型的性能进行对比分析。通过实验验证表明,局部多项式回归模型的拟合效果与预测能力更好。本文结果为电子商务的可持续发展提供了有力保障。

关键词

电子商务, 局部多项式回归, 信息物理系统, 可靠性模型

Reliability Model of Cyber Physical System Based on Local Polynomial Regression from the Perspective of E-Commerce

Wenqian Zhou^{1,2}, Jianfeng Yang^{2,3}, Jiayue Huang^{1,2}

¹School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

²Guizhou University and Guian Kechuang Supercomputing Power Algorithm Laboratory, Guiyang Guizhou

³School of Mathematics and Statistics, Nanning Normal University, Nanning Guangxi

Received: Aug. 30th, 2024; accepted: Sep. 24th, 2024; published: Nov. 28th, 2024

文章引用: 周文倩, 杨剑锋, 黄嘉悦. 电子商务视角下基于局部多项式回归的信息物理系统可靠性模型[J]. 电子商务评论, 2024, 13(4): 5587-5594. DOI: 10.12677/ec.2024.1341796

Abstract

With the high development of the Internet, the reliability of the cyber physical system is critical for e-commerce, as these systems underpin online transactions, data management and user experience. Under the perspective of e-commerce, this paper presents a reliability model of cyber physical system based on local polynomial regression, and compares it with two traditional cyber physical system reliability models. In addition, the least squares estimation method is used to estimate the parameters in the model. Finally, based on the real system failure data of supercomputer BlueGene/L, jupyter notebook is used to compare and analyze the performance of these three types of reliability models of cyber physical system. The experimental results show that the local polynomial regression model has better fitting effect and prediction ability. The results of this paper provide a strong guarantee for the sustainable development of e-commerce.

Keywords

E-Commerce, Local Polynomial Regression, Cyber Physical System, Reliability Model

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着互联网技术的飞速发展，电子商务作为一种新型的商务模式，在全球范围内得到了广泛的应用与普及。电子商务通过网络通信技术实现物品交换和资金转账，极大地改变了传统的商业模式和消费习惯，成为推动全球经济发展的重要力量。在中国，电子商务经历了从无到有、从小到大的发展历程，逐渐形成了庞大的市场规模和完善的产业链。据调查，我国电子商务用户数量持续增长，交易规模逐年扩大，电子商务已成为推动中国经济转型升级的重要引擎。鉴于电子商务在全球经济中的重要地位，深入研究电子商务的系统可靠性具有重要的现实意义和理论价值。

在信息技术与物理世界深度融合的今天，信息物理系统(cyber physical system, CPS)作为连接两者的重要桥梁，在电子商务领域发挥着日益重要的作用。电子商务的顺畅运行不仅依赖于网络通信技术，还高度依赖于 CPS 的可靠性，这一特性直接关系到电子商务的交易效率、客户体验以及整体竞争力。通过确保系统的稳定性、数据准确性和安全性，可以降低交易失败、数据丢失和安全漏洞的风险，从而提升电子商务平台的整体性能和用户信任。因此，在电子商务的发展过程中，应高度重视 CPS 的可靠性建设，不断提升系统的稳定性和安全性，为电子商务的可持续发展提供有力保障。

CPS 已成为现代科技和工业应用的核心，作为智能系统，CPS 将通信、控制和计算深度融合在一起 [1]-[3]。然而，CPS 的复杂性和动态特性使其面临着极高的可靠性要求。研究 CPS 的可靠性变得尤为重要，因为系统的稳定性和安全性直接影响到其性能、用户安全以及经济效益。确保 CPS 在各种运行条件下的可靠性，不仅能够提升系统的整体效率，还能预防潜在的故障和风险，保障相关应用的正常运行。CPS 的可靠性研究不仅涵盖了对系统组件和网络的全面评估，还包括对系统在极端条件下表现的预测和分析。这项研究可以帮助识别潜在的薄弱环节，制定有效的故障预防和修复策略，减少系统停机时间，提高生产效率。此外，CPS 对异构工程系统的关键基础设施具有全新的计算、通信和控制能力，系统的可靠性直接关联到公共安全和社会稳定。因此，深入研究 CPS 的可靠性不仅具有技术层面的重要性，更

