

# АУТОМАТСКА ВАЛИДАЦИЈА ПОДАТАКА У КАНАЛИЗАЦИЈИ

## AUTOMATIC VALIDATION OF SEWAGE DATA

НЕМАЊА БРАНИСАВЉЕВИЋ<sup>1</sup>

ДУШАН ПРОДАНОВИЋ<sup>2</sup>

**Резиме:** Развој технологије мерења и преноса података омогућио је лако и ефикасно прикупљање великог броја података о канализационој инфраструктури, параметрима управљања, хидрауличким параметрима и параметрима квалитета. Проблем поузданости се у случају великог броја аутоматски прикупљених података појављује као кључни фактор при употреби података. У овом раду се разматра поступак аутоматске валидације који се може применити на великом броју података. У описаном поступку се проблем неодређености валидационих метода решава класификацијом података на основу њиховог контекста, а проблем грешака у подацима, који се користе као улазне вредности валидационих метода, осредњавањем резултата више метода. На овај начин је смањен ниво регуларних података регистрованих као података са аномалијама.

**Кључне речи:** квалитет података у канализацији, валидација података

**Abstract:** Advances in measuring techniques and data transfer methods made feasible to collect vast amount of data about the conditions in sewage network, sewage management parameters, hydraulic and water quality parameters. With the increasing amount of data automatically collected, data reliability is becoming a key factor of its usability. In this paper is presented a automatic validation that may be applied on great amount of collected data. The problem of validation methods uncertainty is treated by classifying the data values into corresponding context, and the problem of errors in validation method's input values is treated by averaging the results of several validation methods used to validate data values of one variable. The main effect of the presented approach is expressed in substantial reduction of false alarms.

**Key words:** sewer data quality, data validation

### 1. Увод

Осматрање канализационог система подразумева прикупљање података о стању мреже, параметрима и ефектима управљања, стању хидрауличких

---

<sup>1,2</sup> Институт за хидротехнику и водно-еколошко инжењерство, Грађевински факултет, Универзитет у Београду, Булевар краља Александра 73, 11000 Београд, Србија

параметара и параметара квалитета воде. Ниво поузданости информација које се могу добити из прикупљених података у највећој мери зависи од поузданости самих података [4]. Употребом релација између мерених података и/или расположивих информација о осматраном систему, могуће је спровести проверу мерених података и детектовати присуство потенцијалне грешке у податку која умањује његову поузданост. На тај начин се потенцијално скупа и сложена превентива (нпр. увођење додатних мерила или одржавање система мерача) замењује рачунским алгоритмима након што су подаци прикупљени.

Податке прикупљене у канализационом систему (податке о стању мреже, хидрауличке податке и податке о квалитету воде) одликују неизвесност и присуство грешке.

$$x = (x_i + err) \pm unc$$

Неизвесност (*unc*), као пратилац сваког податка, се може исказати интервалима, расплутим скуповима или статистичким расподелама. Присуство грешке (*err*), са друге стране, остаје скривено, док се не утврди процесом валидације податка. У овом раду је приказан поступак у коме се неизвесност математичких модела који се при валидацији користе смањује стављањем податка у одређени контекст, док се грешка третира као случајна променљива и уклања употребом више типова математичких модела који представљају исту релацију између података.

