

ІНФОРМАТИКА, ОБЧИСЛЮВАЛЬНА ТЕХНІКА ТА АВТОМАТИЗАЦІЯ

УДК 681.5.015

DOI <https://doi.org/10.32782/2663-5941/2025.6.2/01>

Авраменко А.А.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Варакута І.Є.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Савула А.А.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Коротинський А.П.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ВПЛИВ ПОПЕРЕДНЬОГО ЗГЛАДЖУВАННЯ ДАНИХ НА ТОЧНІСТЬ ІДЕНТИФІКОВАНИХ МАТЕМАТИЧНИХ МОДЕЛЕЙ КОЛОНИ СИНТЕЗУ МЕТАНОЛУ

У роботі досліджується вплив попереднього згладжування вимірних сигналів на результати ідентифікації математичних моделей багатовимірних промислових технологічних об'єктів. Як об'єкт розглядається колона синтезу метанолу, представлена у вигляді МІМО-системи з двома входами – витрата сировини та теплова потужність охолодження, та двома виходами – продуктивність за метанолом та температура на виході. Вихідні сигнали є повільними нестационарними процесами з явно вираженим трендом, на який накладаються високочастотні шумові коливання, зумовлені похибками вимірювання та збуреннями. Метою роботи є кількісно оцінити, яким чином згладжування таких даних впливає на структуру та точність отриманих моделей і, відповідно, на доцільність та ефективність у застосуванні таких моделей у подальшому синтезі систем автоматичного керування. Експериментальні дані роботи обладнання розділено на ідентифікаційну – 75% та валідаційну – 25% вибірки. Для обох випадків – сирих та згладжених сигналів – виконується ідентифікація математичних моделей з повним сітковим перебором порядків за входами та виходами. Якість отриманих моделей оцінюється за відсотковою збіжністю між вимірними й змодельованими сигналами. Окремо досліджується залежність точності отриманої математичної моделі від розміру вікна ковзного середнього. Показано, що помірне згладжування, зокрема вікном 5 відліків, істотно підвищує збіжність моделей як на навчальній, так і на валідаційній вибірках, без втрати динамічної інформації, а саме 98% в порівнянні з 91%. Отримані результати демонструють доцільність цілеспрямованого попереднього опрацювання експериментальних даних перед ідентифікацією та можуть бути використані для побудови більш надійних моделей при синтезі систем автоматичного керування хімічно-технологічними процесами.

Ключові слова: ідентифікація, обробка сигналів, математичне моделювання, згладжування сигналів.

Постановка проблеми. Сучасні системи автоматичного керування розробляються на основі математичних моделей об'єкта керування, які в свою чергу можуть бути ідентифіковані за результатами експериментальних досліджень або експлуатаційних даних. Для синтезу ефективних регуляторів необхідні адекватні моделі, які коректно відображають процеси в об'єкті керування. У розрізі ідентифікації такі моделі, як правило, будуються на основі «сирих» часових рядів вимірюваних сигналів, що містять шум, викиди, дрейфи та інші негативні складові, зумовлені обмеженою точністю вимірювальних елементів або зовнішніми збуреннями.

Поширеним підходом до підвищення якості сигналу є попередня його обробка, а саме: згладжування, фільтрація шумів, видалення викидів, масштабування тощо. Такі операції для ідентифікації математичних моделей часто виконуються за емпіричними правилами, без формального аналізу їхнього впливу на реальні властивості отриманих моделей. У результаті модель може добре узгоджуватися з поточними даними, що використовуються для ідентифікації, але давати некоректні результати на новий даних, а відтак не може бути ефективно використана при синтезі систем автоматичного керування.

Отже, наукова проблема цієї роботи полягає у кількісному оцінюванні впливу попередньої обробки експериментальних даних на результати ідентифікації математичних моделей об'єкта керування, зокрема на точність відтворення реальної динаміки та придатність моделі для подальшого синтезу систем автоматичного керування. Розв'язання цієї задачі дозволить сформулювати формалізовану методику попередньої обробки вимірюваних сигналів для задач ідентифікації, що мінімізує ризик отримання ефективних за метриками, але непридатних для практичного керування моделей.

Аналіз останніх досліджень і публікацій. Сучасна теорія ідентифікації динамічних систем інтенсивно розвивається в напрямку побудови математичних моделей, орієнтованих на задачі керування технологічними процесами. У роботах [1–3] наголошується, що якість експериментальних даних (рівень шуму, наявність дрейфів, пропусків, викидів) є одним з ключових факторів, що визначає точність оцінених моделей та їх придатність для синтезу регуляторів і прогнозування. Показано, що для промислових об'єктів широко застосовуються лінійні ARX/ARMAX-моделі, які поєднують достатню гнучкість з від-

носною простотою ідентифікації та інтерпретації параметрів [1–3].

