

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПОШУКУ МОЖЛИВИХ ДЖЕРЕЛ ПІДВИЩЕНОГО ЗАБРУДНЕННЯ РІЧКИ З ВИКОРИСТАННЯМ МОДЕЛІ PROPHET

Керівник МКР:
д.т.н., професор Мокін В. Б.
Розробив:
студент гр. 2ІСТ-19м Шмундяк Д.О.

Вступ

Стан сучасного клімату та його постійні зміни призводять до зменшення кількості води. До цього всього, обсяги забруднення вод стрімко зростають і, через зменшення кількості води, концентрація забруднюючих речовин також зростає. Тому пріоритетом є заходи для підтримки екологічного стану водних ресурсів, які спрямовані на контроль та виявлення джерел можливого забруднення води.

В багатьох країнах Європи, в тому числі й Україні, мережа моніторингу води не завжди функціонує достатньо добре. Існують великі прогалини у наявних даних, що ускладнює процес знаходження та локалізації джерел забруднення води і, як результат, реалізацію кроків проведення політики регулювання стану води.

Отже, актуальним є розроблення та реалізація інформаційної технології пошуку можливих джерел підвищеного забруднення річки, яка дозволить на основі відкритих даних знайти та локалізувати джерела забруднення річки по роках та локаціях.

Мета роботи – створення інформаційної технології пошуку можливих джерел підвищеного антропогенного навантаження на річку за відкритими даними державного ресурсу спостережень якості води.

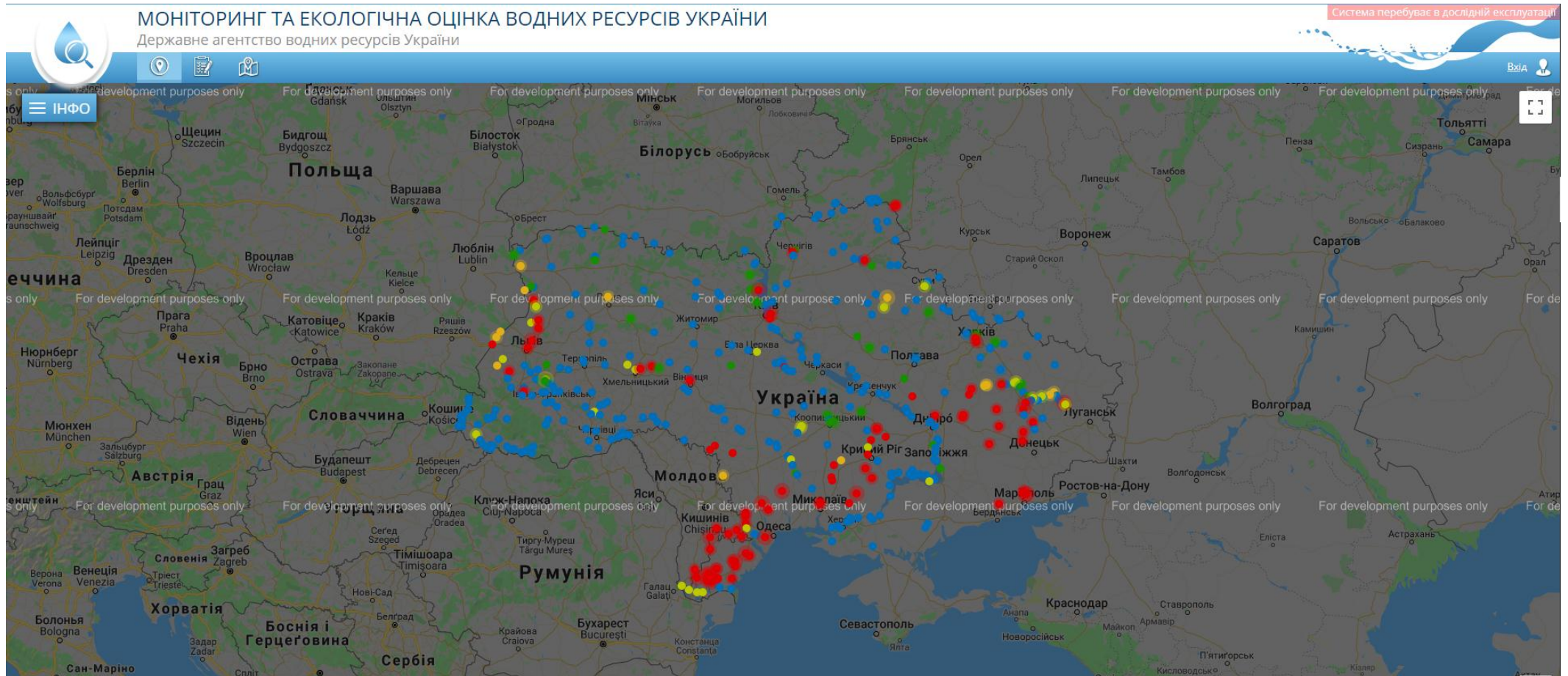
Об'єкт досліджень – процес автоматизації пошуку можливих джерел підвищеного антропогенного навантаження на річку за відкритими даними спостережень якості води.