具有社会和经济层面的广泛意义。

现有大量文献主要围绕 CPS 的网络攻击和防御方法,然而,关注 CPS 可靠性建模的文献很少。从具有不确定性攻击的角度出发,Wang [4]等通过引入 IT2 T-S 模糊模型,研究了基于观测器的具有不确定性和混合攻击的多通道信息物理系统的模型预测控制问题;Ge [5]等对 CPS 的安全问题进行研究,通过对 DoS、信息泄露、隐身攻击和重放攻击等多源网络攻击过程的分析,建立了具有不确定性的统一系统模型。从 CPS 防御策略的角度出发,Lu [6]等为了量化攻击损失并选择最优策略防御网络攻击,提出了一种基于相互依存网络理论的综合信息物理损失计算模型;Bo [7]等针对虚假数据注入攻击的安全威胁,从多时空视角对电力 CPS 干扰的主动防御方法进行了总结和分析。从检测 CPS 中不同攻击的角度出发,Li [8]等考虑了 CPS 中远程状态估计的安全问题,提出一种具有不同触发阈值的随机检测机制;Zhang [9]等采用一种移动窗口子空间辨识方法,构造了一种线性离散时变 CPS 模型;Ahmed [10]等提出了一种实用的物理水印方法来检测 CPS 中的重放攻击。

为了进一步建模和定量评估对 CPS 至关重要的可靠性,Babiceanu [11]等提出了面向集成制造物理网络世界的可信度解决方案,使用弹性系统框架对网络安全组件进行建模。Alemayehu [12]等采用马尔可夫链对 CPS 的组件可靠性进行建模和分析,并提出相应的恢复技术,以保证系统的高可靠性,从而保证系统运行的连续性。Lalropuia [13]等通过建立半马尔可夫模型并得到可靠性、可用性和保密性等不同的可靠性度量,解决了系统的可靠性问题。Lu [14]等结合高斯过程和蒙特卡罗仿真,设计了一种考虑依赖性的新学习策略,并在此基础上提出了一种新的自适应代理模型方法。Dui [15]等通过多目标粒子群优化算法,使系统可靠性最大化,维护成本最小化,进一步提高智能家居系统的可靠性。Wang [16]等提出了一个由可信度本体和循证方法支撑的 NAMAC 可信度建模和评估框架。Wang [17]等重点研究了基于 Wiener 过程退化数据的随机漂移参数的准确可靠性推断,提出了一个精确程序来检验是否存在群体异质性。Xing [18]等提出了一种新的行为驱动的可靠性建模方法,用于分析基于 WSN 的复杂智能系统的可靠性。

综上所述,关于利用非参数模型对 CPS 进行可靠性建模的文献较少。本文通过非参数模型中的局部多项式回归对 CPS 的累计失效数进行建模研究。其余部分主要内容如下:第二节主要介绍基于局部多项式回归的 CPS 可靠性模型;第三节针对 CPS 的真实失效数据集进行案例分析;第四节根据分析结果得出结论。

2. 基于局部多项式回归的 CPS 可靠性模型

2.1. 非参数回归模型

非参数回归[19] [20]用于在参数形式未知的情况下探索变量之间的关系。由于 CPS 的失效数据的分布未知,以及 CPS 可靠性模型不依赖于固定的函数形式,因此本文采用非参数回归模型对失效数进行建模和拟合。

假设现有一组 CPS 累计失效数据,记为 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 需要研究 CPS 累计失效数 Y 与时间 X 之间的关系,则可表示为如下非参数回归模型的形式:

$$y_i = f(x_i) + \varepsilon_i, E(\varepsilon_i) = 0, i = 1, 2, \dots, n \quad (1)$$

其中, $f(\cdot)$ 为回归函数, $f(x)$ 的估计表示为 $\hat{f}_n(x)$, $Var(\varepsilon_i) = \sigma^2$ 。

2.2. 局部多项式回归模型

局部多项式回归(local polynomial regression, LPR)在每个点的局部范围内,利用多项式函数逼近回归函数,并基于加权最小二乘法得到回归函数在各点的估计值。这种方法不依赖于全局的模型结构,而是