Окремий напрям становлять дослідження, присвячені попередній обробці часових рядів перед ідентифікацією. В оглядових роботах [6, 7] розглядаються згладжування, цифрова фільтрація, видалення викидів, декомпозиція тренду та шумової складової, а також аналізується їхній вплив на похибки оцінювання параметрів моделей. Показано, що застосування фільтрації дозволяє зменшити дисперсію оцінок, однак надмірне згладжування може призводити до систематичних зміщень параметрів та втрати швидкодії моделі, особливо для об'єктів із вираженою динамікою у середньочастотному діапазоні [6, 7]. При цьому вибір параметрів фільтра часто здійснюється емпірично, без формалізованих критеріїв оптимального ступеня згладжування саме з погляду задачі ідентифікації.

У публікаціях з цифрової обробки сигналів [4, 5] значна увага приділяється простим за структурою методам фільтрації, зокрема ковзному середньому, рекурсивним НЧ-фільтрам та фільтрам зі скінченною імпульсною характеристикою. Наголошується на їхній обчислювальній простоті, стійкості до реалізації в мікроконтролерах та промислових ПЛК, а також придатності для онлайн-обробки вимірювальних сигналів.

Для хімічних та нафтопереробних виробництв у працях [8] наведені приклади ідентифікації ректифікаційних та реакційних колон на основі експлуатаційних даних, побудови лінійних та нелінійних моделей, а також їхнього використання у MPC-регуляторах і системах оптимального керування. Однак етап попередньої обробки сигналів у більшості таких робіт описується лише стисло, на рівні загальних рекомендацій (усереднення, фільтрація, нормалізація).

Таким чином, наявні публікації [1–8] підтверджують критичну роль якості вимірювальних сигналів та процедур їх попередньої обробки в задачах ідентифікації, проте питання вибору параметрів згладжування, зокрема розміру вікна ковзного середнього, здебільшого вирішується інтуїтивно. Це зумовлює актуальність та наукову новизну даної роботи, спрямованої на кількісну оцінку компромісу між придушенням шуму та збереженням динамічних властивостей моделі для подальшого синтезу систем автоматичного керування.

Постановка завдання. Метою роботи є кількісно оцінити вплив попередньої обробки даних, а саме згладжування експериментальних даних на

результати ідентифікації математичних моделей багатовимірного об'єкта керування. Для досягнення поставленої мети необхідно провести серію чисельних експериментів, у межах яких виконати ідентифікацію математичних моделей на сирих та згладжених даних.

Виклад основного матеріалу. Об'єкт дослідження та експериментальні дані. У якості об'єкта ідентифікації розглядається колона синтезу метанолу, описана у вигляді багатовимірного (MIMO – multiple-input and multiple-output) об'єкту з двома входами і двома виходами.

Як керувальні дії використовуються:

- u_1 – витрата сировини;
- u_2 – теплова потужність охолодження.

Вихідні змінні:

- y_1 – продуктивність колони за метанолом;
- y_2 – температура на виході з колони.

Обидва виходи є повільними нестационарними процесами з вираженим трендом, на якому присутні високочастотні коливання (див рис. 1). Останні зумовлені шумами первинних перетворювачів, квантуванням та збуреннями.

В такому випадку, виміряний сигнал можна розглядати як сума корисної складової $s(t)$, що відповідає реальній динаміці об'єкта, та шуму $n(t)$:

$$y_{\text{raw}}(t) = s(t) + n(t).$$

Для виділення корисної складової було побудовано згладжений сигнал $y_{lp}(t)$, отриманий шляхом цифрової низькочастотної фільтрації $y_{\text{raw}}(t)$. Шумову складову визначали як різницю:

$$n(t) = y_{\text{raw}}(t) - y_{lp}(t)$$

На рисунку 2 представлено структури шуму як функція $n(t)$ та відповідний розподіл амплітуд шуму.

Методика проведення експерименту. Для кількісної оцінки впливу згладжування даних на результати ідентифікації було розділено початкові дані на вибірку для ідентифікації – 75% та вибірку для валідації отриманої математичної моделі – 25%, причому для вихідних сигналів валідаційної вибірки застосовувався цифровий НЧ-фільтр для перевірки ідентифікації як методу отримання математичного опису саме динаміки процесу а не шуму.

Для кожного етапу дослідження виконується MIMO-ідентифікація моделей типу ARX на основі сирих та згладжених даних. Проводиться повний сітковий перебір порядків $n_a \in [1; 29]$, $n_b \in [1; 29]$, де n_a – порядок за виходами, n_b – порядок за входами. Для кожної пари (n_a, n_b) обчислюються метрика відсоток збіжності окремо на ідентифікаційній та валідаційних вибірках:

$$FIT_j = 100 \left(1 - \frac{\|y_j - \hat{y}_j\|}{\|y_j - y_j^-\|} \right),$$

де y_j та $\hat{y}_j$ – вектори виміряних та змодельованих значень j -го виходу, y_j^- – їх середнє значення.

У якості основного методу попереднього згладжування даних для ідентифікації використано ковзке середнє. Такий підхід є лінійним, інтуїтивно зрозумілим і широко застосовується в інженерній практиці завдяки простоті реалізації та невибагливості до налаштувань. Ковзке середнє ефективно пригнічує високочастотний шум, водночас зберігаючи повільну складову сигналу й основну форму тренду, що є критичним для коректної оцінки динамічних характеристик об'єкта.

