

Instytut Badań Systemowych  
Polskiej Akademii Nauk

**mgr Piotr Kopka**

Streszczenie rozprawy doktorskiej

**Zastosowanie metodologii  
bayesowskiej w stochastycznej  
rekonstrukcji zdarzeń na  
przykładzie uwolnień  
niebezpiecznych gazów do  
atmosfery**

Promotor:

**dr hab. inż. Piotr Holnicki-Szulc, prof. IBS PAN**

Promotor pomocniczy:

**dr Anna Wawrzyńczak-Szaban**

WARSZAWA 2019



# 1. Streszczenie rozprawy doktorskiej

## 1.1. Zarys tematyki

Przechowywanie niebezpiecznych substancji oraz ich transport drogą kolejową, morską lub innymi szlakami komunikacyjnymi powoduje potencjalne niebezpieczeństwo. Pomimo rozwoju technologii oraz wprowadzania coraz bardziej restrykcyjnych przepisów, wielu awarii nie udało się uniknąć i prawdopodobnie nie unikniemy ich w przyszłości. Zagrożenia związane z awariami przemysłowymi oraz ich skala wymuszają podejmowanie działań, nie tylko na poziomie lokalnym, ale także na szczeblu międzynarodowym. Odnosi się to nie tylko do organów zarządzających na terenie, na którym do takowych wypadków doszło, ale do całego regionu, ponieważ granice administracyjne w żaden sposób nie chronią przed potencjalnymi skutkami poważnych awarii. Jednym ze znaczących zagrożeń związanym z awariami przemysłowymi jest uwolnienie niebezpiecznych substancji: toksycznych, łatwopalnych lub radioaktywnych, które mogą zagrażać zdrowiu ludzi znajdujących się w pobliżu. Nieoczekiwane wycieki materiałów szkodliwych mają również dalekosiężne skutki dla środowiska naturalnego. Kluczową kwestią we wczesnej fazie zarządzania następstwami takich zdarzeń jest identyfikacja źródła uwolnienia niebezpiecznych substancji, gdyż nie zawsze miejsce awarii da się jednoznacznie określić. Wszelkie informacje o potencjalnym miejscu wycieku pozwalają na podjęcie skutecznych działań oraz zastosowanie adekwatnych procedur chroniących zdrowie i życie ludzi znajdujących się w obszarze narażenia.

Jedną z grup incydentów związanych z uwolnieniem materiałów chemicznych są awarie skutkujące rozprzestrzenieniem się gazów w powietrzu. Głównym niebezpieczeństwem związanym z nieoczekiwanymi uwolnieniami substancji lotnych jest trudny do określenia obszar oddziaływania. W tego typu przypadkach poznanie skali niebezpieczeństwa jest możliwe dzięki symulacji transportu gazu z wykorzystaniem modelu dyspersyjnego. Techniki komputerowego wspomaganie decyzji umożliwiają szybką i bezpieczną identyfikację obszarów zagrożenia. Znajomość charakterystyk źródła oraz podstawowych parametrów meteorologicznych pozwala na w miarę dokładną symulację potencjalnego zdarzenia. Pomiary uwolnionej substancji zazwyczaj pochodzą z sieci sensorów rozmieszczonych na monitorowanym obszarze. Określenie parametrów źródła uwolnienia w oparciu jedynie o stężenia zmierzone w kilku punktach przestrzeni jest zadaniem bardzo trudnym. Ze względu na niepewność, która jest powiązana z danymi pomiarowymi oraz z modelem

dyspersyjnym, rozpatrywane zagadnienie w wielu przypadkach nie ma jednoznacznego rozwiązania i musi być sformułowane w kategoriach probabilistycznych. Identyfikację parametrów źródła możemy rozważyć jako zagadnienie odwrotne (ang. *Inverse Problem, IP*). W literaturze zagadnienie identyfikacji parametrów źródła uwolnienia gazu w oparciu o dane pomiarowe zostało określone mianem estymacji parametrów źródła (ang. *Source Term Estimation, STE*), a metoda rozwiązywania tego zagadnienia nosi nazwę stochastycznej rekonstrukcji zdarzeń (ang. *Stochastic Event Reconstruction, SER*).

## 1.2. Cel i teza pracy

Celem niniejszej rozprawy doktorskiej jest opracowanie efektywnych algorytmów próbkowania oraz zaproponowanie modelu pozwalającego na dynamiczną identyfikację parametrów źródeł uwolnienia gazów do atmosfery na podstawie stężeń uwolnionej substancji pochodzących z sieci sensorów z zastosowaniem metodologii bayesowskiej.

## 1.3. Postawiony problem badawczy

Niniejsza rozprawa obejmuje następujące problemy badawcze:

1. Analiza dotychczas opublikowanych i dostępnych algorytmów oraz systemów zastosowanych w problemie stochastycznej rekonstrukcji zdarzeń.
2. Zbudowanie własnego modelu rekonstrukcji parametrów źródła opierającego się na metodach bayesowskich oraz wykorzystaniu modeli dyspersji gazu jako modeli „w przód”.
3. Zastosowanie w problematyce SER zaawansowanych algorytmów próbkowania oraz metod, takich jak sekwencyjne metody Monte Carlo (ang. *Sequential Monte Carlo, SMC*) oraz przybliżone obliczenia bayesowskie (ang. *Approximate Bayesian Computation, ABC*).
4. Opracowanie własnych modyfikacji do wyżej wymienionych algorytmów dedykowanych problemowi estymacji parametrów źródła uwolnienia gazu.
5. Zwiększenie dokładności wyników rekonstrukcji poprzez zastosowanie dynamicznych modeli dyspersji w metodologii SER, takich jak: *Second-order Closure Integrated Puff - SCIPUFF* (Sykes et al., 2000), *Quick Urban and Industrial Complex Dispersion Modeling System - QUIC* (Williams et al., 2004).
6. Analiza działania opracowanych modeli SER w wielu wariantach dla nierekonstruowanych dotychczas eksperymentów:
  - 6.1 Porównanie efektywności wybranych algorytmów stochastycznej rekonstrukcji zdarzeń z zastosowaniem rzeczywistego zbioru danych z eksperymentu uwolnienia gazu  $SF_6$  przeprowadzonego podczas kampanii *Copenhagen Tracer Experiments - CTE* (Gryning & Lyck, 1984).

- 6.2 Identyfikacja charakterystyk źródła poruszającego się w czasie emisji na podstawie danych z eksperymentu *Over-land atmospheric dispersion* – OLAD (Biltoft & Turley, 2000),
- 6.3 Lokalizacja źródła uwolnienia gazu w środowisku silnie zurbanizowanym w oparciu o dane pochodzące z serii eksperymentów *Dispersion of Air Pollution and its Penetration into the Local Environment* – DAPPLE (Arnold et al., 2004).

## 1.4. Wzór Bayesa w metodologii stochastycznej rekonstrukcji zdarzeń

Twierdzenia Bayesa dostosowane do problemu stochastycznej rekonstrukcji zdarzeń występuje w wielu publikacjach w następującej formie:

$$\underbrace{p(\boldsymbol{\theta}|\mathbf{d}_{obs},I)}_{\text{A posteriori}} = \frac{\overbrace{p(\boldsymbol{\theta}|I)}^{\text{A priori}} \overbrace{p(\mathbf{d}_{obs}|\boldsymbol{\theta},I)}^{\text{Wiarygodność}}}{\underbrace{p(\mathbf{d}_{obs}|I)}_{\text{Czynnik normalizujący}}} \propto p(\boldsymbol{\theta}|I)p(\mathbf{d}_{obs}|\boldsymbol{\theta},I), \quad (1.1)$$

gdzie  $\boldsymbol{\theta}$  jest wektorem parametrów modelu, natomiast  $\mathbf{d}_{obs}$  jest zbiorem danych eksperymentalnych. Celem wprowadzenia danych  $\mathbf{d}_{obs}$  w równaniu (1.1) jest dążenie do zwiększenia naszej wiedzy o parametrach modelu. Element  $I$  oznacza wszystkie dodatkowe informacje, które mogą być użyte w modelu. W wielu interpretacjach wzoru (1.1) kontekst zewnątrz  $I$  jest pomijany i przypisywany bezpośrednio do ustawień modelu.

