



U. Quass, T. A. J. Kuhlbusch



# Anforderungen an Messdaten zur Auswertung von Dioxin/PCB- Profilen mittels multivariater Verfahren zur Quellenidentifizierung

*Institut für Energie- und  
Umwelttechnik*

*Luftreinhaltung und  
Nachhaltige Nanotechnologie*

## **UFOPLAN-Projekt:**

**Zuordnung und Quantifizierung der Dioxineinträge  
auf dem Luftpfad mittels Betrachtung der immissionsseitigen  
und emissionsseitigen Kongenerenmuster (FKZ 3712 65 407 2)**

***Kann man die trotz der erheblichen Emissionsminderungen noch immer bestehenden PCDD/F und PCB-Belastungen von Biota und Lebensmitteln auch auf Einträge über den Luftpfad zurückführen,  
und wenn ja, welche bekannten oder neuen Quellen sind dafür verantwortlich?***

### **Verfahren zur Quellenidentifizierung und –quantifizierung**

***Nutzung räumlicher, zeitlicher und Dateninhärenter Information  
Anforderungen an die auszuwertenden Daten***

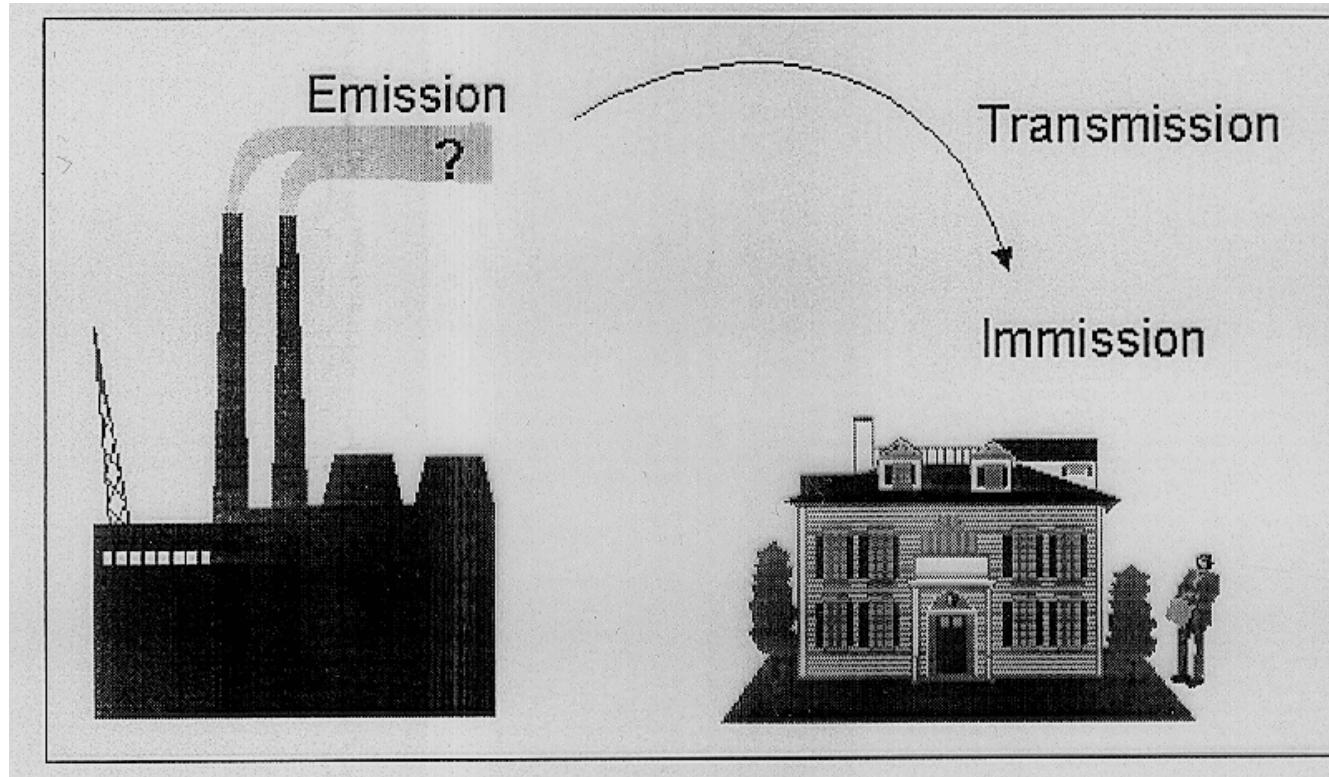
### **Für das Projekt verfügbare Daten (Immissionen/Depositionen/Emissionen)**

***Überblick zum Datenbestand  
Probleme und deren Behebung***

### **Ein Fallbeispiel**

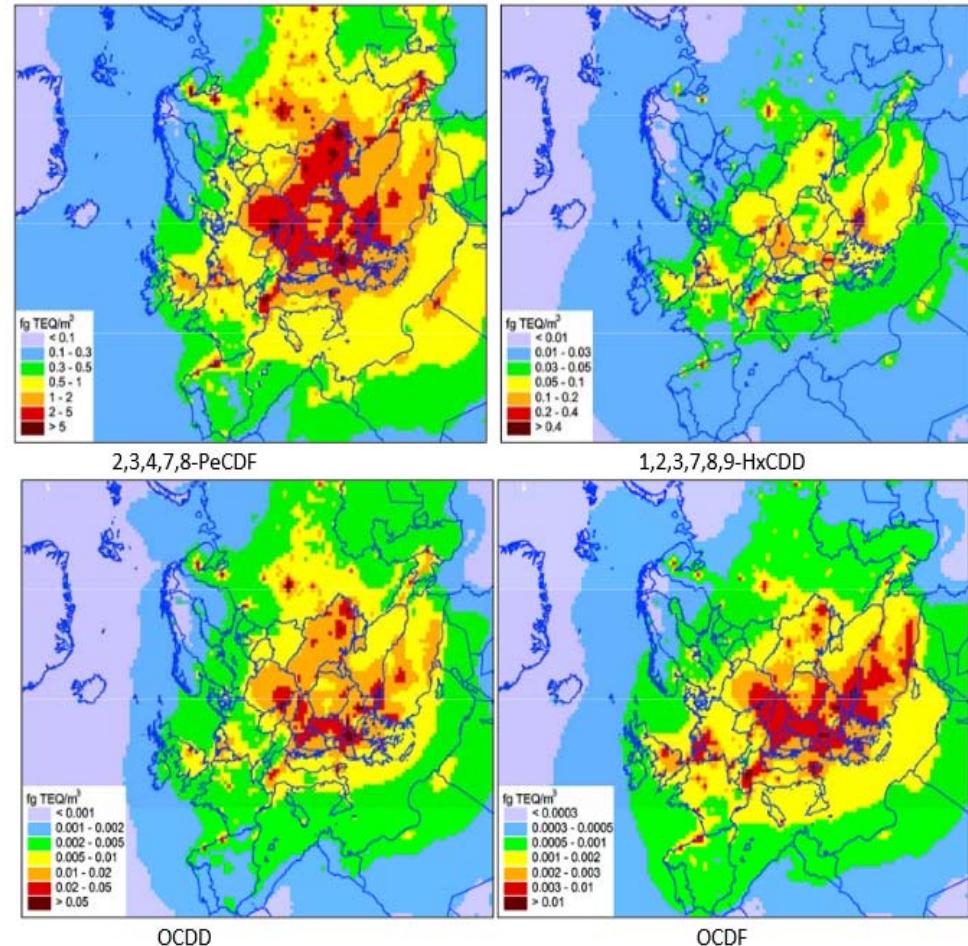
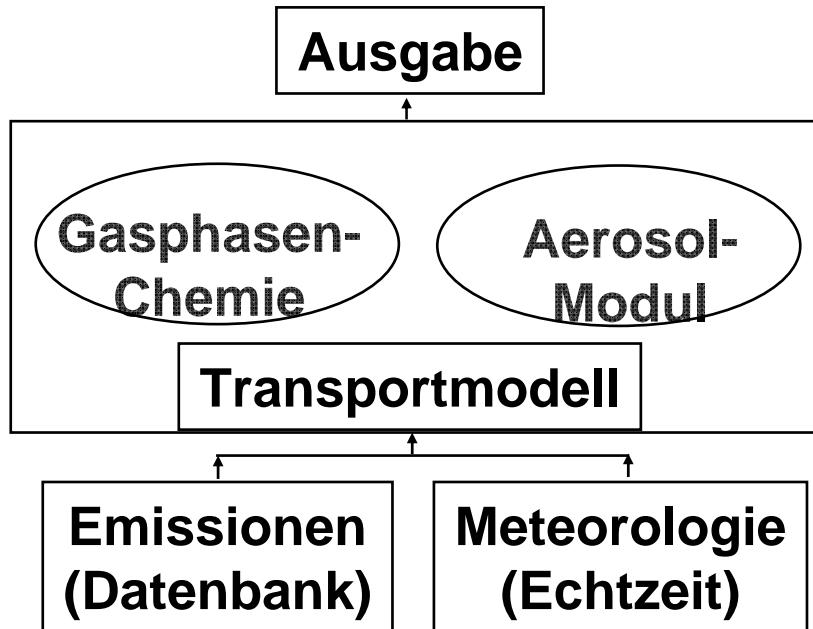
### **Schlussfolgerungen**