Предмет дослідження – інформаційна технологія пошуку можливих джерел підвищеного антропогенного навантаження на річку за відкритими даними спостережень якості води.

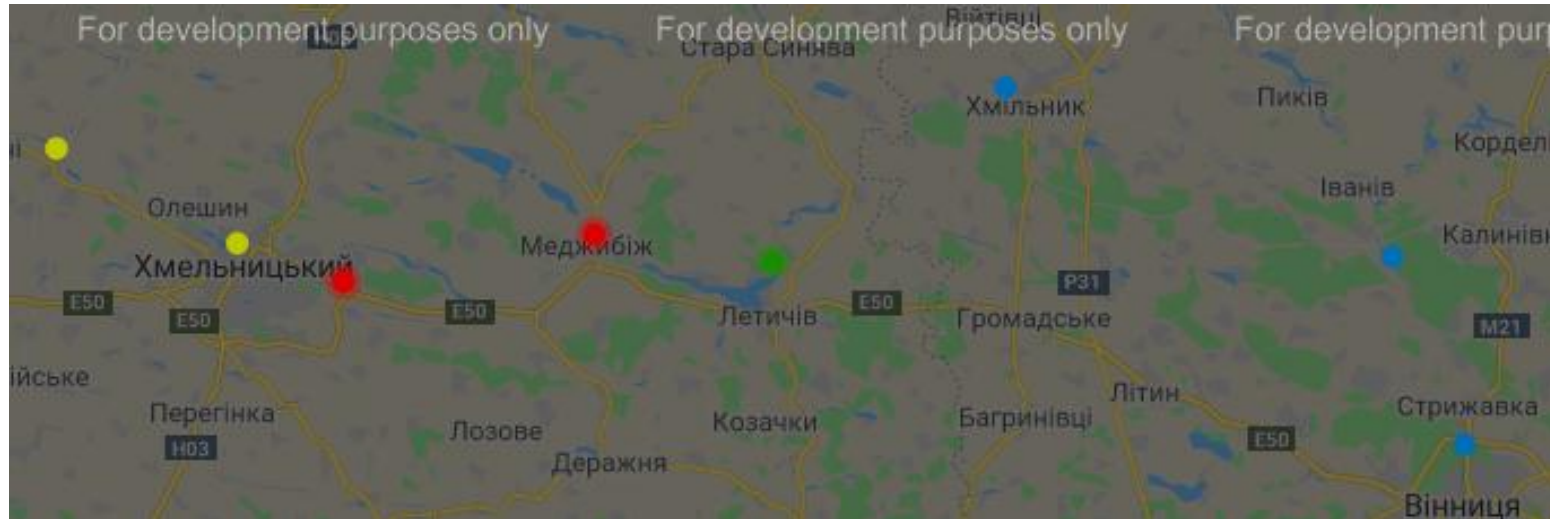
Основні задачі дослідження:

- Проаналізувати предметну область дослідження;
- Вибір моделі та методології розв'язання задачі;
- Розроблення математичного та алгоритмічного апарату виявлення моментів початку сталого актуального забруднення річки;
- Розроблення інформаційної технології пошуку місць та дат можливого антропогенного навантаження на річку за даними моніторингу якості вод;
- Тестування розробленої інформаційної технології на реальних даних.

Веб-портал «Моніторинг та екологічна оцінка водних ресурсів України»



Пости моніторингу води на річці Південний Буг від витоків річки до міста Вінниця



Огляд даних

Повна назва ПС: р. Південний Буг, 582 км, м. Вінниця, Сабарівське вдсх, питний в/з міста, вище міста
Район річкового басейну: Південний Буг
Назва головної лабораторії, що виконує ІЛВ: Лабораторія моніторингу вод та ґрунтів БУВР річки Південний Буг
Остання дата спостереження: 14.01.2020
[Дані моніторингу](#)

Показник	Фактичне значення	ГДК (ОБУВ)	Перевищення нормативу, раз
Азот загальний, мг/дм ³			
Біохімічне споживання кисню за 5 діб, мгО/дм ³	5,3	3	1,77
Завислі (суспендовані) речовини, мг/дм ³	8	15	Немає
Кисень розчинений, мгО ₂ /дм ³	10,5	4	Немає
Сульфат-іони, мг/дм ³	34,9	100	Немає
Хлорид-іони, мг/дм ³	30,5	300	Немає
Амоній-іони, мг/дм ³	0,18	0,5	Немає
Нітрат-іони, мг/дм ³	2,3	40	Немає
Нітрит-іони, мг/дм ³	0,05	0,08	Немає
Фосфат-іони (поліфосфати), мг/дм ³	0,094		