Na potrzeby opisywanego problemu badawczego przyjmijmy następujące założenia:

- $\boldsymbol{\theta}$  jest wektorem parametrów modelu opisującym źródło uwolnienia gazu, np. pozycję źródła, ilość emitowanej substancji itp.
- dane  $\mathbf{d}_{obs}$  są pomiarami stężenia substancji, które pochodzą z sieci sensorów,
- informacja  $I$  odnosi się do danych zewnętrznych, które mogą być użyte w modelu dyspersji, takie jak kierunek wiatru, temperatura, itp.

W formule (1.1) możemy wyróżnić następujące składowe:

- **rozkład a priori**  $p(\boldsymbol{\theta}|I)$  wyraża stan wiedzy na temat parametrów przed uwzględnieniem danych  $\mathbf{d}_{obs}$ . Jeśli nie posiadamy informacji o początkowych wartościach parametrów modelu, wybór rozkładu a priori powinien odzwierciedlać ten stan wiedzy;

- **funkcja wiarygodności**  $p(\mathbf{d}_{obs}|\boldsymbol{\theta}, I)$  określa powiązanie między danymi  $\mathbf{d}_{obs}$  a wektorem parametrów modelu  $\boldsymbol{\theta}$ . Dla rozważanego modelu, wiarygodność bayesowska nie jest stricte rozkładem prawdopodobieństwa ze względu na to, że dane  $\mathbf{d}_{obs}$  zostały fizycznie zaobserwowane (ustalone);
- **rozkład a posteriori**  $p(\boldsymbol{\theta}|\mathbf{d}_{obs}, I)$  wyraża wartości prawdopodobieństwa parametrów  $\boldsymbol{\theta}$  po uwzględnieniu danych obserwowalnych  $\mathbf{d}_{obs}$  oraz zewnętrznych informacji  $I$ .

### 1.5. Porównanie efektywności algorytmów próbkowania w zagadnieniu stochastycznej rekonstrukcji zdarzeń z wykorzystaniem danych z eksperymentu Copenhagen Tracer Experiments

Podczas cyklu eksperymentów kopenhaskich (CTE) uwalniano gaz znakujący  $SF_6$  (heksafluorek siarki) ze źródła umieszczonego na wysokości  $115m$ . Obszar kampanii obejmował teren miejski o stosunkowo niskiej zabudowie, o określonym współczynniku szorstkości równym  $0.6m$ . Stężenie gazu zmierzono na poziomie od  $2m$  do  $3m$  w trzech łukach sensorów oznaczonych kolejno  $ARC1$ ,  $ARC2$  oraz  $ARC3$ . Sensory umieszczono w odległości od  $2km$  do  $6km$  od źródła. W każdych z dziesięciu doświadczeń przeprowadzono trzy pomiary o czasie trwania równym 20 minut, co pozwoliło na uzyskanie całkowitego godzinnego czasu pojedynczego eksperymentu. Stężenia są wartościami średnimi ze wszystkich pomiarów zebranych w ciągu 20 minut. Dane meteorologiczne uzyskane podczas eksperymentów składały się z pomiarów wiatru oraz temperatury zmierzonych na różnych wysokościach wieży, na której umieszczono źródło. Szczegółowe informacje dotyczące technicznych i naukowych aspektów eksperymentów CTE podano w raportach (Gryning & Lyck, 1984). Na rysunku 1.1 przedstawiono mapę domeny kampanii kopenhaskiej. Lokalizację źródła oraz miejsca pomiarów meteorologicznych oznaczono czerwonym punktem z symbolem  $\times$ . Lokalizacje wszystkich sensorów zilustrowane są za pomocą zielonych punktów. Na rysunku 1.1 sensory, na których zarejestrowano istotne pomiary w dniu 19 października mają przypisane numery zgodne z oznaczeniami z dokumentacji (Gryning, 1981).

Dane doświadczalne posłużyły do przeprowadzania studium porównawczego algorytmów:

- **MCMC** : Algorytm Markow Chain Monte Carlo z zastosowanym klasycznego algorytmu Metropolisa-Hastingsa
- **MCMC-SEQ**: Algorytm Metropolisa-Hastingsa wraz z mechanizmem sekwencyjnego przekazywania informacji



Rysunek 1.1: Mapa prezentująca obszar eksperymentu Copenhagen Tracer Experiments. Czerwony punkt z symbolem  $\times$  oznacza lokalizację źródła oraz stacji metodologicznej. Zielone punkty oznaczają miejsca ulokowania sensorów rozmieszczonych w trzech łukach *ARC1*, *ARC2* oraz *ARC3*.

- **SMC**: Sekwencyjny algorytm Monte Carlo
- **ABC-SMC-AW**: Sekwencyjny algorytm przybliżonych obliczeń bayesowskich z wykorzystaniem adaptacyjnych wag

Jako model progresywny posłużył model dyspersji gazu SCIPUFF, z uwzględnieniem danych zewnętrznych  $I^C$ . Poszukiwane charakterystyki źródła obejmowały cztery parametry  $(x, y, z, q)$ . Początek układu współrzędnych  $(x, y)$  został umiejscowiony w rzeczywistej lokalizacji źródła. Parametr  $z$  jest wysokością wyniesienia źródła nad poziomem gruntu. Poziomą emisję gazu  $SF_6$  określa parametr  $q$  wyrażony w  $[kg/s]$ . W tabeli 1.1 zostało przedstawione sumaryczne zestawienie wartości najważniejszych statystyk uzyskanych poprzez uruchomienie wszystkich testowych algorytmów.

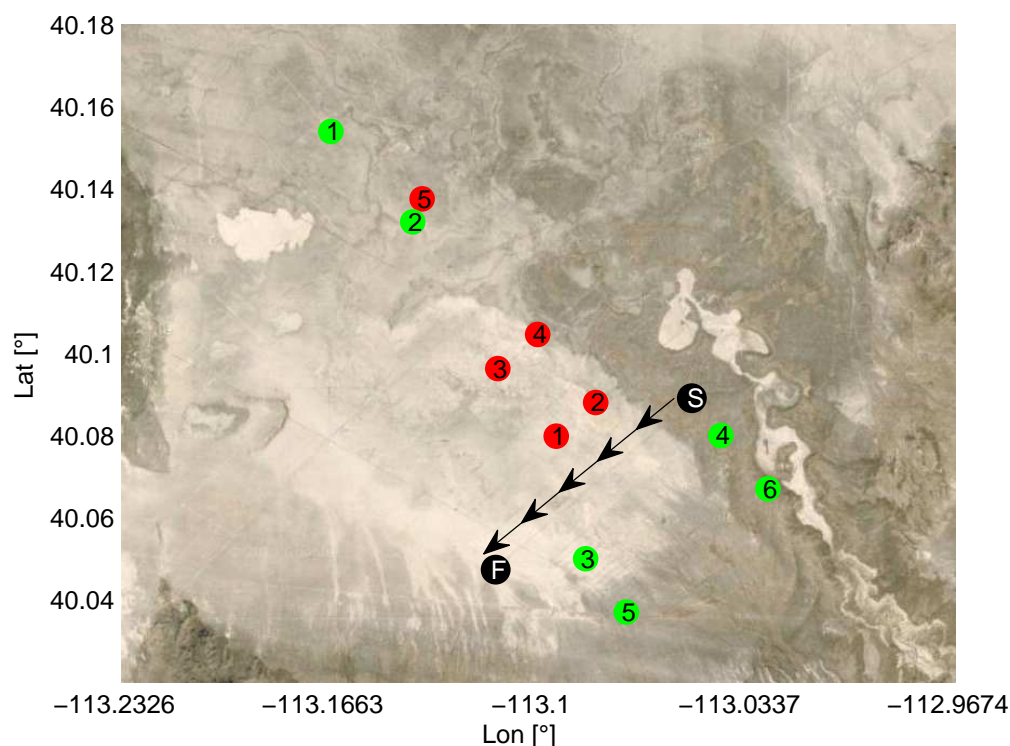
Niniejsza analiza potwierdziła przewagę nowoczesnych technik opartych na