## 2. Методологија

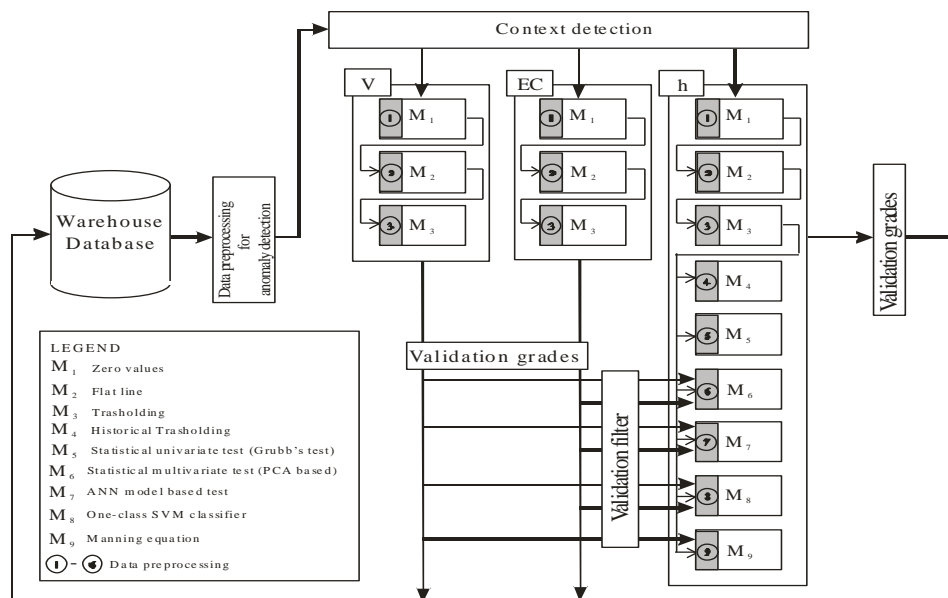
Општа шема која се користи за валидацију података је приказана у Бранисављевић и ост. (2009). Пример шеме валидације података, која одговара дубини воде на мерној локацији Вишњица у Београду [6], је приказана на слици 1. Може се приметити да су у процесу валидације укључени и мерени подаци о брзини воде и електропроводности.

### 2.1 Класификација према контексту

Математички модели у оквиру метода које се користе у процесу валидације морају бити адекватно калибрисани. Да би се побољшао учинак модела, тј. смањила његова неодређеност, подаци се класификују према контекстима [1]. Неки контексти који се могу дефинисати у течењу у у општом канализационом систему су:

- *Климатски*: кишно/суво
- *Хидраулички*: течење под успором / слободно течење
- *Пумпни*: on/off
- *Социјални*: радни дан/викенд
- *Сезонски*: Пролеће/Лето/Јесен/Зима

- *Временски:* 0/.../23
- *У погледу доба дана:* Ноћ/Јутро/Дан/Вече
- *Социјални догађаји:* регуларни дан/празник/утакмица/...



Слика 1: Имплементација система за валидацију

Валидационе методе су дизајниране и калибрисане према контексту у коме се податак налази. На тај начин је смањена неодређеност и омогућено је да се прецизније дефинишу релације између података математичким моделима у склопу методе. На жалост, понекад је класификација података у контексте изузетно захтеван процес који се ослања на сложене математичке алгоритме класификације и груписања.

## 2.2 Методе за валидацију

Методе за валидацију су одабране према искуству и резултатима који се од њих могу очекивати (слика 1). Из те групе метода, посебно се издвајају методе за откривање грубих грешака, чији резултат има елиминациони карактер: податак за који је детектована грешка једном од тих метода, елиминише се из даље анализе и употребе. Код примера наведеног у овом раду, то су методе  $M_1$ ,  $M_2$  и  $M_3$ . На слици 1 се може приметити да су ове методе примењене на све величине које учествују у валидацији дубине воде.

Методе које се користе у систему за валидацију дубине воде су:

- $M_1$  – Детекција нула или неких других предефинисаних вредности.
- $M_2$  – Детекција равних линија.
- $M_3$  – Мин/Мах према геометрији колектора у пресеку мерног места.

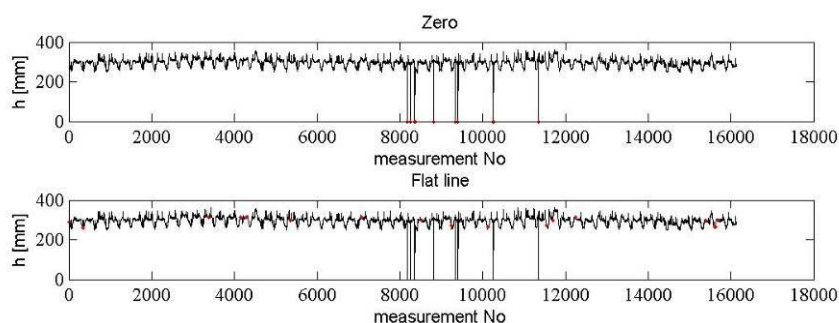
- $M_4$  – Мин/Мах према историјским подацима.
- $M_5$  – Статистички Grubb-ов тест [2, 3]. Овај тест претпоставља да су улазне вредности дубине у исто време (различитих дана) јављају по Студентовој  $t$  расподели [1, 5].
- $M_6$  – Principal Component Analysis (PCA) – метода базирана на статистичкој апроксимацији података и одређивању аномалија у подацима на основу Hotelling-овог  $t^2$  параметра [2, 3, 5].
- $M_7$  – Метода базирана на Artificial Neural Networks (ANN) нелинеарној регресији између података.
- $M_8$  – Откривање аномалија у подацима класификационом методом One-class SVM, са Gauss-овим кернелом [1].
- $M_9$  – Метода базирана на Chézy-Manning-овој једначини