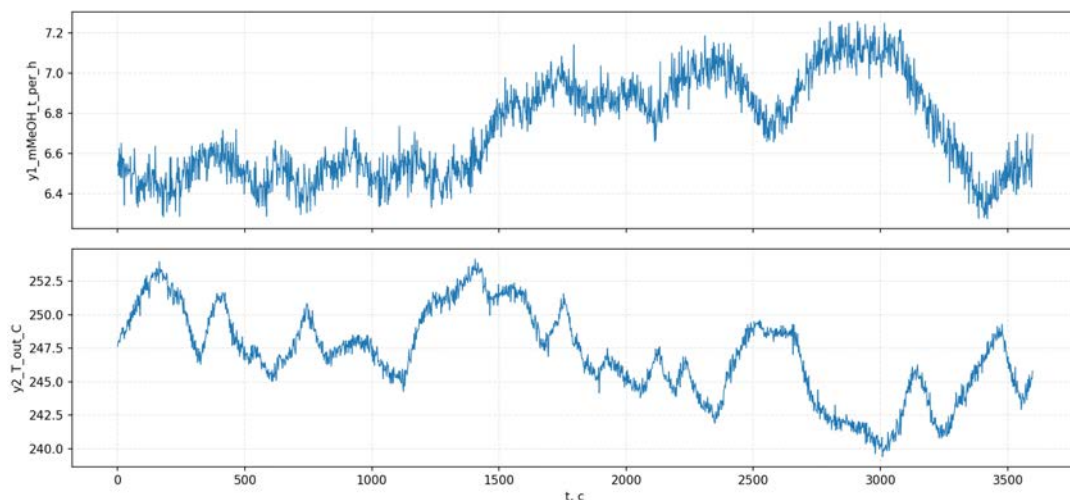


Рис. 1. Графік зміни вихідних змінних об'єкта

Результати ідентифікації на сирих експериментальних даних. За результатами пошуку найкращою виявилась ARX-модель з порядками $n_a = 29$ та $n_b = 29$. Для цієї моделі на ідентифікаційних даних отримано такі показники якості:

- для виходу y_1 : $FIT_{train}(y_1) = 73.21\%$;
- для виходу y_2 : $FIT_{train}(y_2) = 85.38\%$.

Порівняльний графік сирих даних та результатів моделювання наведено на рис. 3–4. Добре видно, що модель високого порядку не лише відтворює повільні тренди, а й досить точно повторює височастотні коливання сигналів. Це свідчить про те, що частина параметрів моделі

фактично ідентифікує шумову складову вимірювань, що небажано з точки зору подальшого синтезу системи керування.

Якість моделі була перевірена на незалежній валідаційній вибірці, яка не використовувалась при ідентифікації. Для цієї ж моделі отримано такі результати:

- для виходу y_1 : $FIT_{test}(y_1) = 91.10\%$;
- для виходу y_2 : $FIT_{test}(y_2) = 91.76\%$.

Порівняння реальних валідаційних даних з розрахованими за моделлю наведено на рис. 5–6. В обох випадках помітна висока точність апроксимації: модель детально повторює форму перехідних про-

