Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE DE LA DÉFENSE

METHODOLOGY FOR STATISTICAL EVALUATION OF ATMOSPHERIC DISPERSION MODELS IN A RISK ASSESSMENT CONTEXT

Bertrand Sapolin[1], Gilles Bergametti[2], Philippe Bouteilloux[1], Alain Dutot[2]

13th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes

[1] DGA Maîtrise NRBC, Vert-le-Petit, France

[2] Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA), Créteil, France

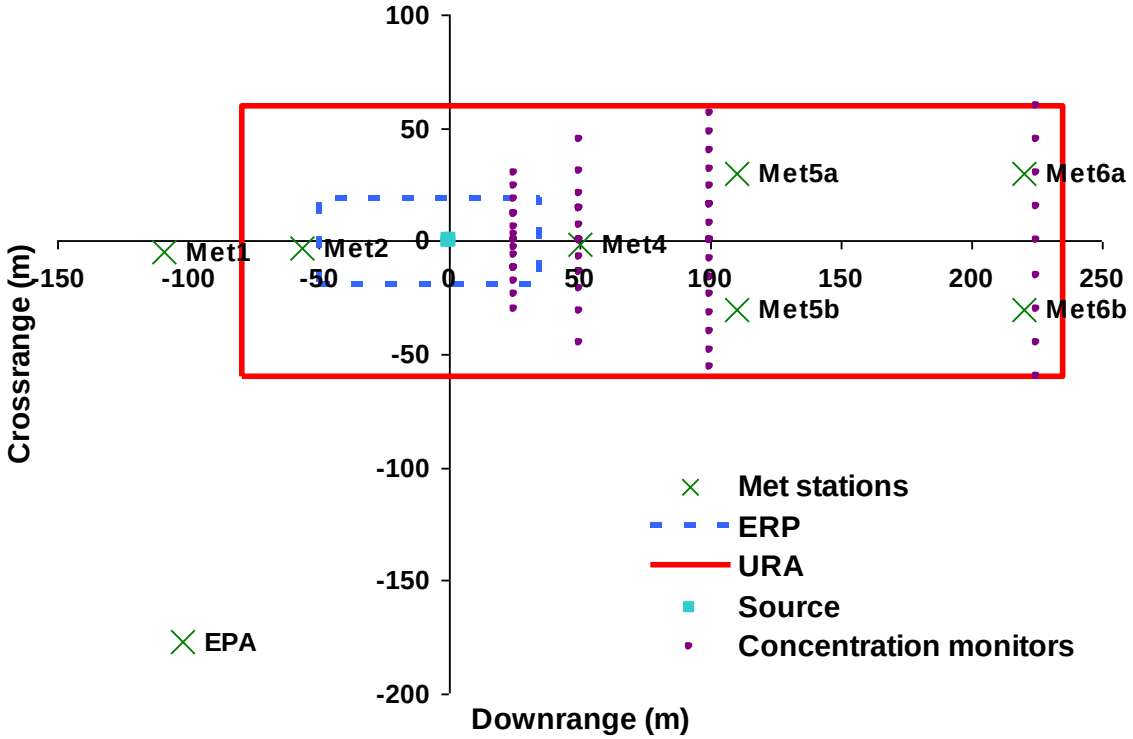
DGA
DIRECTION GÉNÉRALE DE L'ARMEMENT

Background

- Chemical, Biological, Radiological (CBR) risk assessment
  - Evaluate potential consequences of accidental or deliberate releases of toxic substances into the atmosphere
  - Use transport and dispersion models
  - Output: predicted effect on the population
- Scenarios
  - Short term releases
  - Non-stationary transport and diffusion
  - Acute inhalation toxicity
- Focus of the study:
  - Statistical evaluation against experimental data
    - Kit Fox: representative of risk assessment scenarios interesting the French MoD
    - Model: HPAC
  - Chemical risk assessment

Experimental data: Kit Fox

- US DoE Nevada Test Site
- Flat desert area artificially roughened
  - URA (Uniform Roughness Array): z_0 ~ 0.02m
  - ERP (Equivalent Roughness Pattern): z_0 ~ 0.2m
- 52 dense gas CO_2 releases
  - ERP&URA: 13 instantaneous, 6 continuous
  - URA alone: 21 instantaneous, 12 continuous
- 77 concentration samplers
  - 4 downwind distances: 25, 50, 100, 225m
  - Time resolution: 1s
- Met data
  - Local met stations
  - Time resolution: 1-10s
  - Neutral to stable conditions