Приклад вікна показників моніторингу на порталі «Моніторинг та екологічна оцінка водних ресурсів України»

Звітність


- Звіти загального користування
 - Дані моніторингу (за адміністративно-територіальним принципом)
 - Дані моніторингу (за водогосподарською організацією)
 - Дані моніторингу (за ознакою транскордонного створу)
 - Дані моніторингу (за районом річкового басейну або суббасейну)

Звіти загального користування. Дані м


Дата з: *

01.01.2002 

Дата до: *

30.11.2020 

Басейн: *

Південний Буг 

Суббасейн: *

Всі 

Показник: *

Всі 

Виконати

Очистити

Закрити

Експорт даних моніторингу



Дані моніторингу (за районом річкового басейну або суббасейну)
з 01.01.2002 до 31.12.2020

Район річкового басейну: Південний Буг
Суббасейн: Всі

1. Пост: Балка Ташлик, 6 км. Ташлицьке вдсх., став-охолоджувач ПУ АЕС (міст через Ташлицьке вдсх., південь м. Південно-Українськ)

Значення	Показник								
	Амоній-іони, мг/дм3	Біохімічне споживання кисню за 5 діб, мГО2/дм3	Завислі (суспендовані) речовини, мг/дм3	Кисень розчинений, мГО2/дм3	Нітрат-іони, мг/дм3	Нітрит-іони, мг/дм3	Сульфат-іони, мг/дм3	Фосфат-іони (попіфосфати), мг/дм3	Хлорид-іони, мг/дм3
Дата									
09.01.2002	0.000	1.770	12.000	12.040	3.780	0.032	317.200	0.315	168.400
09.04.2002	0.093	1.020	5.600	11.100	5.900	0.043	352.000	0.170	143.600
10.07.2002	0.000	2.510	24.400	8.290	1.260	0.034	339.200	0.035	146.200
01.10.2002	0.180	1.760	2.000	9.670	1.800	0.038	390.400	0.080	144.800
13.01.2003	0.200	1.190	16.000	9.520	5.760	0.010	340.000	0.370	134.000
10.06.2003	0.050	1.800	12.800	10.940	2.350	0.018	371.200	0.010	128.600
12.08.2003	0.000	1.780	65.500	9.660	0.725	0.013	358.400	0.000	131.400
13.11.2003	0.040	1.570	2.200	12.710	5.180	0.032	358.000	0.092	143.400
12.02.2004	0.000	1.090	14.400	12.310	4.530	0.015	382.600	0.220	140.600
09.06.2004	0.000	0.380	22.800	12.010	1.640	0.053	412.500	0.010	140.380
27.09.2004	0.001	5.650	14.400	8.060	1.160	0.015	393.100	0.013	145.300
02.11.2004	0.000	1.160	7.200	8.070	1.120	0.010	412.900	0.110	136.120
13.01.2005	0.000	1.780	14.400	9.090	3.850	0.010	501.600	0.300	147.500
07.04.2005	0.028	1.340	17.600	9.390	3.480	0.038	498.240	0.110	125.130
20.07.2005	0.000	1.290	40.800	7.600	0.300	0.069	472.300	0.008	144.610
25.10.2005	0.021	1.240	12.800	7.810	1.250	0.044	435.200	0.014	137.600
09.02.2006	0.032	1.580	21.200	8.880	5.230	0.012	390.400	0.320	134.800
05.04.2006	0.000	1.300	25.200	9.140	5.810	0.021	441.600	0.200	112.020
28.08.2006	0.004	1.430	38.000	8.100	1.200	0.028	467.200	0.180	132.800
12.10.2006	0.027	1.040	7.600	8.820	0.000	0.000	384.000	0.006	126.200
14.03.2007	0.003	1.330	9.200	11.260	3.430	0.026	371.200	0.280	127.620
15.06.2007	0.000	1.110	4.000	7.940	1.340	0.027	390.400	0.020	128.300
10.09.2007	0.000	0.700	0.000	7.790	0.910	0.026	384.000	0.000	132.400
04.10.2007	0.004	1.100	4.800	7.920	1.240	0.000	390.000	0.121	130.100
13.03.2008	0.000	0.900	4.800	8.690	2.350	0.028	345.600	0.290	133.800
04.06.2008	0.004	0.630	20.400	8.400	1.480	0.012	403.200	0.022	134.700
02.09.2008	0.006	1.810	35.200	8.740	0.790	0.011	377.600	0.022	136.100
12.11.2008	0.046	2.520	10.400	8.520	1.950	0.009	358.400	0.177	126.700
04.03.2009	0.000	1.650	11.200	13.240	3.550	0.006	320.000	0.252	112.700
19.05.2009	0.062	2.440	10.000	14.120	3.630	0.029	313.600	0.026	118.900
01.09.2009	0.000	1.100	10.000	7.950	0.330	0.029	307.200	0.013	125.090
03.11.2009	0.000	1.010	6.400	8.020	2.130	0.028	294.400	0.032	126.200
10.02.2010	0.000	1.110	9.200	9.670	1.640	0.007	275.200	0.300	121.800
12.05.2010	0.180	0.500	8.000	10.330	3.760	0.300	345.600	0.057	121.700

