Uncertainty representation and propagation in chemical networks

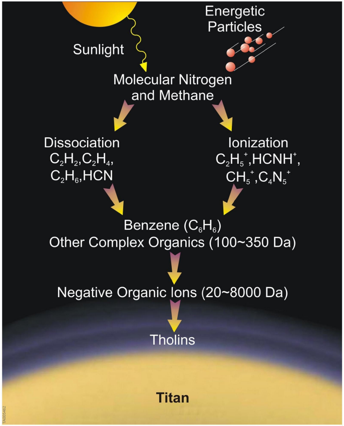
P. Pernot

Laboratoire de Chimie Physique, Orsay

Plan

1. Introduction
2. Uncertainty propagation
3. Treatment of uncertainties for chemical rates
4. Conclusions

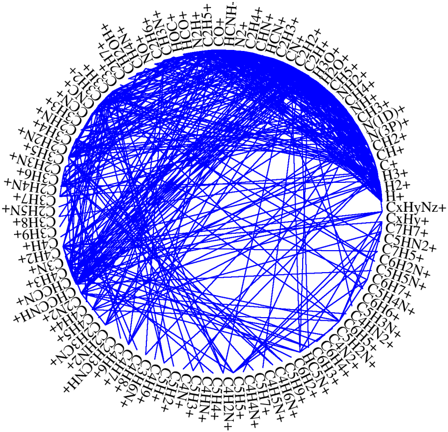
Uncertainties in a chemical model



Tholin formation in Titan's upper atmosphere

- Structural uncertainties
  - Model incompleteness
- Parametric uncertainties
  - transport
  - excitation processes
  - **reaction rates**
    - neutral-neutral
    - ion-molecule

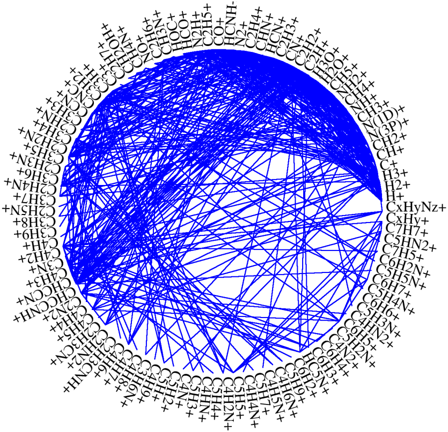
Uncertainties in a chemical model



ca. 680 Ion-molecule reactions in database

- Structural uncertainties
  - Model incompleteness
- Parametric uncertainties
  - transport
  - excitation processes
  - **reaction rates**
    - neutral-neutral
    - ion-molecule

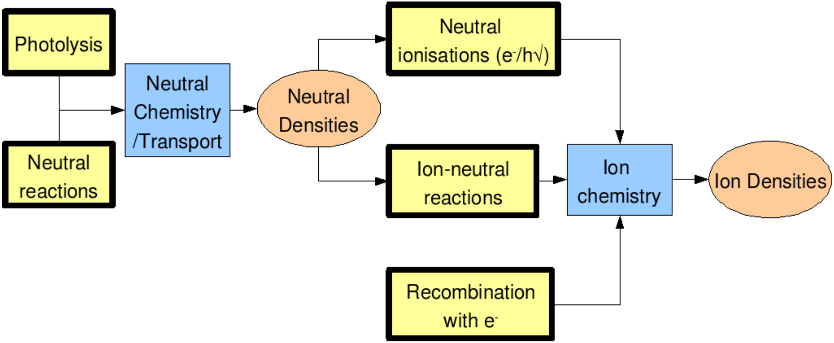
Uncertainties in a chemical model



ca. 680 Ion-molecule reactions in database

- Structural uncertainties
  - Model incompleteness
- Parametric uncertainties
  - transport
  - excitation processes
  - **reaction rates**
    - neutral-neutral
    - ion-molecule

Parametric uncertainty sources



Uncertainty propagation as a tool

- Model predictions to be compared with observational data
  $\longrightarrow$ model predictions as "virtual measurements" : value + incert.
  - *Model assessment* : significative discrepancies can be identified (model improvement) ;
  - *Sensitivity analysis* : major prediction uncertainties can be analyzed to improve input parameters (new lab. experiments...).

Uncertainty propagation as a tool

- Model predictions to be compared with observational data
  $\longrightarrow$ model predictions as "virtual measurements" : value + incert.
  - *Model assessment* : significative discrepancies can be identified (model improvement) ;
  - *Sensitivity analysis* : major prediction uncertainties can be analyzed to improve input parameters (new lab. experiments...).

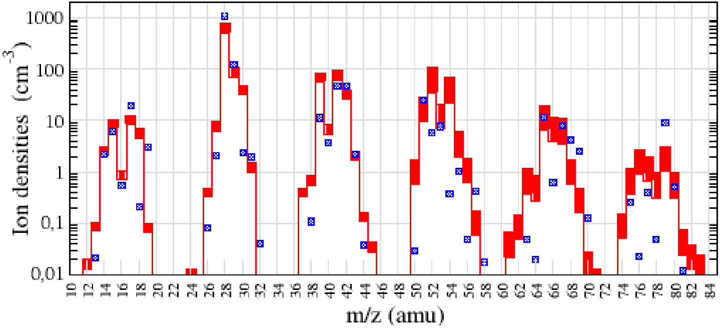
Uncertainty propagation as a tool

- Model predictions to be compared with observational data
  $\longrightarrow$ model predictions as "virtual measurements" : value + incert.

