

# PODSTAWY OPTIMALNEGO SZACOWANIA ŁADUNKÓW SUBSTANCJI ROZPUSZCZONYCH W ODPLYWIE RZECZNYM

Alfred Stach

**Stach A., 2003:** Podstawy optymalnego szacowania ładunków substancji rozpuszczonych w odpływie rzeczonym (*Fundamentals of optimal solute substances loads estimation in river outflow*), Regionalny Monitoring Środowiska Przyrodniczego nr 4, s. 57-71, Kieleckie Towarzystwo Naukowe, Kielce.

**Zarys treści:** Głównym celem studium jest porównanie szerokiego zakresu metod estymacji ładunków substancji rozpuszczonych w odpływających wodach rzecznych, przy wykorzystaniu parametrów populacji i bieżących wartości. Zbiór danych analizowanych dotyczył dwóch lat codziennych pomiarów w dwóch kontrastowych zlewniach: nizinnej rzeki Parsęty i górskiej rzeki Bystrzanki. Każdy zbiór danych był porównywalny w siedmiu seriach pomiarowych w tygodniowych interwałach przesuniętych o jeden dzień. Stwierdzono, że żadna z metod testowanych nie może być uważana za uniwersalną i najlepszą w każdych warunkach. Wyraźną poprawę jakości wyników uzyskuje się po zastosowaniu procedur KED, cOCK, CSOL, cCSPL, w mniejszym stopniu IL, OK, LRSPL. Najgorszą metodą do uzupełniania ciągów pomiarów przepływów fizykochemicznych wody jest regresja w stosunku do przepływów. Zaskakująco wysokie miejsce w rankingu dokładności i precyzji zajmuje interpolacja liniowa. Do rutynowej estymacji stężeń składników roztworu rzecznoego zaleca się sześciennne funkcje krzywiznowe (CSPL), dające nieco lepsze wyniki od IL. W końcowym zestawieniu dokładności i precyzji najlepszy jest kolokacyjny kokriging (cOCK), nadający się do generowania „syntetycznych” ciągów stężeń i jako podstawa selekcji optymalnej metody obliczeń składowych roztworu rzecznoego.

*Alfred Stach*, Uniwersytet im. Adama Mickiewicza, Instytut Badań Czwartorzędu i Geoekologii,  
ul. Fredry 10, 61-710 Poznań, frdstach@amu.edu.pl

## 1. Wprowadzenie

Szacunek chwilowych i okresowych stężeń i ładunków substancji rozpuszczonych w wodach rzecznych to podstawa wielu prac hydrologicznych, geomorfologicznych i ekologicznych. Bazują one najczęściej nie na ciągłej rejestracji, ale na wynikach analiz próbek wody pobieranych z bardzo różną frekwencją. Częstość próbkowania nie jest zazwyczaj dostosowana do specyfiki hydrologicznej badanego obiektu, ale jest adaptacją typowych metodyk do możliwości finansowych, technicznych i organizacyjnych (Littlewood 1995).

Depozycja atmosferyczna monitorowana jest w sposób ciągły lub kumulacyjny. Można przyjąć – szczególnie w przypadku małej zlewni nizinnej, gdzie zmienność przestrzenna jest znikoma – że ten sposób

pomiaru gwarantuje potencjalnie maksymalną możliwą dokładność jej estymacji. Inaczej jest w przypadku odpływu. Próbkki wody z ciekłu do oznaczeń makro- i mikroskładników pobierane są w odstępach tygodniowych, dwutygodniowych lub nawet miesięcznych. Dokładność określenia średnich stężeń i ładunków składników chemicznych w odpływie rzecznoym jest więc trudna do oszacowania, a na dodatek zmienna w czasie, bo uzależniona od dynamiki przepływów i specyfiki konkretnej zlewni. Średnie okresowe stężenia oblicza się najczęściej metodą średniej ważonej w stosunku do przepływu. Do estymacji chwilowych wartości, a w konsekwencji także do obliczenia średniej okresowej – stosuje się najczęściej funkcję regresji pomiędzy przepływem a stężeniem. Podobnie jest z ładunkami. Ograniczając się do stosowania jedynie regresji ignorujemy istnienie silnej autokorelacji czasowej stężeń i przepływów.

Zasadniczym problemem oceny jakości różnych metod obliczania ładunku substancji rozpuszczonych transportowanych w rzece jest brak odpowiednich danych referencyjnych (Webb et al. 1997, 2000). W przypadku ładunku zawiesin jest on łatwiejszy do rozwiązania, ponieważ istnieje możliwość wykorzystania stosunkowo dokładnych, ciągłych nefelometrycznych pomiarów mętności wody (Webb et al. 1997). Webb i in. (2000) przedstawili propozycję takiej oceny dla metod kalkulacji ładunku składników roztworu rzeczno opartą o syntetyczne, generowane „sztucznie” ciągi pomiarowe stężeń. Idea jest niewątpliwie dobra, lecz do jej realizacji można mieć nieco zastrzeżeń. Podstawą algorytmu generującego ciągi stężeń jest bowiem regresja w stosunku do przepływu. Tak wyliczone wartości stężeń były dość arbitralnie, jak wynika z opisu, modyfikowane, aby uwzględnić opóźnienia „chemogramu” w stosunku do hydrogramu w trakcie wezbrań.

W niniejszym opracowaniu proponuje się wykorzystanie do estymacji stężeń chwilowych metod geostatystycznych (geostatystyka – inaczej Teoria Zmiannych Regionalnych, zob: Goovaerts 1997; Wackernagel 1998) i funkcji krzywkowych (de Boor 1978). Zostały one skonfrontowane z tradycyjnie stosowanymi: interpolacją liniową i regresją w stosunku do przepływu. Do testowania ich skuteczności posłużyły wyniki codziennych pomiarów hydrologicznych z kontrastowych zlewni: nizinnej i górskiej.

## **2. Szacowanie ładunków materiału transportowanego przez rzeki**

Do szacowania ładunków materiału transportowanego w odpływie rzeczno zaproponowano już wiele formuł obliczeniowych (tab. 1). Obejmują one między innymi metody interpolacyjne oparte na założeniu, że stężenie zmierzone w próbce wody jest reprezentatywne dla okresu pomiędzy kolejnymi próbkowaniami. Drugą klasę stanowią metody ekstrapolacyjne oparte o zależności regresyjne pomiędzy stężeniem a chwilowym przepływem w momencie poboru próbki. Dysponując danymi z rejestracji przepływów i ustalonym równaniem regresji możemy na tej podstawie łatwo obliczyć ładunki (odpływ substancji). Ponieważ zależności pomiędzy przepływem a stężeniem różnych składników roztworu rzeczno mają najczęściej charakter potęgowy (bo zmienne charakteryzują się rozkładem logarytmiczno-normalnym), dane pomiarowe są przed obliczeniami regresji loga-

rytmowane. Procedura logarytmowania danych i antylogarytmowania wyników powoduje, że szacowane stężenia są obciążone pewnym błędem. W niektórych metodach ekstrapolacyjnych (tab. 1 – metoda 8 i 9) stosowane są zatem procedury korekcyjne, mające na celu eliminację tych błędów.

Wzory 2 i 5 z tabeli 1 są na przykład zalecane przez Komisję Paryską zajmującą się dopływem zanieczyszczeń do Morza Północnego (Littlewood 1995). W Wielkiej Brytanii w monitoringu jakości wód prowadzonym przez U.K. Department of the Environment stosowany jest wzór 2.

W niektórych pracach wskazywano, że ładunek roztworów może być określony stosunkowo dokładnie na podstawie próbek pobieranych w odstępach tygodniowych lub nawet rzadszych (Webb et al. 2000). Dotyczy to zwłaszcza makroskładników w stosunkowo słabo zanieczyszczonych rzekach nizinnych. Z drugiej jednakże strony standardowe metody dają wyniki znacznie mniej dokładne i precyzyjne w przypadku składników roztworu wykazujących znaczne i niesystematyczne wahania stężeń w trakcie wezbrań, oraz takich, których transport jest związany z silną adsorbcją przez zawiesinę. Wyniki wstępnego opracowania Webba i in. (1997) wskazywały, że w rzekach wschodniej Anglii, w zależności od użytej procedury obliczeniowej, dla 20 z 36 analizowanych składników roztworu rzeczno różnice między minimalną, a maksymalną oceną ładunków przekraczały 50%, a 5 – 100% (w tym zwłaszcza bardzo toksyczny arsen).

Obiektywna selekcja optymalnej metody kalkulacji ładunków składników roztworu rzeczno dla konkretnego przekroju wodowskazowego wiązałaby się z koniecznością wykonania stosunkowo długiej serii (minimum 1 rok) pomiarów o wysokiej frekwencji. Częstość pobierania próbek wody musiałaby być co najmniej dobową, z dodatkowym opróbowaniem wysokich przepływów wezbraniowych. Być może okazałoby się, że dla różnych składników roztworu należy używać różnych procedur obliczania ładunku, nie wykluczone, że także zależałoby to od pory roku. W większości programów monitoringowych, czy też badawczych nie ma możliwości finansowych i organizacyjnych aby zastosować wyżej omówione podejście.