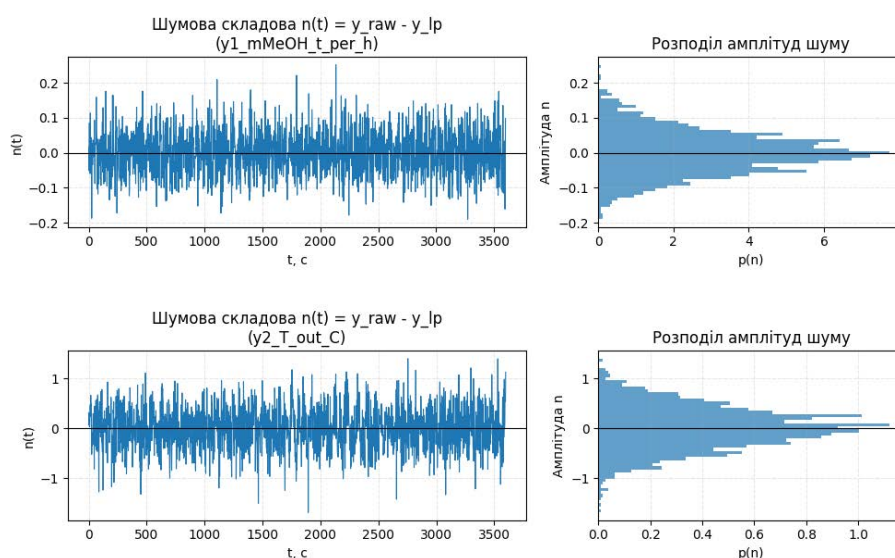


Рис. 2. Розподіл амплітуд шуму

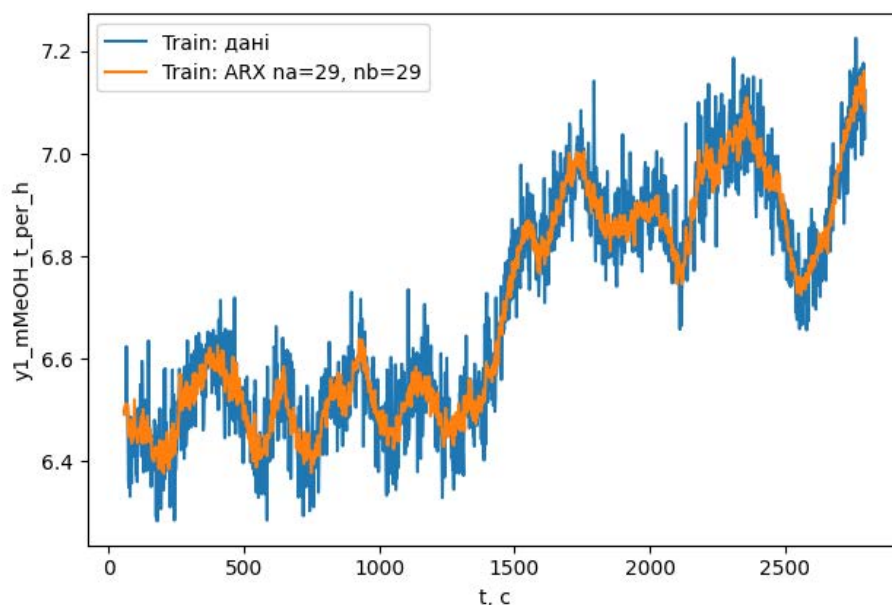
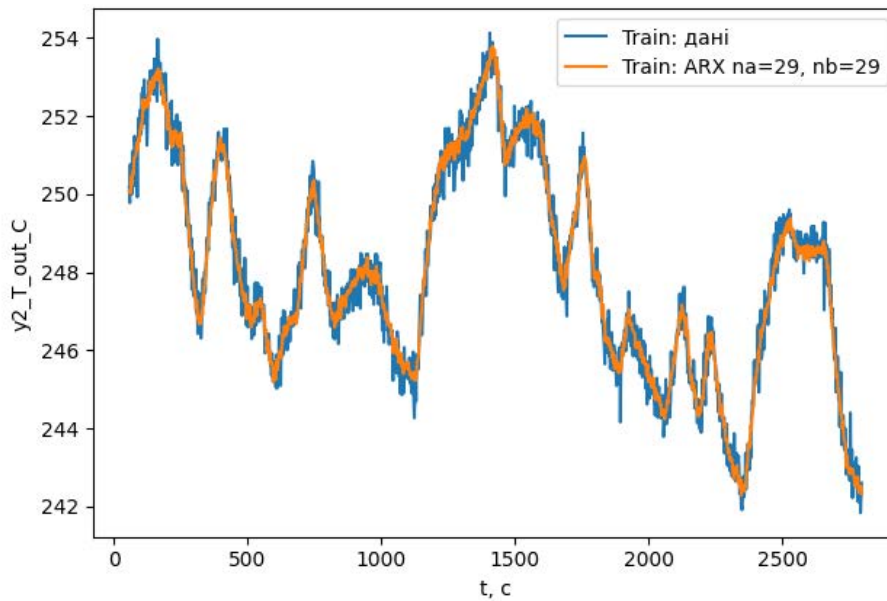
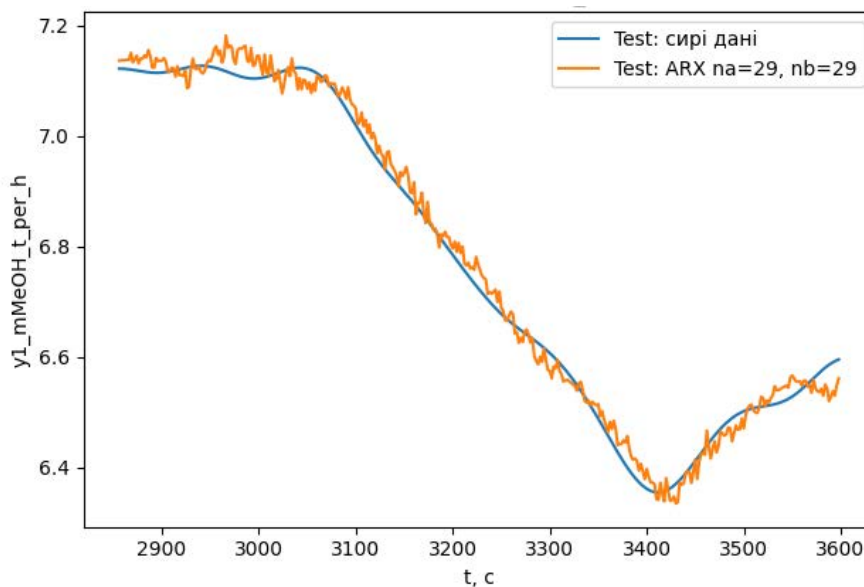