  - *Model assessment* : significative discrepancies can be identified (model improvement) ;

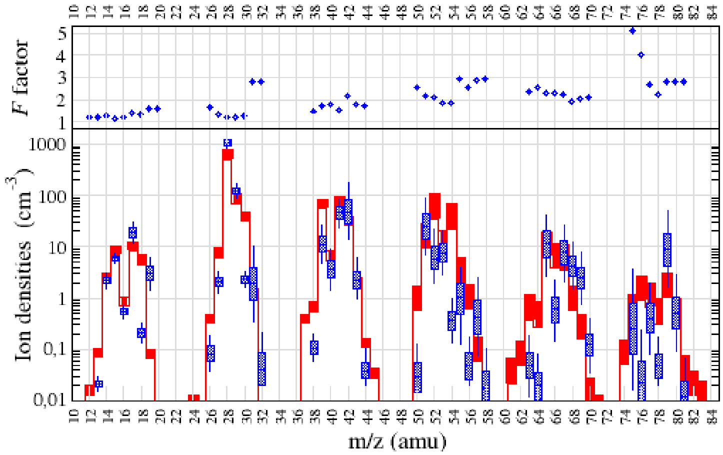
  - *Sensitivity analysis* : major prediction uncertainties can be analyzed to improve input parameters (new lab. experiments...).

Example : Cassini INMS



Predicted ion mass spectrum with all uncertainty sources vs. Cassini's INMS (T5@1200km)

Example : Cassini INMS



Predicted ion mass spectrum with all uncertainty sources vs. Cassini's INMS (T5@1200km)

Local Uncertainty Propagation

1. **List all sources of uncertainty**

2. **Characterize uncertainty sources and estimate standard uncertainties**

   1. *Type A* : statistical analysis of a sample
   2. *Type B* : all the rest
      u_x from designed pdf

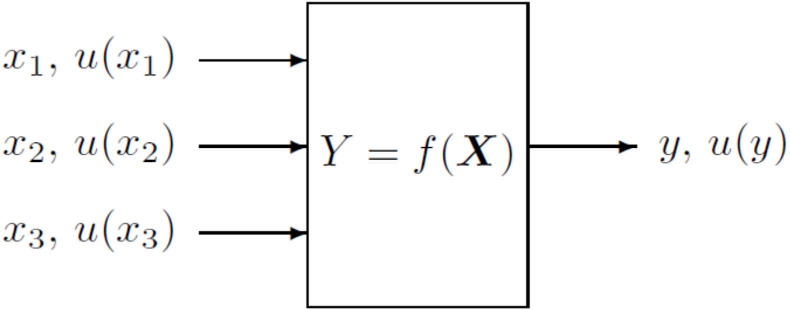
3. **Combine standard uncertainties**

$$\hat{y} = F(\hat{x}_1, \hat{x}_2, ...)$$

$$u_y^2 = \sum_i \left(\frac{\partial y}{\partial x_i}\right)^2_{\hat{x}_i} u_{x_i}^2 + \sum_{i \neq j} \left(\frac{\partial y}{\partial x_i}\right)_{\hat{x}_i} \left(\frac{\partial y}{\partial x_j}\right)_{\hat{x}_j} cov(x_i, x_j)$$

Guide to the expression of **U**ncertainty in **M**easurement (BIPM *et al.*, 1995)

Local Uncertainty Propagation



Guide to the expression of **U**ncertainty in **M**easurement (BIPM *et al.*, 1995)

Global Uncertainty Propagation

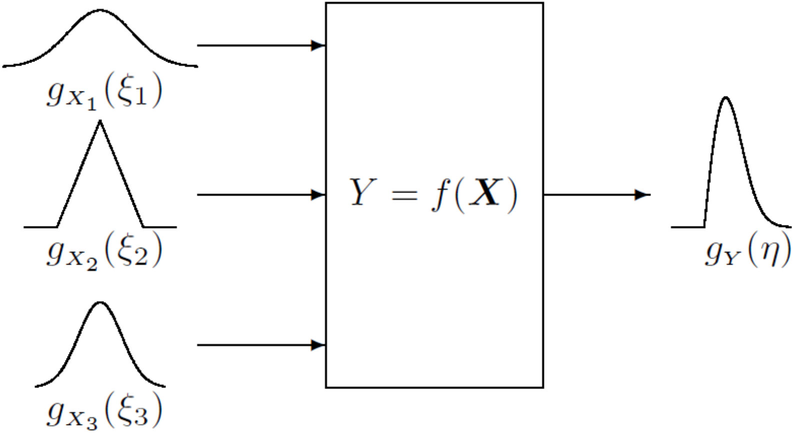
1. **List all sources of uncertainty**
2. **Represent uncertainty sources by pdf** $g_{\{x_i\}}(\{\xi_i\})$
3. **Perform uncertainty propagation**

$$g_y(\eta) = \int d\,\{\xi_i\}\;\delta(\eta - f(\{\xi_i\}))\,g_{\{x_i\}}(\{\xi_i\})$$

4. **Estimate uncertainties of model outputs from** $g_y(\eta)$

Evaluation of measurement data - Supplement 1 to the GUM (BIPM *et al.*, 2006)

Global Uncertainty Propagation



Evaluation of measurement data - Supplement 1 to the GUM (BIPM *et al.*, 2006)

Limitations of local and global UP

Local approach, pb. with :

- asymmetrical pdfs (ex. lognormal) ; $\exists$ improved versions of standard formula
- nonlinear models (large uncertainties) ;
- nonlinear correlations (ex : prescribed sum).