W takiej sytuacji rozsądną alternatywą jest rozwiązanie zaproponowane przez Webba i in. (2000), a oparte na generowaniu „syntetycznych” serii pomiarów stężeń, próbkowanych następnie w różnych interwałach czasowych. Dzięki temu można by teore-

tycznie dla każdego stanowiska i składnika chemicznego oszacować optymalną frekwencję pomiarów, i najlepszą, bo dającą najmniejszy błąd, procedurę obliczania ładunków. Dane wyjściowe do symulacji stanowiły 15-minutowe wartości z ciągłej rejestracji przepływów, oraz stężenia składników roztworu rzecznoego z 5-letniej serii tygodniowych lub rzadszych poborów próbek. Pierwszym etapem obliczeń było ustalenie sezonowych (zima/lato) funkcji regresji pomiędzy zlogarytmowanymi danymi przepływów i stężeń. Funkcje te były modyfikowane interaktywnie przez operatora programu komputerowego dla każdego wezbrania przez podanie czasu opóźnienia/wyprzedzenia chemogramu w stosunku do hydro-

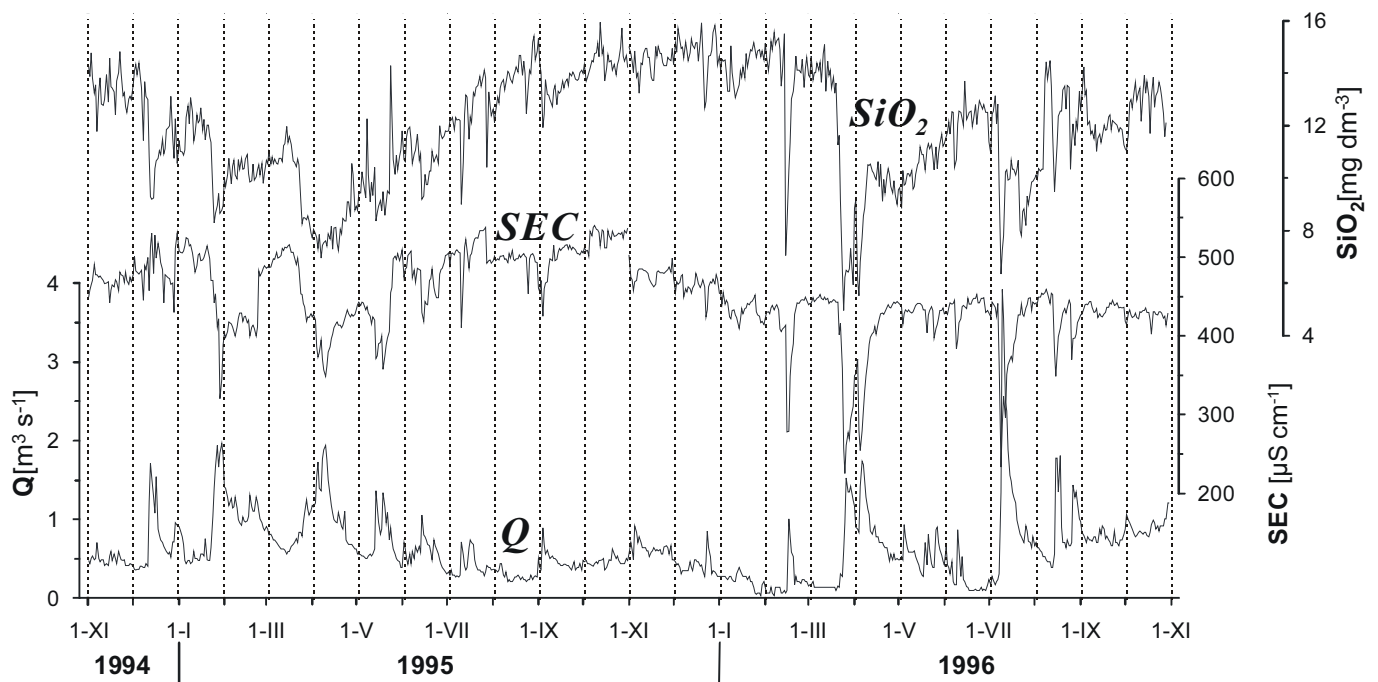
gramu. Na tej podstawie oszacowano przebieg stężeń analizowanych parametrów co 15 minut. Najślabszym ogniwem tego niewątpliwie bardzo ciekawego pomysłu są arbitralnie ustalone wartości czasu opóźnień i wyprzedzeń między chemogramem i hydrogramem. Można również mieć wątpliwości, czy najlepszym sposobem estymacji stężeń jest użycie równań regresji.

### 3. Obszar badań i analizowane dane

Opracowane dane obejmują dwuletnie serie dobowych pomiarów przepływu i wybranych parametrów fizykochemicznych dla zlewni górnej Parsęty (73,4 km<sup>2</sup>, ryc. 1) i Bystrzanki (13,0 km<sup>2</sup>, ryc. 2). Pierwsza

Tab. 1. Metody obliczania ładunku (TL) substancji transportowanych przez rzekę (Webb i in. 2000)  
Table 1. Methods for calculating river loads (fluxes) (Webb et al. 2000)

Metoda	Wzór	Objaśnienie
Interpolacja		
1	$TL = K \left( \sum_{i=1}^{ns} \frac{C_i}{ns} \right) \left( \sum_{i=1}^{ns} \frac{Q_i}{ns} \right)$	<p><math>K</math> to współczynnik przeliczeniowy związany z długością okresu pomiarowego,  <math>\bar{Q}_r</math> to średni przepływ w okresie pomiarowym,  <math>\bar{Q}_p</math> to średni przepływ w okresie pomiędzy poborem próbek,  <math>C_i</math> i <math>Q_i</math> to stężenie (koncentracja) mierzonego parametru w próbce, oraz przepływ w momencie poboru próbki,  <math>ns</math> to ilość próbek</p>
2	$TL = K \sum_{i=1}^{ns} \left( \frac{C_i \cdot Q_i}{ns} \right)$	
3	$TL = K \sum_{i=1}^{ns} (C_i \cdot \bar{Q}_p)$	
4	$TL = K \left( \sum_{i=1}^{ns} \frac{C_i}{ns} \right) \bar{Q}_r$	
5	$TL = K \frac{\sum_{i=1}^{ns} C_i \cdot Q_i}{\sum_{i=1}^{ns} Q_i} \bar{Q}_r$	
6 – Estymator relacji Beale’go (CF3)	$CF3 = \left( \frac{1 + \frac{1}{ns} \frac{S_{lq}}{\bar{l} \cdot \bar{q}}}{1 + \frac{1}{ns} \frac{S_q^2}{\bar{q}^2}} \right)$ $S_{lq} = \left( \frac{1}{ns-1} \right) \left( \sum_{i=1}^{ns} (Q_i^2 \cdot C_i) - ns \cdot \bar{q} \cdot \bar{l} \right)$ $S_q^2 = \left( \frac{1}{ns-1} \right) \left( \sum_{i=1}^{ns} (Q_i^2) - ns \cdot \bar{q}^2 \right)$	<p><math>C_i</math>, <math>Q_i</math> i <math>ns</math> oznaczają to samo co powyżej,  <math>\bar{l}</math> to średni ładunek dla momentów w których pobierano próbki, a <math>\bar{q}</math> to średni przepływ dla tych samych momentów</p>
Ekstrapolacja		
7 – regresja danych zlogarytmowanych	$C_i = a \cdot Q_i^b$	$C_i$ i $Q_i$ to chwilowe wartości stężeń i przepływów w momentach próbkowania; całkowity ładunek liczony jest poprzez zastosowanie równania regresji do ciągłych (częstych) danych przepływu
8 – skorygowana zależność regresyjna (CF1) wg Fergusona	$CF1 = \exp(2,651 \cdot s^2) a \cdot Q_i^b$	$s$ to błąd standardowy estymacji równaniem regresji wyrażony w wartościach logarytmicznych
9 – współczynnik korekcyjny (CF2) Duana	$CF2 = \left( \frac{1}{n} \right) \sum_{i=1}^n 10^{e_i}$ $e_i = \log(C_i) - \log(Ce_i)$	Wyliczone z równania regresji wartości stężenia są mnożone przez $CF2$ , $C_i$ to stężenie w konkretnej próbce, $Ce_i$ to stężenie estymowane z regresji przy zastosowaniu wartości przepływu z momentu poboru próbki, $n$ to ilość pomiarów



Ryc. 1. Analizowane serie dobowych pomiarów przepływu ( $Q$ ), przewodnictwa elektrolitycznego wody ( $SEC$ ) i stężenia zjonizowanej krzemionki ( $SiO_2$ ) w latach hydrologicznych 1995-96 ze zlewni górnej Parsęty  
 Fig. 1. Analysed daily measurements data of discharge ( $Q$ ), water electrical conductivity ( $SEC$ ) and concentration of ionised silica ( $SiO_2$ ) from upper Parsęta catchment in 1995-96 hydrological years

z nich jest typowa dla obszarów pojezierzy Polski północno-zachodniej, druga dla Karpat Fliszowych. Szczegółowe charakterystyki obu zlewni zawarte są w publikacjach Kostrzewskiego i in. (1994a, b), Soi (1981), Starkla i Gila (1994) oraz Welca (1985). Dla górnej Parsęty analizowano przewodnictwo elektryczne wody ( $SEC$ ) i stężenie zjonizowanej krzemionki ( $SiO_2$ ), dla Bystrzanki –  $SEC$  i stężenie jonów wodorowych (odczyn wody – pH). Wszystkie wymienione powyżej dane pomiarowe zostały zebrane w ramach Zintegrowanego Monitoringu Środowiska Przyrodniczego zgromadzone w Centralnej Bazie Danych ZMŚP.

#### 4. Podstawy teoretyczne

Jeśli celem pracy jest oszacowanie wartości ciągłej cechy (parametru)  $z$  w nieopróbowanym momencie czasu  $\mathbf{u}$  przy wykorzystaniu istniejących danych pomiarowych  $z\{\mathbf{u}_\alpha, \alpha=1, \dots, n\}$ , to wszystkie estymatory geostatystyczne są wariantami podstawowej formuły regresji liniowej  $Z^*(\mathbf{u})$  zdefiniowanej w następujący sposób (Goovaerts 1997, wzór 1):

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) [Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)] \quad (1)$$

gdzie  $\lambda_\alpha(\mathbf{u})$  to waga przyporządkowana pomiarowi  $z(\mathbf{u}_\alpha)$  interpretowanemu jako realizacja zmiennej loso-

wej  $Z(\mathbf{u}_\alpha)$  i zlokalizowana w określonym przedziale czasu  $W(\mathbf{u})$  ze środkiem w  $\mathbf{u}$ . Wagi  $n(\mathbf{u})$  są tak obliczane, aby zminimalizować szacowaną wartość bądź wariancję błędu  $\sigma_E^2(\mathbf{u}) = Var\{Z^*(\mathbf{u}) - Z(\mathbf{u})\}$  przy założeniu braku obciążenia estymatora. Wagi te są uzyskiwane poprzez rozwiązanie układu równań liniowych znanych jako „kriging system” (2). Zasadnicze znaczenie mają w nim wyliczone z modelu wartości semiwariancji  $\gamma$  dla dowolnego odstępu czasu między pomiarem wykonanym a szacowanym (ryc. 3 i 4).