根据数据在局部的表现来构建回归模型，因此能够更灵活地适应数据的局部特性。

具体来说，假设在 x_0 的领域内，CPS 累计失效数 Y 与时间变量 X 的关系可以用一个 d 次多项式近似表示：

$$y_i \approx \beta_0 + \beta_1(x_1 - x_0) + \beta_2(x_1 - x_0)^2 + \cdots + \beta_d(x_1 - x_0)^d + \varepsilon_i \quad (2)$$

其中， $\beta_0, \beta_1, \dots, \beta_d$ 是多项式的系数， ε_i 是误差项。

然后，通过使用加权最小二乘法估计这些系数。基于 Nadaraya-Watson 核估计，选择一个估计量 $r \equiv \hat{f}_n(x)$ 使得平方和 $\sum_{i=1}^n (y_i - r)^2$ 最小。令权函数为 $w_i(x) = K((x_i - x)/h)$ ，选择 $r \equiv \hat{f}_n(x)$ 使得下式的加权平方和最小：

$$\sum_{i=1}^n w_i(x)(y_i - r)^2 \quad (3)$$

其中， $K(\cdot)$ 是一个核， $h > 0$ ，为带宽。求解得到系数后，即可用多项式函数来估计 x_0 处的累计失效数。

3. 案例分析

3.1. 累计失效数据

BlueGene/L 是 IBM 公司设计并制造的一款超级计算机，它在全球超级计算机领域具有重要地位。本文所研究的超级计算机系统是信息物理系统的一种特殊情况。BlueGene/L 自发布以来取得了多项成就，包括在 TOP 500 榜单上多次排名第一、创造多项计算性能纪录等。它不仅推动了超级计算机技术的发展，还为科学研究提供了强大的计算支持。

本文将 BlueGene/L 的系统日志作为研究对象，以日为单位进行数据整合，提取了从 2005 年 6 月 1 日到 2006 年 3 月 1 日(提取间隔为 1 天)，共计 160 组数据，累计失效数据如表 1 所示。

Table 1. Cumulative failure data for BlueGene/L
表 1. BlueGene/L 的累计失效数据

时间/日	累计失效数	时间/日	累计失效数	时间/日	累计失效数	时间/日	累计失效数	时间/日	累计失效数
1	3	33	10,190	65	22,074	97	54,437	129	73,792
2	2049	34	10,554	66	22,762	98	54,512	130	73,881
3	2751	35	10,600	67	23,066	99	54,751	131	74,141
4	2753	36	10,999	68	24,139	100	55,513	132	74,283
5	2844	37	11,052	69	24,519	101	55,559	133	74,499
6	3095	38	11,203	70	24,621	102	55,609	134	74,652
7	3191	39	11,362	71	27,450	103	55,839	135	74,868
8	3237	40	11,904	72	28,548	104	56,065	136	74,921
9	3239	41	11,950	73	28,802	105	56,216	137	76,091
10	4385	42	12,870	74	29,844	106	56,778	138	76,204
11	5751	43	13,141	75	29,890	107	57,085	139	76,396
12	5762	44	13,230	76	29,941	108	57,131	140	76,627
13	7925	45	13,309	77	29,951	109	61,509	141	76,765
14	7974	46	13,355	78	30,750	110	62,962	142	76,949

续表

15	8021	47	13,454	79	30,785	111	63,173	143	77,376
16	8067	48	13,530	80	32,714	112	64,609	144	77,534
17	9004	49	13,638	81	34,186	113	65,675	145	77,587
18	9148	50	13,731	82	36,288	114	66,097	146	77,609
19	9159	51	13,763	83	37,764	115	66,328	147	77,716
20	9218	52	13,790	84	39,208	116	68,325	148	77,796
21	9250	53	14,163	85	40,630	117	68,396	149	77,904
22	9305	54	14,403	86	42,067	118	68,480	150	78,006
23	9324	55	14,822	87	42,251	119	68,601	151	78,151
24	9379	56	14,912	88	43,721	120	70,075	152	78,197
25	9480	57	15,616	89	43,999	121	70,088	153	78,403
26	9548	58	15,666	90	45,640	122	70,269	154	78,556
27	9568	59	15,943	91	47,303	123	70,656	155	78,715
28	9668	60	16,069	92	49,000	124	71,908	156	78,869
29	9715	61	19,588	93	52,358	125	72,054	157	81,178
30	9791	62	20,288	94	52,375	126	73,284	158	82,122
31	9878	63	21,119	95	53,513	127	73,450	159	82,252
32	9925	64	21,628	96	54,345	128	73,577	160	82,606