Dispersion model

- HPAC (US DTRA)
  - Dispersion: SCIPUFF (Lagrangian Puff Model)
  - Version 4.04 SP4
- Kit Fox simulations
  - URA/ERP: 42x42 grid cells
  - Modelling domain: 420x420m
  - Source term: stack release (stack height = 0m)
  - Met data: all stations and vertical levels, 20s averaging time
  - Concentration output time step: 1s
- Note
  - Same configuration for the 52 trials (no “case by case adjustment”)
  - The purpose is not to evaluate model performance but rather use the evaluation results to investigate new methodologies for model evaluation

Comparison HPAC / Kit Fox with the MVK

- Model Validation Kit (MVK) protocol: arc max concentrations
- Example of results (FAC2 with 95% confidence intervals)

		Instantaneous concentration	20s moving average concentration
Block results	ERP puff	63.5 [49-76.4]	50 [35.8-64.2]
	ERP continuous	54.2 [32.8-74.4]	45.8 [22.1-63.4]
	URA puff	65.5 [54.3-75.5]	66.7 [55.5-76.6]
	URA continuous	45.8 [29.5-58.8]	41.7 [27.6-56.8]
Overall results		**59.2 [52.1-65.9]**	**54.3 [46.8-60.8]**

- MVK protocol:
  - Arc max value not appropriate => risk assessment more interested in values on the borders of toxic clouds
  - Concentration cannot be directly related to toxic effect

=> Need for a risk oriented evaluation methodology

Guidelines for a risk oriented evaluation methodology

Effect-related variables (1/3)

- Acute inhalation toxicity is a non linear function of concentration (C) and time (t)
  - Dosage: $d(t) = \int_0^t C(\xi)\, d\xi$
  - Toxic load TL: $TL(t) = \int_0^t C(\xi)^n d\xi$
  - Exponent n depends on the toxic substance
- Toxicological law: a given effect on an individual is reached by a fixed value of toxic load:

$$TL(t) = k \quad (eq.\ 1)$$

- Variability of population response to a given TL
  - Level k has a statistical meaning
  - Statistical distribution of population response is usually lognormal
  - $eq.\ 1$ can be extended to a Cumulative Distribution Function of the population response

$$\Phi(TL) = \frac{1}{2}\left[1 + erf\left(\frac{a.\ln(TL) + b}{\sqrt{2}}\right)\right]$$

a, b: constants associated to the toxic agent

Fraction of the population suffering adverse effect as a function of toxic load

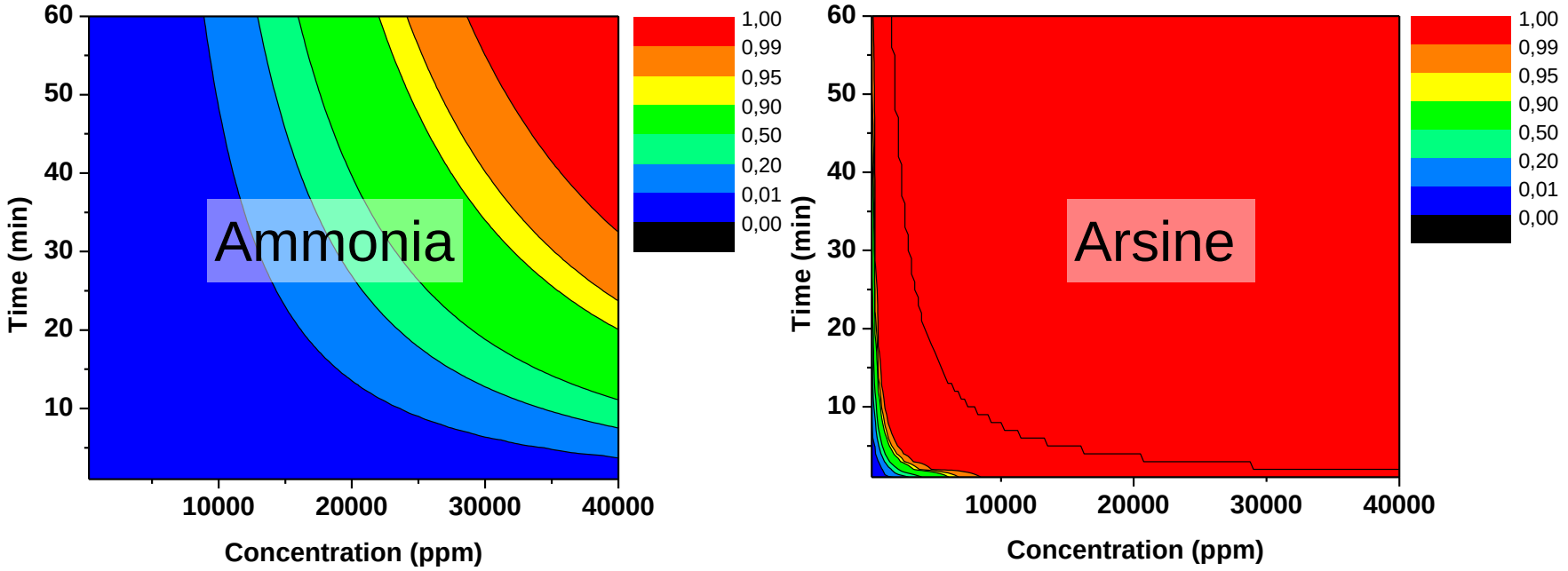
Effect-related variables (2/3)