Рис. 3. Графік зміни вихідної змінної y_1 сирих даних та ідентифікованої моделі на основі сирих даних

Рис. 4. Графік зміни вихідної змінної y_2 сирих даних та ідентифікованої моделі на основі сирих данихРис. 5. Графік зміни вихідної змінної y_1 валідаційних даних та ідентифікованої моделі на основі сирих даних

цесів як на ідентифікаційних, так і на валідаційних даних. Разом із цим візуально добре видно, що для виходу y_1 траєкторія моделі повторює значну частину високочастотних коливань, притаманних сирим вимірюванням, тобто оцінена структура фактично інтерпретує випадковий шум як «динаміку» об'єкта.

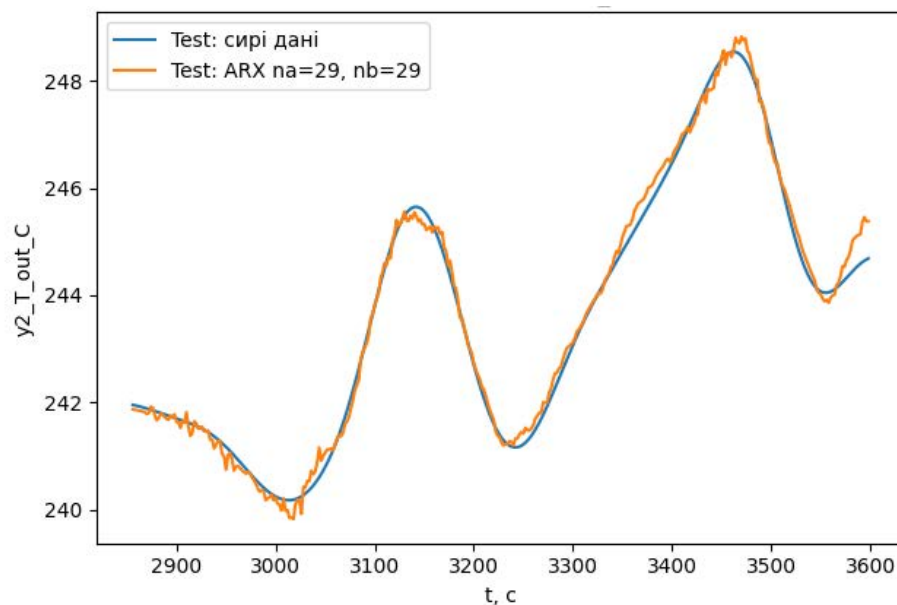
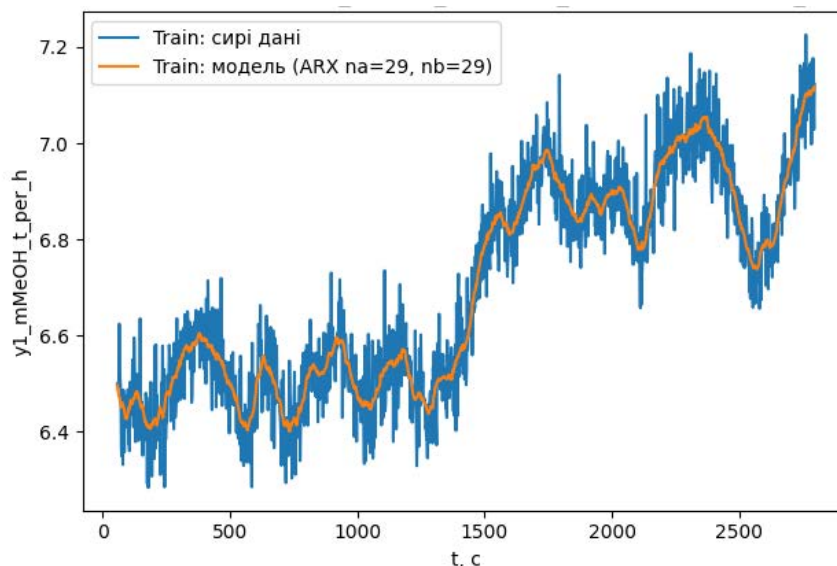
Результати ідентифікації після згладжування ковзним середнім. На наступному етапі до вихідних сигналів було застосовано цифрове згладжування ковзним середнім з вікном $M=5$. Така фільтрація зменшує високочастотну шумову складову вимірювань, зберігаючи повільні зміни

режиму роботи колони. Відповідні часові реалізації виходів y_1 та y_2 до і після згладжування показують, що фільтровані сигнали стають більш гладкими, а випадкові піки навколо тренду практично зникають, при цьому форма перехідних процесів зберігається (див. рис. 7–8).

Для згладженого випадку отримано такі показники на ідентифікаційній вибірці:

- для y_1 : $FIT_{test} = 98,12\%$;
- для y_2 : $FIT_{test} = 98,22\%$.