# Ermittlung von Quellbeiträgen zur Immissionsbelastung

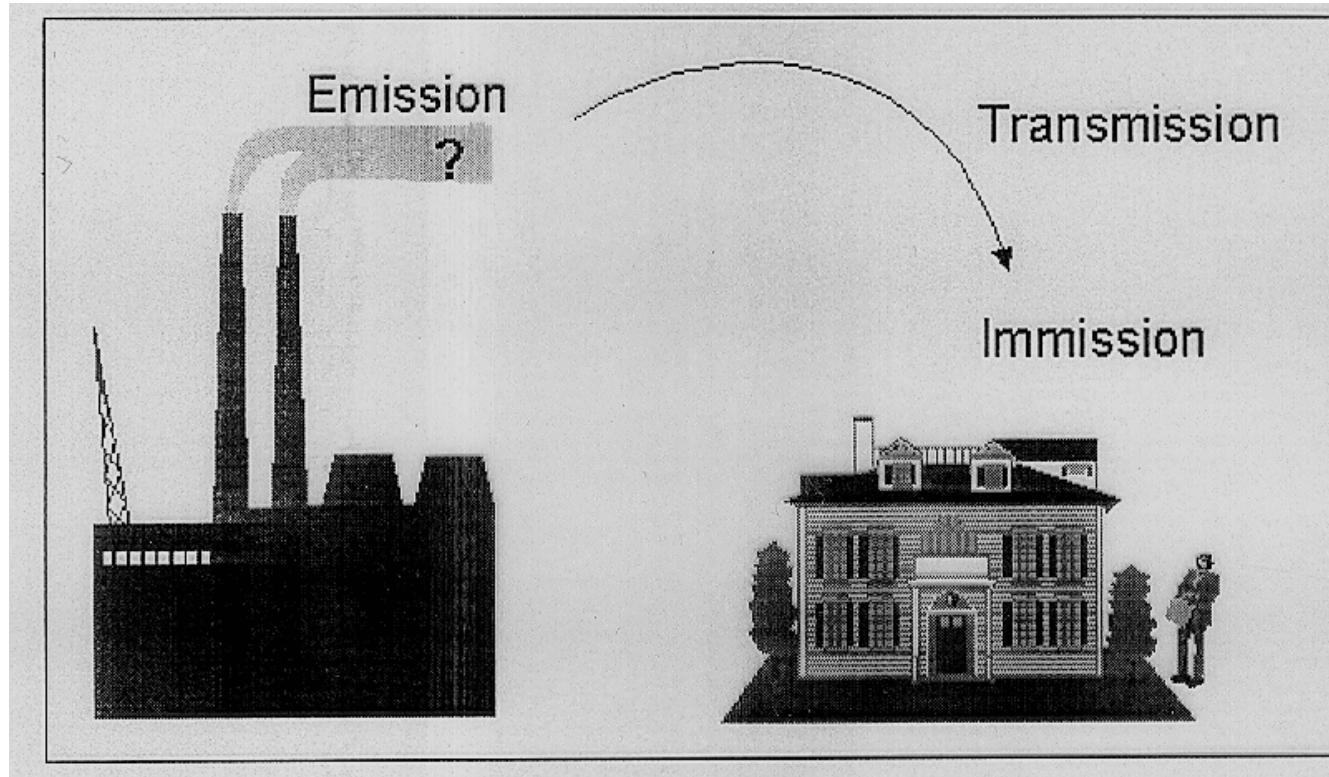


Ausbreitungsrechnung →

# Ausbreitungsrechnung



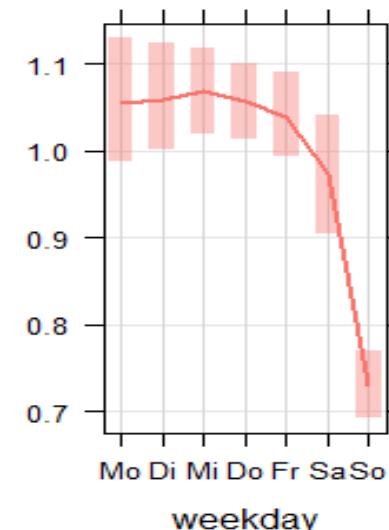
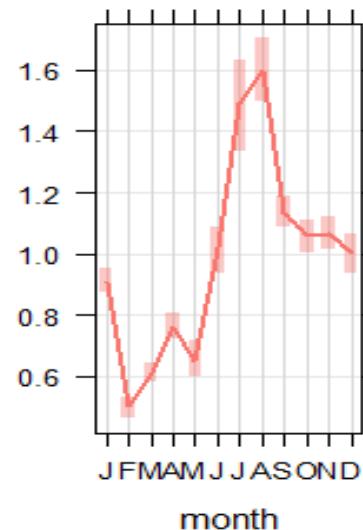
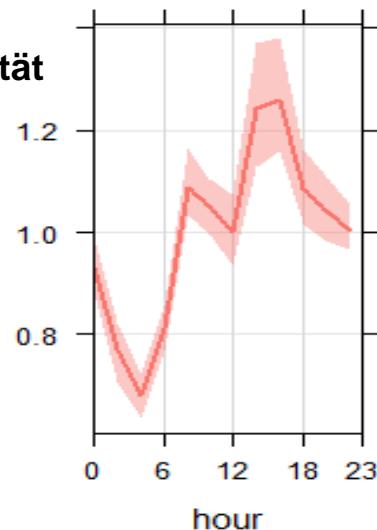
Shatalov et al. – Atmospheric Pollution Research 3 (2012) 408-416



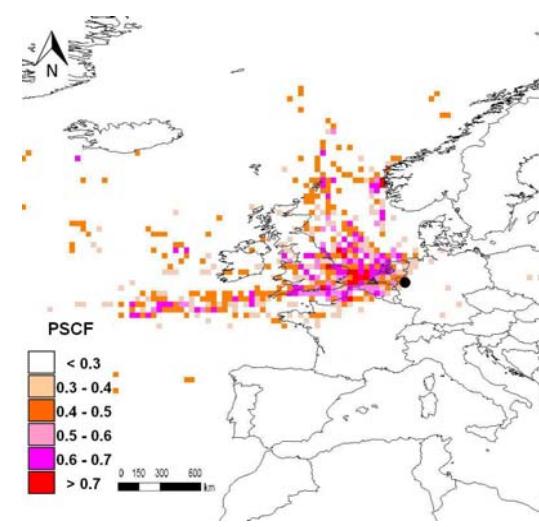
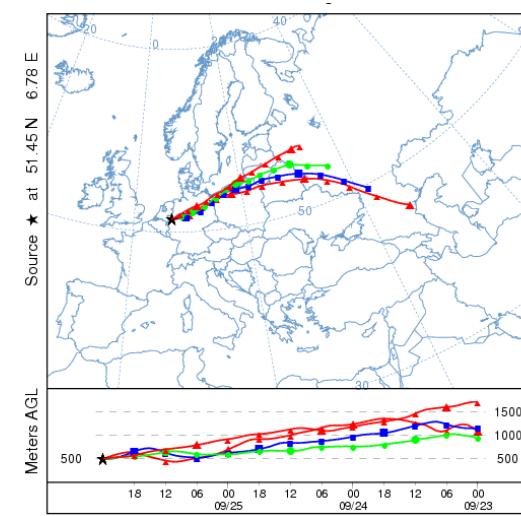
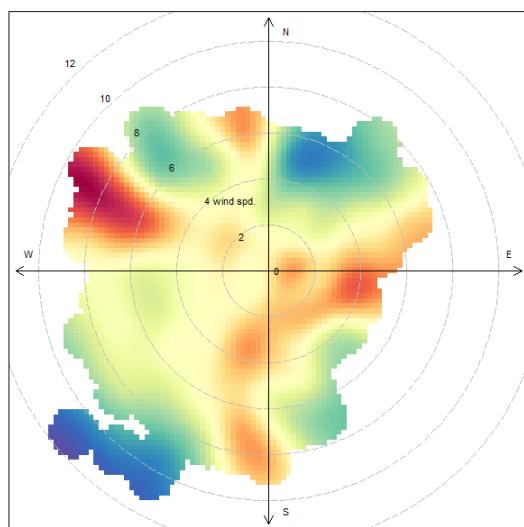
Ausbreitungsrechnung →  
← Datenanalysen