TABELA 1.1: Podsumowanie najważniejszych statystyk uzyskanych z wyników procedury rekonstrukcji parametrów źródła z eksperymentu 19Oct-CTE. Rzeczywiste wartości poszczególnych parametrów oznaczone są symbolem  $\theta^*$ . Wartości dominanty rozkładu a posteriori oznaczono symbolem  $\theta^{MAP}$ . Wartości, które występują po symbolu  $\pm$  oznaczają połowę szerokość jednego słupka estymowanego histogramu. Statystyka  $\bar{\theta}$  oznacza średnia arytmetyczną z odchyleniem standardowym równym  $\sigma(\theta)$ . Przedziały ufności 50%, 90% w statystyce bayesowskiej przedstawione są za pomocą wskaźników  $CI^{50\%}(\theta)$  oraz  $CI^{90\%}(\theta)$ .

Parametr	Algorytm	$x[m]$	$y[m]$	$z[m]$	$q[kg/s]$
$\theta^*$		0	0	115.00	0.0032
$\theta^{MAP}$	MCMC	1156.25 $\pm$ 53.00	-406.25 $\pm$ 52.41	188.72 $\pm$ 3.25	0.0006 $\pm$ 0.000001
	MCMC-SEQ	1093.75 $\pm$ 53.00	-406.25 $\pm$ 52.41	143.59 $\pm$ 3.25	0.0010 $\pm$ 0.000001
	SMC	514.45 $\pm$ 53.00	57.80 $\pm$ 52.41	176.39 $\pm$ 3.25	0.0012 $\pm$ 0.000001
	ABC-SMC-AW	616.28 $\pm$ 53.00	58.14 $\pm$ 52.41	183.43 $\pm$ 3.25	0.0023 $\pm$ 0.000001
$\bar{\theta}$	MCMC	1514.39	-168.93	130.05	0.0077
	MCMC-SEQ	924.45	502.95	118.84	0.0064
	SMC	417.51	344.69	147.75	0.0031
	ABC-SMC-AW	3.47	555.16	118.65	0.0055
$\sigma(\theta)$	MCMC	2423.08	1766.32	48.50	0.0058
	MCMC-SEQ	2199.88	1575.01	51.31	0.0059
	SMC	1687.05	1316.31	41.08	0.0038
	ABC-SMC-AW	1410.01	1062.69	52.55	0.0047
$CI^{90\%}(\theta)$	MCMC	[-1904.29 7123.12]	[-4234.67 2501.15]	[58.88 199.54]	[0.00010 0.01619]
	MCMC-SEQ	[-1894.11 4250.13]	[-1465.67 2651.15]	[47.85 199.54]	[0.0001 0.0157]
	SMC	[-1913.67 1786.68]	[-1271.09 2601.55]	[69.68 200.06]	[0.0002 0.0132]
	ABC-SMC-AW	[-1825.30 848.79]	[-174.42 2267.72]	[26.78 198.27]	[0.0047 0.0008]
$CI^{50\%}(\theta)$	MCMC	[-996.20 1607.07]	[-574.82 1348.44]	[102.22 196.95]	[0.0001 0.0117]
	MCMC-SEQ	[-156.20 1532.17]	[-561.66 1228.49]	[109.22 198.95]	[0.0001 0.0021]
	SMC	[-468.01 919.27]	[-346.21 751.66]	[152.36 193.58]	[0.0004 0.0019]
	ABC-SMC-AW	[-81.07 732.28]	[-58.14 523.37]	[84.38 198.76]	[0.0014 0.0037]

metodach aproksymacyjnych w porównaniu z metodami częściej spotykanymi w publikacjach z dziedziny SER. W szczególności, widać to w przypadku parametru o dużej czułości, jakim jest poziom emisji gazu  $q$ . Wyniki otrzymane z zastosowaniem metody ABC-SMC-AW charakteryzują się najmniejszymi uzyskanymi obszarami  $CR^{90\%}$ , w porównaniu z innymi algorytmami. Uzyskane wyniki oraz przedstawione wnioski stanowią dobry punkt wyjścia do dalszych rekonstrukcji, dla znacznie bardziej złożonych eksperymentów przedstawionych w dalszej części rozprawy.

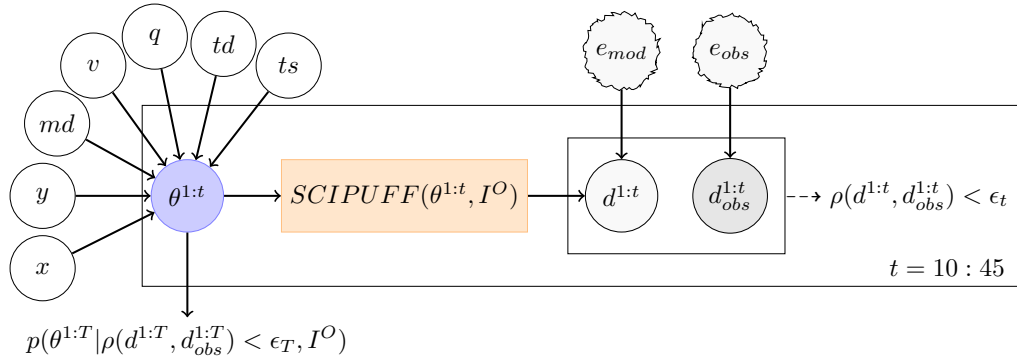




Rysunek 1.2: Mapa prezentująca obszar eksperymentu OLAD. Czarny punkt (S) – początkowa lokalizacja źródła; czarny punkt (F) – końcowa lokalizacja źródła; strzałki – kierunek ruchu pojazdu; czerwone punkty – sensory; zielone punkty – przenośne meteorologiczne stacje pomiarowe.

## 1.6. Rekonstrukcja parametrów mobilnego źródła dla eksperymentu Over-Land Alongwind Dispersion

Eksperyment Over-Land Alongwind Dispersion został przeprowadzony między 8-25 września 1997 roku na terenie poligonu armii USA (Watson, 2000). W procedurze rekonstrukcji wykorzystujemy dane z drugiego testu OLAD oznaczonego w raporcie jako *Trial252 – 2* przeprowadzonego w dniu 9 września 1997 r. Uwolnienie gazu  $SF_6$  nastąpiło między 6:45 a 6:53 ( $8min$ ) z ciężarówki poruszającej się po drodze *Bravo*. Do wyboru zestawu *Trial252 – 2*, jako testowych danych w procedurze rekonstrukcji, przyczyniły się stosunkowo stabilne warunki pogodowe (wiatr z kierunku Pd.-Wsch. z prędkością  $\approx 5m/s$ ). Stężenie gazu osiągnęło szczyt na pierwszej linii sensorów po około  $6min$  od uwolnienia, na drugiej po  $14min$ . Maksymalne wartości stężenia zarejestrowały sensory umieszczone na wschodnim i zachodnim końcu linii 3, odpowiednio w 26 i 27 $min$ . Całkowity czas trwania pomiarów wynosił 45 minut.