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta(\mathbf{u}) \gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu(\mathbf{u}) = \gamma(\mathbf{u}_\alpha - \mathbf{u}), & \alpha = 1, \dots, n(\mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta(\mathbf{u}) = 1 \end{cases}$$

(2)

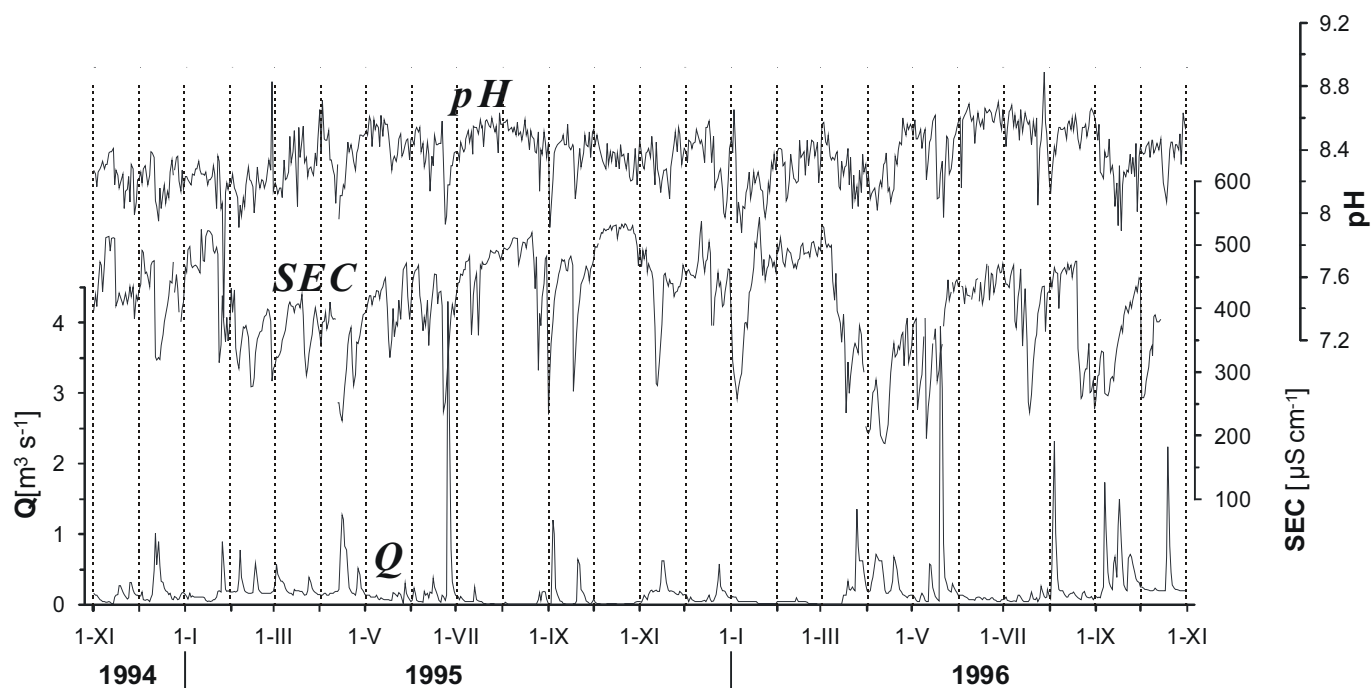
Różnice pomiędzy wariantami krigingu jednej zmiennej (*univariate kriging*) związane są z różnymi modelami trendu  $m(\mathbf{u})$  we wzorze 1. Przywołamy tutaj dwa z nich:

1) prosty kriging (*simple kriging*, SK) opiera się na założeniu, że średnia wartość  $m(\mathbf{u})$  jest znana i stała w całym analizowanym przedziale czasu  $\mathcal{A}$ :

$$m(\mathbf{u}) = m \quad \text{znane } \forall \mathbf{u} \in \mathcal{A}$$

2) zwykły kriging (*ordinary kriging*, OK) uwzględnia okresowe fluktuacje zmiennej ograniczając domenną stacjonarności średniej do przedziału czasu (sąsiedztwa)  $W(\mathbf{u})$ :

$$m(\mathbf{u}') = \text{constant} \quad \text{lecz nieznanne } \forall \mathbf{u}' \in W(\mathbf{u})$$



Ryc. 2. Analizowane serie dobowych pomiarów przepływu ( $Q$ ), przewodnictwa elektrolitycznego ( $SEC$ ) i odczynu wody ( $pH$ ) w latach hydrologicznych 1995-96 ze zlewni Bystrzanki

Fig. 2. Analysed daily measurements data of discharge ( $Q$ ), water electrical conductivity ( $SEC$ ) and reaction ( $pH$ ) from upper Bystrzanka catchment in 1995-96 hydrological years

w przeciwieństwie do prostego krigingu średnia jest tutaj uważana za nieznaną.

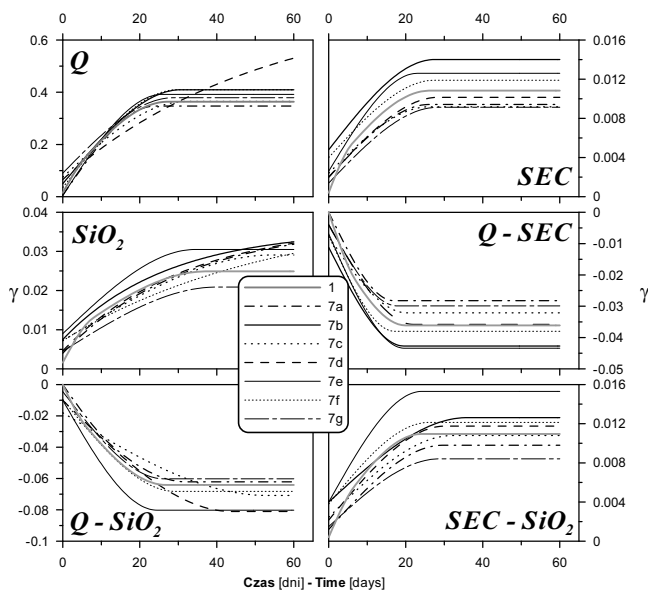
Kiedy wyników pomiarów jest niewiele i/lub są słabo ze sobą skorelowane, szacunek wartości interesującego nas parametru (cechy) może być znacznie bliższy rzeczywistości kiedy uwzględnimy dodatkowe informacje z pomiarów innego, wykazującego podobny przebieg czasowy (skorelowanego) parametru. Przy analizach transportu fluwialnego najczęściej dysponujemy wynikami „ciągłych” (z samopisów) lub bardzo częstych pomiarów przepływu i wykonywanych znacznie rzadziej pomiarów koncentracji transportowanej przez wodę substancji. Bywa również tak, że często mierzymy in situ (czasami automatycznie) „proste” parametry fizykochemiczne wody jak przewodnictwo elektrolityczne czy  $pH$ , natomiast próbki wody przeznaczone do przeprowadzenia wyrafinowanych laboratoryjnych analiz składu chemicznego pobierane są ze znacznie mniejszą częstotliwością. Reasumując, dysponujemy zazwyczaj wynikami pomiarów dodatkowych parametrów, które pochodzą zarówno z terminów pomiarów parametru szacowanego, jak i ze wszystkich terminów, dla których zamierzamy go oszacować. Geostatystyka oferuje w takich sytuacjach szereg metod umożliwiających wykorzystanie w pełni posiadanych dodatkowych danych pomiarowych, i to zarówno o charakterze ilościowym, jak i jakościowym. W niniejszym opracowaniu wykorzy-

stane zostaną trzy z nich: „prosty kriging ze zmiennymi średnimi lokalnymi” (*simple kriging with varying local means*) w skrócie SKlm, „kriging z zewnętrznym trendem” (*kriging with external drift*, KED) i „kolokacyjny zwykły kokriging” (*colocated ordinary cokriging*, cOCK).

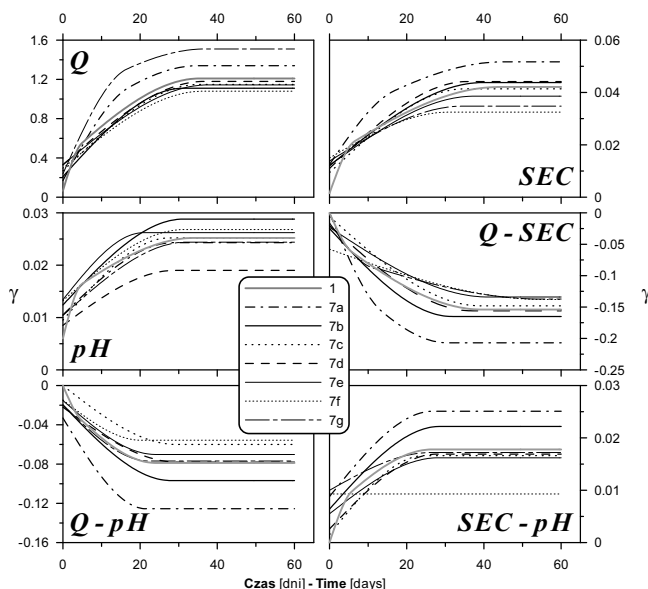
W pierwszej z wyżej wymienionych metod globalna średnia prostego krigingu zastąpiona zostaje lokalnymi średnimi uzyskanymi na podstawie kalibracji dodatkowymi danymi pomiarowymi. Jeśli dodatkowa zmienna jest ciągła, lokalne średnie uzyskujemy z regresji pomiędzy nią a zmienną szacowaną. W praktyce procedura jest następująca:

- 1) obliczamy równanie regresji pomiędzy danymi zmiennej szacowanej ( $Y$ ) a odpowiednimi danymi skorelowanej zmiennej dodatkowej ( $X$ ),
- 2) dla każdej zmierzonej wartości zmiennej szacowanej określamy resztę z regresji (wartość prognozowana z regresji – wartość zmierzona),
- 3) obliczamy semiwariancje empiryczne reszt z regresji oraz dopasowujemy do nich model semiwariancji,
- 4) dla każdego momentu, dla którego chcemy uzyskać estymację interesującej nas zmiennej obliczamy wartość z regresji oraz wartość reszty z prostego krigingu,
- 5) sumujemy obie wartości (predykcje z regresji i szacunek reszty). Istota metody polega zatem na „odfiltrowaniu” z analizowanej serii pomiarowej

składowej deterministycznej i wykonaniu obliczeń geostatystycznych na wykazującej autokorelację czasową części składowej losowej.