- Remarks
  - Effect-related variables are built from concentration time series (observed / predicted)
  - Model performance depends on the substance
- Choice of substances
  - Risk assessment: numerous substances covering a large toxicity range
  - Impossible to test all of them => choose representative substances
    - Toxicity range cut into 4 classes: low, moderate, high & very high toxicity
    - Criterion: AEGL-3 thresholds, exposure time = 10min
    - 1 representative substance in each class

Classes			Benchmark agents			
Rank	Toxicity	AEGL-3 10 min range (mg/m^3)	Agent name	Probit parameters (C in ppm, t in min)		
				a	b	n
I	Low	AEGL-3>500	**Ammonia NH_3**	2.17	-47.4	1.83
II	Moderate	50<AEGL-3<500	**Hydrogen fluoride HF**	2.63	-29.9	1
III	High	5<AEGL-3<50	**Phosphine PH_3**	16.81	-120.89	0.5
IV	Very high	AEGL-3<5	**Arsine AsH_3**	2.65	-26.08	1.18

Effect-related variables (3/3)

- Compared toxicity
  - Class I: ammonia (“low” toxicity)
  - Class IV: arsine (very high toxicity)
- Fraction of fatalities as a function of concentration and exposure duration



Comparisons based on effect-related variables

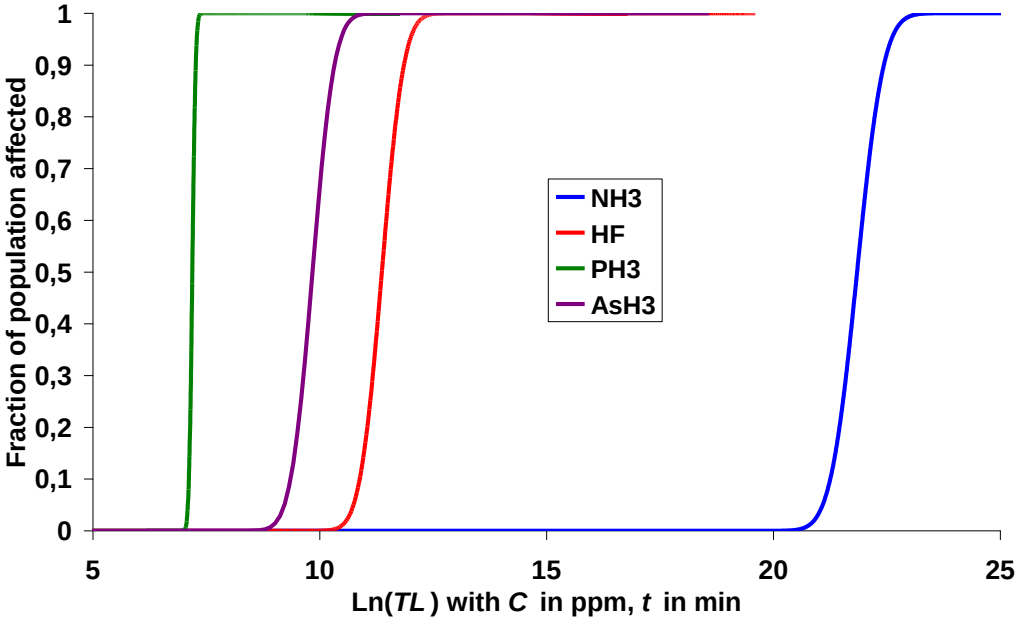
- Point to point comparisons
- Variables: dosage, toxic load
- Results (FAC2)