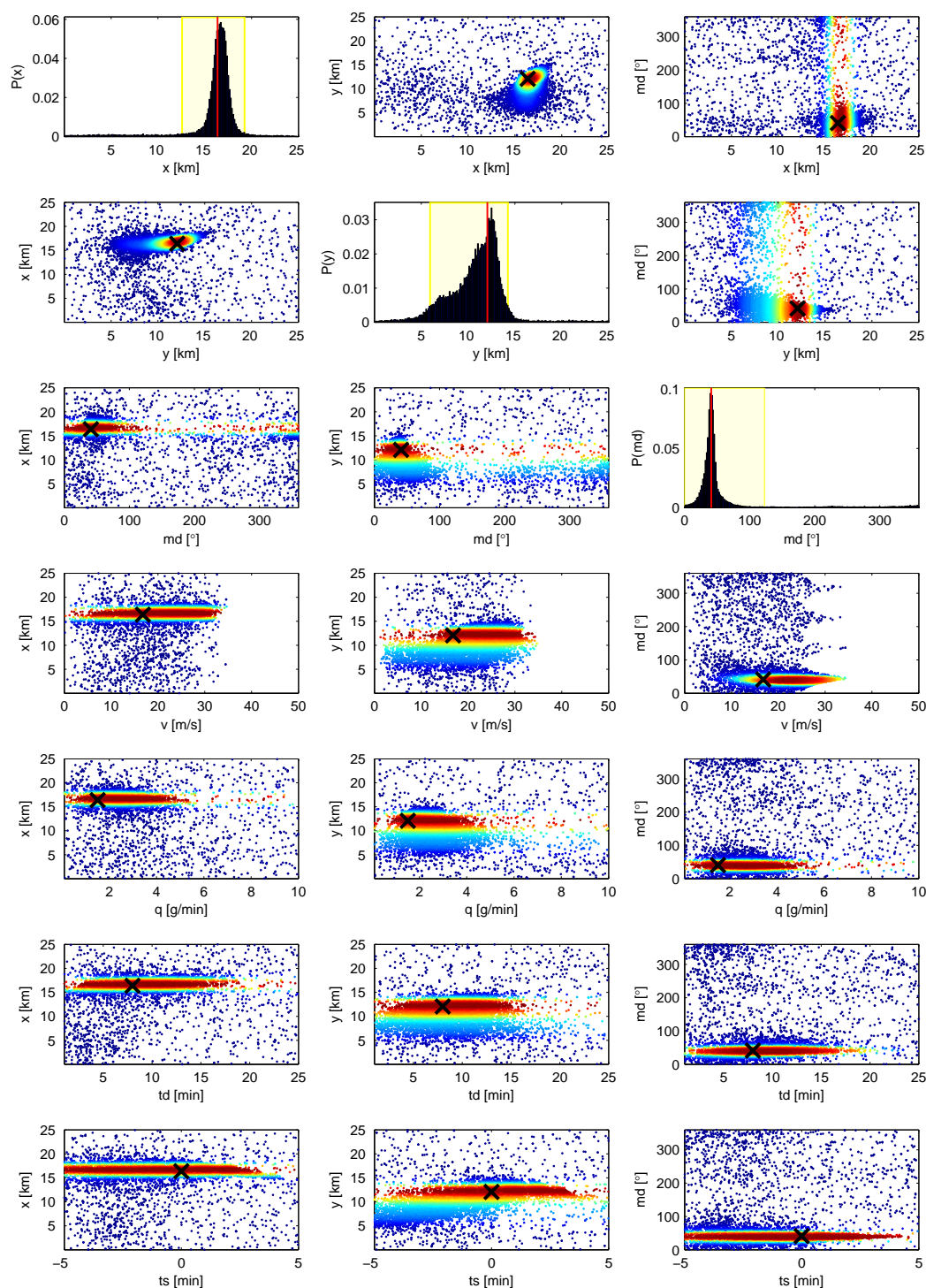


Rysunek 1.3: Graficzny model przedstawiający dynamiczną bayesowską procedurę wnioskowania w problematyce stochastycznie rekonstrukcji zdarzeń dla eksperymentu OLAD. Element  $SCIPUFF(\theta^{1:t}, I^O)$  oznacza model dyspersji  $SCIPUFF$ , gdzie  $I^O$  odnosi się do danych zewnętrznych. Strzałki oznaczają zależności między elementami modelu.

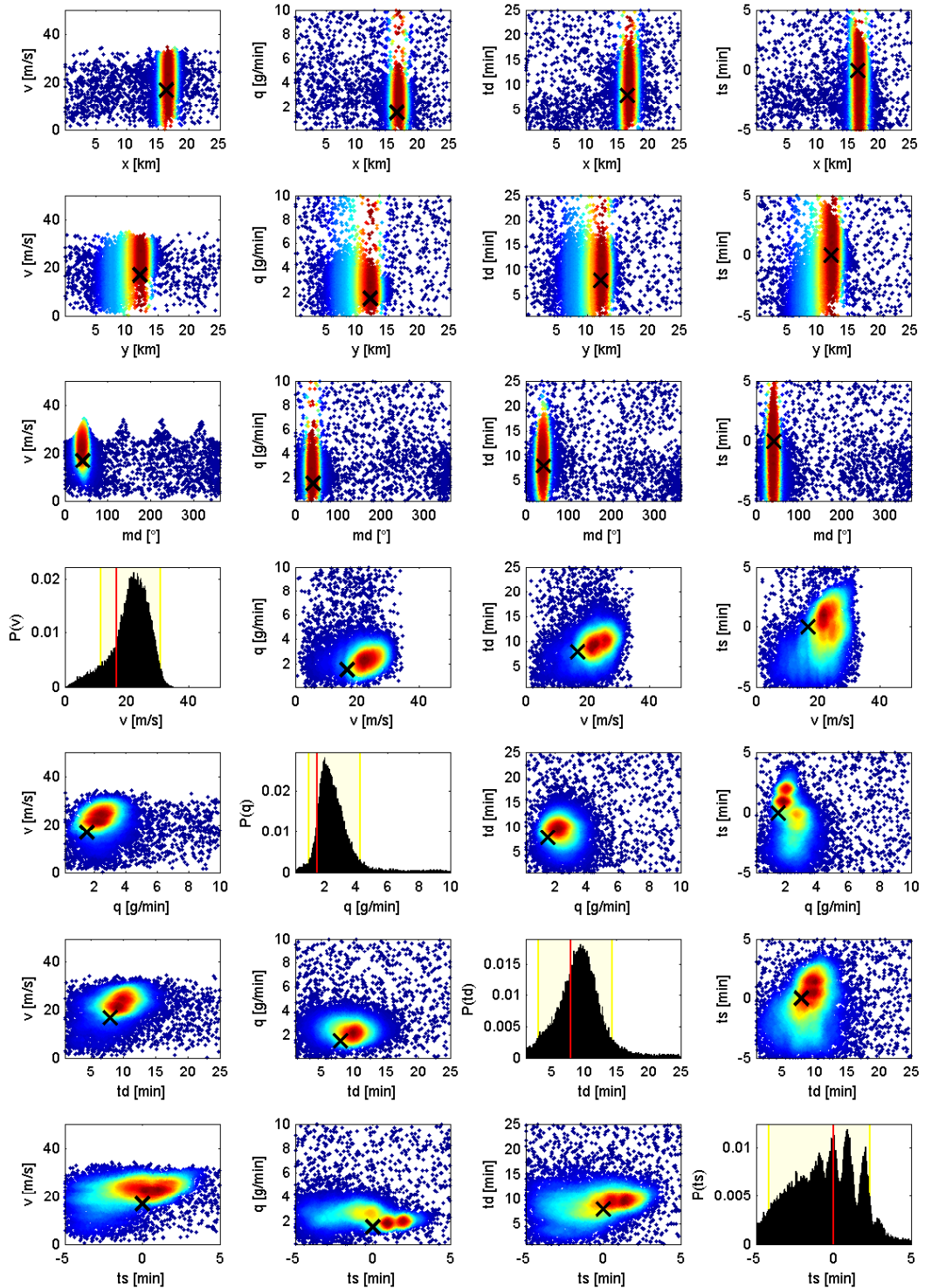
Rekonstrukcja opierająca się na danych z eksperymentu OLAD obejmowała analizę 7 parametrów charakteryzujących źródło uwolnienia:  $(x, y)$  punkt startowy;  $md$  kierunek poruszania się źródła;  $v$  prędkość źródła,  $q$  poziom emisji;  $td$  całkowity czas uwolnienia gazu;  $ts$  czas rozpoczęcia uwolnienia. Na rysunku 1.3 przedstawiono model procedury rekonstrukcji parametrów mobilnego źródła z eksperymentu OLAD Trial 252-2, wraz z relacjami zachodzącymi między elementami zaproponowanego modelu odwrotnego.