Ryc. 3. Modele semiwariogramów i krossemiwarigramów dobowych (1) i tygodniowych (7a – 7g) dla analizowanych serii pomiarowych z górnej Parsęty  
 Fig. 3. Semivariograms and cross semivariograms models of daily (1) and weekly (7a – 7g) data from upper Parsęta river



Ryc. 4. Modele semiwariogramów i krossemiwarigramów dobowych (1) i tygodniowych (7a – 7g) dla analizowanych serii pomiarowych z Bystrzanki  
 Fig. 4. Semivariogram and cross semivariogram models of daily (1) and weekly (7a – 7g) data on the analyzed measuring series of Bystrzanka river

W krigingu z zewnętrznym trendem przyjmuje się, że nieznaną lokalną średnią  $m(\mathbf{u}')$  zmienia się stopniowo wewnątrz każdego lokalnego przedziału czasu (sąsiedztwa)  $W(\mathbf{u})$ , a jej trend jest modelowany jako liniowa funkcja dodatkowej (zewnętrznej) skorelowanej zmiennej  $y(\mathbf{u})$  (3):

$$m(\mathbf{u}) = a_0(\mathbf{u}) + a_1(\mathbf{u})y(\mathbf{u}) \quad (3)$$

Różnica pomiędzy oboma metodami polega na tym, że w SKlm uwzględniamy globalny trend analizowanej zmiennej w stosunku do zmiennej dodatkowej (zewnętrznej), natomiast w KED – tylko trend lokalny w obrębie przedziału czasu (sąsiedztwa) zdefiniowanego najczęściej jako zasięg autokorelacji. Z punktu widzenia przyrodniczego można przypuszczać, że metoda KED powinna dawać lepsze estymacje wartości rzeczywistych. Niestety, ze względu na problemy numeryczne algorytm KED uwzględnia jedynie liniową zależność między zmienną pierwotną i wtórną, podczas gdy przy SKlm możemy użyć dowolnej, optymalnej funkcji regresji, również wielokrotnej.

Pełnym wielozmiennym rozszerzeniem krigingu jest kokriging, gdzie szacowaną wartość oblicza się według wzoru (4):

$$Z_1^*(\mathbf{u}) - m_1(\mathbf{u}) = \sum_{i=1}^{N_1} \sum_{\alpha_i=1}^{n_1(\mathbf{u})} \lambda_{\alpha_i}(\mathbf{u}) [Z_i(\mathbf{u}_{\alpha_i}) - m_i(\mathbf{u}_{\alpha_i})] \quad (4)$$

gdzie  $\lambda_{\alpha_i}(\mathbf{u})$  jest wagą przypisaną do danej pierwotnej  $z_1(\mathbf{u}_{\alpha_i})$  a  $\lambda_{\alpha_i}(\mathbf{u})$  gdy  $i > 1$  jest wagą przypisaną do danej wtórnej (zmiennej dodatkowej)  $z_i(\mathbf{u}_{\alpha_i})$ .

W zwykłym kokrigingu nie jest konieczne aby zmienna(e) wtórna (dodatkowa) była określona dla każdego terminu, dla którego wykonuje się estymację.

Kokriging jest znacznie trudniejszą numerycznie procedurą niż kriging, ponieważ należy obliczyć  $N_1(N_1+1)/2$  semiwariogramów i krossemiwarigramów empirycznych, które muszą być jednocześnie modelowane oraz rozwiązać wielki układ równań liniowych. Wyraźną poprawę estymacji w stosunku do krigingu osiąga się jedynie wówczas, kiedy ilość pomiarów zmiennej szacowanej jest znacznie mniejsza niż zmiennej(ych) dodatkowej, i kiedy istnieje między nimi silna korelacja.

Wariant kokrigingu, kiedy zmienna(e) dodatkowa jest określona we wszystkich terminach, dla których szacuje się zmienną pierwotną nazywa się koloka-

cyjnym kokrigingiem (cOCK). Z teoretycznego punktu widzenia cOCK ma istotne zalety w stosunku do KED. W tym drugim algorytmie informacje uzyskane ze zmiennej wtórnej służą jedynie do oceny trendu zmiennej pierwotnej. Kokriging wykorzystuje je szerzej poprzez bezpośrednie włączenie wartości zmiennej wtórnej do obliczeń i pomiar stopnia czasowej synchronizacji ze zmienną pierwotną poprzez krossemiariogram.

Oprócz metod geostatystycznych do estymacji stężeń użyto dwóch tradycyjnych sposobów – interpolacji liniowej (IL) i regresji potęgowej (REG) w stosunku do przepływu, oraz funkcje krzywkowe (*spline*) w trzech wariantach: sześciennie (*cubic spline*, CSPL), sześciennie ograniczone (*constrained cubic spline*, cCSPL) i lokalnej regresji (*local regression spline*, LRSPL)\*. Jednym z podstawowych źródeł bibliograficznych informacji o funkcjach krzywkowych jest publikacja de Boora (1978).

Sześcienna funkcja krzywkowa interpoluje pomiędzy danymi za pomocą wielomianów 3 stopnia. Algorytm zapewnia, że estymowana krzywa przechodzi przez punkty danych a jej pierwsza i druga pochodna jest ciągła w węzłach (punktach przegięcia krzywej). Ponieważ nieciągłości trzeciej i wyższych pochodnych nie są widoczne, krzywe CSPL mają łagodne, „estetycznie” wyglądające, kształty. Ich wygląd, podobny do linii rysowanych za pomocą tradycyjnego krzywika kreślarskiego, był podstawą stosowanego nazewnictwa. Ograniczona sześcienna funkcja krzywkowa (cCSPL) jest niewielką modyfikacją podstawowego algorytmu, redukującą znaczne skoki standardowej krzywej CSPL występujące przy dopasowaniu do bardzo chaotycznych danych. *Local regression spline* (LRSPL) łączy dopasowanie lokalnie ważoną metodą najmniejszych kwadratów, które wygląda dane w sposób nieciągły, z wygładzaniem ciągłym typu B-spline. Nie jest to wierny interpolator (krzywa nie przechodzi przez punkty danych), lecz dobrze ekstrahuje regularne struktury nawet z bardzo chaotycznych serii pomiarowych. Algorytm przeprowadza dopasowanie w ruchomym oknie o ustalonej szerokości wzdłuż całej serii danych. W niniejszym opracowaniu stosowano okno 6-elementowe, rozkład wag gaussowski, wielomian trzeciego stopnia oraz wygładzającą sześcienną funkcję krzywkową.

W kategoriach geostatystycznych funkcje krzywkowe polegają na odfiltrowaniu składowej nugge-

towej na podstawie niestacjonarnego liniowego modelu regionalizacji. Podstawowa różnica polega na tym, że w metodach geostatystycznych parametry estymacji uzyskuje się poprzez interaktywny proces obliczania semiwariogramu empirycznego i jego modelowania, natomiast w metodzie funkcji krzywkowych są one uzyskiwane automatycznie za pomocą zgeneralizowanej krosvalidacji (*generalized cross validation*, GCV). W wielu pracach cytowanych przez Wackernagla (1998) wykazywano duży stopień podobieństwa estymacji krigingowej i funkcjami krzywkowymi.

Ze względu na poziom komplikacji numerycznej wykorzystanych w niniejszym opracowaniu metod, kluczowe znaczenie dla ich szerszego stosowania w praktyce ma dostępność ich komputerowych implementacji. Bardzo korzystne warunki stwarza fakt, że większość obliczeń wykonano dzięki publicznie dostępnym, bezpłatnym programom komputerowym działającym na zwykłych komputerach PC. Semiwariancje i krossemiariancje empiryczne obliczono oraz ich modelowanie przeprowadzono za pomocą programu Varwin 2.21 (Pannatier 1996). Większość estymacji geostatystycznych przeprowadzono za pomocą pakietu GSLIB 2 (Deutsch, Journel 1998). Jedynie obliczenia metodą cOCK wykonano komercyjnym programem ISATIS (Bleines et al. 2001). Ostatnio jednakże Ma i Journel (1999) opublikowali opis publicznie dostępnego programu komputerowego będącego uzupełnieniem biblioteki GSLIB i obejmującego także algorytm kolokacyjnego kokrigingu. Regresje oraz funkcje sklepane (*spline*) estymowano korzystając z bezpłatnej wersji testowej programu KyPlot 2.0 (<<http://www.kyence.com/>>). Niestety, aktualnie zmienił on swój status na komercyjny. Większość jednak powszechnie dostępnego oprogramowania matematyczno-statystycznego posiada wyżej wymienione procedury obliczeniowe.

## 5. Wyniki

W monitoringu jakości wód w ciekach stosuje się najczęściej tygodniowy interwał czasowy. Z analizowanych dobowych serii czasowych usunięto zatem odpowiednią część danych tak, aby uzyskać ekwiwalent takiego schematu próbkowania. Podstawą oceny jakości jakiegokolwiek metody pomiarowej jest jej dokładność, czyli zgodność uzyskanego wyniku z wartością rzeczywistą, oraz precyzja, czyli zmienność wyników powtarzalnych pomiarów. Jeżeli tak jak w tym

\* W polskiej literaturze spotyka się czasem termin funkcje „sklejane”

przypadku możemy porównać na różne sposoby wartości estymowane z wynikami rzeczywistych pomiarów, to ocena dokładności nie nastęca żadnych problemów. Gorzej jest z porównaniem precyzji analizowanych metod szacowania stężeń składników roztworu rzecznoego. Aby uzyskać jakiegokolwiek informacji na ten temat zdecydowano się za każdym razem obliczenia wykonać siedmiokrotnie w oparciu o serie tygodniowych pomiarów przesuniętych o jeden dzień (pomiarzy poniedziałkowe, wtorkowe... itp.). Siedem wartości to nieco za mało, aby uzyskać wiarygodne miary statystyczne rozrzutu, jednak wystarczająco dużo, aby dokonać jakościowych porównań.

Bezpośrednie obliczanie statystyk parametru, który zapisywany jest w jednostkach logarytmicznych, tak jak pH, jest oczywiście błędne. W tym jednak przypadku zdecydowano się tak uczynić z dwóch względów. Po pierwsze zmienność odczynu wody w Bystrzance jest niewielka – tylko 0,25 pH odstepu międzykwartylowego. Po drugie antylogarytmowanie wprowadza dodatkowy bład, który przy tak niewielkiej zmienności pH miałby istotne znaczenie. W rzeczywistości różnica między średnią obliczoną z wartości pH, a uzyskaną ze stężeń  $H^+$  wynosi 0,03 pH.

Większe znaczenie ma fakt, że operując statystykami parametrów fizykochemicznych wody w cieku najczęściej posługujemy się wartościami ważonymi w stosunku do przepływu. W niniejszym opracowaniu, gdzie główny nacisk położony został na estymację wartości chwilowych owych parametrów, ma to jednak znaczenie drugorzędne.

Szersze omówienie wyników analizy strukturalnej (semiwariancji i krossemiwanacji) opracowywanych serii pomiarowych przepływu i właściwości fizykochemicznych wody w obu ciekach znajdzie czytelnik w publikacji Stacha z 2002 roku. Na rycinach 3 i 4 przedstawione zostało jedynie porównanie modeli semiwariancji i krossemiwanacji referencyjnych danych dobowych, z wyliczonymi dla 7 serii danych tygodniowych.