Most of the above apply to chemical networks of interest here...

Limitations of local and global UP

Local approach, pb. with :

- asymmetrical pdfs (ex. lognormal) ; $\exists$ improved versions of standard formula
- nonlinear models (large uncertainties) ;
- nonlinear correlations (ex : prescribed sum).

Most of the above apply to chemical networks of interest here...

Global approach

- more complex (pdf design)
- grid-based methods (polynomial chaos, Galerkin...)
  - curse of dimensionality
  - pb. with positivity constraints
- sampling based methods
  - pb. if model $y = F(x)$ computer intensive

Monte Carlo, Latin Hypercube Sampling (LHS)...

$$g_{\{x_1,\ldots,x_n\}}(\xi_1,\xi_2,\ldots,\xi_n)$$

$$\downarrow$$

$$\overbrace{\begin{array}{cccc} \xi_1^{(1)} & \xi_2^{(1)} & \ldots & \xi_n^{(1)} \\ \xi_1^{(2)} & \xi_2^{(2)} & \ldots & \xi_n^{(2)} \\ \vdots & \vdots & & \vdots \\ \xi_1^{(m)} & \xi_2^{(m)} & \ldots & \xi_n^{(m)} \end{array}}$$

Monte Carlo, Latin Hypercube Sampling (LHS)...

$$g_{\{x_1,\dots,x_n\}}(\xi_1, \xi_2, \dots, \xi_n)$$
$$\downarrow$$
$$\begin{array}{cccccc}
\multicolumn{4}{c}{\overbrace{\hspace{12em}}} & F & \\
\hline
\xi_1^{(1)} & \xi_2^{(1)} & \dots & \xi_n^{(1)} & \longrightarrow & \eta^{(1)} \\
\hline
\xi_1^{(2)} & \xi_2^{(2)} & \dots & \xi_n^{(2)} & & \\
\vdots & \vdots & & \vdots & & \\
\xi_1^{(m)} & \xi_2^{(m)} & \dots & \xi_n^{(m)} & &
\end{array}$$

Sampling based uncertainty propagation

Monte Carlo, Latin Hypercube Sampling (LHS)...

$$g_{\{x_1,\dots,x_n\}}(\xi_1, \xi_2, \dots, \xi_n)$$

$$\downarrow$$

$$\begin{array}{cccccc}
\multicolumn{4}{c}{\overbrace{\hphantom{\xi_1^{(1)} \quad \xi_2^{(1)} \quad \dots \quad \xi_n^{(1)}}}} & F & \\
\xi_1^{(1)} & \xi_2^{(1)} & \dots & \xi_n^{(1)} & \longrightarrow & \eta^{(1)} \\
\hline
\xi_1^{(2)} & \xi_2^{(2)} & \dots & \xi_n^{(2)} & \longrightarrow & \eta^{(2)} \\
\hline
\vdots & \vdots & & \vdots & & \\
\xi_1^{(m)} & \xi_2^{(m)} & \dots & \xi_n^{(m)} & &
\end{array}$$

Sampling based uncertainty propagation

Monte Carlo, Latin Hypercube Sampling (LHS)...

$$g_{\{x_1,\dots,x_n\}}(\xi_1, \xi_2, \dots, \xi_n)$$

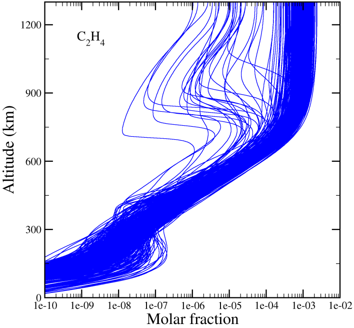
$$\downarrow$$

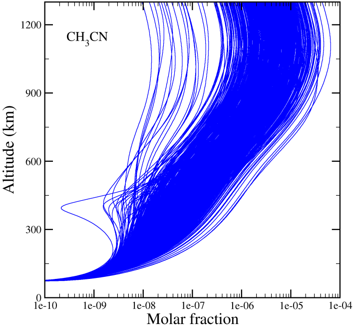
$$\begin{array}{cccccc}
\multicolumn{4}{c}{\overbrace{\hspace{10em}}} & F & \\
\xi_1^{(1)} & \xi_2^{(1)} & \dots & \xi_n^{(1)} & \rightarrow & \eta^{(1)} \\
\xi_1^{(2)} & \xi_2^{(2)} & \dots & \xi_n^{(2)} & \rightarrow & \eta^{(2)} \\
\vdots & \vdots & & \vdots & & \vdots \\
\hline
\xi_1^{(m)} & \xi_2^{(m)} & \dots & \xi_n^{(m)} & \rightarrow & \eta^{(m)} \\
\hline
\end{array}$$