3.2. 模型评估标准

基于真实的累计失效数据集对 Goel-Okumoto (GO)模型、Delayed S-shaped (DSS)模型，以及本文提出的 LPR 模型进行对比分析，选用 MSE 和 AIC 作为评价模型性能的指标。

均方误差(MSE)是一个在多个领域中广泛应用的统计量，主要用于衡量模型预测值与实际观测值之间的差异程度。MSE 是预测值与实际观测值之差的平方的平均值。其计算公式为：

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{4}$$

其中， n 是样本数量， y_i 是第 i 个观测值， $\hat{y}_i$ 是第 i 个观测值的预测值。

赤池信息准则(AIC)是一种衡量统计模型拟合优良性的标准。AIC 准则在模型选择中十分重要，它旨在找到既能很好地解释数据又不过于复杂的模型。AIC 值越小，表示模型在拟合度和复杂度之间达到了更好的平衡，因此更可能是最优的模型。其计算公式为：

$$AIC = 2k - 2\ln(L) \tag{5}$$

其中， k 是模型中未知参数的个数，代表了模型的复杂度。 L 是模型的对数似然函数值，衡量了模型拟合数据的优良性。对数似然函数值越大，说明模型拟合数据的效果越好。

3.3. 模型性能对比分析

本文基于表 1 中 BlueGene/L 的系统累计失效数据，通过最小二乘估计法求解得到 GO 模型、DSS 模型的参数估计结果。三个可靠性模型的参数估计结果和性能对比结果如表 2 所示。从表可知，本文提出

的 LPR 模型的 MSE 和 AIC 的值比另外两个模型的 MSE 和 AIC 的值都要小，且 MSE 的值要小得多。另外，DSS 模型的 MSE 和 AIC 的值又比 GO 模型的值小。因此，从两个拟合优度评估指标来看，LPR 模型的拟合效果最好，其次为 DSS 模型，而 GO 模型的拟合效果最差。

Table 2. The results of parameter estimation and performance comparison of the model
表 2. 模型的参数估计结果和性能对比结果

模型名称	参数估计结果		MSE (1×10^4)	AIC (1×10^3)
	a (1×10^3)	b (1×10^{-2})		
GO 模型	2522.1390	0.0207	4977.2816	2.8390
DSS 模型	160.1008	1.1496	1822.6402	2.6780
LPR 模型	-	-	4.9973	1.7351

三种模型的累计失效数拟合图以及三种模型的拟合对比图如图 1 所示。从图可以明显看出，LPR 模型的拟合效果最好，拟合值与实际观测值较吻合，而 GO 模型拟合效果最差，该模型拟合值与实际观测值有较大差距。