		Ct	C^nt			
			NH_3	HF	PH_3	AsH_3
Block results	ERP puff	21.1[18.2-24]	13.8[11.4-16.4]	21.6[18.6-24.6]	33.9[30.4-37.4]	19.2[16.4-22.1]
	ERP cont.	22.9[18.7-27.1]	13.1[9.7-16.6]	23.3[19.2-27.8]	34.9[30.1-39.8]	22[17.8-26.2]
	URA puff	29.5[26.6-32.5]	18.3[15.8-21]	30.4[27.4-33.5]	55[51.4-58.3]	26.1[23.2-29]
	URA cont.	35.5[32-39.1	20.4[17.4-23.4]	36[32.5-39.6]	61.2[57.3-64.7]	29.9[26.5-33.3]
Overall results		**27.8[26.1-29.5]**	**16.9[15.5-18.3]**	**28.4[26.7-30.1]**	**47.8[45.9-49.7]**	**24.6[23-26.2]**

- Poor performance
  - Point to point comparisons
  - n > 1 gives more weight to the uncertain variable => FAC2 decreases as n increases

Suggested use of effect-related variables (1/3)

- Population response =*f*(toxic load)



- Same pattern for all the substances
  - A plateau “nobody affected”
  - A plateau “everybody affected”
  - A narrow sloping part
- A same measure / prediction difference does not have the same impact whether the difference covers or not the sloping part of the response curve
- Large measure / prediction differences in the steady parts are unimportant

- TL95% / TL05%

Agent	***r* = *TL95%/TL05%* = *C95%/C05%***
NH_3	**2.29**
HF	**3.48**
PH_3	**1.47**
AsH_3	**2.85**

- Population response increases only on a very narrow range of toxic load
- r small => FAC2 inappropriate
- Non linear population response => criteria emphasizing amplitude of model errors are inappropriate (FB, NMSE…)

Suggested use of effect-related variables (2/3)

- Suggestion
  - **Compare fractions of population affected instead of toxic load**
  - Choose an incidence level & count the monitors where this level is exceeded
  - Event = the fixed incidence level is exceeded
  - Contingency table

Event predicted? \ Event observed?	Yes	No	Total
Yes	A	D	$A+D$
No	C	B	$C+B$
Total	$A+C$	$D+B$	$N = A+B+C+D$

- Criteria
  - False positive rate $R_{fp} = \frac{D}{D+B}$
  - False negative rate $R_{fn} = \frac{C}{A+C}$
  - Detection rate $R_{d} = \frac{A}{A+C}$
  - Good analysis rate $R_{ga} = \frac{A+B}{N}$
  - Bad analysis rate $R_{ba} = \frac{C+D}{N}$
- Similarity with the Measures of Effectiveness (MOE, Warner, Platt et al 2001)

$$MOE1 = \frac{A_{ov}}{A_{ov} + C_{FN}A_{FN} + C_{FP}A_{FP}} = \frac{A}{A + C_{FN}C + C_{FP}D}$$

Suggested use of effect-related variables (3/3)

- Results
  - Detection rates > 70%
  - False negative rates < 30%
  - False positive rates < 20%
  - Good analysis rates > 75%
  - Bad analysis rates < 25%

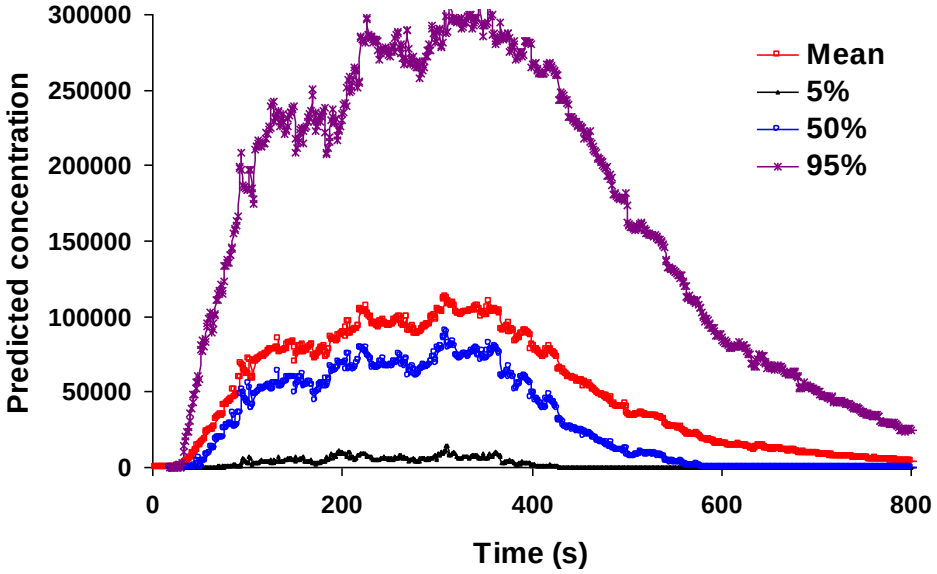
Agent	R_d	R_{fn}	R_{fp}	R_{ga}	R_{ba}
NH_3	n.s.	n.s.	n.s.	n.s.	n.s.
HF	82%	18%	6%	93%	7%
PH_3	80%	20%	17%	98%	2%
AsH_3	72%	28%	19%	79%	21%