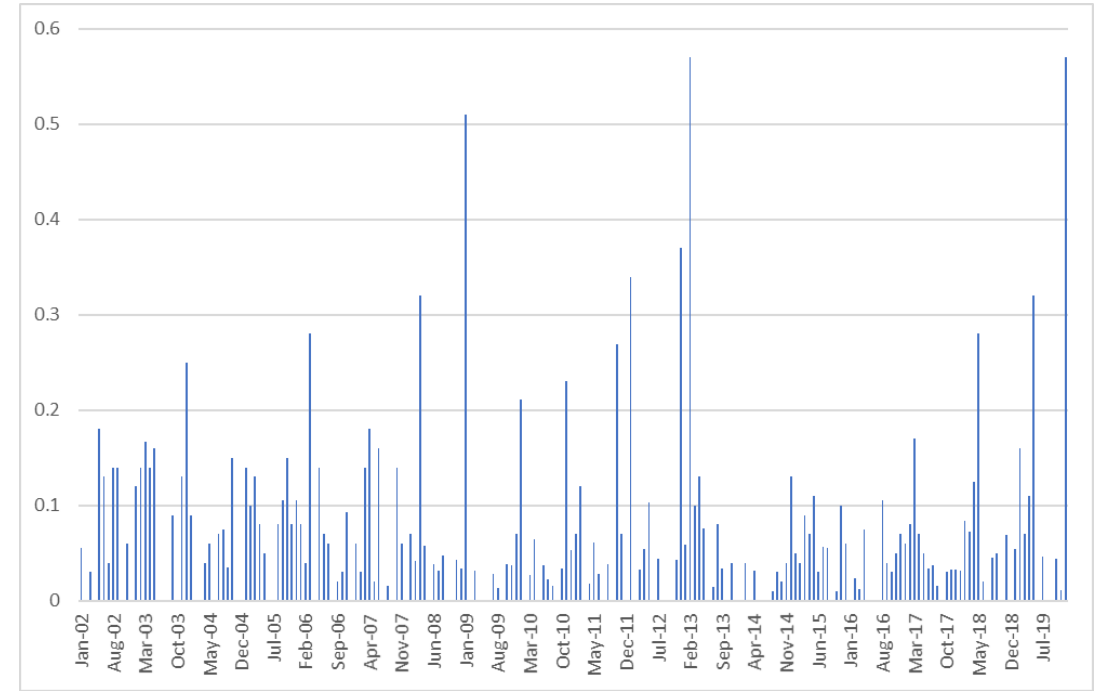
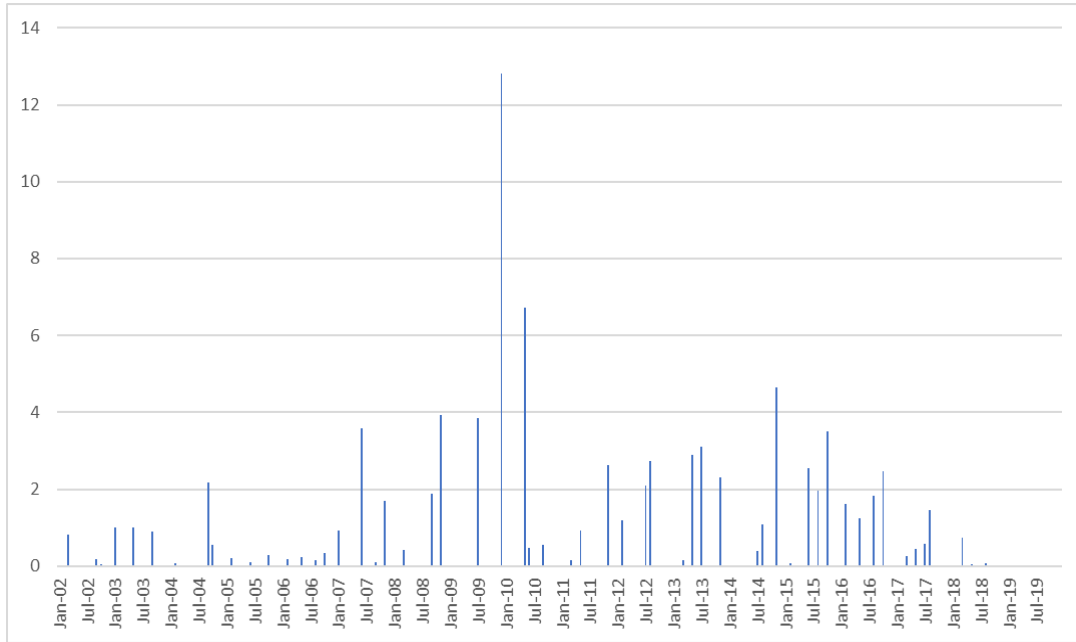
Огляд даних