Tabela 2 zawiera porównanie statystyk referencyjnych dobowych serii danych z ich rozrzutem w seriach tygodniowych: minimum, maksimum oraz odchylenie standardowe. Zestawienie to ma na celu stwierdzenie czy uzupełnienie różnymi metodami brakujących danych dobowych w seriach tygodniowych może wpłynąć na polepszenie jakości typowych miar statystycznych populacji. W tabeli 3 przedsta-

Tab. 2. Porównanie charakterystyk statystycznych codziennych danych pomiarowych i wyliczonych z 7 serii o tygodniowym interwale pomiarów  
Miary statystyczne: Śred. – średnia arytmetyczna, Med. – mediana, Min. – minimum, Maks. – maksimum, Q1 i Q3 – pierwszy i trzeci kwartyl, SD – odchylenie standardowe, Sk – skośność, K – kurtoza

Table 2. Comparison of statistical parameters of the daily measurements data with calculated for 7 chosen subsets with weekly intervals  
Parameters: Śred. – arithmetic mean, Med. – median, Min. – minimum, Maks. – maximum, Q1 and Q3 – first and third quartile, SD – standard deviation, Sk – skewness, K – kurtosis

		Śred.	Med.	Min.	Maks.	Q1	Q3	SD	Sk	K
Parsęta										
SEC-1		450,5	445	225	540	427	487	48,96	-1,003	2,511
SEC-7	Min.	449,6	443,0	225	532	424,0	481,0	46,37	-1,420	0,544
	Maks.	453,1	448,0	306	540	430,5	490,0	51,86	-0,566	4,908
	SD	1,224	1,676	25,158	2,752	2,464	3,092	2,260	0,307	1,405
SiO <sub>2</sub> -1		12,00	12,23	4,93	15,93	10,38	13,85	2,190	-0,474	-0,491
SiO <sub>2</sub> -7	Min.	11,97	12,11	4,93	15,23	10,21	13,75	2,070	-0,597	-0,796
	Maks.	12,10	12,35	7,31	15,93	10,57	13,99	2,265	-0,376	-0,234
	SD	0,048	0,084	0,757	0,261	0,130	0,085	0,066	0,070	0,200
Bystrzanka										
SEC-1		408,3	422	187	543	358	468	77,07	-0,605	-0,306
SEC-7	Min.	402,0	419,0	187	528	348,5	463,0	72,94	-0,755	-0,568
	Maks.	412,9	426,0	206	543	374,5	473,0	80,13	-0,419	0,181
	SD	3,324	2,474	7,274	5,255	9,431	4,113	2,353	0,121	0,257
pH-1		8,353	8,37	7,34	8,89	8,23	8,48	0,174	-0,483	0,929
pH-7	Min.	8,335	8,350	7,34	8,63	8,19	8,46	0,163	-1,474	-0,669
	Maks.	8,365	8,385	7,98	8,89	8,26	8,52	0,189	-0,198	6,435
	SD	0,012	0,014	0,221	0,098	0,022	0,022	0,010	0,464	2,586



wiono wszechstronne zestawienie wyników estymacji wykonanych różnymi metodami z referencyjnymi danymi pomiarowymi. Umożliwia ono zarówno ocenę ich jakości z punktu widzenia wierności odwzorowania miar centralnych analizowanej próby (średnia i mediana), miar rozrzutu/rozkładu (odchylenie standardowe, kwartyle, minimum i maksimum, skośność i kurtoza) jak i poszczególnych wartości (średni błąd –  $ME$  – wzór 5, średni błąd kwadratowy –  $MSE$  – wzór 6, i współczynnik korelacji). Dla każdej metody i każdej miary jakości estymacji podano wartość średnią i odchylenie standardowe dla siedmiu tygodniowych serii danych. Za pomocą funkcji krzywkowych nie można ekstrapolować. Skrajnych terminów, poza pierwszym i ostatnim w podzbiorach obliczeniowych, nie uwzględniono zatem przy kalkulacji parametrów oceniających jakość estymacji. Tabela 3 jest bardzo obszerna i trudno bez czasochłonnego jej studiowania porównać obiektywnie i wszechstronnie jakość anali-

zowanych metod. Zamieszczono zatem również syntetyczne ryciny zestawiające wybrane parametry estymacji za pomocą wykorzystanych w niniejszym opracowaniu metod (ryc. 5–8). Przeprowadzono również ich ranking zarówno w odniesieniu do dokładności, jak i precyzji, i jego wyniki zestawiono w tabeli 4. Dla każdej metody policzono również średnią rangę estymacji miar centralnych, miar rozrzutu/rozkładu i wartości chwilowych oraz średnią rangę globalną.

$$ME = \frac{\sum_{i=1}^n (x_i^m - x_i^e)}{n} \quad (5)$$

$$MSE = \frac{\sum_{i=1}^n (x_i^m - x_i^e)^2}{n} \quad (6)$$

gdzie:  $x^m$  to wartość zmierzona,  $x^e$  – szacowana.

Tab. 3. Porównanie charakterystyk statystycznych codziennych danych pomiarowych z estymacjami opartymi na 7 seriach pomiarów o tygodniowym interwale.

Miary statystyczne jak w tabeli 1 oraz:  $ME$  – średni błąd,  $MSE$  – średni błąd kwadratowy,  $r$  – współczynnik korelacji liniowej Pearsona. Metody estymacji:  $IL$  – interpolacja liniowa,  $REG$  – regresja w stosunku do przepływu,  $OK$  – zwykły kriging,  $SKlm$  – prosty kriging z lokalnymi średnimi,  $KED$  – kriging z zewnętrznym trendem,  $cOCK$  – kolokacyjny zwykły kokriging z przepływem jako zmienną dodatkową,  $cOCK2$  – kolokacyjny zwykły kokriging z  $SEC$  jako zmienną dodatkową,  $CSPL$  – sześcienna funkcja krzywkowa,  $cCSPL$  – ograniczona sześcienna funkcja krzywkowa,  $LRSPL$  – funkcja krzywkowa lokalnej regresji. Górna wartość to średnia z siedmiu estymacji, dolna – odchylenie standardowe

Table 3. Comparison of statistical parameters of the daily measurements data with calculated for estimated daily sets based for 7 chosen subsets with weekly intervals.

Statistical parameters like in tab. 1 and:  $ME$  – mean error,  $MSE$  – mean square error,  $r$  – Pearson correlation coefficient. Estimation methods:  $IL$  – linear interpolation,  $REG$  – rating relationship with discharge,  $OK$  – ordinary kriging,  $SKlm$  – simple kriging with varying local means,  $KED$  – kriging with external drift,  $cOCK$  – collocated ordinary cokriging with discharge as a secondary variable,  $cOCK2$  – collocated ordinary cokriging with  $SEC$  as a secondary variable,  $CSPL$  – cubic spline,  $cCSPL$  – constrained cubic spline,  $LRSPL$  – local regression spline. Upper value for each method is a mean for seven estimations, lower one – standard deviation

	Śred.	Med.	Min.	Maks.	Q1	Q3	SD	Sk	K	ME	MSE	r
Parsęta – SEC												
Dane	450,5	445	225	540	427	487	48,96	-1,003	2,511			
$IL$	450,6 1,079	444,8 1,303	265,7 25,16	536,3 2,752	426,9 2,419	486,2 2,391	46,17 1,983	-0,751 0,244	1,503 0,895	-0,059 1,138	393,3 26,186	0,916 0,007
$REG$	454,6 1,967	464,7 2,492	141,0 47,72	471,5 3,826	453,3 1,997	470,2 3,538	29,90 5,62	-4,618 0,529	33,570 7,992	-4,087 1,968	1546,9 30,410	0,610 0,009
$OK$	449,8 1,220	440,7 2,105	305,8 13,73	527,7 3,011	425,5 2,459	484,3 2,407	42,57 1,679	-0,474 0,204	0,513 0,556	0,813 1,228	464,3 40,670	0,899 0,009
$SKlm$	450,6 1,406	444,2 1,867	135,0 64,69	536,3 2,752	427,9 2,098	486,1 2,610	47,31 2,966	-1,352 0,369	5,518 2,415	0,053 1,415	364,6 57,640	0,924 0,011
$KED$	451,2 1,016	448,6 1,964	264,9 24,26	547,0 9,894	426,7 1,561	486,3 1,605	47,09 1,598	-0,773 0,189	1,386 0,645	-0,629 1,025	296,3 43,290	0,937 0,009
$cOCK$	450,7 0,880	445,9 1,340	223,9 39,30	543,6 10,500	426,4 2,070	487,4 2,370	47,70 1,890	-0,908 0,198	2,159 0,892	-0,122 0,888	230,5 27,240	0,952 0,006
$CSPL$	450,6 1,067	445,1 1,172	259,3 23,18	536,3 2,729	427,0 2,586	486,6 2,137	47,30 2,079	-0,831 0,264	1,811 1,031	-0,038 1,126	405,4 24,360	0,914 0,006
$cCSPL$	450,6 1,064	445,9 1,625	255,1 22,72	539,1 2,838	426,8 2,068	487,0 2,153	48,26 2,192	-0,867 0,275	1,883 1,066	-0,072 1,121	439,4 31,790	0,908 0,007
$LRSPL$	450,6 1,066	446,0 1,734	268,7 19,12	535,4 2,498	427,0 2,516	486,7 2,727	46,82 2,032	-0,763 0,247	1,469 0,884	-0,035 1,123	412,9 26,170	0,912 0,006

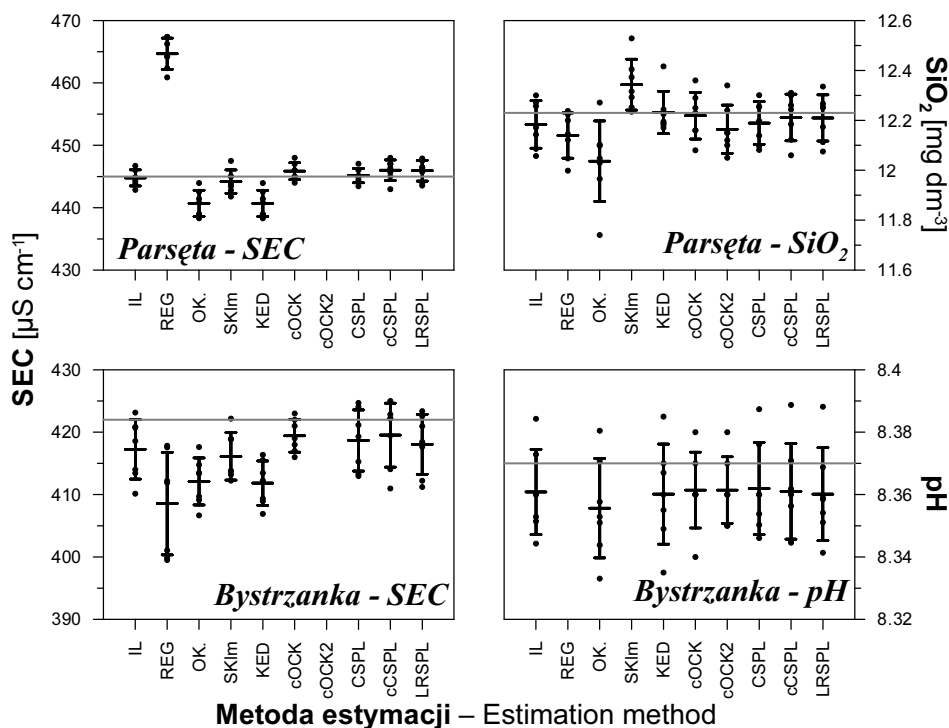
Tab. 3. – cd.