HPAC vs 52 Kit Fox trials – n.s.: not significant

- Analysis
  - Better results
  - Suggested methodology
    - Focus on the end-user variable of interest (evaluation objective = risk assessment)
    - Measured / predicted toxic load differences without impact on the population response do not penalize the model

Concentration fluctuations (1/3)

- The suggested methodology has been applied to ensemble average model results.
- The methodology could be extended to include inherent uncertainties
  - Model result ≠ measure
  - Model result = ensemble average, measure = one realization of the ensemble => part of measure / prediction discrepancies may not be ascribed to the model
  - Need for a model able to predict inherent uncertainties
- SCIPUFF
  - Mean concentration + variance of fluctuation + integral timescale for concentration fluctuations (autocorrelation)
  - Theoretical distribution for concentration (clipped normal, left-shifted and clipped gamma…)
  - => uncertainties in the concentration time series



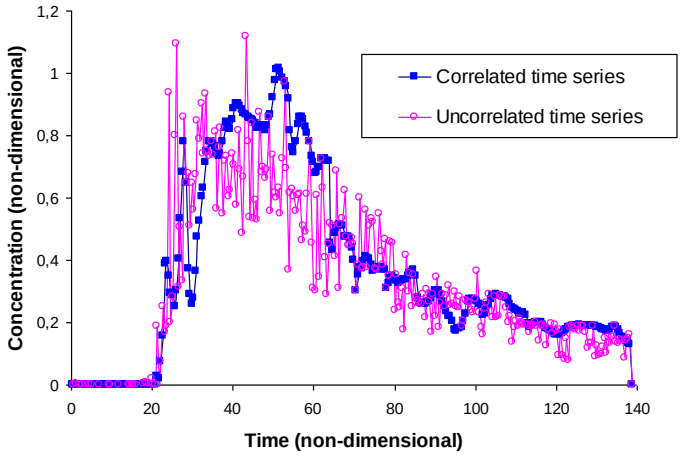
SCIPUFF: time series of concentration distribution (left-shifted and clipped gamma model)

Concentration fluctuations (2/3)

- Suggestion
  - 1) Use SCIPUFF to build modelled distributions of toxic load
  - 2) Compare the modelled distributions to measures
- How to build modelled toxic load distributions?
  - Generate many synthetic concentration time series from SCIPUFF results
  - For each time series, calculate toxic load
  - Build empirical toxic load distribution
- How to generate synthetic time series?
  - Sampling one concentration value at each time step produces uncorrelated time series
  - In reality, time series are correlated
  - Is it a conservative assumption to build toxic load distributions without considering time correlations?

Concentration fluctuations (3/3)

- Wind tunnel experiments (Hall et al, 2000)
  - Several repeats of instantaneous gas release
  - Concentration time series measured at several locations
  - At each location, measured time series (correlated) were used to calculate “natural” mean & variance of toxic load
  - Time series were then artificially decorrelated and used to calculate “artificial” mean & variance of toxic load



C(

	Correlated (“natural”)	Uncorrelated (“artificial”)
Mean	45.34 [43.04-47.63]	45.34 [44.9-45.78]
Standard deviation	s_1 = 3.21 [2.2-5.85]	s_2 = 0.62 [0.42-1.13]
Null hypothesis s1=s2 rejected at the 5% significance level		

Toxic load distribution (toxic load exponent =1), using correlated or uncorrelated time series

- Conclusion
  - Ignoring time series correlations amounts to
    - underestimating statistical variance of toxic load
    - underestimating upper percentiles of toxic load => not a conservative error
  - => **<u>Synthetic time series must include autocorrelations</u>**

Conclusion

- Risk-oriented methodology
  - Effect-related variables: toxic load + response distribution => fraction of population affected
  - Compare fraction of population instead of toxic load => release some useless constraints in model evaluation
  - Point to point comparisons
  - Contour thresholds
- Future work: extend the methodology to include inherent uncertainties
  - Develop a method to build statistical distribution of toxic load / population response
  - The methodology could be applied to probabilistic models (first & second moments of concentration distribution)