Приклад сформованого звіту для басейну річки Південний Буг

Огляд даних

	A	B	C	D	E	F	G	H	I	J	K
1		NO2	NO2	NO2	NO2	NO2	NO2	NO2	NO2		
2	ID_Station	27	28	29	30	31	32	33	34		
3	Date_Month										
4	2002-1	0.055	0.085	0.08							
5	2002-10						0.0415	0.0745	0.155		
6	2002-11	0.06	0.075	0.066667							
7	2002-12				0.137333	0.39					
8	2002-3	0.03	0.05				0.81	0.04	0.03		
9	2002-4							0.05	0.069		
10	2002-5	0.18	0.153333	0.193333	0.02	0.66					
11	2002-6	0.13	0.12	0.13							
12	2002-7	0.04	0.11	0.08							
13	2002-8	0.14	0.18	0.06							
14	2002-9	0.14	0.5	0.52	0.11	0.04	0.19				
15	2003-1	0.12	0.08								
16	2003-11	0.13		0.09	0.08	0.1					
17	2003-12	0.25	0.12	0.07	0.09	0.14					
18	2003-2	0.14	0.145	0.095	0.1		1				
19	2003-3	0.166667	0.12	0.246667		0.42		0.11	0.09		
20	2003-4	0.14	0.19	0.115							
21	2003-5	0.16	0.08	0.05			1	0.09	0.1		

Приклад сформованого набору даних для значення показника NO₂ для станцій вздовж р. Південний Буг



Порівняння значень NO₂ для постів «Хмельницький» та «Вінниця»

Prophet

Prophet – пакет програм для мов Python та R, розроблений компанією Facebook. Він дозволяє працювати з періодичними та неперіодичними часовими рядами, виконувати прогнозування для заданого інтервалу тощо. Основними перевагами є:

- працює з періодичними і неперіодичними рядами даних;
- допускає чималу кількість пропущених даних;
- дає можливість додавати довільні складові сезонності;
- враховує «holidays» (свята) і вікна їх впливу, тобто дати аномальних значень;
- використовує і лінійну, і логістичну модель тренду на різних інтервалах, які обмежуються точками зміни тренду(changepoints);
- здійснює прогнозування даних на заданий інтервал з побудовою зони невизначеності.



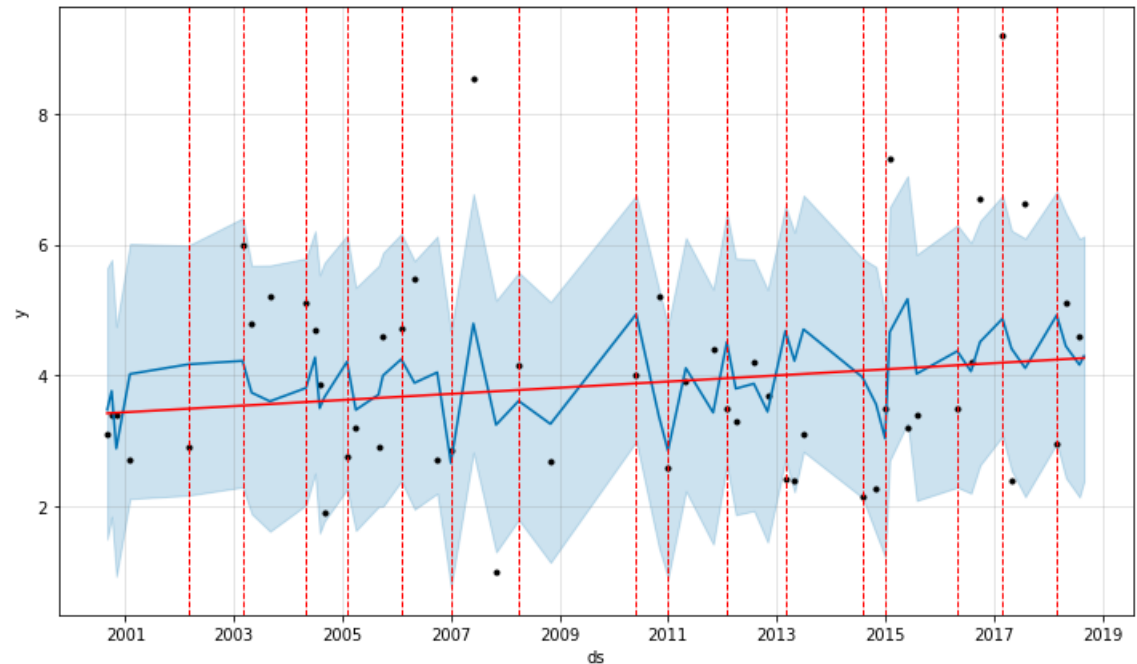
Запропонована методологія розв'язання задачі