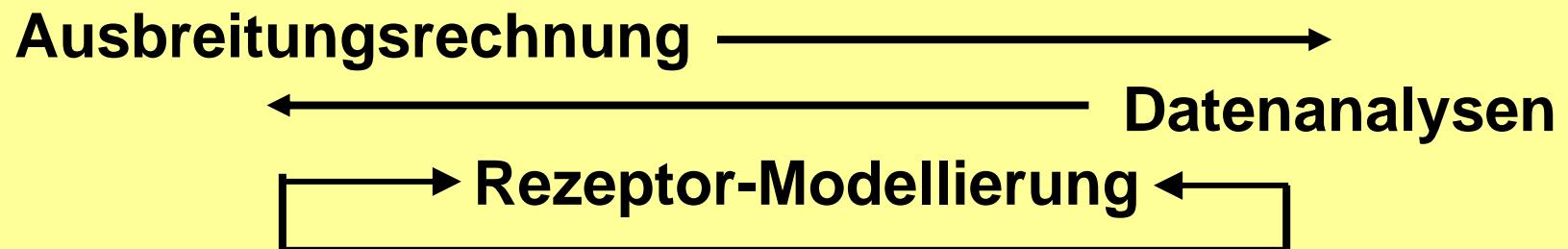
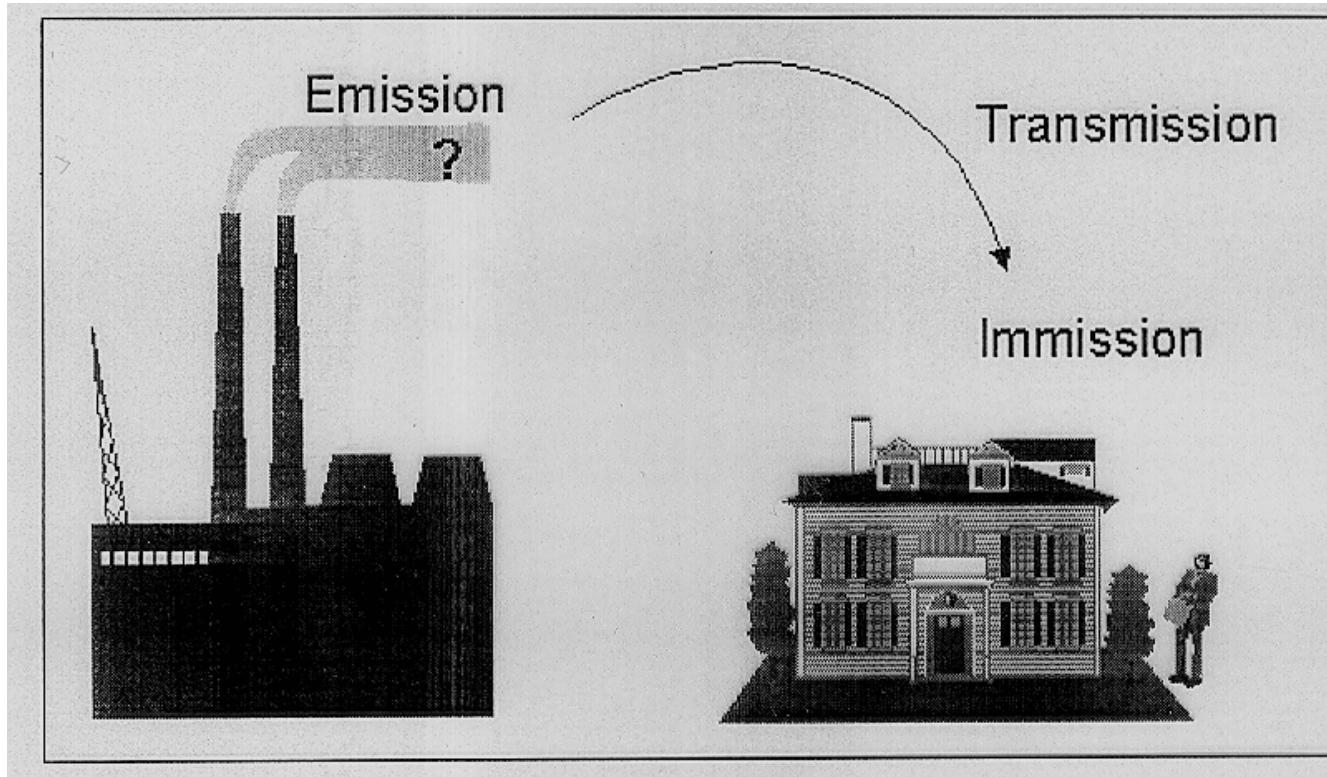
# Datenanalysen

## Zeitliche Variabilität

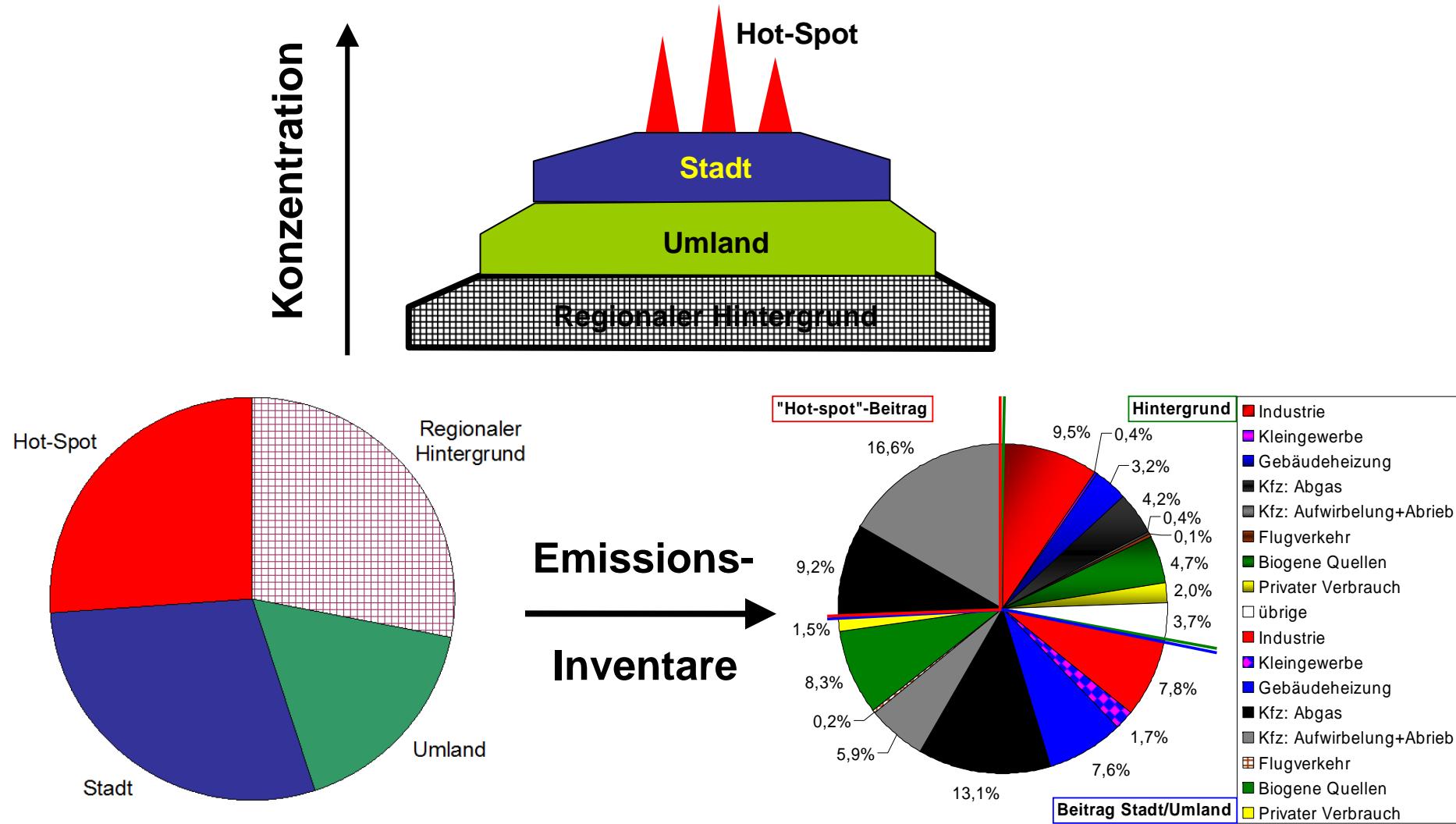


## Räumliche Variabilität/Koinzidenz



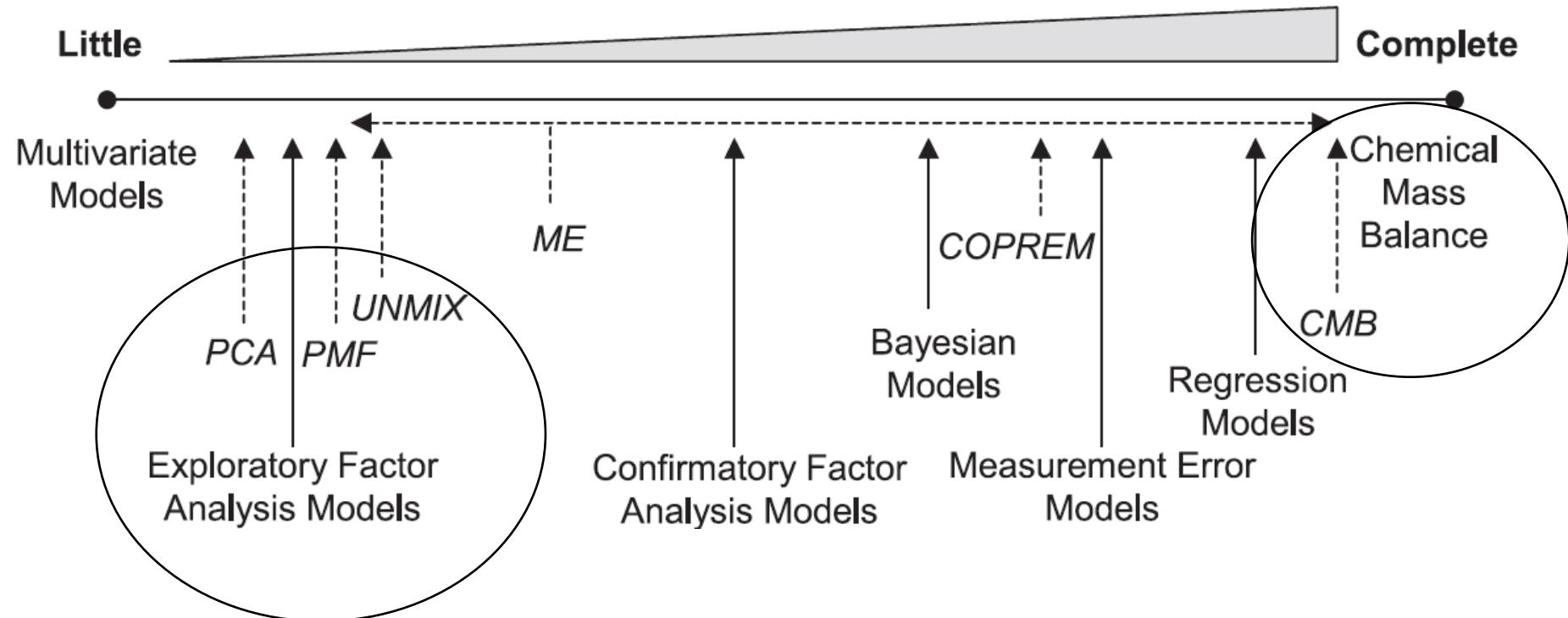


# Räumliche Gradienten: „Lenschow-Ansatz“



(Lenschow et. al, Atmospheric Environment, Volume 35, Number 1, 2001 , pp. 23-33(11)