Порівняно з сирими даними ($FIT_{test} \approx 91,10\%$ та $91,76\%$ відповідно для y_1 та y_2), відносна похибка

Рис. 6. Графік зміни вихідної змінної y_2 валідаційних даних та ідентифікованої моделі на основі сирих данихРис. 7. Графік зміни вихідної змінної y_1 ідентифікаційних даних та ідентифікованої моделі на основі згладжених даних

моделі на валідації суттєво зменшується, а криві моделі практично збігаються з експериментальними реалізаціями (рис. 9–10). Це означає, що попереднє згладжування ковзним середнім $M=5$ не призводить до спотворення динаміки об'єкта, а навпаки – усуває високочастотну шумову складову і дозволяє побудувати модель, яка відтворює основний тренд керованої системи як на ідентифікаційній, так і на валідаційній вибірках.

Вплив розміру вікна ковзного середнього на якість моделі. Для узагальнення результатів було проаналізовано вплив параметра згладжування, а саме розміру вікна ковзного середнього на

якість отриманих моделей. На основі сирих даних формувалися згладжені вибірки для значень у діапазоні $Me[2; 600]$. Для кожного значення виконувалася ідентифікація МІМО-моделі на згладженій ідентифікаційній вибірці, після чого модель перевірялась на незалежній валідаційній вибірці, сформованій із сирих даних. Як інтегральний показник якості використовувалося середнє значення метрики FIT по виходах об'єкта.

Залежність середнього FIT від розміру вікна ковзного середнього наведено на рис. X.

При малих значеннях розміру вікна (2–5 відліків) спостерігається різке зростання якості

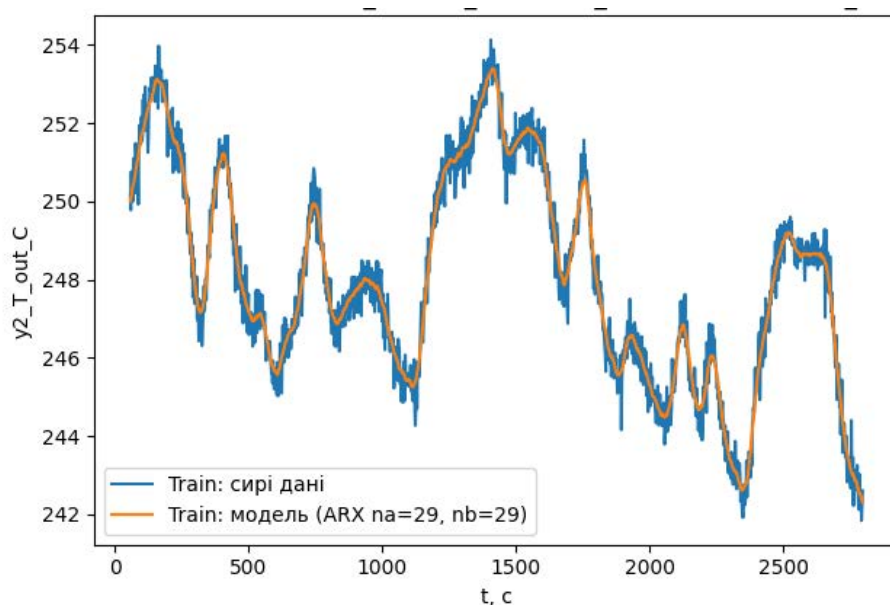


Рис. 8. Графік зміни вихідної змінної y_2 ідентифікаційних даних та ідентифікованої моделі на основі згладжених даних

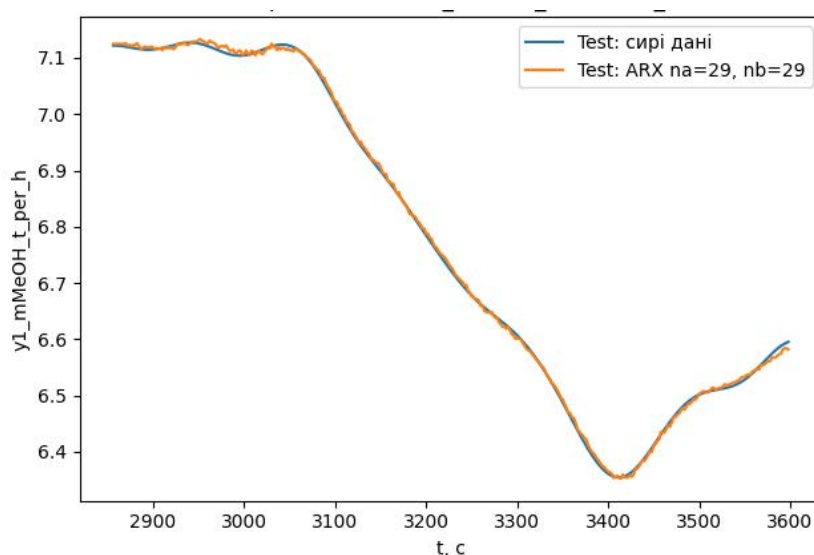


Рис. 9. Графік зміни вихідної змінної y_1 валідаційних даних та ідентифікованої моделі на основі згладжених даних

моделі: придушення високочастотного шуму дозволяє зменшити похибку прогнозу без суттєвого спотворення динамічної складової сигналу. Подальше збільшення розміру вікна призводить до поступового насичення кривої: починаючи орієнтовно з $M \approx 80-100$, середній FIT досягає плато на рівні близько 99,7–99,8 і практично не змінюється при подальшому збільшенні розміру вікна.