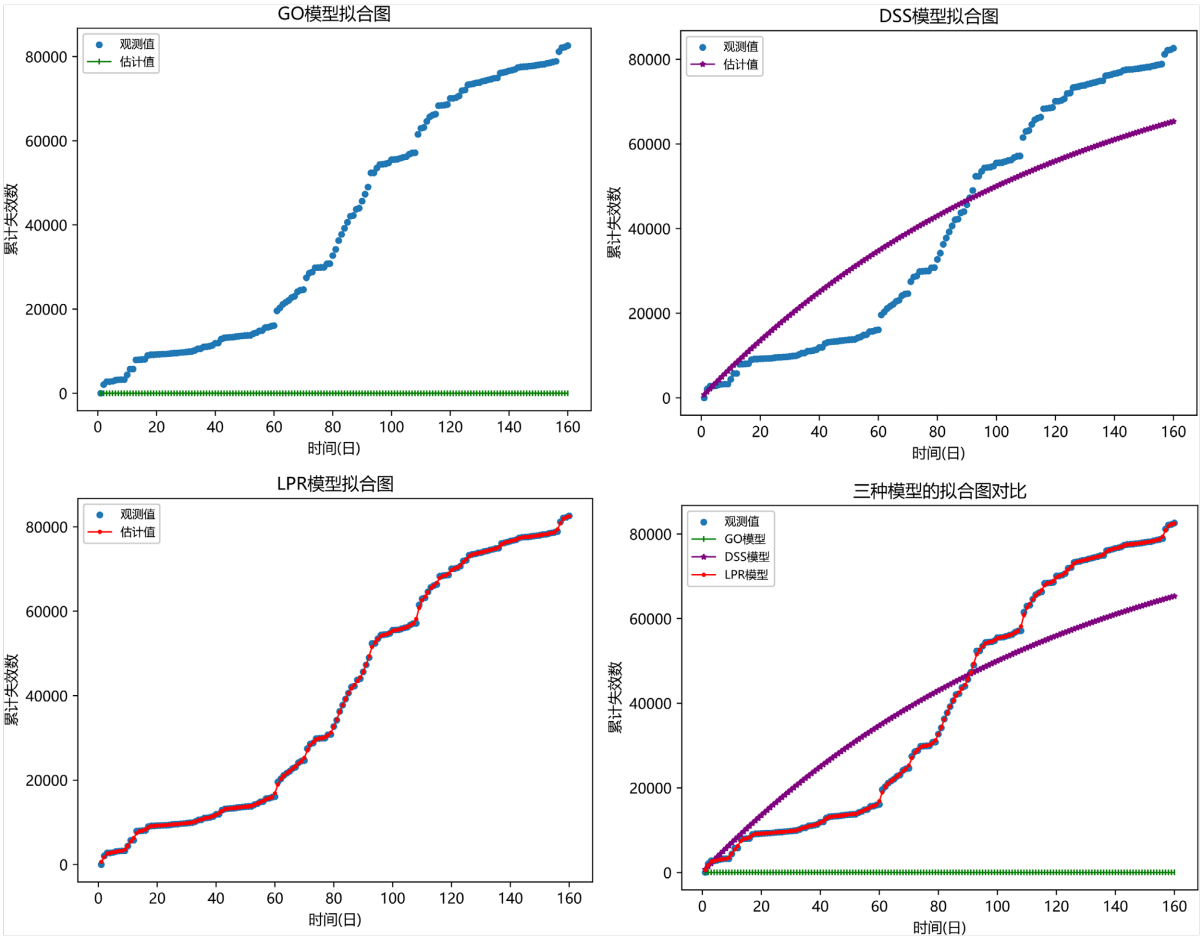


Figure 1. The fitting graph of the cumulative failure data of each model and the fitting graph comparison
图 1. 各模型的累计失效数拟合图以及拟合图对比

3.4. 预测性能对比分析

为了判断三种模型预测结果的稳定性和可靠性,本文对三种模型的相对误差(RE)曲线进行绘制并对比分析。RE 曲线是通过将一系列测量结果的相对误差(即测量值与真值之差占真值的百分比)绘制在坐标系上而形成的曲线。若 RE 曲线越接近于 0,则表明预测性能越好。

通过 BlueGene/L 的系统累计失效数据进行预测,展现了模型对未来累计失效数的预测能力。从图 2 可知:(1) LPR 模型的 RE 值最趋于 0,因此 LPR 模型的预测效果在这三种模型中最好。其次为 DSS 模型,随着时间推移,DSS 模型的 RE 值在一定的上下起伏波动后逐渐接近于 0。GO 模型的预测效果最差,在 $t = 80$ 之前,其 RE 值波动程度较大,但最后也趋于 0;(2) 在测试的初始阶段,LPR 模型的 RE 值趋于 0 的速度最快,其次为 DSS 模型,表明 LPR 模型对累计失效数据的拟合适应性比 DSS 模型强。