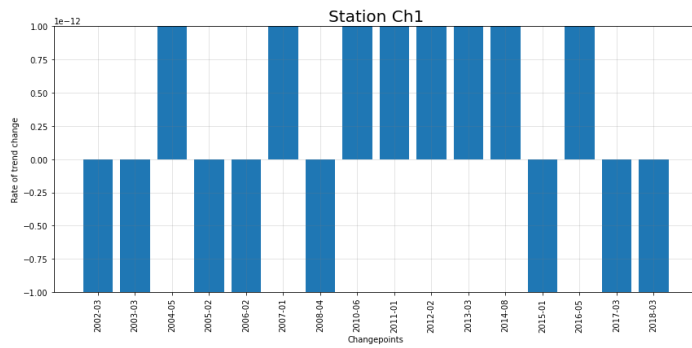
1. Вибираємо параметри моделі Prophet
2. Встановлюємо точки зміни тренду («changepoints», далі «CP») на 1 січня кожного року і будуємо лінійну модель Prophet між ними для кожного показника і кожної станції окремо, таким чином, щоб вона дозволяла виявляти початок забруднюючого впливу джерел із тренувальної вибірки.
 - 2.1. Аналізуємо нахил тренду між кожними «CP» та шукаємо серед них найбільші прирости, але такі які демонструють монотонне зростання. Результатом аналізу є приріст забруднення у дату, з якої все почалось, решта приростів, прирівнюються до нуля.
 - 2.2. Усім місяцям, які потрапляють в діапазон між «CP», ставиться у відповідність значення цього імпульсу. Зазвичай, по різних показниках після моделювання CP будуть мати різні значення. Тому такий підхід дозволить їх співставлення з точністю до місяця.
 - 2.3. Масштабуємо у діапазон $[0, 1]$ і додаємо ці виявлені забруднюючі імпульси по різних показниках по кожному посту.

Результат роботи моделі Prophet

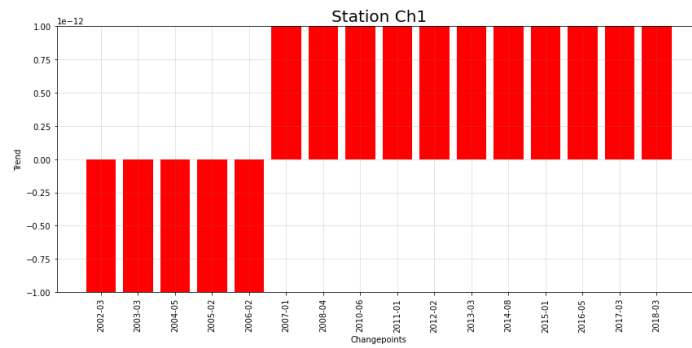
$$\hat{y}_i = (k + \Delta k_i) d_i, i = \overline{1, m},$$

де \hat{y}_i – стандартизоване значення тренду в i -й точці зміни d_i , тобто масштабоване між мінімальним і максимальним значеннями та поділене на середньоквадратичне відхилення – це стандартне позначення пакету програм Prophet