Однак така збіжність досягається ціною втрати частини динамічної інформації та досить висом порядком моделі. Для великих розмірів вікна ковзке середнє починає згладжувати не лише шум, а й реальні швидкі зміни процесу, що призводить

до спрощення часових залежностей та зменшення ефективного порядку динаміки.

Вибір параметра згладжування має здійснюватися як компроміс між придушенням шуму та збереженням інформативної складової сигналів. У подальших розрахунках як робоче значення було обрано вікно $M=5$, яке забезпечує істотне покращення якості ідентифікації порівняно з сирими даними, водночас зберігаючи характерні коливання та тренди вихідних змінних.

Висновки. У роботі показано, що використання сирих експериментальних даних для ідентифікації багатовимірної моделі колони синтезу

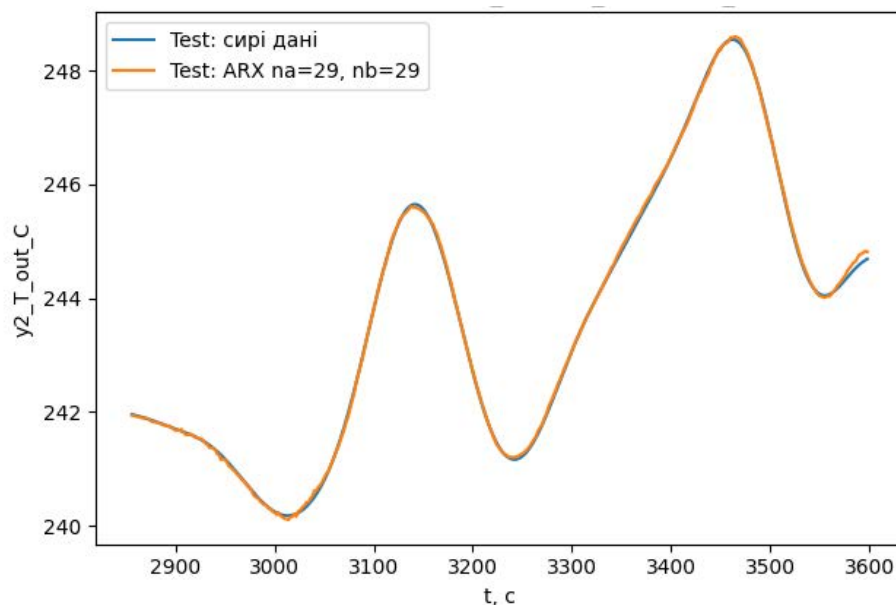
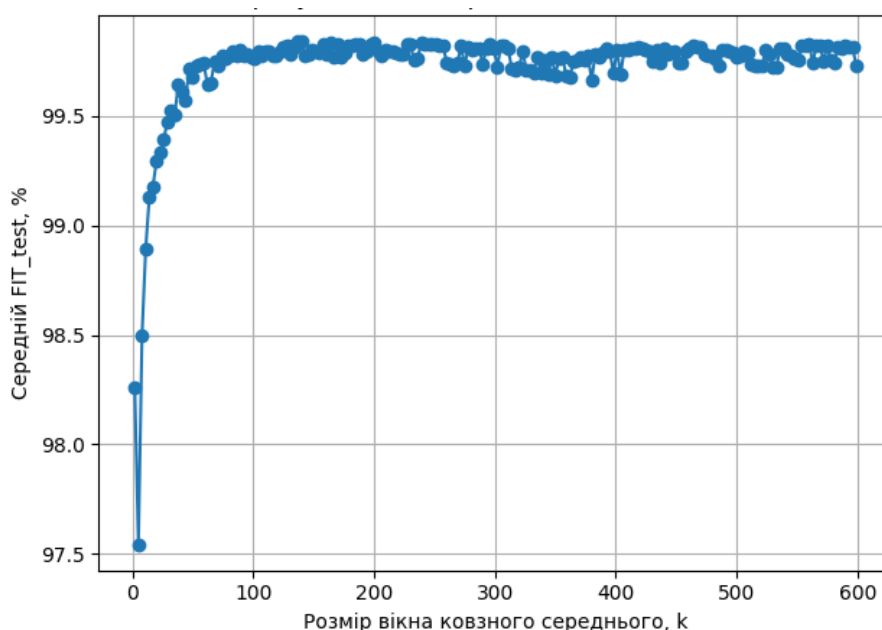


Рис. 10. Графік зміни вихідної змінної y_2 валідаційних даних та ідентифікованої моделі на основі згладжених даних



метанолу призводить до того, що частина параметрів моделі фактично описує високочастотну шумову складову вимірювань. У результаті отримана модель високого порядку демонструє високий відсоток збіжності як на ідентифікаційній, так і на валідаційній вибірках, проте відтворює не лише динаміку об'єкта, а й випадкові коливання.