Na rysunkach 1.4 oraz 1.5 przedstawiono rozkłady a posteriori dla wszystkich parametrów. Do określenia rozkładów znajdujących się na głównej przekątnej obliczono rozkłady brzegowe. Na pozostałych rysunkach widzimy kombinacje par wszystkich parametrów. Rozkłady te przedstawione są z pomocą „map ciepła”, gdzie bardziej czerwone obszary oznaczają większe wartości gęstości prawdopodobieństwa. Czerwona pionowa linia na rysunkach na głównej przekątnej określa rzeczywiste wartości parametrów  $\theta^*$ , co odpowiada czarnemu  $\times$  na podstałych rysunkach.

Dzięki zastosowaniu zaawansowanego modelu dyspersji  $SCIPUFF$  wraz z nowoczesnymi technikami pobierania próbek, skutecznie zbudowano rozkład a posteriori w siedmiowymiarowej przestrzeni parametrów źródła. Największy wpływ na czas obliczeniowy niezbędny dla efektywnego działania procedury SER ma model dyspersji, który musiał być uruchamiany w każdej iteracji. Program PC- $SCIPUFF$  1.22 został zastosowany jako model typu „black-box” i zainstalowany w systemie maszyn wirtualnych o następujących parametrach Intel Core i7-3820 CPU 3,60 GHz 3,80 GHz 64 GB. Zastosowana procedura rekonstrukcji w połączeniu z systemem pięciu instancji modeli działających równolegle i niezależnie, pozwoliła osiągnąć opisane



Rysunek 1.4: Rozkłady a posteriori dla poszczególnych kombinacji par oraz pojedynczych parametrów uzyskane poprzez zastosowanie procedury odwrotnej oraz danych z eksperymentu OLAD *Trial 252-2* z 35 kroków czasowym od  $t = 10$  do  $t = 45$ . Czerwona linia oraz czarny krzyż określają szukane wartości parametrów  $\theta^*$ . Dla parametru  $md$  - kierunek wschód-zachód odpowiada za  $0^\circ$ , natomiast orientacja jest przeciwna do kierunku wskazówek zegara. Rysunek stanowi lewą część całej macierzy wykresów, której dopełnieniem jest rysunek 1.5.



Rysunek 1.5: Rozkłady a posteriori dla poszczególnych kombinacji par oraz pojedynczych parametrów uzyskane poprzez zastosowanie procedury odwrotnej oraz danych z eksperymentu OLAD *Trial 252-2* z 35 kroków czasowym od  $t = 10$  do  $t = 45$ . Czerwona linia oraz czarny krzyż określają szukane wartości parametrów  $\theta^*$ . Dla parametru  $ts$  - wartość 0 oznacza rzeczywisty moment rozpoczęcia uwolnienia. Rysunek stanowi prawą część całej macierzy wykresów, której dopełnieniem jest rysunek 1.4.

wyniki w ciągu około 2,5 godziny. Bez stosowania skutecznych technik pobierania próbek, takich jak ABC-SMC-AW, czas potrzebny do osiągnięcia wyników wzrasta wraz z wzrostem wymiaru wektora parametrów źródła przeszukiwanej przestrzeni. Możliwe jest przyspieszenie tego procesu poprzez zastosowanie większej liczby niezależnych maszyn wirtualnych z zainstalowanym programem PC-SCIPUFF 1.2. Proponowana metodologia okazuje się być skuteczna w problemie identyfikacji źródła opisanego siedmioma charakterystykami. Co więcej, jednostajny ruch źródła emitera nie powodował dodatkowych problemów i był oceniany podobnie jak inne parametry.

### 1.7. Rekonstrukcja parametrów skażenia w terenie silnie urbanizowanym dla danych z eksperymentu Dispersion of Air Pollution and its Penetration into the Local Environment

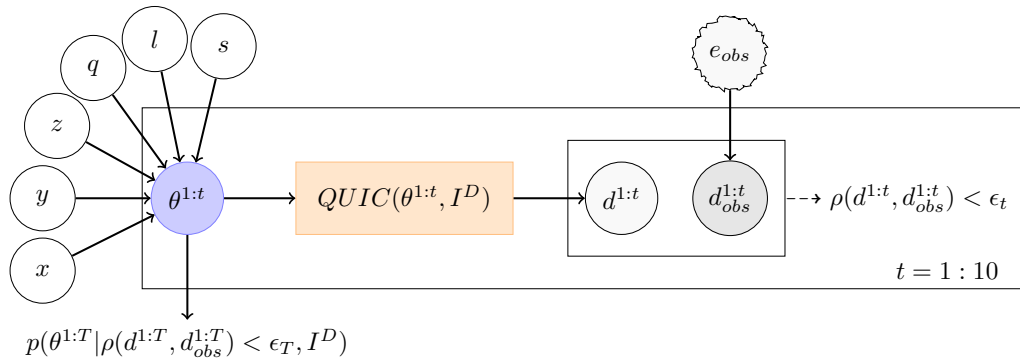
Kampania eksperymentów DAPPLE (ang. *Dispersion of Air Pollution and its Penetration into the Local Environment*) to nazwa dla serii badań polowych polegających na kontrolowanej emisji gazu w centralnej części Londynu (Wood, Barlow, Belcher, & Dobre, 2009). Średnia wysokość budynków w tej lokalizacji wynosi około 21.6m (od 10m do 64m). Interesujące, z punktu widzenia dynamicznej rekonstrukcji zdarzeń, było uwolnienie z dnia 28 czerwca 2007 roku. Uzyskane dane pochodzą z sieci złożonej z 18 sensorów, które przez 30 minut mierzyły stężenie uwolnionego gazu. Dla każdego czujnika dostępne są informacje o dziesięciu uśrednionych pomiarach, które zostały uzyskane w cyklach trwających 150 sekund. Między czasem próbkowania występowały 30 sekundowe przerwy, co daje sumaryczny czas 180 sekund przypadający na jeden krok czasowy. Lokalizacje źródła oraz przyrządów pomiarowych zostały ustalone na podstawie prognozy pogody, która była dostępna w dniu poprzedzającym właściwy eksperyment. Dokładną mapę analizowanego obszaru przedstawiono na rysunku 1.6, na której miejsce uwolnienia zaznaczono zielonym punktem z symbolem  $\times$ . Całkowita masa emitowana ze źródła punktowego wynosiła 323mg gazu *perfluoromethyl-cyclohexane* (PMCH, C7F14).

Pozycję emitera opisuje para współrzędnych  $(x,y)$ , w układzie o początku leżącym w lewym dolnym rogu domeny. Parametr  $z$  jest wysokością wyniesienia źródła nad poziomem gruntu. Całkowitą masę uwolnionej substancji określa parametr  $q$ . Czas trwania emisji oraz czas rozpoczęcia uwolnienia określają atrybuty  $l$  oraz  $s$ . Na rysunku 1.7 przedstawiono model procedury rekonstrukcji parametrów źródła z eksperymentu DAPPLE wraz z relacjami zachodzącymi między elementami zaproponowanego modelu rekonstrukcji.