Monte Carlo, Latin Hypercube Sampling (LHS)...

$$
\begin{array}{c}
g_{\{x_1,\ldots,x_n\}}(\xi_1,\xi_2,\ldots,\xi_n) \\
\downarrow \\
\overbrace{
\begin{array}{cccc}
\xi_1^{(1)} & \xi_2^{(1)} & \ldots & \xi_n^{(1)} \\
\xi_1^{(2)} & \xi_2^{(2)} & \ldots & \xi_n^{(2)} \\
\vdots & \vdots & & \vdots \\
\xi_1^{(m)} & \xi_2^{(m)} & \ldots & \xi_n^{(m)}
\end{array}}
\end{array}
\begin{array}{c}
F \\
\longrightarrow \\
\longrightarrow \\
\\
\longrightarrow \\
\\
\\
\end{array}
\left|
\begin{array}{c}
\\
\eta^{(1)} \\
\eta^{(2)} \\
\vdots \\
\eta^{(m)} \\
\downarrow \\
g_y(\eta)
\end{array}
\right|
$$

Global UP : Titan photochemistry





Bimodality of outputs PDF : out of reach of local UP

Global UP provides extensive exploration of parameter space

Hébrard. *et al.*, *JPPC* (2006), *PSS* (2007)

PDF DESIGN

- type of PDF depends on the nature of the parameters
  - discrete vs. continuous
  - interval of definition
    $]-\infty, +\infty[$, $[0, +\infty[$, $[a, b]$...
- PDF of outputs depends on PFDs of inputs,
  tempered by "Central Limit Theorems"
- the joint PDF of all uncertain parameters can be factorized in groups of independent parameters

$$g_{\{x_i\}}\left(\{\xi_i\}\right) = \prod_k g_{\{x_j\}_{j\in k}}\left(\{\xi_j\}_{j\in k}\right)$$

PDF DESIGN

- type of PDF depends on the nature of the parameters
  - discrete vs. continuous
  - interval of definition
    $]-\infty, +\infty[$, $[0, +\infty[$, $[a, b]$...
- PDF of outputs depends on PFDs of inputs,
  tempered by "Central Limit Theorems"
- the joint PDF of all uncertain parameters can be factorized in groups of independent parameters

$$g_{\{x_i\}}\left(\{\xi_i\}\right) = \prod_k g_{\{x_j\}_{j\in k}}\left(\{\xi_j\}_{j\in k}\right)$$

PDF DESIGN

- type of PDF depends on the nature of the parameters
  - discrete vs. continuous
  - interval of definition
    $]-\infty,+\infty[$, $[0,+\infty[$, $[a,b]$...

- PDF of outputs depends on PFDs of inputs,
  tempered by "Central Limit Theorems"

- the joint PDF of all uncertain parameters can be factorized in groups of independent parameters

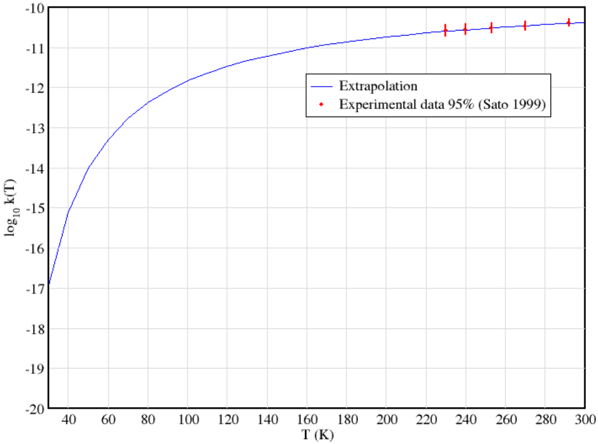
$$g_{\{x_i\}}\left(\{\xi_i\}\right)=\prod_k g_{\{x_j\}_{j\in k}}\left(\{\xi_j\}_{j\in k}\right)$$

Parametric uncertainties of reaction rates

Analysis of measured rate constants

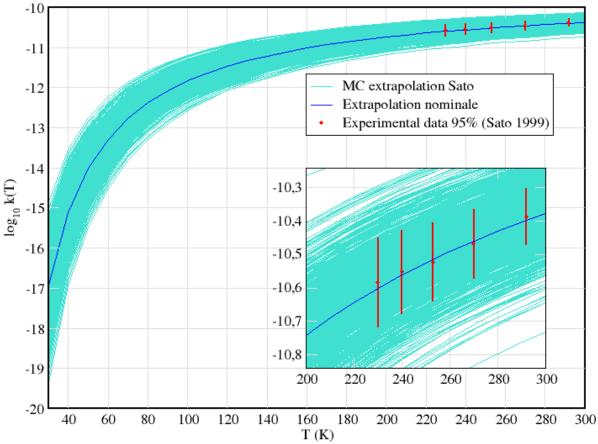
- Example. *Arrhenius* $\ln k = \ln A - \vartheta / T$
  - Data analysis provides $\{\ln A, \vartheta\} \sim \mathrm{Norm}\left(\left\{\overline{\ln A}, \overline{\vartheta}\right\}, \Sigma\right)$
    Σ : variance/covariance matrix
  - sample of k from $\{\ln A, \vartheta\}$ and stat. analysis
    or
    $u_{\ln k}^2 = u_{\ln A}^2 + u_{\vartheta}^2 / T^2 - 2/T * \mathrm{Cov}(\ln A, \vartheta)$
  - Note : $\mathrm{Cov}(\ln A, \vartheta)$ generally large (>0.9), very important for UP !