Запропонований підхід попереднього згладжування вихідних сигналів за допомогою ковзного середнього дає змогу суттєво зменшити рівень високочастотного шуму без помітного спотворення повільної складової та форми перехідних

процесів. Результати чисельних експериментів свідчать, що ідентифікація моделей на згладжених даних забезпечує помітне покращення показників відсоткової збіжності на незалежній валідаційній вибірці у порівнянні з моделлю, побудованою на сирих даних. При цьому траєкторії моделі добре узгоджуються з очищеними вихідними сигналами, що вказує на коректне відтворення саме динаміки об'єкта, а не шумових компонент вимірювань.

Отримані результати підтверджують доцільність цілеспрямованої попередньої обробки екс-

периментальних даних перед ідентифікацією, зокрема застосування простих лінійних методів згладжування, таких як ковзке середнє. Разом з тим показано, що параметри згладжування (розмір

вікна, характеристики фільтра) мають добиратися з урахуванням спектральних властивостей сигналів, щоб уникнути надмірного згладжування, яке може призвести до втрати інформативної динаміки.

Список літератури:

1. Simpkins A. System Identification: Theory for the User, 2nd Edition (Ljung, L.; 1999) [On the Shelf]. *IEEE Robotics & Automation Magazine*. 2012. 19(2). pp.95–96. doi:<https://doi.org/10.1109/mra.2012.2192817>.
2. Isermann R., Münchhof M. *Identification of Dynamic Systems*. 2010. doi: <https://doi.org/10.1007/978-3-540-78879-9>.
3. Schoukens J., Ljung L. Nonlinear system identification: A user-oriented road map. *IEEE Control Systems Magazine*. 2019. Vol. 39, No. 6. P. 28–99.
4. Rodrigues E.P., Santos D.S. dos and Buiochi F. Comparison of Three Linear Digital Filters Applied to Improve the Quality of the Measured Acoustic Field. *Eng*. 2023. 4(4). pp.2582–2600. doi:<https://doi.org/10.3390/eng4040148>.
5. Sunny M. N. M., Amin M. M., Hossain K. M. S., et al. Design of Digital Filters for Noise Reduction in RF Communication Systems. *International Journal of Novel Research in Electrical and Mechanical Engineering*. 2024.
6. Arici M., Daosud W., Vargan J., Fikar M. Multi-Model Predictive Control of a Distillation Column. *Systems and Control Transactions*. 2025. 4. pp.999–1004. doi:<https://doi.org/10.69997/sct.180258>.
7. Fortuna L., Graziani S., Rizzo A., Maria Gabriella Xibilia. *Soft Sensors for Monitoring and Control of Industrial Processes*. Springer eBooks. 2007. Springer Nature. doi:<https://doi.org/10.1007/978-1-84628-480-9>.
8. Wahid A., Brillianto Z. H. Multivariable model predictive control (4x4) of methanol–water separation in dimethyl ether production. *AIP Conference Proceedings*. 2020. Vol. 2255. Art. 020056.

Avramenko A.A., Varakuta I.Ye., Savula A.A., Korotynskyi A.P. THE INFLUENCE OF PRELIMINARY DATA SMOOTHING ON ACCURACY OF IDENTIFIED MATHEMATICAL MODELS OF THE METHANOL SYNTHESIS COLUMN

The paper investigates the influence of preliminary smoothing of measured signals on the results of identification of mathematical models of multidimensional industrial technological objects. The object under consideration is a methanol synthesis column, represented as a MIMO system with two inputs – raw material consumption and cooling thermal power – and two outputs – methanol productivity and outlet temperature. The output signals are slow non-stationary processes with a clearly expressed trend, overlaid with high-frequency noise fluctuations caused by measurement errors and disturbances. The aim of the work is to quantitatively assess how smoothing such data affects the structure and accuracy of the models obtained and, accordingly, the feasibility and effectiveness of using such models in the further synthesis of automatic control systems. The experimental data of the equipment operation is divided into an identification sample (75%) and a validation sample (25%). In both cases – raw and smoothed signals – the identification of mathematical models is performed with a complete grid search of orders for inputs and outputs. The quality of the obtained models is evaluated by the percentage of convergence between the measured and modeled signals. The dependence of the accuracy of the obtained mathematical model on the size of the moving average window is studied separately. It is shown that moderate smoothing, in particular with a window of 5 readings, significantly increases the convergence of models on both training and validation samples without losing dynamic information, namely 98% compared to 91%. The results obtained demonstrate the feasibility of targeted preliminary processing of experimental data prior to identification and can be used to build more reliable models in the synthesis of automatic control systems for chemical and technological processes.

Key words: identification, signal processing, mathematical modeling, signal smoothing.

Дата надходження статті: 23.11.2025

Дата прийняття статті: 12.12.2025

Опубліковано: 30.12.2025