Poniżej przedstawione są wyniki procedury rekonstrukcji parametrów źródła uwolnienia gazu dla danych DAPPLE z dnia 28 czerwca 2007. Sumaryczny



Rysunek 1.6: Mapa prezentująca obszar eksperymentu DAPPLE. Zielony punkt z symbolem  $\times$  oznacza lokalizację źródła. Żółte ponumerowane punkty oznaczają lokalizację 18 sensorów. Czerwona gwiazdka przedstawia miejsce ulokowania stacji meteorologicznej na dachu budynku WCC.



Rysunek 1.7: Graficzny model przedstawiający dynamiczną bayesowską procedurę wnioskowania w problematyce stochastycznie rekonstrukcji zdarzeń dla eksperymentu DAPPLE. Element  $QUIC(\theta^{1:t}, I^D)$  oznacza model dyspersji QUIC-PLUME, gdzie  $I^D$  odnosi się do danych zewnętrznych - Dane WCC. Strzałki oznaczają zależności między elementami modelu.

TABELA 1.2: Podsumowanie najważniejszych statystyk uzyskanych z wyników procedury rekonstrukcji parametrów źródła z eksperymentu DAPPLE. Właściwe wartości dla poszczególnych parametrów oznaczone są poprzez  $\theta^*$ . Statystyka  $\theta^{MAP}$  oznacza wartości parametrów o największych wartościach gęstości prawdopodobieństwa a posteriori. Statystyka  $\bar{\theta}$  oznacza średnią arytmetyczną z odchyleniem standardowym równym  $\sigma(\theta)$  obliczoną na podstawie próbek. Przedziały ufności 90% w statystyce bayesowskiej przedstawione są za pomocą  $CI^{90\%}(\theta)$ .

Parametry	$x[m]$	$y[m]$	$z[m]$	$q[mg]$	$l[s]$	$s[s]$
$\theta^*$	243.3	282.8	1.5	323.0	900.0	0.0
$\theta^{MAP}$	$203.5 \pm 3.2$	$291.8 \pm 0.5$	$5.4 \pm 0.06$	$265.7 \pm 10.4$	$881.0 \pm 18.9$	$71.0 \pm 9.5$
$\bar{\theta}$	243.9	290.9	7.6	280.0	903.3	72.4
$\sigma(\theta)$	124.6	84.3	4.1	118.7	352.3	41.8
$CI^{90\%}(\theta)$	[26.0 604.2]	[142.0 502.6]	[1.8 13.1]	[56.9 494.6]	[331.6 1620.0]	[14.2 137.4]

zbiór wartości najważniejszych statystyk uzyskanych z próbek wygenerowanych w ramach zaproponowanej procedury jest przedstawiony w tabeli 1.2.

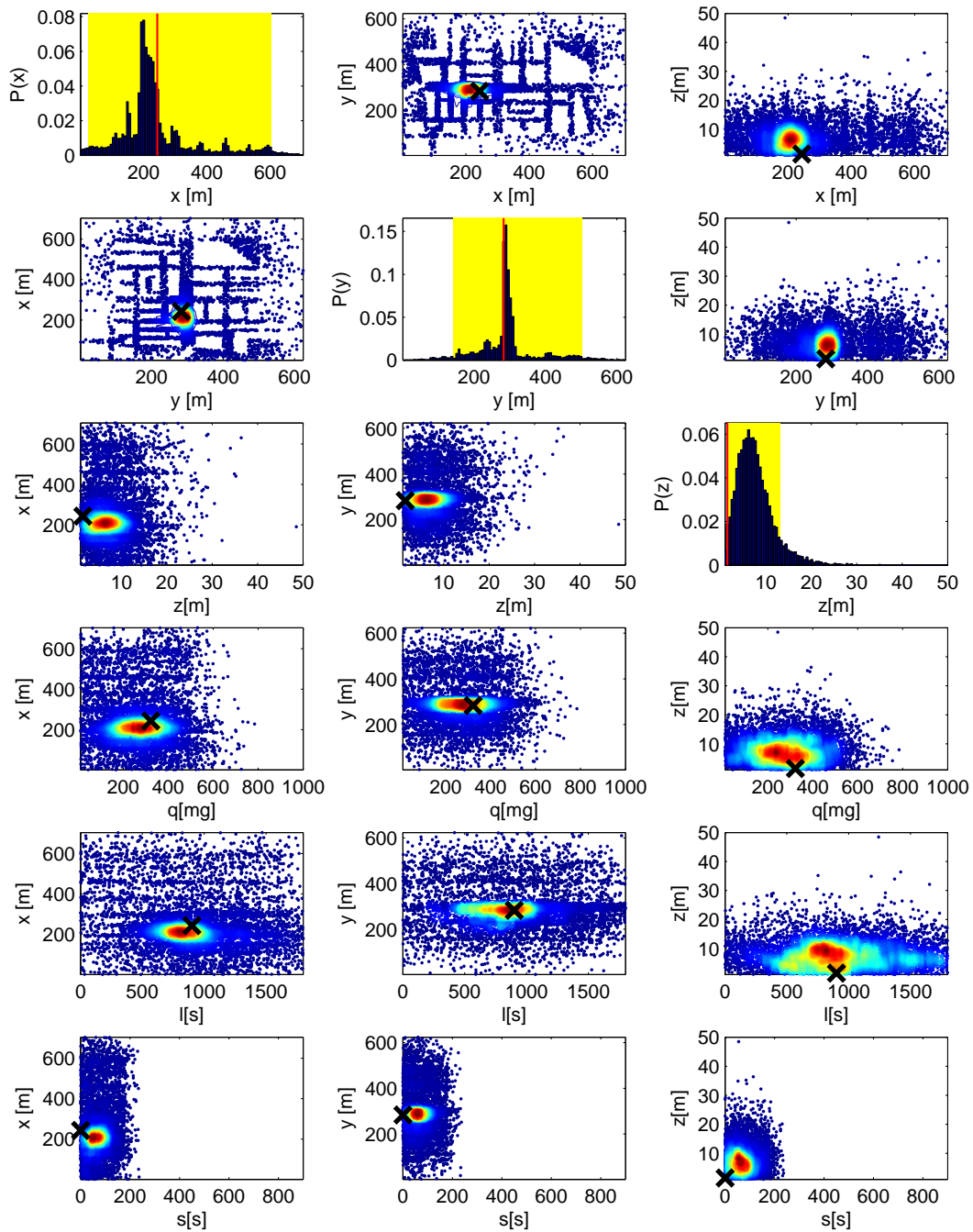
## 1.8. Podsumowanie i wnioski

W niniejszej pracy opracowano modele oraz algorytmy, które pozwoliły na identyfikację parametrów źródeł uwolnienia gazów do atmosfery na podstawie stężeń uwolnionej substancji pochodzących z sieci sensorów. W pracy przedstawiona została argumentacja świadcząca o rozwiązaniu postawionych problemów badawczych. Problemy badawcze oraz wnioski płynące z ich rozwiązania obejmowały następujące punkty:

1. Analiza dotychczas opublikowanych i dostępnych algorytmów oraz systemów zastosowanych w problemie stochastycznej rekonstrukcji zdarzeń.