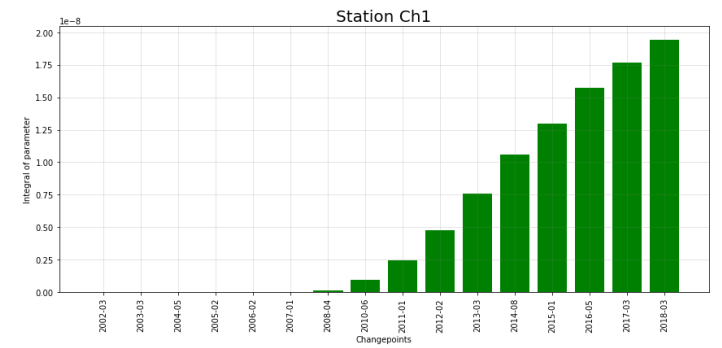




Δk_i тренду між точками зміни d_i



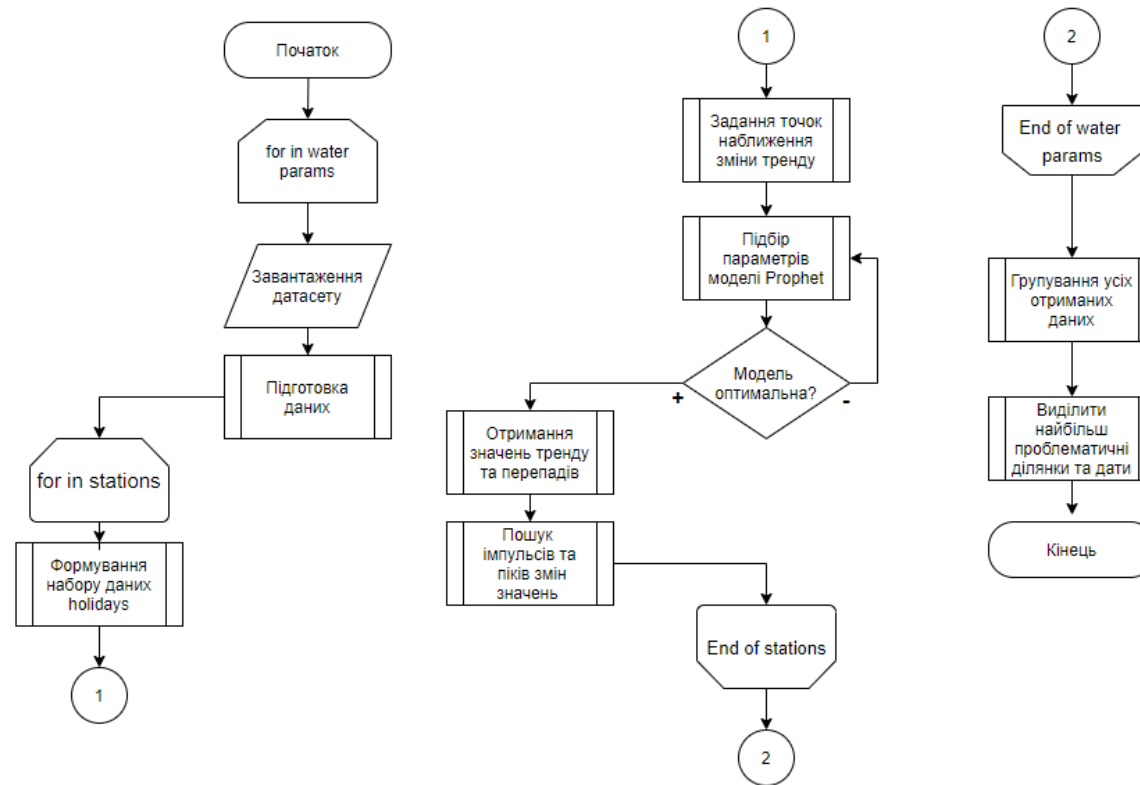
$$S = \sum_{i=1}^m \Delta k_i,$$



$$J = \sum_{j=i}^m S_i = \sum_{q=j}^m \sum_{j=i}^m \Delta k_{qj},$$

Пошук змін тренду

Блок схема алгоритму роботи інформаційної технології



Набір імпульсів змін тренду по датах

ds	V_total	K_total	ChV_total	Ch4_total	Ch3_total	Ch2_total	Ch1_total	Ch0_total
2002-01-31	0.019463	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2002-02-28	0.019463	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2002-03-31	0.019463	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2002-04-30	0.019463	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2002-05-31	0.019463	0.000000	0.000000	0.000000	0.005274	0.000000	0.000000	0.000000
...
2019-03-31	0.327736	0.182022	0.025933	0.089293	0.386242	0.155707	0.562512	0.090579
2019-04-30	0.327736	0.182022	0.025933	0.089293	0.386242	0.155707	0.562512	0.090579
2019-05-31	0.327736	0.182022	0.025933	0.089293	0.386242	0.155707	0.562512	0.090579
2019-06-30	0.327736	0.182022	0.025933	0.089293	0.386242	0.155707	0.562512	0.090579
2019-07-31	0.327736	0.182022	0.025933	0.089293	0.386242	0.155707	0.562512	0.090579