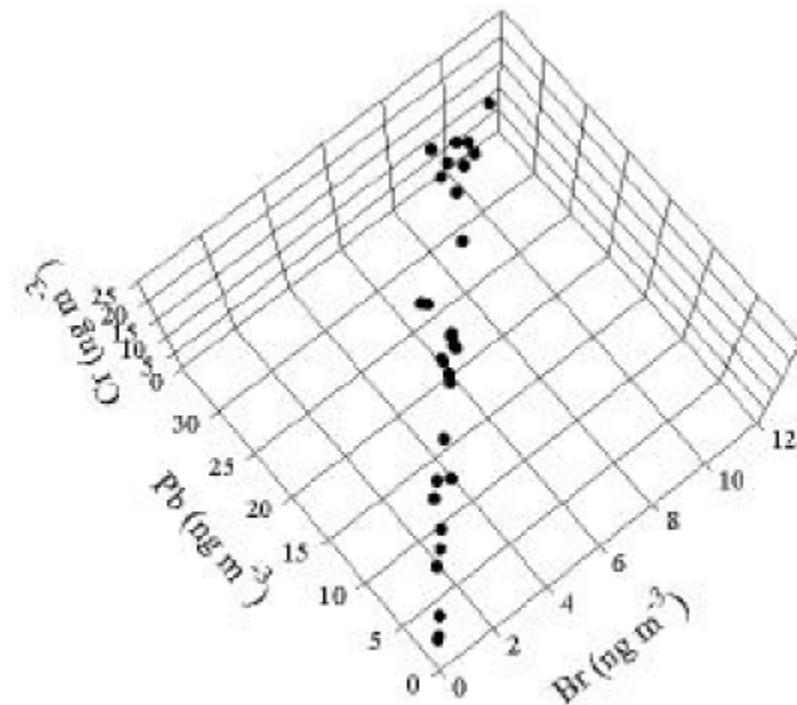
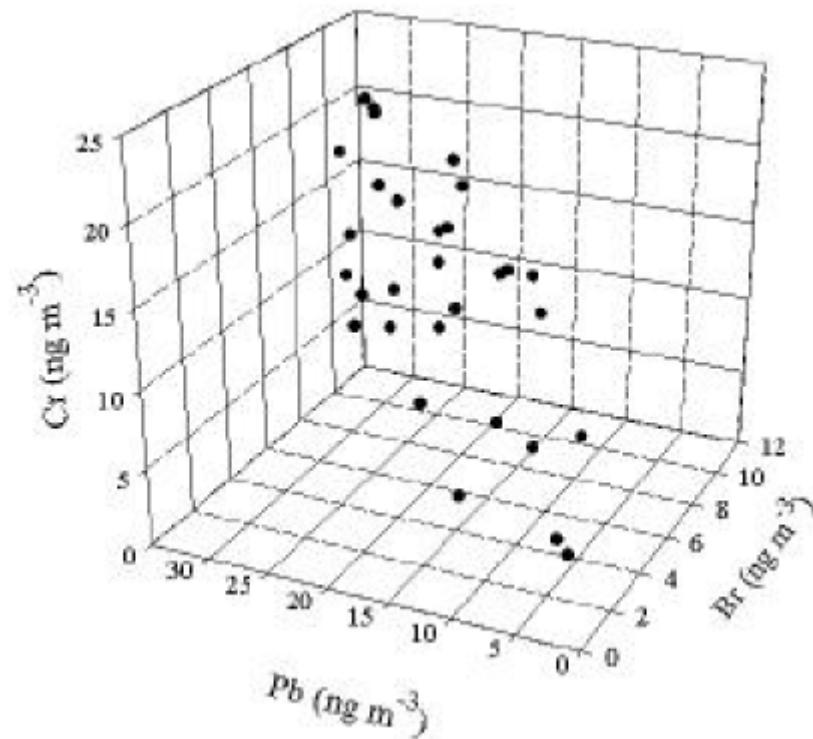
## Benötigtes Vorwissen über Emissionsquellen



M. Viana et al. / Aerosol Science 39 (2008) 827–849

**Ziele: Reduktion der Variablenanzahl**

**Identifikation der maßgeblichen latenten Einflussgrößen  
„Faktoren“) und ihrer Struktur bzgl. der ursprünglichen Variablen**



P.K. Hopke: A guide to Positive Matrix Factorisation

# Positive Matrix Factorisation (PMF)



$$X = G \cdot F + E$$

$$x_{i,j} = \sum_{k=1}^p g_{i,k} \cdot f_{k,j} + e_{i,j}$$

**Messwerte  
(Unsicherheiten)**

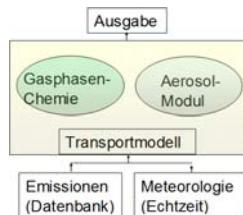
Probe	PM10 µg/m³	Summe F1-F8								
			F1	F2	F3	F4	F5	F6	F7	F8
23.4.2002	55,8	59,9	8,8	6,9	5,3	2,5	3,2	0,8	4,1	28,1
24.4.2002	49,9	46,9	1,7	7,3	3,0	1,5	4,1	8,8	3,0	17,6
25.4.2002	63,6	64,6	5,4	4,3	15,0	1,5	4,0	1,0	2,2	31,3
26.4.2002	59,1	57,2	1,8	3,0	13,7	0,7	2,7	2,8	4,0	28,5
27.4.2002	28,4	19,3	2,7	3,2	2,3	0,3	1,3	4,7	0,0	4,7
28.4.2002	21,1	13,4	1,7	0,5	2,3	0,3	0,4	2,7	1,8	3,8
29.4.2002	14,0	12,7	1,0	3,3	0,9	0,8	1,1	4,4	1,2	0,0
30.4.2002	18,3	17,5	0,1	3,4	0,6	1,3	0,9	8,3	2,6	0,3
1.5.2002	17,3	12,3	3,2	0,9	0,6	0,5	2,0	3,6	0,8	0,8
2.5.2002	12,3	13,8	5,6	1,3	1,5	0,2	1,1	1,6	0,1	2,5
Mittelwert	33,1	31,8	8,3	1,8	5,1	1,6	1,4	2,0	3,0	8,7

# Quellbeiträge

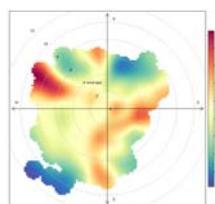
Faktor	PM10	PM2,5	PM10-2,5	OM	Nitr	Sulfat	Ammonium	Chlорид	Natrium	Magnesium	Kalium	Caizum	Al203	Fe203	PbO	C203	NiO	Cr2O3	CuO	ZnO
1	1,00	0,25	0,75	0,38	0,00	0,14	0,00	0,00	0,00	0,0000	0,0006	0,0000	0,00	0,00	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
2	1,00	0,00	1,00	0,21	0,00	0,00	0,00	0,00	0,00	0,0003	0,0005	0,01	0,00	0,00	0,0004	0,0001	0,00017	0,00008	0,0040	0,0000
3	1,00	0,00	0,02	0,00	0,69	0,16	0,00	0,00	0,00	0,0000	0,0000	0,00	0,00	0,00	0,0001	0,0001	0,0001	0,0001	0,0001	0,0000
4	1,00	0,00	0,14	0,01	0,00	0,00	0,00	0,00	0,00	0,0005	0,01	0,00	0,00	0,00	0,0001	0,0001	0,0001	0,00044	0,0140	0,0000
5	1,00	0,00	0,26	0,15	0,00	0,00	0,00	0,00	0,08	0,037	0,02	0,01	0,00	0,00	0,0000	0,0000	0,0001	0,0002	0,0020	0,0140
6	1,00	0,00	0,11	0,07	0,11	0,00	0,16	0,16	0,00	0,018	0,009	0,01	0,00	0,00	0,0001	0,0001	0,0000	0,0004	0,0001	0,0000
7	1,00	0,14	0,42	0,00	0,00	0,06	0,02	0,00	0,00	0,0000	0,016	0,01	0,03	0,01	0,0017	0,0008	0,0002	0,00005	0,0008	0,0040
8	1,00	0,01	0,02	0,62	0,00	0,20	0,00	0,00	0,00	0,0000	0,001	0,00	0,00	0,00	0,0002	0,0000	0,0000	0,0001	0,0001	0,0000

$$Q = \sum_{i=1}^n \sum_{j=1}^m \left[ \frac{x_{ij} - \sum_{k=1}^p g_{ik} f_{kj}}{u_{ij}} \right]$$