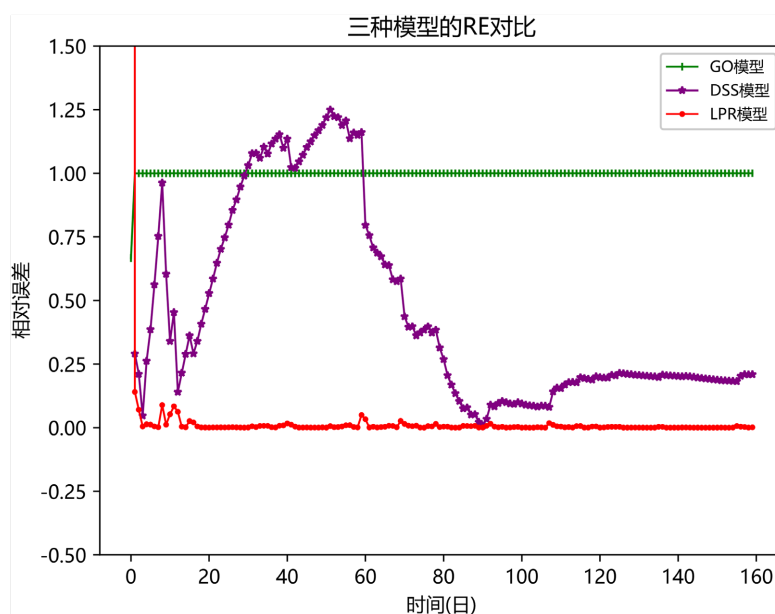


Figure 2. Comparison of the relative error curves of the three models

图 2. 三种模型的相对误差曲线对比

4. 结论

本文在电子商务视角下,将 BlueGene/L 的系统累计失效数据作为研究对象,提出一种基于局部多项式回归的信息物理系统可靠性模型,并将该非参数回归可靠性模型与传统的 GO 模型和 DSS 模型进行比较。最后,通过计算三种模型的拟合优度评估指标以及绘制拟合图、RE 曲线图进行对比分析,得出结论:基于局部多项式回归的信息物理系统可靠性模型的拟合效果和预测效果更好,对数据适应性较强。本文的研究结果为电子商务的稳定运行和高效发展提供了坚实的理论支撑。

基金项目

国家自然科学基金(编号: 72361008, 71901078); 贵州省科技计划项目(编号: 黔科合支撑[2023]一般 268)。

参考文献

- [1] Negi, N. and Chakraborty, A. (2020) Sparsity-Promoting Optimal Control of Cyber-Physical Systems over Shared