Arrhenius UP example : $N(^2D) + C_2H_4$



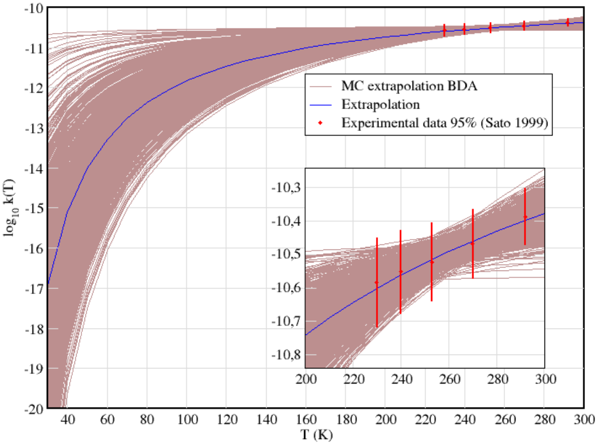
	$\ln A$	E_a/R (K)	Correl
Bayesian	-22.222 $\pm$ 0.66	504 $\pm$ 170	0.996

ARRHENIUS UP EXAMPLE : $N(^2D) + C_2H_4$



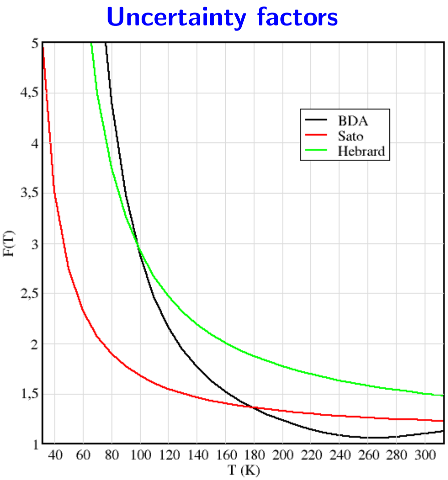
	$\ln A$	E_a/R (K)	Correl
Sato *et al.* (1999)	-22.193 $\pm$ 0.13	503 $\pm$ 50	n/a

Arrhenius UP example : $N(^2D) + C_2H_4$



	$\ln A$	E_a/R (K)	Correl
Bayesian	-22.222 $\pm$ 0.66	504 $\pm$ 170	0.996

Arrhenius UP example : $N(^2D) + C_2H_4$



Choice of pdf for evaluated rates

Distributions implementing positivity constraint

- **Preferred value** k_i^0, estimated uncertainty factor F_i^0

$$k_i \sim \text{Lognormal}(k_i^0, F_i^0)$$

  - $\text{P}(k_i^0/(2F_i^0) < k_i < k_i^0 * 2F_i^0) \simeq 0.95$
  - k_i^0 is *median* of pdf, not *mean* (Stewart and Thompson, 1994)
  - **modelers truncate to $2 * F_i^0$ or $3 * F_i^0$ to avoid "exotic" rates**

- **Preferred interval**

$$k_i \sim \text{Loguniform}(k_i^{min}, k_i^{max})$$

  - no preferred value within an interval (dispersed experimental data)
  - $k_i^{min} = k_i^0/(2F_i^0)$, $k_i^{max} = k_i^0 * (2F_i^0)$

Choice of pdf for evaluated rates

Distributions implementing positivity constraint

- **Preferred value** k_i^0, estimated uncertainty factor F_i^0

$$k_i \sim \text{Lognormal}(k_i^0, F_i^0)$$

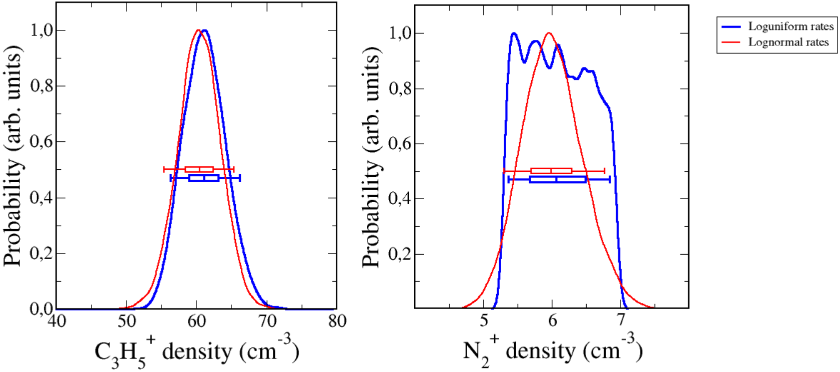
  - $\text{P}(k_i^0/(2F_i^0) < k_i < k_i^0 * 2F_i^0) \simeq 0.95$
  - k_i^0 is *median* of pdf, not *mean* (Stewart and Thompson, 1994)
  - **modelers truncate to** $2 * F_i^0$ **or** $3 * F_i^0$ **to avoid "exotic" rates**