Изабране валидационе методе су подешене (калибрисани су математички модели у склопу методе) према унапред одабраном и валидованом скупу мерених података.

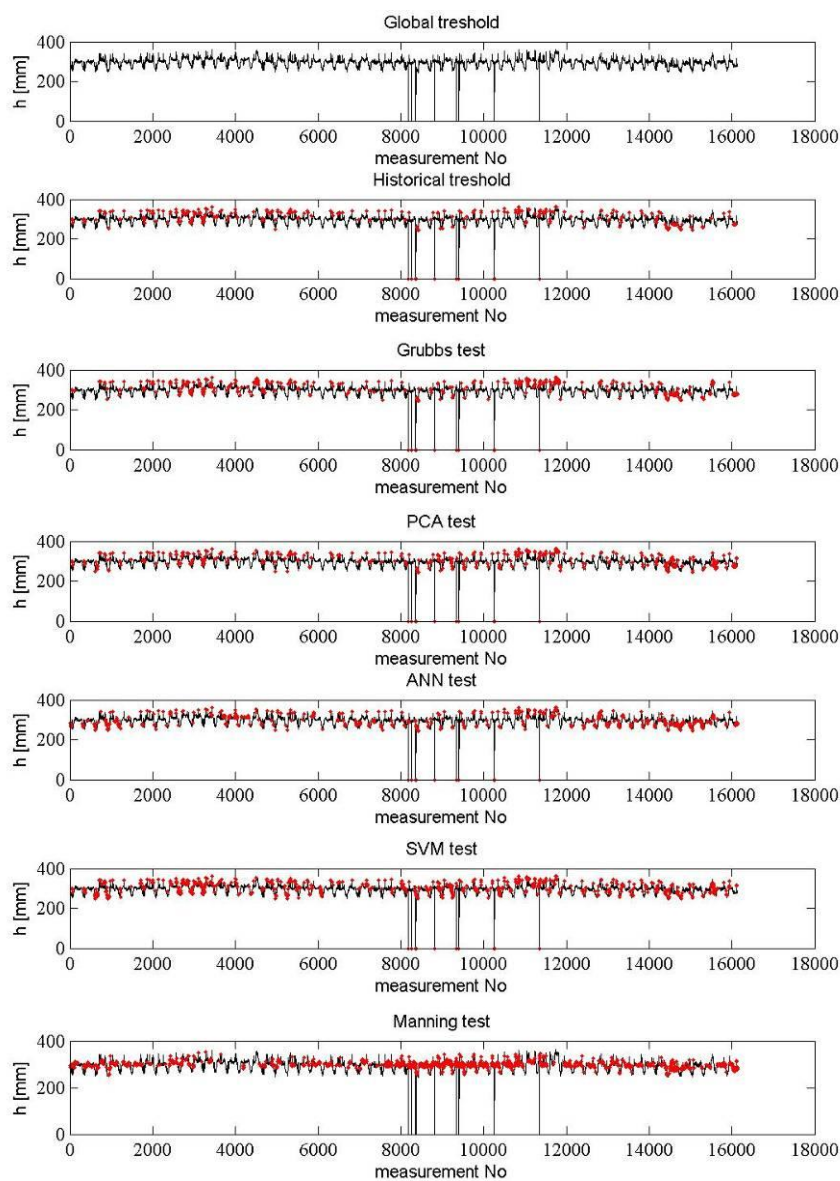
### 3. Резултати и дискусија

Анализом података издвојене су два најзначајнија контекста у које су подаци класификовани: 1) доба дана (јутро/дан/вече/ноћ) и 2) суво/кишно време. Искусствени подаци о карактеристикама тока воде у различито доба дана су коришћени за класификацију у први контекст, док је за други контекст коришћена класификациона метода на бази карактеристика течења у суво време [1].