	V_total	K_total	ChV_total	Ch4_total	Ch3_total	Ch2_total	Ch1_total	Ch0_total
ds								
2002	0.233558	0.000000	0.000000	0.000000	0.042192	0.000000	0.000000	0.000000
2003	0.314363	0.000000	0.000000	0.000000	0.086598	0.000000	0.000000	0.000000
2004	0.402775	0.000000	0.000000	0.000000	0.167494	0.000000	0.000000	0.000000
2005	0.402775	0.000000	0.000000	0.000000	0.359179	0.000000	0.067918	0.000000
2006	0.402775	0.000000	0.000000	0.000000	0.583785	0.000000	0.074092	0.000000
2007	0.700619	0.000000	0.311196	0.000000	0.604204	0.000000	0.115861	0.000000
2008	1.782258	0.000000	0.311196	0.000000	0.604204	0.000000	0.208265	0.000000
2009	2.042791	0.385680	0.311196	0.000000	0.604204	0.035482	0.239066	0.000000
2010	2.295605	0.410290	0.311196	0.000000	0.848374	0.085157	0.492680	0.000000
2011	3.296659	0.938343	0.311196	0.458333	1.269685	0.514657	1.473536	0.000000
2012	3.570756	1.193577	0.311196	1.023895	1.543780	0.844403	1.916383	0.000000
2013	3.932827	1.650828	0.311196	1.071522	2.544332	1.258741	2.437039	0.000000
2014	3.932827	2.139814	0.311196	1.071522	2.919797	1.602051	2.790137	0.000000
2015	3.932827	2.184267	0.311196	1.071522	3.594352	1.868489	4.754254	1.086947
2016	3.932827	2.184267	0.311196	1.071522	3.960554	1.868489	5.852971	1.086947
2017	3.932827	2.184267	0.311196	1.071522	4.549823	1.868489	6.665042	1.086947
2018	3.932827	2.184267	0.311196	1.071522	4.634899	1.868489	6.750139	1.086947
2019	2.294149	1.274156	0.181531	0.625054	2.703691	1.089952	3.937581	0.634053

Згруповані по роках імпульси змін тренду

Висновки

Під час виконання магістерської кваліфікаційної роботи розглянуто проблему пошуку основних антропогенних джерел забруднення водойм за регулярними багаторічними даними моніторингу вод України.

Проаналізовано наявні відкриті дані державної системи моніторингу України за 1920-2020 рр. і було виявлено наявність чималих пропусків та аномальних даних. Для вирішення поставленої задачі було обрано модель Prophet, передусім, через її можливості роботи із зашумленими даними з певною періодичністю.

Розроблено й охарактеризовано математичний і алгоритмічний апарат запропонованої технології та створена за рахунок використання моделі Prophet з кусково-лінійною апроксимацією тренду для локалізації в часі та просторі забруднюючого впливу по похідній цього тренду. Доведено, що особливістю запропонованої технології є її стійкість до аномальних завад з різкою зміною якості чи кількості води.

Створена програма для реалізації цієї технології на Python. Проведено її успішне випробування на ділянці р. Південний Буг на реальних даних 8 постів державної системи моніторингу вод від витoku р. Південний Буг до м. Вінниці за даними 2002—2019 рр. по показниках БСК5 (біохімічне споживання кисню за 5 днів), концентрація нітратів NO₂ та концентрація азоту амонійного NH₄, за якими останніми роками часто фіксуються наднормативні значення.

Наукова новизна

Наукова новизна, отримана в результаті проведеного дослідження, полягає у тому, що дістала подальшого розвитку інформаційна технологія пошуку можливих джерел підвищеного антропогенного навантаження на річку за відкритими даними за рахунок використання моделі Prophet з кусково-лінійною апроксимацією тренду для локалізації в часі та просторі забруднюючого впливу по похідній цього тренду. Особливістю запропонованої технології є її стійкість до аномальних завад з різкою зміною якості чи кількості води (аварій, межені, повені тощо), які спеціальним чином враховує блок параметрів «holidays» моделі Prophet, та зміни водності річки, за рахунок аналізу багаторічних даних та виявлення сталих трендів, що мають місце тривалий час.

Практичне значення

Практичне значення одержаних результатів полягає у наступному:

- розроблено математичний і алгоритмічний апарат інформаційної технології пошуку можливих джерел підвищеного забруднення річки;;
- за допомогою пакету програм Prophet реалізовано програмний інструментарій, що базується на розробленому алгоритму;
- проведено успішне тестування на реальних даних 8 постів державної системи спостереження ділянки р. Південний Буг, від її витoku до м. Вінниця за 2002-2019 роки;

За результатами магістерської кваліфікаційної роботи опубліковано статтю в науковому фаховому періодичному виданні «Вісник Вінницького політехнічного інституту» категорії Б зі спеціальності 126 Інформаційні системи та технології, яке входить в Index Copernicus.

Дякую за увагу