- **Preferred interval**

$$k_i \sim \text{Loguniform}(k_i^{min}, k_i^{max})$$

  - no preferred value within an interval (dispersed experimental data)
  - $k_i^{min} = k_i^0/(2F_i^0)$, $k_i^{max} = k_i^0 * (2F_i^0)$

Lognormal vs. Loguniform



Output pdf depends on the place of the species in the network
Might have an impact on sensitivity analysis

Carrasco. *et al.*, *PSS* (2007)

Parametric uncertainties of branching ratios

Multi-pathway reactions

$$k_{i,j} = k_i * b_{i,j}; \sum_j b_{i,j} = 1$$

- **Reaction rates and branching ratios are mostly measured by different experiments/techniques**
  - larger uncertainties for branching ratios (more difficult to measure than rates).

- **In such cases, it is better to keep an explicit separation of uncertainty sources**
  - more pertinent sensitivity analysis (key parameters) ;
  - easier to manage the sum rule wrt. uncertainties ;
  - T-dependence of k_i different from $b_{i,j}$
  - safer update of databases when new branching ratios or rate available.

Parametric uncertainties of branching ratios

Multi-pathway reactions

$$k_{i,j} = k_i * b_{i,j}; \sum_j b_{i,j} = 1$$

- **Reaction rates and branching ratios are mostly measured by different experiments/techniques**
  - larger uncertainties for branching ratios (more difficult to measure than rates).

- **In such cases, it is better to keep an explicit separation of uncertainty sources**
  - more pertinent sensitivity analysis (key parameters) ;
  - easier to manage the sum rule wrt. uncertainties ;
  - T-dependence of k_i different from $b_{i,j}$
  - safer update of databases when new branching ratios or rate available.

PDFs for branching ratios

Distributions implementing the sum rule

- **Preferred values and precision**

$$\{b_{i,j}\} \sim \mathrm{Diri}\left(\{\alpha_{i,j}\}\right) \quad \propto \prod_j b_{i,j}^{\alpha_{i,j}-1}$$

- **No preference**

$$\{b_{i,j}\} \sim \mathrm{Diri}\left(1, 1, ..., 1\right)$$

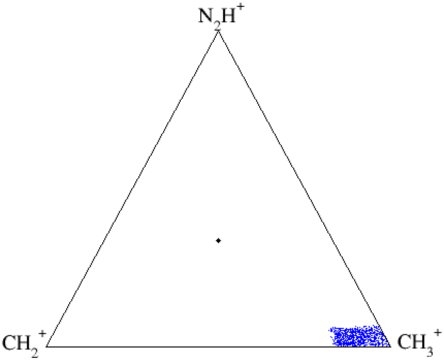
ex : low-T extrapolation for ion-molecule reactions

- **Preferred intervals**

$$\{b_{i,j}\} \sim \mathrm{Diun}\left(\left\{b_{i,j}^{min}, b_{i,j}^{max}\right\}\right)$$

Carrasco *et al.*, *PSS* (2007)

EXAMPLE : $N_2^+ + CH_4$



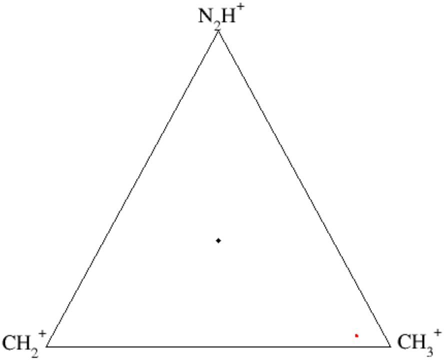
CH_3^+	CH_2^+	N_2H^+
0.89	0.08	0.03
10%	100%	100%

Standard elicitation

$$1 \leq 10\ \% \leq 0.5$$
$$0.5 < 30\ \% \leq 0.1$$
$$0.1 < 100\ \% \leq 0$$

Carrasco *et al.*, *PSS* (2007)

Example : $N_2^+ + CH_4$

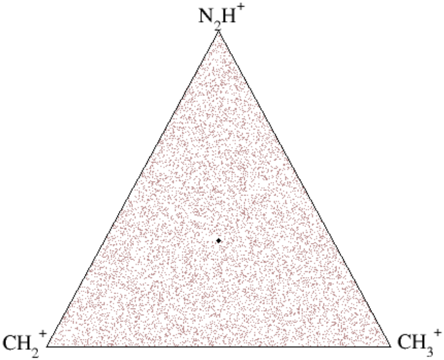


CH_3^+	CH_2^+	N_2H^+
0.89	0.08	0.03
2%	10%	10%

Improved elicitation
according to Nicolas (PhD Thesis, 2002)

Carrasco *et al.*, *PSS* (2007)

Example : $N_2^+ + CH_4$



CH_3^+	CH_2^+	N_2H^+
1/3	1/3	1/3
100%	100%	100%

Full uncertainty

Carrasco *et al.*, *PSS* (2007)