Резултати свих метода посебно су приказани на сликама 2 и 3. Може се приметити да су неке од метода као резултат пружиле много “лажних аларма”. Ови лажни аларми, између осталог, потичу од грешака у подацима о брзинама и електропроводности који су употребљени као улазни подаци у математичким моделима у склопу метода за валидацију.

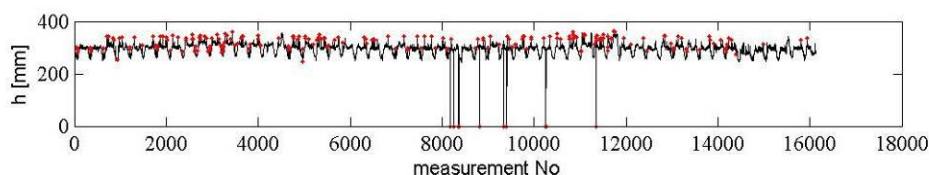


Слика 2. Резултати валидације података изабраним методама на примеру мерених дубина воде на мерном месту Вишњица (наставка на слици 3)



Слика 3: Резултати валидације података изабраним методама на примеру мерених дубина воде на мерном месту Вишњица (наставак са слике 2)

Уколико се претпостави да грешке у резултатима метода за валидацију имају карактеристике случајних вредности, оне се могу уклонити осредњавањем резултата више метода. Обзиром да су подаци са аномалијама регистрованим методама за откривање грубих грешака искључене из даље анализе, осредњени су само резултати метода 4, 5, 6, 7, 8 и 9. Након осредњавања, резултати су показали да се ниво лажних аларма значајно смањило. Резултат осредњавања је приказан на слици 4.



Слика 4: Резултати валидације података након осредњавања

Тестирање успешности поступка валидације је спроведена помоћу статистичког показатеља, индикатора успешности валидационе методе приказаног у [2]. Серија за тестирање је ручно валидована и детектовано је 244 податка са аномалијама. Методологијом која је презентована у овом раду је детектовано 195 података са аномалијама, са 58 лажних аларма, што представља бољи резултат од било које методе посебно:

$$p = \frac{N_{registered}}{N_{anomalies} + N_{missed} + N_{registered\ nonanomalies}} = \frac{195}{244 + 49 + 58} = 0.56$$

#### 4. Закључак

Савремена употреба SCADA система за осматрање омогућава лако прикупљање великог броја података о мереним величинама у канализационом систему. У овом раду је приказана методологија аутоматске валидације података примењене на мерења спроведена на канализационом систему у Београду. Резултати охрабрују и упућују на закључак, да са адекватном применом метода за валидацију и анализом њихових резултата поузданост података може бити значајно унапређена.

#### 5. Литература

- [1] Branislavljevic N., Kapelan Z. and Prodanovic D., "Improved Real-time Data Anomaly Detection using Context Classification", *Journal of Hydroinformatics (accepted)*
- [2] Branislavljević N., Prodanović D. and Pavlović D., "Automatic, Semi-automatic and Manual Validation of Urban Drainage Data", *The 8th Conference on Urban Drainage Modelling*, Tokyo-Japan, (2009)
- [3] Branislavljević N., Kapelan Z. and Prodanović D., "Online time data series pre-processing for the improved performance of anomaly detection methods", *CCWI-2009: Integrating Water Systems*, Taylor & Francis Group, London, (2009), pp 99-103.
- [4] Fletcher T. and Deletic A. (Editors), "*Data requirements for integrated urban water management*", Urban Water Series – UNESCO IHP, Teylor & Francis, (2008)
- [5] Kottogoda N. and Rosso R., "*Statistics, Probability and Reliability for Civil and Environmental Engineers*", McGraw-Hill, International Editions, (1998)
- [6] Продановић Д., Павловић Д., Огњановић В., Иветић М. и Костић Д., "Мерења на Београдском канализационом систему", Грађевински факултет, (2007)