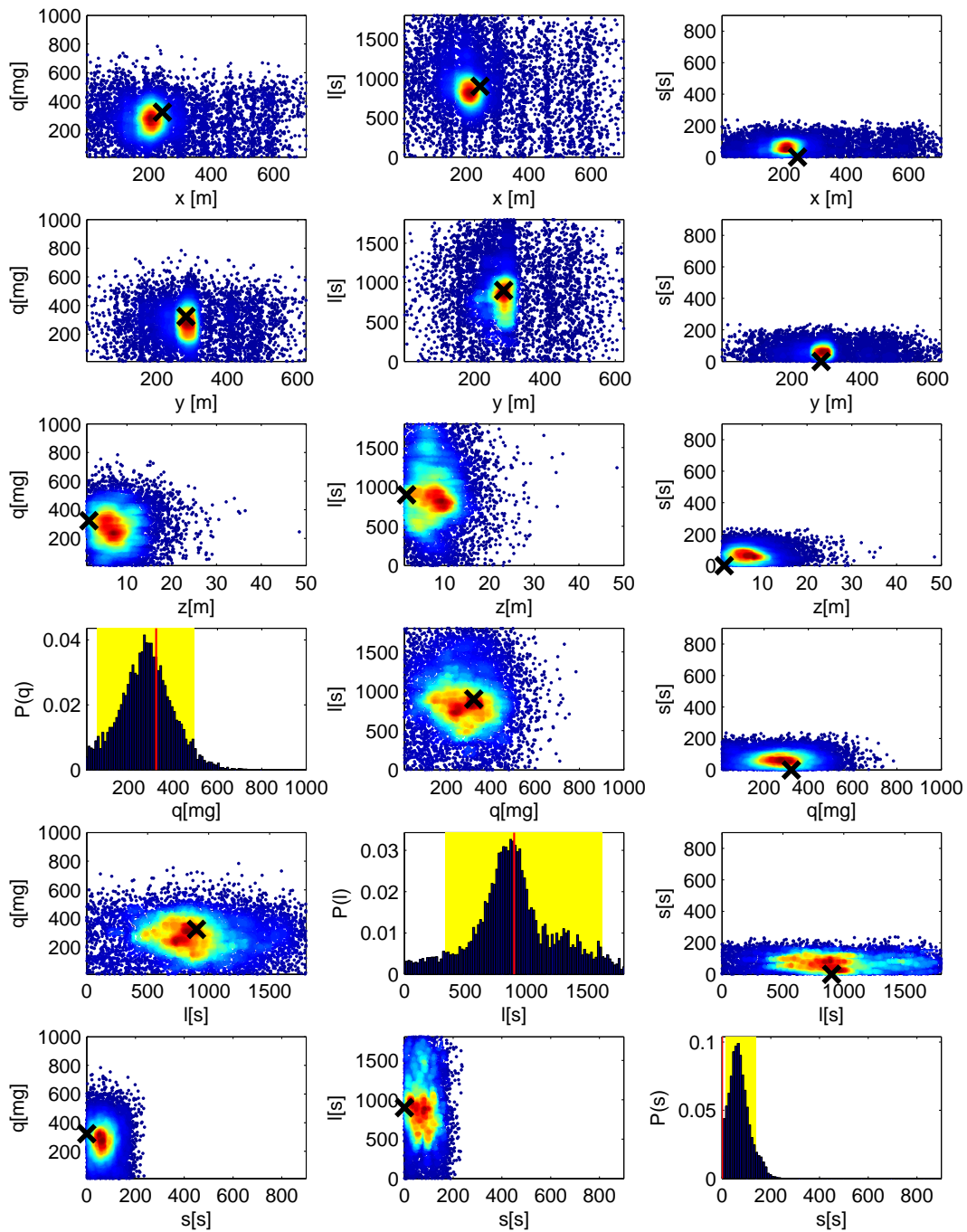
W niniejszej rozprawie przedstawiono szeroki przegląd literatury oraz metod z dziedziny stochastycznej rekonstrukcji zdarzeń. W szczególności, w rozdziale 1 przytoczono wszystkie najważniejsze publikacje wraz z odniesieniami do poszczególnych elementów modelu rekonstrukcji parametrów. Jednym z najważniejszych celów niniejszej rozprawy była dekompozycja modelu SER na poszczególne elementy i precyzyjny ich opis, co zostało zrealizowane w rozdziale 1. Wyodrębnienie takich komponentów jak zbiory danych obserwowalnych, rozkłady a priori parametrów, progresywny model dyspersji oraz algorytmy próbkujące, pozwoliło na formalny opis problemu badawczego.

2. Zbudowanie własnego modelu rekonstrukcji parametrów źródła opierającego się na metodach bayesowskich oraz wykorzystaniu modeli dyspersji gazu jako modeli „w przód”.



Rysunek 1.8: Rozkłady a posteriori dla poszczególnych kombinacji par oraz pojedynczych parametrów, uzyskane poprzez zastosowanie procedury rekonstrukcji z wykorzystaniem danych z eksperymentu DAPPLE. Na głównej przekątnej pokazane są brzegowe rozkłady a posteriori dla wszystkich parametrów. Czerwona linia na wykresach diagonalnych określa prawdziwą wartości parametrów  $\theta^*$ . Na pozostałych wykresach pokazane są rozkłady brzegowe dla wszystkich kombinacji dwóch parametrów. Czarny krzyż oznacza prawdziwą wartość. Rozkłady zostały wyznaczone na podstawie próbek wygenerowanych po 10 krokach czasowych. Rysunek stanowi lewą część całej macierzy wykresów, której dopełnieniem jest rysunek 1.9.





Rysunek 1.9: Rozkłady a posteriori dla poszczególnych kombinacji par oraz pojedynczych parametrów, uzyskane poprzez zastosowanie procedury rekonstrukcji z wykorzystaniem danych z eksperymentu DAPPLE. Na głównej przekątnej pokazane są brzegowe rozkłady a posteriori dla wszystkich parametrów. Czerwona linia na wykresach diagonalnych określa prawdziwą wartości parametrów  $\theta^*$ . Na pozostałych wykresach pokazane są rozkłady brzegowe dla wszystkich kombinacji dwóch parametrów. Czarny krzyż oznacza prawdziwą wartość. Rozkłady zostały wyznaczone na podstawie próbek wygenerowanych po 10 krokach czasowych. Rysunek stanowi prawą część całej macierzy wykresów, której dopełnieniem jest rysunek 1.8

W rozdziale 2 zaproponowano model bayesowski oraz mechanizm implementujący dowolny model dyspersji wewnątrz procedury SER. Zastosowanie metodologii bayesowskiej pozwoliło na zupełnie naturalne połączenie wielu elementów o charakterze probabilistycznym. W podejściach opierających się na metodach próbkowania z reguły występuje wysokie zapotrzebowanie na moce obliczeniowe. Zredukowane je poprzez zastosowanie zaawansowanych algorytmów próbkowania, które zostały opisane w rozdziale 3. Dodatkowo, wykorzystano dynamiczny proces dostarczania danych do modelu odwrotnego. Zaproponowany model „online” danych obserwowanych stanowi duży krok w przód, zwłaszcza w kontekście zastosowania metodologii SER w systemie operującym w czasie rzeczywistym.

3. Zastosowanie w problematyce SER zaawansowanych algorytmów próbkowania oraz metod, takich jak sekwencyjne metody Monte Carlo (ang. *Sequential Monte Carlo*, SMC) oraz przybliżone obliczenia bayesowskie (ang. *Approximate Bayesian Computation*, ABC).

Własna implementacja zaawansowanych algorytmów próbkowania oraz ich integracja z nowymi modelami dyspersji jest ważnym elementem niniejszej rozprawy. Na przestrzeni całej pracy badawczej zastosowano zmodyfikowany algorytm Metropolisa-Hastingsa, sekwencyjny algorytm Monte Carlo (ang. *Sequential Monte Carlo*, SMC) oraz algorytm przybliżonych obliczeń bayesowskich (ang. *Approximate Bayesian Computation*, ABC). Przydatność wszystkich algorytmów w zagadnieniach STE została zweryfikowana poprzez zastosowanie ich wyłącznie na danych pochodzących z realnych eksperymentów. Szczególnie metody przybliżonych obliczeń bayesowskich stanowią zupełną nowość w dziedzinie estymacji parametrów źródła uwolnienia gazu do atmosfery.

4. Opracowanie własnych modyfikacji do wyżej wymienionych algorytmów, dedykowanych problemowi estymacji parametrów źródła uwolnienia gazu.