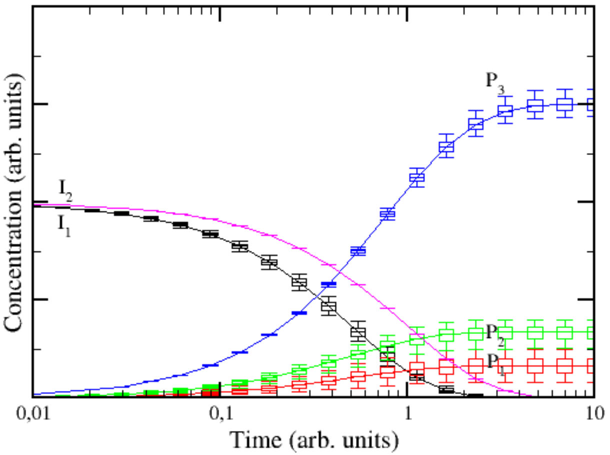
Branching ratios and the sum rule

$$I_1 + M_1 \longrightarrow P_1\ ;\ \ k_1,\ b_{11}$$
$$I_1 + M_1 \longrightarrow P_2\ ;\ \ k_1,\ b_{12}$$
$$I_1 + M_2 \longrightarrow P_3\ ;\ \ k_2$$
$$I_2 + M_2 \longrightarrow P_3\ ;\ \ k_3$$

$$[M_i] \gg [I_i]$$

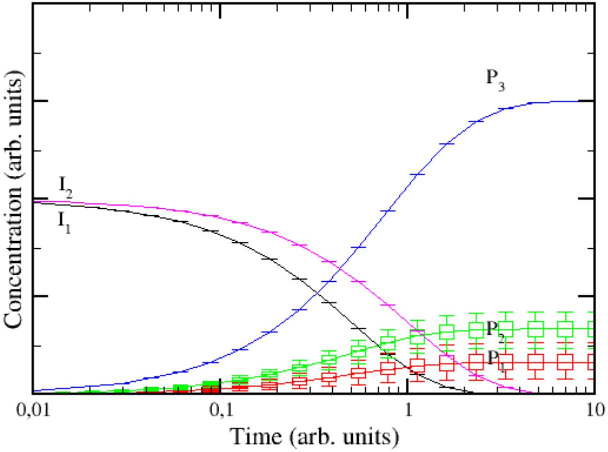
$$F_k \ll F_b$$

Branching ratios and the sum rule



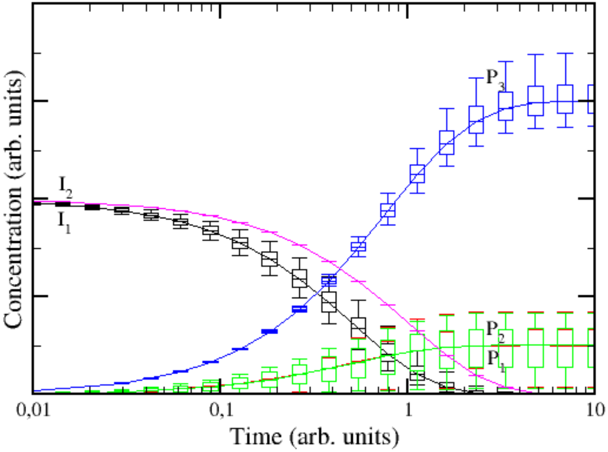
Uncorrelated partial rates : $b_{11} = 0.33 \pm 0.12$, $b_{12} = 0.67 \pm 0.12$

Branching ratios and the sum rule



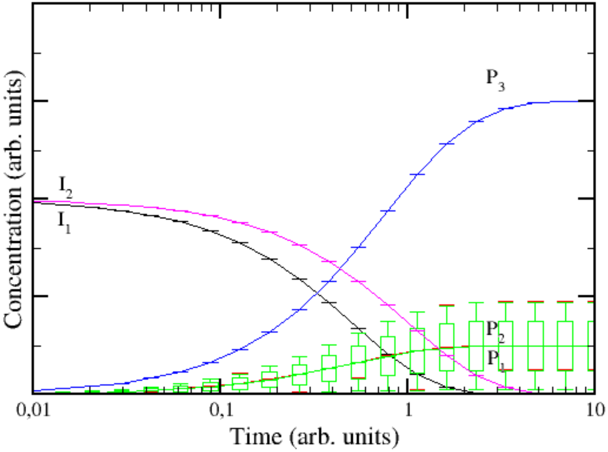
Correlated partial rates : $\{b_{11}, b_{12}\} \sim \mathrm{Diri}\,(15, 30)$

Branching ratios and the sum rule



Uncorrelated partial rates : $b_{11} \sim b_{12} \sim \mathrm{Unif}\,(0, 1)$

Branching ratios and the sum rule



Correlated partial rates : $\{b_{11}, b_{12}\} \sim \mathrm{Diri}\,(1,1)$

Shopping for rate uncertainty

What have existing databases to offer to MCUP-aware modelers ?