	Śred.	Med.	Min.	Maks.	Q1	Q3	SD	Sk	K	ME	MSE	r
Parsęta – SiO <sub>2</sub>												
Dane	12,00	12,23	4,93	15,93	10,38	13,85	2,190	-0,474	-0,491			
IL	12,00 0,049	12,17 0,095	6,06 0,757	15,54 0,264	10,40 0,086	13,83 0,087	2,095 0,062	-0,412 0,053	-0,651 0,158	0,001 0,048	0,794 0,056	0,916 0,006
REG	12,00 0,039	12,14 0,091	4,80 1,577	14,85 0,142	11,13 0,066	13,01 0,076	1,520 0,077	-0,773 0,336	1,297 0,945	0,005 0,039	2,576 0,011	0,681 0,001
OK	11,95 0,057	12,04 0,161	7,65 0,464	15,00 0,278	10,41 0,036	13,69 0,107	1,928 0,079	-0,259 0,105	-0,936 0,152	0,059 0,057	0,928 0,130	0,899 0,015
SKIm	12,05 0,037	12,34 0,102	4,31 1,077	15,75 0,255	10,38 0,086	13,81 0,076	2,149 0,064	-0,555 0,086	-0,306 0,320	0,007 0,037	0,557 0,077	0,941 0,008
KED	12,02 0,043	12,23 0,085	4,94 0,228	15,73 0,130	10,43 0,079	13,85 0,079	2,113 0,058	-0,509 0,044	-0,347 0,168	-0,002 0,034	0,613 0,071	0,936 0,011
cOCK	12,00 0,041	12,22 0,093	5,54 1,143	15,56 0,259	10,42 0,054	13,83 0,073	2,117 0,064	-0,501 0,055	-0,469 0,236	0,007 0,040	0,573 0,071	0,939 0,007
cOCK2	12,00 0,031	12,16 0,097	5,55 0,581	15,66 0,291	10,45 0,077	13,79 0,052	2,120 0,058	-0,464 0,056	-0,489 0,155	0,003 0,031	0,556 0,059	0,941 0,006
CSPL	12,00 0,048	12,18 0,086	5,83 0,767	15,58 0,254	10,38 0,121	13,86 0,091	2,133 0,063	-0,442 0,053	-0,574 0,188	0,001 0,047	0,820 0,060	0,914 0,006
cCSPL	12,00 0,048	12,20 0,093	5,60 0,828	15,72 0,322	10,40 0,125	13,85 0,085	2,161 0,064	-0,443 0,046	-0,524 0,198	0,001 0,047	0,907 0,056	0,906 0,006
LRSPL	12,00 0,047	12,19 0,095	6,03 0,700	15,47 0,237	10,39 0,127	13,83 0,100	2,115 0,064	-0,424 0,051	-0,622 0,192	0,001 0,046	0,837 0,061	0,912 0,006
Bystrzanka – SEC												
Dane	408,3	422	187	543	358	468	77,07	-0,605	-0,306			
IL	408,1 2,75	418,3 4,07	196,7 7,27	533,4 5,26	360,5 5,67	463,9 4,07	70,80 2,537	-0,509 0,104	-0,348 0,120	-0,235 2,625	1526,9 100,5	0,865 0,009
REG	406,2 2,48	408,6 8,22	178,8 70,48	498,4 21,80	378,5 4,29	453,3 10,39	57,86 10,450	-0,669 0,742	1,068 1,145	1,703 2,486	3008,3 112,0	0,715 0,003
OK	404,6 2,90	412,1 3,77	243,0 9,43	520,3 7,31	361,5 3,72	450,6 3,28	61,38 2,770	-0,397 0,104	-0,435 0,157	3,182 2,835	1843,2 197,9	0,834 0,019
SKIm	407,1 3,153	416,13 3,802	120,61 59,456	545,5 10,89	366,5 4,24	463,5 7,22	74,25 6,768	-0,713 0,299	0,476 0,650	0,649 3,034	1593,4 239,1	0,865 0,019
KED	405,4 2,65	411,8 3,56	158,5 32,10	585,2 43,20	361,5 4,19	462,8 5,83	74,72 3,960	-0,493 0,178	0,032 0,210	1,924 2,654	1622,1 128,8	0,862 0,012
cOCK	407,9 2,13	419,4 2,64	178,6 35,34	533,7 5,12	361,3 4,75	463,3 4,03	71,32 2,034	-0,570 0,151	-0,181 0,398	-0,242 0,197	1368,4 105,2	0,879 0,010
CSPL	408,5 3,22	420,2 4,34	188,4 6,89	534,8 5,57	356,7 4,39	468,3 5,04	75,44 2,573	-0,554 0,096	-0,362 0,173	-0,604 3,058	1802,9 244,4	0,847 0,022
cCSPL	408,1 2,73	419,4 4,65	194,5 7,60	534,4 6,43	358,7 4,52	465,8 4,24	73,39 2,639	-0,534 0,109	-0,349 0,143	-0,172 2,785	1640,5 145,9	0,860 0,009
LRSPL	408,1 2,74	418,0 5,35	207,4 12,97	531,6 2,70	357,3 3,65	465,5 4,58	72,37 2,569	-0,488 0,106	-0,422 0,107	-0,279 2,628	1661,0 128,4	0,854 0,012
Bystrzanka – pH												
Dane	8,353	8,37	7,34	8,89	8,23	8,48	0,174	-0,483	0,929			
IL	8,353 0,012	8,361 0,015	7,840 0,221	8,720 0,098	8,236 0,012	8,477 0,019	0,157 0,009	-0,298 0,366	0,046 1,702	-0,0003 0,012	0,016 0,002	0,713 0,024
OK	8,353 0,012	8,356 0,016	8,062 0,082	8,621 0,030	8,254 0,009	8,444 0,018	0,126 0,010	-0,018 0,249	-0,605 0,339	-0,0001 0,012	0,016 0,001	0,700 0,023
KED	8,349 0,014	8,360 0,016	7,665 0,233	8,731 0,094	8,238 0,014	8,467 0,018	0,159 0,009	0,494 0,418	1,035 2,004	-0,003 0,014	0,016 0,001	0,720 0,013
cOCK	8,354 0,012	8,361 0,012	7,790 0,239	8,720 0,098	8,240 0,013	8,474 0,017	0,154 0,009	-0,302 0,309	0,080 1,472	-0,001 0,012	0,0148 0,001	0,731 0,019
cOCK2	8,353 0,012	8,361 0,011	7,830 0,220	8,730 0,096	8,240 0,015	8,470 0,018	0,151 0,011	-0,283 0,334	0,035 1,497	-0,0006 0,012	0,0128 0,0016	0,771 0,023
CSPL	8,353 0,012	8,362 0,015	7,835 0,221	8,725 0,097	8,234 0,013	8,482 0,019	0,163 0,009	-0,341 0,375	0,155 1,872	-0,0002 0,012	0,017 0,002	0,708 0,026
cCSPL	8,353 0,012	8,361 0,015	7,829 0,219	8,732 0,093	8,231 0,014	8,486 0,018	0,168 0,010	-0,347 0,381	0,142 1,828	-0,0004 0,012	0,018 0,002	0,691 0,032
LRSPL	8,353 0,012	8,360 0,015	7,894 0,181	8,695 0,068	8,235 0,010	8,478 0,020	0,157 0,010	-0,266 0,334	-0,172 1,389	-0,0003 0,012	0,017 0,002	0,702 0,027

Tab. 4. Średnie rangi i odchylenia standardowe rang dokładności i precyzji poszczególnych testowanych metod estymacji stężeń. Podano też sumę średnich globalnych rang dokładności i precyzji

Table 4. Mean ranks and standard deviations of ranks of accuracy and precision for each tested methods of river solute concentration estimations. Sum of mean global deviations of accuracy and precision are also placed

	Miary centralne Central tendency estimation		Miary rozkładu/Rozrzutu Estimation of scatter/ distribution		Miary estymacji punktowej Instant values estimation		Ranga globalna Global rank		
	Śred.	SD	Śred.	SD	Śred.	SD	Śred.	SD	
<b>Dokładność – Accuracy</b>									
IL	2,5	1,69	5,0	1,82	3,2	1,19	3,5	1,57	
REG	6,2	2,93	8,0	2,38	8,3	1,80	7,5	2,37	
OK.	5,4	2,88	7,6	1,10	7,0	2,45	6,7	2,14	
SKIm	4,5	2,74	5,4	2,69	3,3	1,66	4,4	2,36	
KED	4,3	2,38	4,7	2,37	4,2	2,25	4,4	2,33	
cOCK	2,4	0,92	3,6	1,88	2,8	2,08	2,9	1,63	
cOCK2	2,8	2,87	4,6	2,17	2,2	2,04	3,2	2,36	
CSPL	1,8	1,39	2,7	1,44	4,5	1,93	3,0	1,59	
cCSPL	2,3	1,04	2,2	1,25	5,4	2,35	3,3	1,55	
LRSPL	2,8	1,75	4,8	2,07	4,8	1,99	4,1	1,94	
<b>Precyzja – Precision</b>									
IL	4,9	2,70	4,3	1,90	3,2	1,99	4,1	2,20	Dokładność + Precyzja Accuracy + Precision
REG	5,8	3,49	7,2	2,70	3,2	2,59	5,4	2,93	7,6
OK.	6,3	3,28	3,4	2,64	5,4	2,57	5,0	2,83	12,9
SKIm	6,2	2,71	6,7	1,82	6,2	2,22	6,4	2,25	11,7
KED	3,3	1,98	4,1	2,64	3,8	2,12	3,7	2,25	10,8
cOCK	2,1	1,36	4,3	2,57	2,5	1,73	3,0	1,89	8,1
cOCK2	2,8	3,50	3,9	2,18	2,3	0,82	3,0	2,17	5,9
CSPL	4,3	2,96	4,7	2,02	4,5	2,84	4,5	2,61	6,2
cCSPL	4,4	2,26	4,8	2,11	4,0	2,17	4,4	2,18	7,5
LRSPL	4,8	2,12	3,8	2,39	3,6	1,62	4,1	2,04	7,7
									8,2