Zaproponowano modyfikację algorytmu przybliżonych obliczeń bayesowskich, dostosowując go do specyfiki rozpatrywanych zagadnień rekonstrukcji parametrów źródła. Za najważniejsze modyfikacje należy uznać:

- silnie umotywowane rozkłady a priori wynikające z charakterystyk rozpatrywanych parametrów źródła
- zmodyfikowaną miarę rozbieżności opartą na formule, która nominalnie służyła weryfikacji modeli dyspersji gazu.
- procedurę adaptacji harmonogramu progów akceptacji – wstępna modyfikacja
- adaptacyjny mechanizm wag w algorytmie ABC-SMC-AW
- zastosowanie zaawansowanych modeli jądra przejścia dostosowanych do

problemu źródła poruszającego się w czasie emisji oraz emitera osadzonego w środowisku miejskim

5. Zwiększenie dokładności wyników rekonstrukcji poprzez zastosowanie dynamicznych modeli dyspersji w metodologii SER, takich jak: *Second-order Closure Integrated Puff* - SCIPUFF, *Quick Urban and Industrial Complex Dispersion Modeling System* - QUIC.

Dużym wkładem w rozwój metodologii stochastycznej rekonstrukcji zdarzeń jest zastosowanie zaawansowanych progresywnych modeli dyspersji gazu. Wynikało to bezpośrednio z potrzeby dynamicznej symulacji zaistniałych zdarzeń. Dotychczas takie modele jak QUIC nie były stosowane w zagadnieniach estymacji parametrów źródła uwolnienia gazu. Zastosowanie programów symulujących transport gazu, znacznie bardziej rozwiniętych niż często stosowana formuła Pasquilla, pozwoliło na rozpoznanie zupełnie nowych w tej dziedzinie klas problemów. W efekcie, w niniejszej rozprawie rozpatrzono zagadnienie identyfikacji charakterystyk źródła poruszającego się w czasie emisji oraz problem lokalizacji emitera gazu w środowisku silnie zurbanizowanym.

6. Analiza działania opracowanych modeli SER w wielu wariantach dla nierekonstruowanych dotychczas eksperymentów:
  - 6.1 Porównanie efektywności wybranych algorytmów stochastycznej rekonstrukcji zdarzeń z zastosowaniem rzeczywistego zbioru danych z eksperymentu uwolnienia gazu  $SF_6$  przeprowadzonego podczas kampanii *Copenhagen Tracer Experiments* - CTE.

W rozdziale 5 zaprezentowano studium porównawcze zaimplementowanych algorytmów stochastycznej rekonstrukcji zdarzeń w warunkach rzeczywistego zagadnienia. Porównanie wyników testowanych metod SER w rekonstrukcji tego samego eksperymentu przeprowadzonego w ramach kampanii *Copenhagen Tracer Experiments* pozwoliło wykazać przewagę implementacji algorytmu ABC-SMC-AW nad MCMC oraz SMC. Model ten przy tych samych warunkach zdołał bardziej precyzyjnie zidentyfikować lokalizację źródła  $(x,y)$ , co dało możliwość rozpoznania bardziej wrażliwych parametrów, takich jak poziom emisji gazu  $q$ .

- 6.2 Identyfikacja charakterystyk źródła poruszającego się w czasie emisji na podstawie danych z eksperymentu *Over-land atmospheric dispersion* – OLAD (Biltoft & Turley, 2000),

W rozdziale 6 zaprezentowano model SER pozwalający na skuteczną identyfikację aż siedmiu parametrów charakteryzujących ruchome źródło emisji gazu. Zdefiniowana charakterystyka źródła zawierała pozycję startową

emitera, prędkość oraz kierunek ruchu, czas rozpoczęcia uwolnienia gazu, długość emisji oraz ilość substancji. Działanie zaproponowanej procedury zostało zweryfikowane na podstawie danych z eksperymentu *Over-Land Alongwind Dispersion*. Wyniki oraz przedstawiony algorytm został opublikowany w (Kopka et al., 2015, 2016).

### 6.3 Lokalizacja źródła uwolnienia gazu w środowisku silnie zurbanizowanym w oparciu o dane pochodzące z serii eksperymentów *Dispersion of Air Pollution and its Penetration into the Local Environment* – DAPPLE.

W rozdziale 7 w modelu SER posłużono się miejskim modelem dyspersji gazu QUIC. Umożliwiło to estymację sześciu parametrów charakteryzujących źródło emisji gazu w mieście: pozycję źródła, wysokość wyniesienia emitera, ilość uwolnionej substancji, czas rozpoczęcia oraz długość uwolnienia. W celu weryfikacji poprawności działania zaproponowanej procedury wykorzystano dane eksperymentalne pochodzące z kampanii *Dispersion of Air Pollution and its Penetration into the Local Environment*. Przedstawiona metodologia oraz wyniki procedury rekonstrukcji zostały opublikowane w pracy (Kopka & Wawrzynczak, 2018).

Przedstawione powyżej podpunkty oraz rezultaty zawarte w rozprawie potwierdzają rozstrzygnięcie problemów badawczych oraz realizację celów postawionych w rozdziale 1. Cel, jakim była implementacja algorytmów próbkowania oraz opracowanie dynamicznego modelu rekonstrukcji pozwalającego na identyfikację parametrów źródeł uwolnienia gazów do atmosfery na podstawie stężeń uwolnionej substancji pochodzących z sieci sensorów został zrealizowany. W szczególności, jest to zaprezentowane w badaniach dotyczących złożonych problemów identyfikacji parametrów źródła, gdzie zagadnienie potraktowano jako proces dynamiczny. Najważniejsze wyniki przedstawione w niniejszej rozprawie zostały przedstawione na konferencjach, opublikowane lub zostały wysłane do publikacji w znaczących czasopiśmie.

## Bibliografia

- Arnold, S., ApSimon, H., Barlow, J., Belcher, S., Bell, M., Boddy, J., . . . Colvile, R. (2004). Introduction to the DAPPLE air pollution project. *Science of the Total Environment*, 332(1), 139–153.
- Biltoft, C., & Turley, S. (2000). Final test report of the Over-Land Alongwind Dispersion Field Trials.
- Gryning, S.-E. (1981). *Elevated source SF 6-tracer dispersion experiments in the Copenhagen area* (Tech. Rep.).
- Gryning, S.-E., & Lyck, E. (1984). Atmospheric dispersion from elevated sources in an urban area: comparison between tracer experiments and model calculations. *Journal of climate and applied meteorology*, 23(4), 651–660.
- Kopka, P., & Wawrzynczak, A. (2018). Framework for stochastic identification of atmospheric contamination source in an urban area. *Atmospheric environment*, 195, 63–77.
- Kopka, P., Wawrzynczak, A., & Borysiewicz, M. (2015). The Approximate Bayesian Computation methods in the localization of the atmospheric contamination source. In *Journal of physics: Conference series* (Vol. 633, p. 012118).
- Kopka, P., Wawrzynczak, A., & Borysiewicz, M. (2016). Application of the Approximate Bayesian Computation methods in the stochastic estimation of atmospheric contamination parameters for mobile sources. *Atmospheric Environment*, 145, 201–212.
- Sykes, R., Henn, D., & Parker, S. (2000). PC-SCIPUFF Version 1.3 technical documentation. *Titan Corporation, ARAP Report No, 725*.
- Watson, T. (2000). *The Over-Land Along-Wind Dispersion (OLAD) Field Experiment*. US Department of Commerce, National Oceanic and Atmospheric Administration, Oceanic and Atmospheric Research Laboratories, Air Resources Laboratory.
- Williams, M. D., Brown, M. J., Singh, B., & Boswell, D. (2004). QUIC-PLUME theory guide. *Los Alamos National Laboratory*.
- Wood, C. R., Barlow, J. F., Belcher, S. E., & Dobre, A. (2009). Dispersion experiments in central London: the 2007 DAPPLE project. *Bulletin of the American Meteorological Society*, 90(7), 955–969.