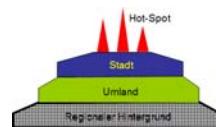
*...zu minimieren*



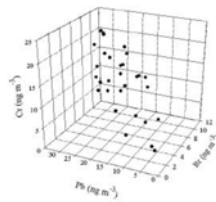
**Meteorologie, orts- und zeitaufgelöst  
Emissionsdaten für alle relevanten Quellen,  
Vorbelastungsdaten,  
Chemisch-phys. Umwandlungsprozesse**



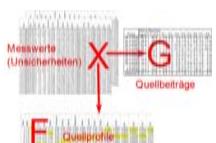
**Zeitaufgelöste Messungen  
(Stunden, Tage)**



**Messdaten (Mittelwerte) für Raumklassen;  
Raumklassen-bezogene Emissionsinventare**



**Matrix mit untereinander teilweise  
korrelierenden Messgrößen  
Keine/wenige Fehldaten bzw. BDLs  
Existenz quellen-indizierender Profile  
und/oder Tracer-Komponenten**



**Wie Faktorenanalyse, zusätzlich Matrix der  
Messunsicherheiten**

**Liefert Orts- und zeitaufgelöste Konzen-  
trations-Felder; Quellenzuordnung durch  
on/off-Simulationen (neuere Modelle  
auch direkt); aufwändig, *Defizite im  
E-Kataster verfälschen Ergebnis***

**Kleinräumige Identifizierung von Quellen;  
Quantifizierung z.B. durch Vergleich von  
Luv-Lee Situationen; *keine umfassende  
Quellenzuordnung***

**Abgrenzung lokaler von regionalen  
Einflüssen; Zuordnung bekannter u.  
registrierter Emissionen. *Defizite im  
E-Kataster verfälschen Ergebnis***

**Identifizierung möglicher Quellmuster;  
Auch für „unbekannte“ Quellen  
*Verifizierung erforderlich; phys. Interpre-  
tation oft unmöglich (Negativität)***

**wie FA, jedoch nicht-negative outputs,  
besseres Handling von Fehldaten/BDLs;  
*Subjektivität bei Interpretation  
von Quellprofilen***

**Masunaga et al. (2003): PCA von Proben versch. Kompartimente mit untersch. Messwert-Kombinationen; → 17K allein schlecht, 17K+8Hom. möglich, besser aber umfassendes Kongener-Spektrum; Quantifizierung über multilineare Regression**

Masunaga et al. Chemosphere 53 (2003), 315

**Götz und Lauer (2003): Cluster-Analyse von Elbe Sedimenten anhand verschiedener Referenzproben; 17K+8Hom., Identifizierung, aber keine Quantifizierung**

\*Götz und Lauer Environ. Sci. Technol. 37(2003), 5559

**Sundqvist et al (2010): PMF von Ostsee-Sedimenten anhand von Referenzproben mit breit analysiertem Kongener-Spektrum.**

Sundqvist et al, Environ. Sci. Technol. 44(2010), 1690

## UBA-POP-Dioxin-Datenbank

	Immission	Deposition	Emission
Erfasster Zeitraum	1992-2010	1990-2010	1991-2002
Anzahl Einzeldaten	88952	52123	4525
Anzahl Datensätze (Proben) insgesamt	2377	1880	225
Proben mit Daten zu:			
--PCDD/F	2161	1690	183 NRW: (1)
--PCB	1820	618	23
--PCDD/F und PCB	1726	433	14
Anzahl Messprogramme	17 <sup>2</sup>	19	6
Anzahl Standorte	86	78	31 (17 Anlagenarten)

Vom LANUV bereitgestellte Daten NRW (Monatswerte, 1995-2011)

PCDD/F	936	936	129
PCB	862	1101	keine vorhanden

## Korrektheit Probenahme (PN)-Zeiten:

- Unklare PN-Dauern (negativ, 0) → falsche bzw. fehlende Datumswerte ersetzt
- Identische PN-Zeiten für mehrere Proben am selben Standort
  - a) jeweils durchgängig 2 Proben, hohe Korrelation → Mittelwerte gebildet
  - b) Unplausibler Wechsel zu Doppelbeprobung → fehlerhafte Standort-Zuordnung... ok
  - c) > 2 Proben → Einsicht Originalberichte → versch. Standorte, Aufteilung PN-Zeit (noch für ein Messprogramm (SN) ungeklärt, sonst ok)

## Korrektheit der Messdaten (PCDD/F, Immission)

- Anzahl Nachkommastellen unterschiedlich, oft nur 3 bei Angabe in pg/m<sup>3</sup>
  - Verlust von Werten < 1 fg/m<sup>3</sup> (z. B. 0,4 fg/m<sup>3</sup> = 0,0004 pg/m<sup>3</sup> => 0,000 pg/m<sup>3</sup>)
  - viele „Nullwerte“ in Datenbank können Artefakte sein
  - mögliche Fehlerreduzierung durch Addition von 0,5 fg/m<sup>3</sup> zu verdächtigen Werten?
- Probeninterne Interne Datenkonsistenz: Homologen >= Kongenere(nsummen)?
  - Ausschluss einiger Datensätze (<5%)

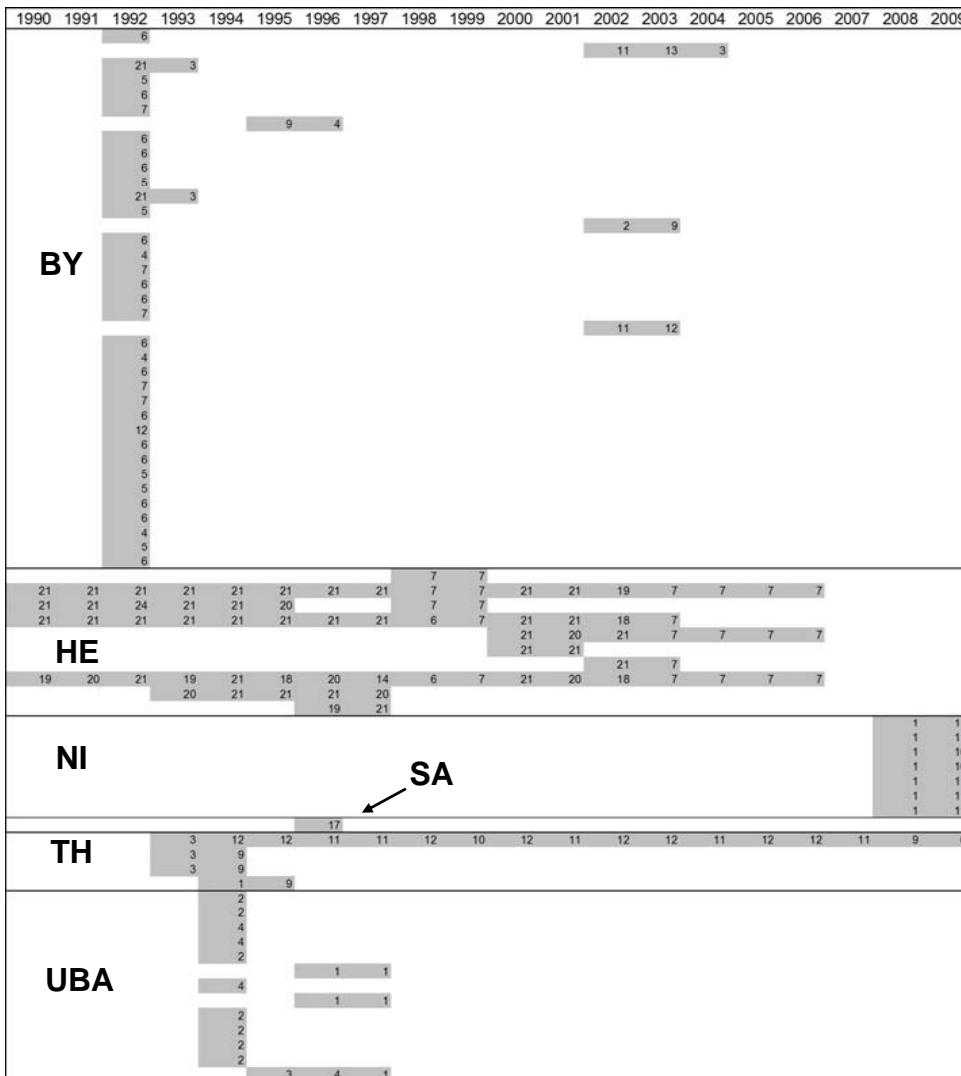
## Korrektheit der Standortangaben

- multiple Standort-IDs für gleichen Messort
- Fehlende/fehlerhafte Koordinaten, geschätzte RH-Werte, falsche Ortszuweisungen,
  - Berichtigung anhand von Berichten und Messnetzdaten
  - Klassifizierung der berichtigten Standorte (Messnetz-Infos oder Luftbild)