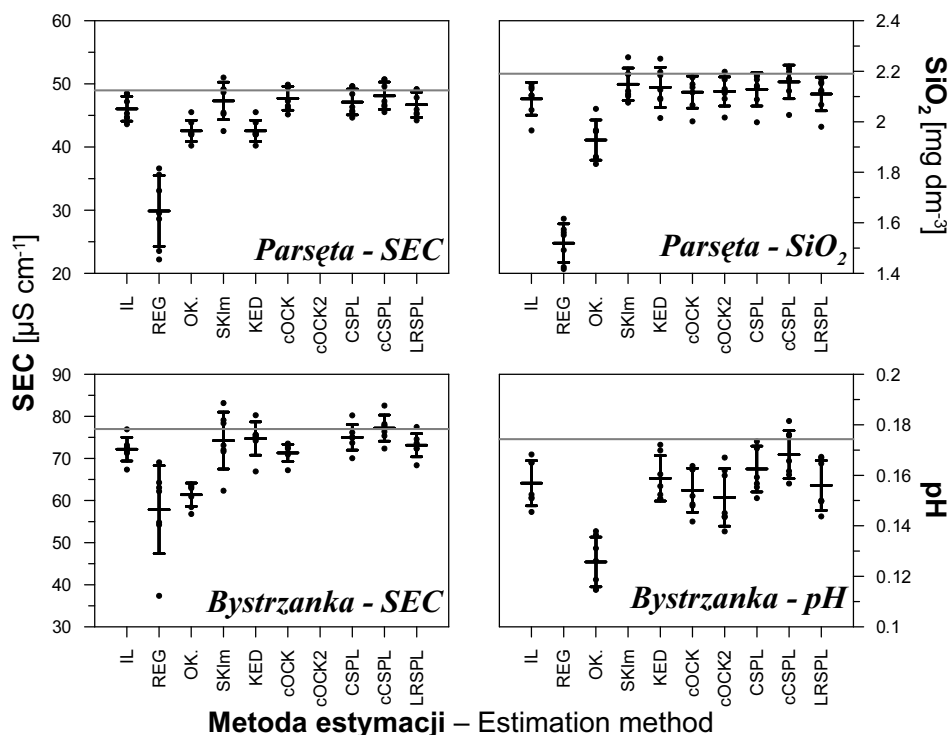


Ryc. 5. Podsumowanie rezultatów estymacji mediany stężeń dobowych za pomocą 10 porównywanych metod.

Kropkami zaznaczono wynik dla pojedynczej serii danych tygodniowych, a liniami średnią i zakres 1 odchylenia standardowego. Pozioma szara linia oznacza rzeczywistą medianę danych referencyjnych.

Fig. 5. Summary of solute daily concentrations median estimation by 10 studied methods.

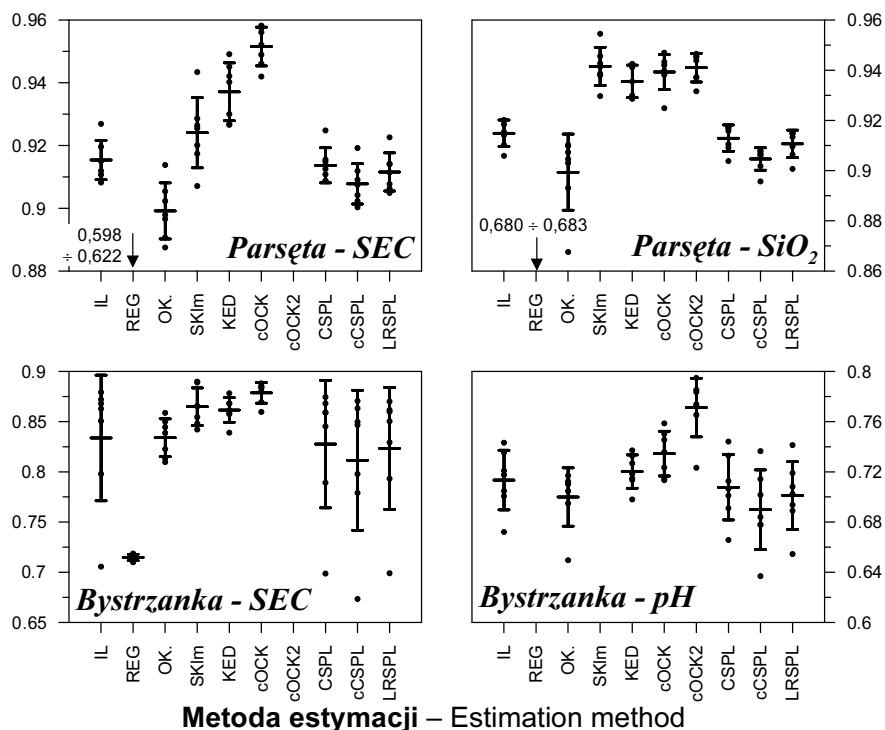
Results for the single series of weekly data was marked by the dots. Error bars shows mean value of estimation plus/minus 1 standard deviation range. Horizontal grey line indicates the real median value of reference data.



Ryc. 6. Podsumowanie rezultatów estymacji odchylenia standardowego stężeń dobowych za pomocą 10 porównywanych metod. Kropkami zaznaczono wynik dla pojedynczej serii danych tygodniowych, a liniami średnią i zakres 1. odchylenia standardowego. Pozioma szara linia oznacza rzeczywiste odchylenie standardowe danych referencyjnych

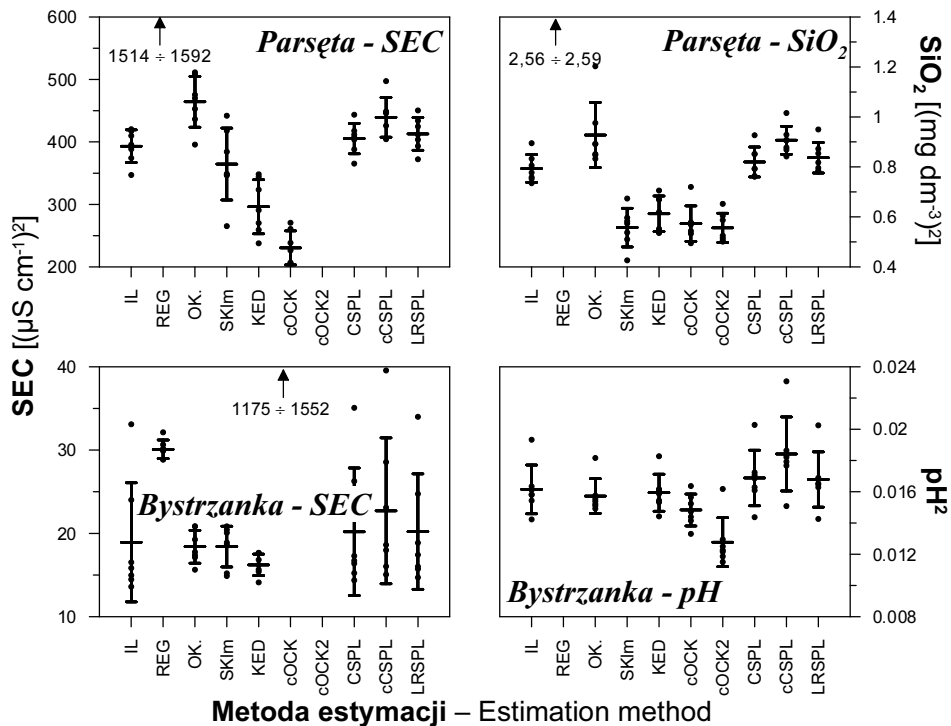
Fig. 6. Summary of solute daily concentrations standard deviation estimation by 10 studied methods.

Results for the single series of weekly data was marked by the dots. Error bars shows mean value of estimation plus/minus 1. standard deviation range. Horizontal grey line indicates the real standard deviation value of reference data



Ryc. 7. Współczynnik korelacji liniowej Pearsona pomiędzy referencyjnymi danymi dobowymi a szacowanymi za pomocą 10 porównywanych metod. Kropkami zaznaczono wynik dla pojedynczej serii danych tygodniowych, a liniami średnią i zakres 1. odchylenia standardowego

Fig. 7. Pearson's linear correlation coefficient between daily solute concentrations reference data and estimated from weekly samples by 10 studied methods. Results for the single series of weekly data was marked by the dots. Error bars shows mean  $r$  value plus/minus 1. standard deviation



Ryc. 8. Średni błąd kwadratowy (MSE) estymacji stężeń dobowych wykonanych za pomocą 10 porównywanych metod. Kropkami zaznaczono wynik dla pojedynczej serii danych tygodniowych, a liniami średnią i zakres 1. odchylenia standardowego  
 Fig. 8. Mean Square Error of daily solute concentrations estimated from weekly samples by 10 studied methods.  
 Results for the single series of weekly data was marked by the dots. Error bars shows mean MSE value plus/minus 1. standard deviation

## 6. Wnioski

Wyniki obliczeń zestawione w tabelach 2–4, oraz na rycinach 5–8, pozwalają na następujące wnioski:

1. Żadna z testowanych metod nie może być uważana za uniwersalną i najlepszą w każdych warunkach. Wyniki testów są nie tylko różne w przypadku poszczególnych parametrów, ale także w sytuacji powtarzalnych serii pomiarów tego samego parametru. Porównanie wyników obliczeń wykonanych na jednym jedynie zbiorze danych referencyjnych może prowadzić do fałszywych uogólnień. Istotny jest również typ charakterystyk statystycznych, które potrzebujemy. Interpolacja liniowa na przykład daje bardzo dobre wyniki, kiedy chcemy oszacować miary centralne analizowanej zmiennej. Lecz w przypadku miar rozkładu/rozrzutu jakości szacunków uzyskanych za pomocą IL jest zdecydowanie gorsza.
2. Zestawienie rozrzutu estymacji parametrów populacji opartych na tygodniowych seriach pomiarowych i tych samych seriach uzupełnionych do ciągów dobowych wykazuje wyraźną poprawę

jakości po zastosowaniu takich procedur, jak KED, cOCK, CSPL, cCSPL, a także w mniejszym stopniu IL, OK, LRSPL.