- Communication Networks. *Automatica*, **122**, Article 109217. <https://doi.org/10.1016/j.automatica.2020.109217>
- [2] Lu, A. and Yang, G. (2020) Observer-Based Control for Cyber-Physical Systems under Denial-of-Service with a Decentralized Event-Triggered Scheme. *IEEE Transactions on Cybernetics*, **50**, 4886-4895. <https://doi.org/10.1109/tcyb.2019.2944956>
 - [3] Yang, Y., Huang, J., Su, X. and Deng, B. (2021) Adaptive Control of Cyber-Physical Systems under Deception and Injection Attacks. *Journal of the Franklin Institute*, **358**, 6174-6194. <https://doi.org/10.1016/j.jfranklin.2021.06.008>
 - [4] Wang, C., Geng, Q., Liu, F. and Han, M. (2022) Interval Type-2 T-S Fuzzy MPC for CPS under Hybrid Attacks over a Multi-Channel Framework. *Journal of the Franklin Institute*, **359**, 9666-9689. <https://doi.org/10.1016/j.jfranklin.2022.09.054>
 - [5] Ge, H., Yue, D., Xie, X., Deng, S. and Dou, C. (2021) A Unified Modeling of Multi-Sources Cyber-Attacks with Uncertainties for CPS Security Control. *Journal of the Franklin Institute*, **358**, 89-113. <https://doi.org/10.1016/j.jfranklin.2019.01.006>
 - [6] Lu, Z., Liu, H., Kong, X., Li, H. and Li, T. (2023) Optimal Defense Strategy Selection Method for CPS Considering Integrated Cyber-Physical Losses. *Sustainable Energy, Grids and Networks*, **36**, Article 101143. <https://doi.org/10.1016/j.segan.2023.101143>
 - [7] Bo, X., Qu, Z., Liu, Y., Dong, Y., Zhang, Z. and Cui, M. (2022) Review of Active Defense Methods against Power CPS False Data Injection Attacks from the Multiple Spatiotemporal Perspective. *Energy Reports*, **8**, 11235-11248. <https://doi.org/10.1016/j.egyr.2022.08.236>
 - [8] Li, Y., Yang, Y., Chai, T. and Chen, T. (2022) Stochastic Detection against Deception Attacks in CPS: Performance Evaluation and Game-Theoretic Analysis. *Automatica*, **144**, Article 110461. <https://doi.org/10.1016/j.automatica.2022.110461>
 - [9] Zhang, Z., Li, M. and Xie, L. (2024) Data-Driven Replay Attack Detection for Unknown Cyber-Physical Systems. *Information Sciences*, **670**, Article 120562. <https://doi.org/10.1016/j.ins.2024.120562>
 - [10] Ahmed, C.M., Palleti, V.R. and Mishra, V.K. (2022) A Practical Physical Watermarking Approach to Detect Replay Attacks in a CPS. *Journal of Process Control*, **116**, 136-146. <https://doi.org/10.1016/j.jprocont.2022.06.002>
 - [11] Babiceanu, R.F. and Seker, R. (2017) Trustworthiness Requirements for Manufacturing Cyber-Physical Systems. *Procedia Manufacturing*, **11**, 973-981. <https://doi.org/10.1016/j.promfg.2017.07.202>
 - [12] Alemayehu, T.S. and Kim, J. (2017) Dependability Analysis of Cyber Physical Systems. *IET Computers & Digital Techniques*, **11**, 231-236. <https://doi.org/10.1049/iet-cdt.2016.0164>
 - [13] Lalropuia, K.C. and Gupta, V. (2019) Modeling Cyber-Physical Attacks Based on Stochastic Game and Markov Processes. *Reliability Engineering & System Safety*, **181**, 28-37. <https://doi.org/10.1016/j.ress.2018.08.014>
 - [14] Lu, N., Li, Y., Huang, H., Mi, J. and Niazi, S.G. (2023) AGP-MCS+D: An Active Learning Reliability Analysis Method Combining Dependent Gaussian Process and Monte Carlo Simulation. *Reliability Engineering & System Safety*, **240**, Article 109541. <https://doi.org/10.1016/j.ress.2023.109541>
 - [15] Dui, H., Wang, X., Dong, X., Zhu, T. and Zhai, Y. (2024) Reliability Model and Emergency Maintenance Strategies for Smart Home Systems. *Reliability Engineering & System Safety*, **251**, Article 110402. <https://doi.org/10.1016/j.ress.2024.110402>
 - [16] Wang, L., Lin, L. and Dinh, N. (2024) Trustworthiness Modeling and Evaluation for a Nearly Autonomous Management and Control System. *Reliability Engineering & System Safety*, **245**, Article 110008. <https://doi.org/10.1016/j.ress.2024.110008>
 - [17] Wang, X., Wang, B.X., Jiang, P.H. and Hong, Y. (2020) Accurate Reliability Inference Based on Wiener Process with Random Effects for Degradation Data. *Reliability Engineering & System Safety*, **193**, Article 106631. <https://doi.org/10.1016/j.ress.2019.106631>
 - [18] Xing, L., Zhao, G., Xiang, Y. and Liu, Q. (2021) A Behavior-Driven Reliability Modeling Method for Complex Smart Systems. *Quality and Reliability Engineering International*, **37**, 2065-2084. <https://doi.org/10.1002/qre.2845>
 - [19] Ding, J. and Zhang, Z. (2021) Statistical Inference on Uncertain Nonparametric Regression Model. *Fuzzy Optimization and Decision Making*, **20**, 451-469. <https://doi.org/10.1007/s10700-021-09353-0>
 - [20] Dong, H., Otsu, T. and Taylor, L. (2023) Bandwidth Selection for Nonparametric Regression with Errors-in-Variables. *Econometric Reviews*, **42**, 393-419. <https://doi.org/10.1080/07474938.2023.2191105>