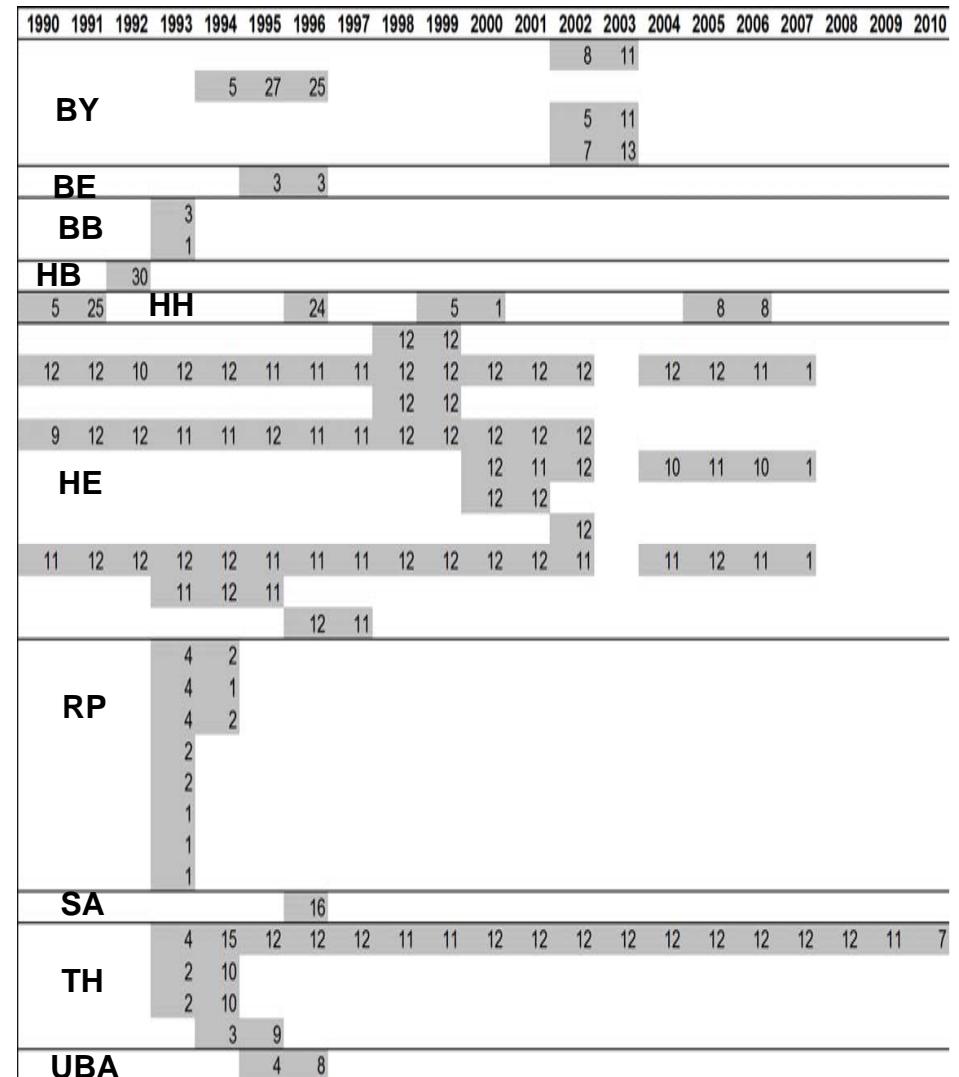
# Überblick Orts-Zeit-Verteilung IMM und DEP (PCDD/F, o. NRW)