3. Generalnie najgorszą metodą do uzupełniania ciągów pomiarów parametrów fizykochemicznych wody w ciekach jest – chyba najczęściej stosowana – regresja w stosunku do przepływów. Rezultaty jej stosowania są zdecydowanie najgorsze w przypadku dokładności. Także, jeśli chodzi o precyzję obliczeń, regresja znajduje się w końcu zestawienia porównywanych metod. Zdecydowanie lepsze wyniki dają metody interpolacyjne. Mówi to wiele o naturze zmienności czasowej składu roztworu rzecznoego. „Bezpośrednia” składowa deterministyczna związana ze zmiennością objętości wody w korycie cieku ma bardzo małe znaczenie. Znacznie ważniejsze są czynniki, które można nazwać „buforowością” hydrologiczną i biogeochemiczną systemu zlewni, wpływające na wysoką autokorelację czasową parametrów fizykochemicznych wody. Można jednakże przypuszczać, że w miarę zmniejszania się frekwencji pomiarów różnice na korzyść interpolacyjnych metod estymacji będą maleć.

4. Zaskakująco wysokie miejsce w rankingu zarówno dokładności, jak i precyzji zajmuje interpolacja liniowa (IL). Lokata ta jest porównywalna, a czasami nawet lepsza niż kilku innych bardziej wyrafinowanych i pracochłonnych metod (OK, SKlm i KED). Wszędzie tam, gdzie nie jest ważna subtelna dokładność, a w cenie jest prostota i szybkość obliczeń, należy stosować interpolację liniową.
5. Nieco lepsze efekty niż interpolacja liniowa dają sześciennie funkcje krzywkowe (CSPL). Można je zalecać do rutynowej estymacji stężeń składników roztworu rzecznoego, bo (1) ich stosowanie nie wymaga żadnego czasochłonnego i wymagającego specjalistycznej wiedzy parametryzowania, a poza tym (2) procedury te występują w wielu łatwo dostępnych, tanich, a czasami nawet bezpłatnych programach matematyczno-statystycznych.
6. Stosunkowo słabą jakość estymacji za pomocą OK i SKlm można interpretować jako efekt globalnej ich parametryzacji. Oparte są one bowiem na wyliczonych dla całych analizowanych serii pomiarowych: modelu semiwariancji (OK) oraz modelu semiwariancji reszt i regresji w stosunku do przepływu (SKlm). Charakterystyki te nie są stacjonarne i zmieniają się sezonowo, a nawet w krótszych odcinkach czasu. Funkcje krzywkowe, które mają wybitnie charakter algorytmu lokalnie adaptacyjnego wykazują tutaj swoją wyższość. Także kriging z zewnętrznym trendem (KED), który choć oparty o globalny model semiwariancji, ale modyfikowany lokalną liniową regresją w stosunku do przepływu dawał lepsze rezultaty niż OK i SKlm. Mankamenty estymacji parametrów fizykochemicznych wody za pomocą OK i SKlm można by w dużym stopniu zmniejszyć stosując do obliczeń lokalne modele semiwariancji liczone dla ruchomego okna. Taki algorytm już istnieje, choć wykorzystywany jest głównie do estymacji zmiennych w przestrzeni dwuwymiarowej (Haas 1990). Być może efekty stosowania OK i SKlm byłyby również lepsze, gdyby zastosować bardziej „dokładne” globalne semiwariogramy liczone z całej serii danych, a nie z każdej z tygodniowych próbek osobno. Zdarza się bowiem, że przez jakiś okres prowadzimy częste pomiary, a później redukujemy ich frekwencję. „Dokładne” semiwariancje obliczone dla serii częstych pomiarów mogą być użyte do podniesienia jakości estymacji zmiennej w nieoprobowanych terminach dla okresu, kiedy frekwencja pomiarów była mniejsza.
7. Zdecydowanie najlepszym w końcowym zestawieniu dokładności i precyzji jest kolokacyjny kokriging (cOCK) zarówno wówczas, kiedy zmienną dodatkową jest przepływ, jak i przewodnictwo elektryczne wody (SEC). Szczególnie zdecydowaną dominację algorytm ten wykazuje przy estymacji punktowej. Pomimo, że do obliczeń używa się „globalnych” modeli semiwariancji i krossemiwanacji, korygujący lokalny wpływ zmiennej dodatkowej jest wystarczający do podniesienia jakości estymacji. Prawdopodobnie, jak wspomniano wyżej, byłaby ona jeszcze lepsza, gdyby zastosować semiwariogramy i krossemiwariogramy liczone z całej serii danych, a nie z każdej z tygodniowych próbek osobno. Metoda cOCK jest również w najmniejszym stopniu obciążona słabością wszystkich metod interpolacyjnych: spłaszczeniem rozkładu estymowanej zmiennej poprzez przeszacowanie wartości minimalnych i niedoszacowanie – maksymalnych.
8. Istotne znaczenie ma możliwość użycia w cOCK, z równie dobrym skutkiem jak przepływ, łatwej do pomiaru skorelowanej zmiennej, jaką jest przewodnictwo elektryczne wody. Jeśli celem badań jest jedynie charakterystyka jakości wody w wybranej lokalizacji, można wybrać strategię pomiarów pomijającą kłopotliwe często pomiary przepływu i obejmującą ciągłe, albo bardzo częste pomiary przewodnictwa, pH, temperatury, mętności wody itp., oraz znacznie rzadsze pobieranie próbek do pełnych analiz składu chemicznego. Za pomocą cCOK można oszacować wówczas, z dużą wiarygodnością, wartości owych rzadko oznaczanych parametrów fizykochemicznych wody dla każdego terminu, dla którego dysponujemy wynikami pomiarów zmiennych dodatkowych. Innym, możliwym kierunkiem polepszenia jakości estymacji za pomocą cOCK, jest rozszerzenie jej z wariantu dwóch zmiennych na wielozmienny.
9. Wydaje się, w świetle powyżej zaprezentowanych danych, że bardzo dobre – bo oddające znacznie lepiej naturę zjawiska – „syntetyczne” ciągi pomiarowe stężeń można by generować za pomocą kolokacyjnego kokrigingu. Mogą one być według sugestii Webb’a i in. (2000) podstawą selekcji optymalnej metody obliczania ładunku składowych roztworu rzecznoego.

## 7. Literatura

- Bleines C., Deraisme J., Geffroy F., Perseval S., Rambert F., Renard D., Touffait Y., 2001:** *ISATIS software manual*. Geovariances & Ecole Des Mines de Paris :1–531.
- De Boor C., 1978:** *A practical guide to splines*. Springer-Verlag, New York.
- Deutsch C.V., Journel, A.G., 1998:** *GSLIB. Geostatistical software library and user's guide. Second Edition*. Oxford University Press :1–369.
- Goovaerts P., 1997:** *Geostatistics for natural resources evaluation*. Oxford University Press, 1–483.
- Haas T.C., 1990:** *Kriging and automated variogram modeling within a moving window*. Atmospheric Environment 24A :1759–1769.
- Kostrzewski A., Mazurek M., Zwoliński Z., 1994a:** *Dynamika transportu fluwialnego górnej Parsęty jako odbicie funkcjonowania systemu zlewni*. Stowarzyszenie Geomorfologów Polskich, Poznań :1–165.
- Kostrzewski A., Mazurek M., Zwoliński Z., 1994b:** *Monitoring hydrologiczny i hydrochemiczny w Stacji Geoekologicznej w Storkowie w latach 1986-1993*. [w:] Zintegrowany Monitoring Środowiska Przyrodniczego. Stacja Bazowa Storkowo, A. Kostrzewski (red.), Biblioteka Monitoringu Środowiska, Warszawa :45–68.
- Littlewood I.G., 1995:** *Hydrological regimes, sampling strategies, and assessment of errors in mass load estimates for United Kingdom rivers*. Environmental International, vol. 21, no. 2 :211–220.
- Ma X., Journel A.G., 1999:** *An expanded GSLIB cokriging program allowing for two Markov models*. Computers and Geosciences, 25 :627–639.
- Pannatier Y., 1996:** *VARIOWIN. Software for Spatial Data Analysis in 2D*. Springer-Verlag, New York :1–91.
- Soja R., 1981:** *Analiza odpływu z fliszowych zlewni Bystrzanki i Ropy (Beskid Niski)*. Dokumentacja Geograficzna IGiPZ PAN, z. 1 :1–91.
- Stach A., 2002:** *Geostatystyczna identyfikacja mechanizmów transportu roztworów w ciekach*. Przegląd Naukowy Inżynierii i Kształtowania Środowiska, 11 (2) :130–143.
- Starkel L., Gil E., (red.), 1994:** *Zintegrowany Monitoring Środowiska Przyrodniczego. Stacja Bazowa Szymbark (Karpaty Fliszowe)*. Biblioteka Monitoringu Środowiska, Warszawa :1–169.
- Wackernagel H., 1998:** *Multivariate geostatistics. 2<sup>nd</sup> edition*. Springer-Verlag, Berlin Heidelberg :1–291.
- Webb B.W., Phillips J.M., Walling D.E., 2000:** *A new approach to deriving "best estimate" chemical fluxes for rivers draining the LOIS study area*. The Science of the Total Environment, 251/252 :45–54.
- Webb B.W., Phillips J.M., Walling D.E., Littlewood I.G., Watts C.D., Leeks G.J.L., 1997:** *Load estimation methodologies for British rivers and their relevance to the LOIS RACS(R) programme*. The Science of the Total Environment, 194/195 :379–389.
- Welc A., 1985:** *Zmienność denudacji chemicznej w Karpatach Fliszowych (na przykładzie zlewni potoku Bystrzanka)*. Dokumentacja Geograficzna IGiPZ PAN, z. 5 :1–102.

## FUNDAMENTALS OF OPTIMAL SOLUTE SUBSTANCES LOADS ESTIMATION IN RIVER OUTFLOW

### Summary

The main aim of the present study is to compare a wide set of methods for estimation solute substance concentrations in river outflow, both statistical parameters of population and instant values. The reference data set consist two years of daily measurements from two catchments: lowland and mountainous. Each data set was modified to simulate seven series of measurements with weekly intervals. This allows to compare both accuracy and in some extent also precision of each tested estimation methods. No one method are always better than others. It depends not only on parameter under consideration but also on particular data subset. Generally the worse choice in terms of both accuracy and precision is to use rating relationships with discharge. Surprisingly very good result was obtained using simple linear interpolation. For routine calculation of river solute concentration from infrequent data, cubic spline was recommended both by it's good performance, simplicity and popularity in relatively chip mathematical-statistical software for PC. The best method in comparison was collocated ordinary cokriging (cOCK). It's performance is similar good both were secondary variable was discharge or specific electric conductivity of water (SEC). It is suggested that cOCK will be particularly useful for generating synthetic river solute concentration data for comparison of different methods for solute loads calculation (eg. Webb et al. 2000).