- **udfa[06]**
  - $k_i(T) = \alpha_i \, (T/300)^{\beta_i} \exp\left(-\gamma_i/T\right)$
  - The accuracy is described by a letter - A, B, C, D, E - where the errors are $< 25\%$, $< 50\%$, within a factor of 2, within an order of magnitude, and highly uncertain, respectively.
  - No T-dependence of uncertainty
  - No pdf proposed

What have existing databases to offer to MCUP-aware modelers ?

- **osu_01_2007**
  - $k_i(T) = \alpha_i \, (T/300)^{\beta_i} \exp\left(-\gamma_i/T\right)$
  - $F_i = 1.25$,1.5, 2.0 or 10.0
  - No T-dependence of uncertainty
  - No pdf proposed

What have existing databases to offer to MCUP-aware modelers ?

- **Anicich** (ion-molecule, JPL 2003)
  - Global rate $k_i \pm f_i$ (f_i in percent)
  - Branching ratios $\{b_{ij}\}_{j=1,N}$
  - No uncertainty on branching ratios
  - No T-dependence on properties and uncertainties
  - No pdf proposed

What have existing databases to offer to MCUP-aware modelers ?

- **IUPAC - NASA/JPL**
  - $k_i(T) = k_i^0 \exp\left(-E_i/T\right)$

    $F_i(T) = F_i^0 \exp\left(g_i \left|\frac{1}{T} - \frac{1}{T^0}\right|\right);\ T^0 = 298\,K$
  - F_i is an **expanded** uncertainty, $CI \simeq 95\%$
  - "The assignment of the uncertainties is a subjective assessment of the evaluators. *They are not determined by a rigorous statistical analysis of the database.*"
  - No pdf proposed

What have existing databases to offer to MCUP-aware modelers ?

- **Hébrard et al.** (*JPPC*, 2006 ; *PSS*, 2007)
  - $k_i(T) = \alpha_i\,(T/300)^{\beta_i}\exp\left(-\gamma_i/T\right)$

    $F_i(T) = F_i(300\,K)\,\exp\left(g_i\left|\frac{1}{T} - \frac{1}{300}\right|\right)$
  - F_i is a **standard** uncertainty, $CI \simeq 67\%$
  - "Both uncertainty factors, $F_i(300\,K)$ and g_i, do not necessarily result from a rigorous statistical analysis of the available data."
  - No pdf proposed, but
    - $\log k_i = \log k_i(T) + \epsilon\,\log F_i(T)$; $\epsilon \sim \mathrm{Norm}(0,1)$, in Hébrard *et al.* (*PSS*, 2007)

Conclusions

- **Global UP is necessary for chemical networks**
  - Monte Carlo UP is now recognized as a standard tool in metrology
- **Utility of MCUP depends on adapted PDFs**, with
  - a correct description of inputs uncertainty
    - amplitude (not too small, not too large...)
    - distribution shape (to a minor degree, but we need more experiences in SA to conclude)
  - a structure reflecting experimental uncertainty sources
    - necessary step to exp.-oriented sensitivity analysis
- **Existing databases have not be designed with MCUP in mind**
  - they can be updated and improved along these proposed lines...

Conclusions

- **Global UP is necessary for chemical networks**
  - Monte Carlo UP is now recognized as a standard tool in metrology
- **Utility of MCUP depends on adapted PDFs**, with
  - a correct description of inputs uncertainty
    - amplitude (not too small, not too large...)
    - distribution shape (to a minor degree, but we need more experiences in SA to conclude)
  - a structure reflecting experimental uncertainty sources
    - necessary step to exp.-oriented sensitivity analysis
- **Existing databases have not be designed with MCUP in mind**
  - they can be updated and improved along these proposed lines...

Conclusions

- **Global UP is necessary for chemical networks**
  - Monte Carlo UP is now recognized as a standard tool in metrology
- **Utility of MCUP depends on adapted PDFs**, with
  - a correct description of inputs uncertainty
    - amplitude (not too small, not too large...)
    - distribution shape (to a minor degree, but we need more experiences in SA to conclude)
  - a structure reflecting experimental uncertainty sources
    - necessary step to exp.-oriented sensitivity analysis
- **Existing databases have not be designed with MCUP in mind**
  - they can be updated and improved along these proposed lines...

Conclusions

- **Global UP is necessary for chemical networks**
  - Monte Carlo UP is now recognized as a standard tool in metrology
- **Utility of MCUP depends on adapted PDFs**, with
  - a correct description of inputs uncertainty
    - amplitude (not too small, not too large...)
    - distribution shape (to a minor degree, but we need more experiences in SA to conclude)
  - a structure reflecting experimental uncertainty sources
    - necessary step to exp.-oriented sensitivity analysis
- **Existing databases have not be designed with MCUP in mind**
  - they can be updated and improved along these proposed lines...

Remerciements

- N. Carrasco (SA, Verrières)
- S. Plessis & Ch. Alcaraz (LCP, Orsay)
- M. Dobrijevic (LAB, Bordeaux)
- E. Hébrard (LISA, Crétail)
- M. Banaszkiewicz (SRC, Warsaw)
- R. Thissen & O. Dutuit (LPG, Grenoble)
- V. Vuitton & R. Yelle (LPL, Tucson)

- CNRS
- CNES
- EuroPlaNet
- Programme National de Planétologie