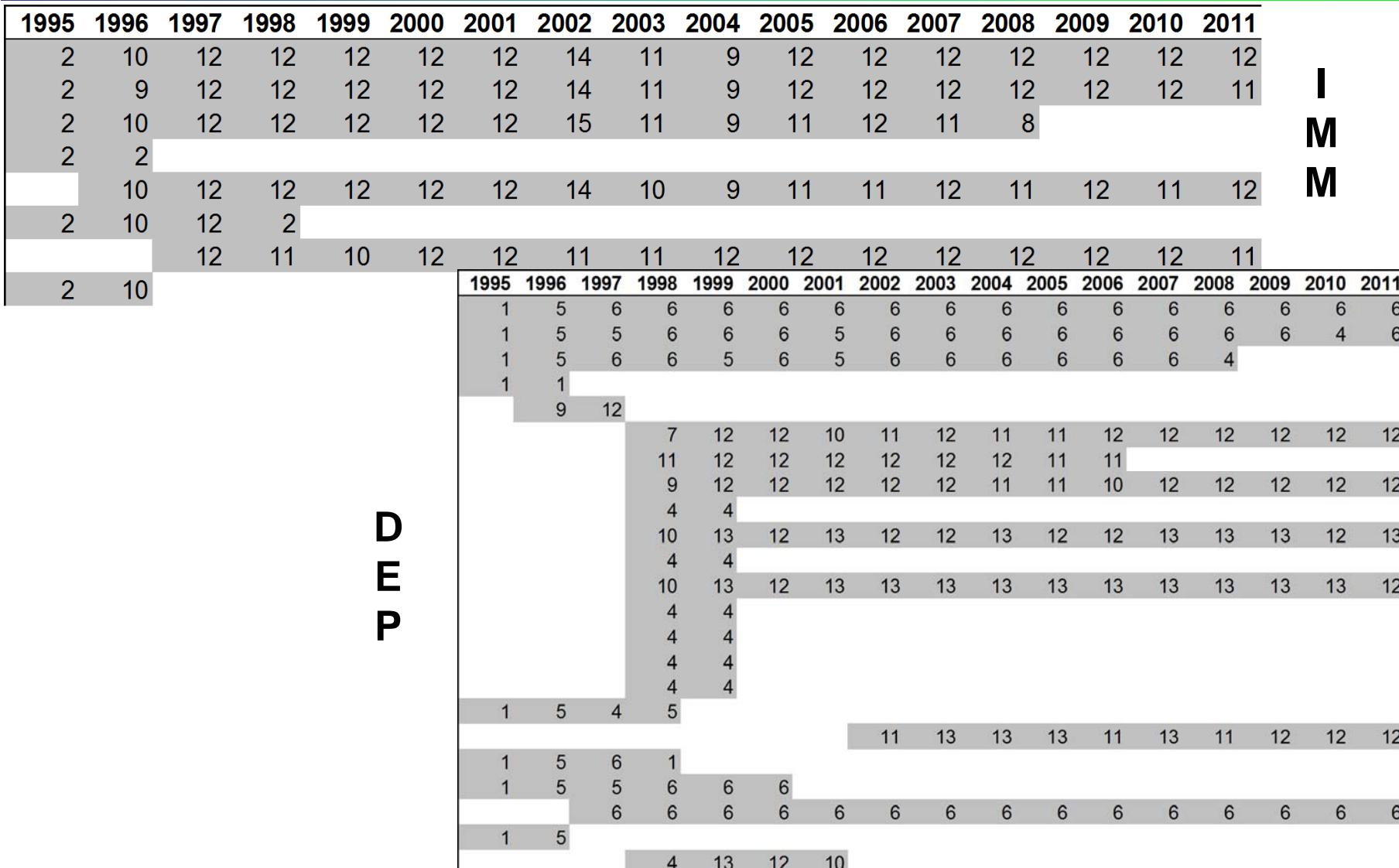
IMM



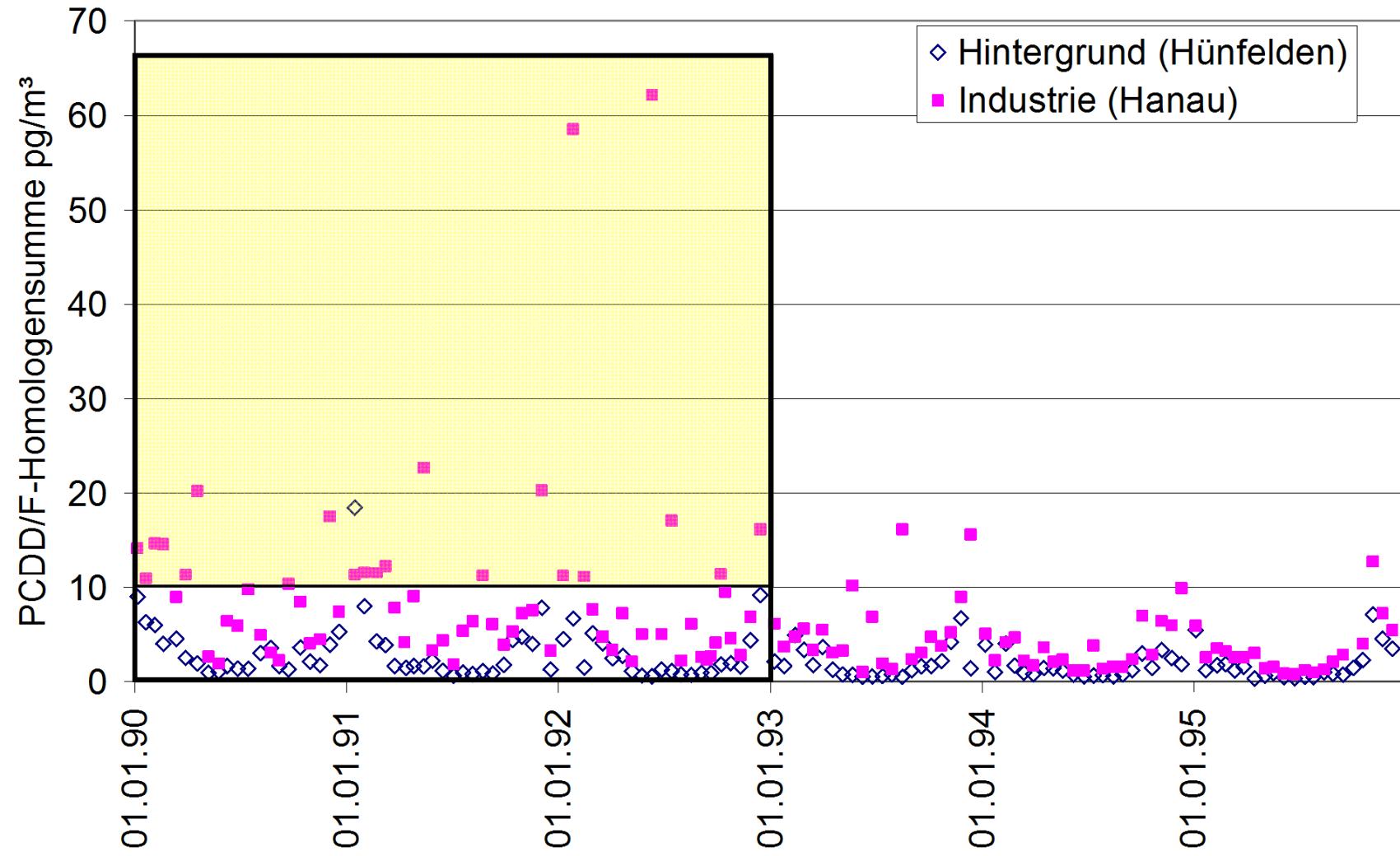
DEP



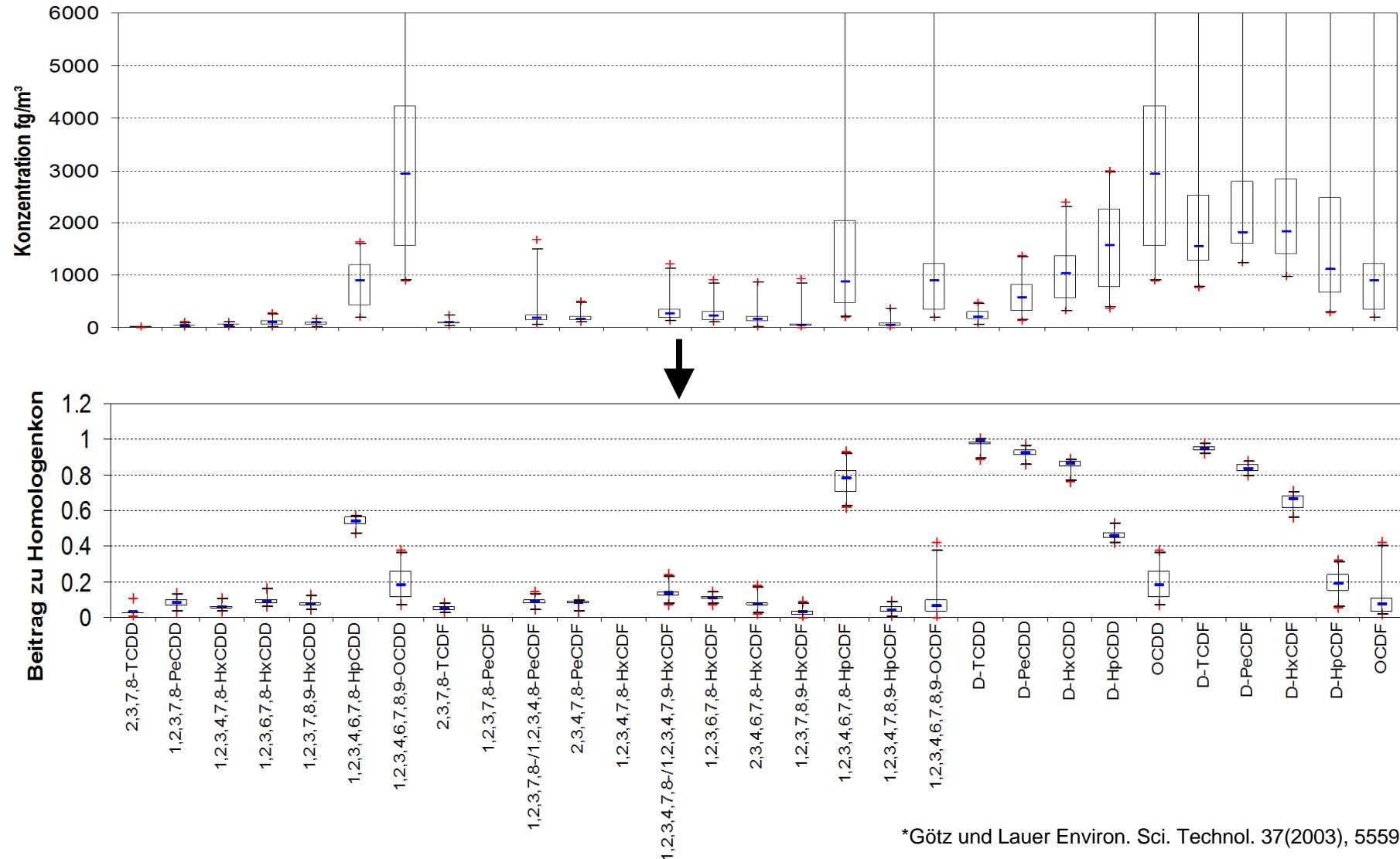
# Überblick Orts-Zeit-Verteilung IMM und DEP (PCDD/F, NRW)



# Testfall: Vergleich Hintergrund / Industriestandort

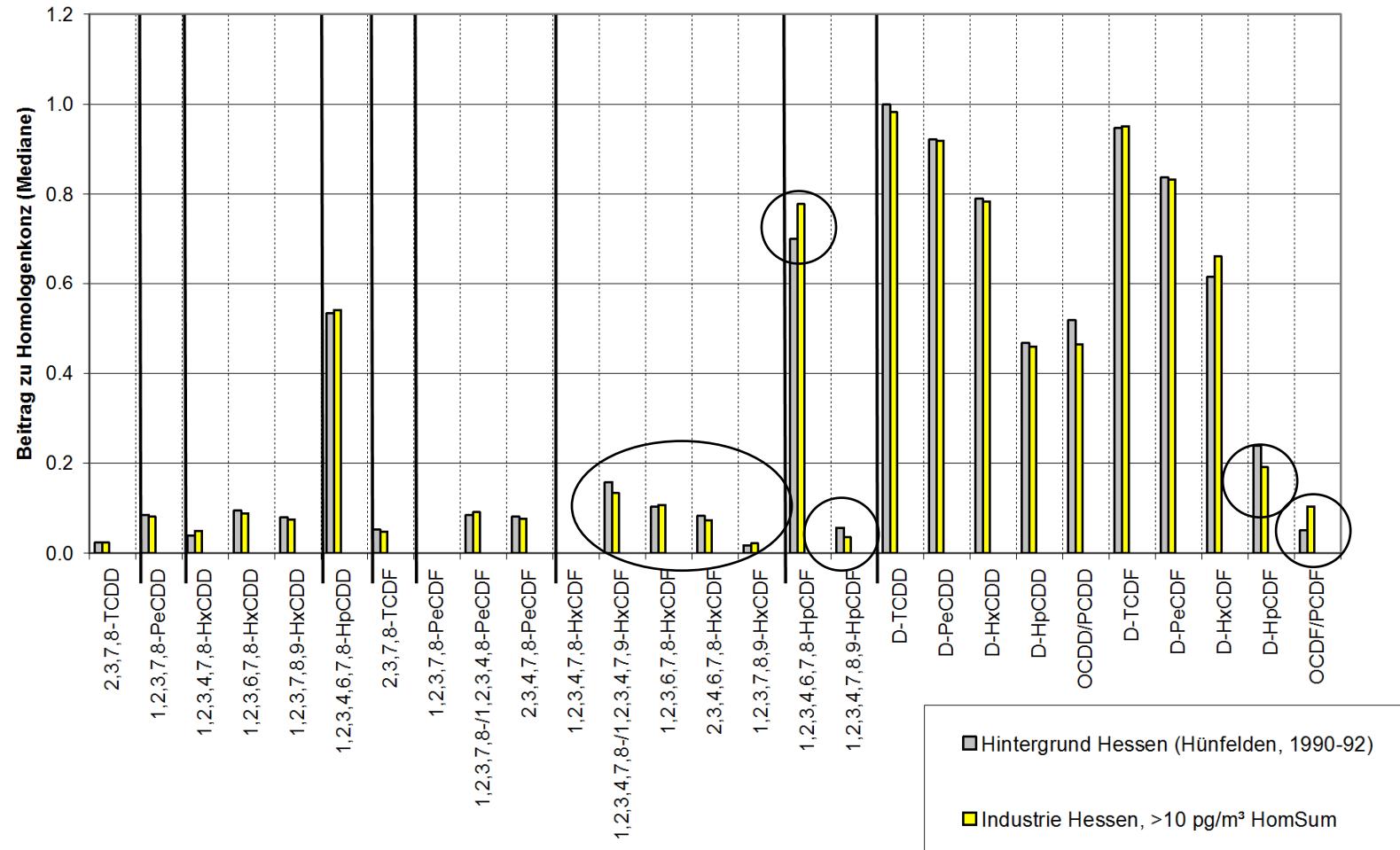


**Normierung \* auf jeweilige Homologen-Konzentration  
(OCDD/F: auf PCDD bzw. PCDF-Summe)**

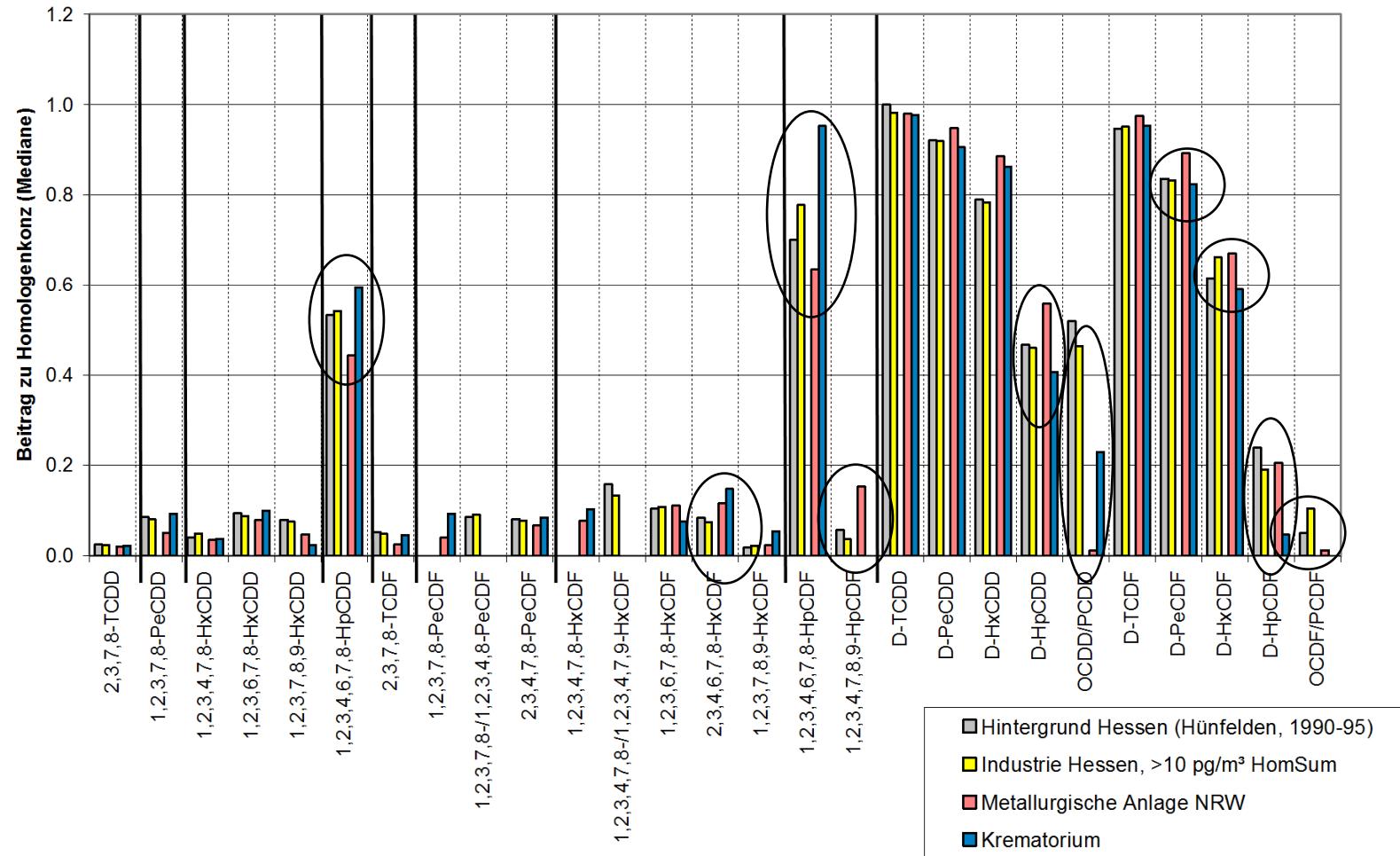


\*Götz und Lauer Environ. Sci. Technol. 37(2003), 5559

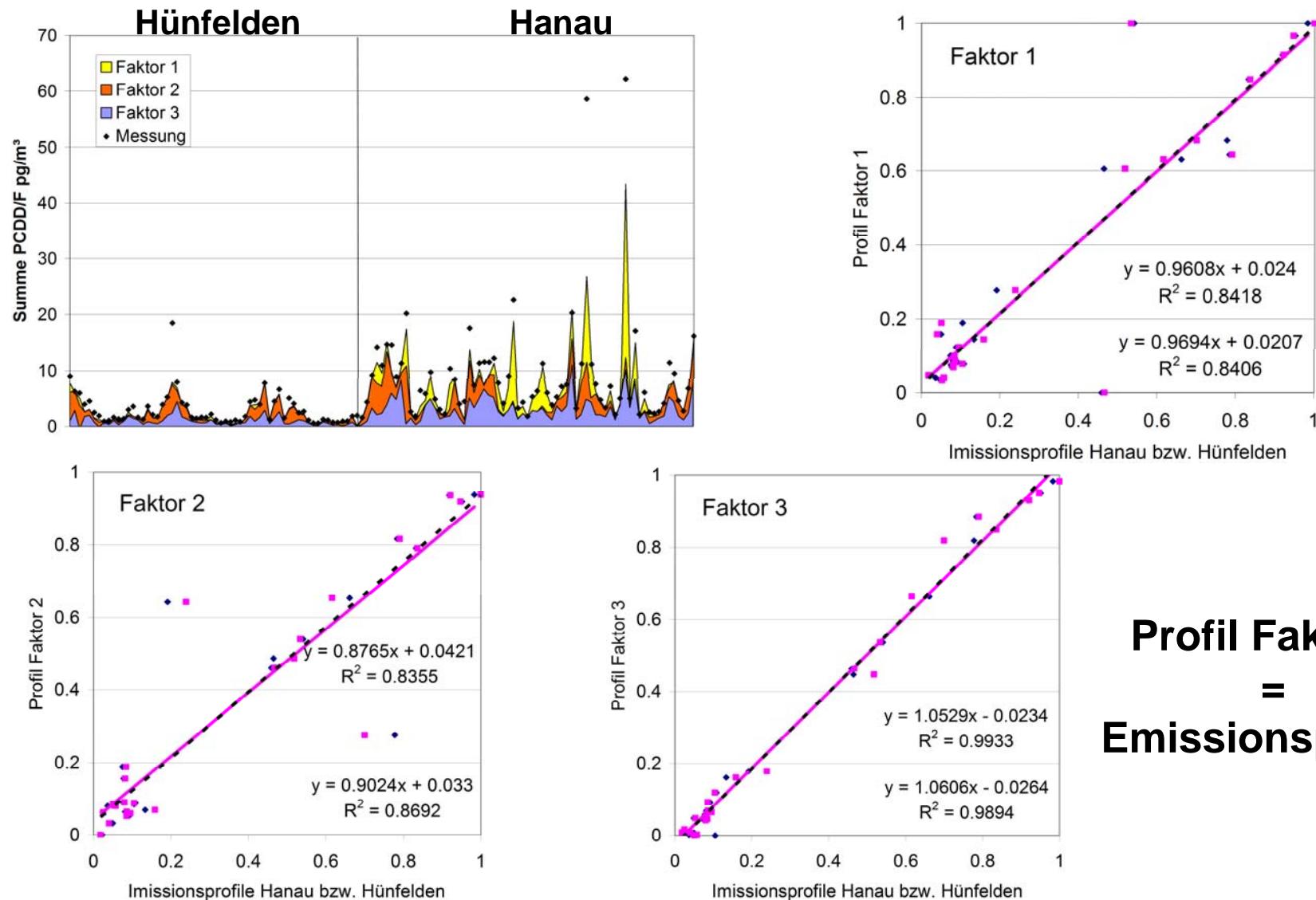
# Profil-Vergleich Hintergrund / Industriestandort



# Vergleich mit zwei (beliebig ausgesuchten) Quellprofilen



# PMF-Testlauf



**Profil Faktor 1**  
=  
**Emissionsprofil?**

## Schlussfolgerungen

---

- Verfahren um Quellen für Luftschadstoffe zu identifizieren und zu quantifizieren sind verfügbar
- Anwendung auf PCDD/F (PCB) Immissionen/Depositionen möglich
- Limitierungen für typische Kongeneren-Spektren und Zeitauflösung der Daten
- Für PDDC/F sind Profilvergleiche aus „17+8“-Datensätzen hilfreich
- Multivariate Verfahren, z.B. PMF, eröffnen zusätzliche Möglichkeiten, erfordern aber zusätzliche Informationen (Messunsicherheiten) und benötigen Datenmatritzen mit minimalen „Fehlstellen“
- Alle von Rezeptordaten ausgehenden Verfahren müssen anhand von Emissionsprofilen verifiziert werden...Ausweitung der Datenbestände hier dringend notwendig
- Nutzung historischer Messdaten erfordert sorgfältige Prüfung der Meta-Informationen und Datenqualität

Danke...



Bundesministerium  
für Umwelt, Naturschutz  
und Reaktorsicherheit

für die Förderung des Vorhabens,

für die Einladung zum Fachgespräch,

...und Ihnen für Ihre Aufmerksamkeit!