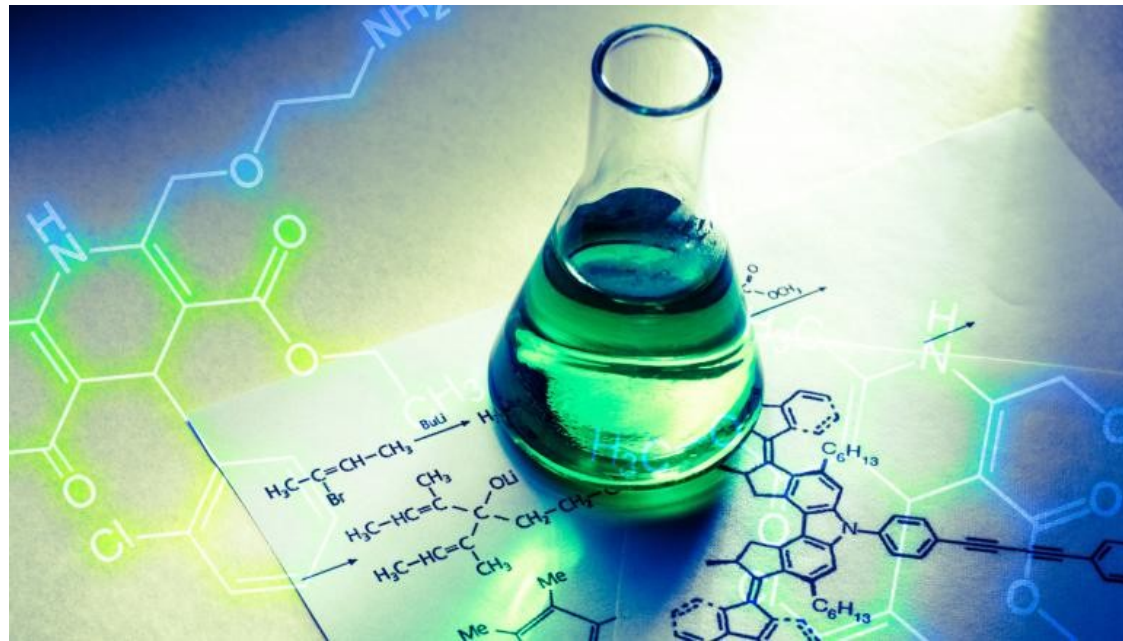


On the usefulness of imprecise Bayesianism in chemical kinetics



Dr. Marc Fischer

Assistant professor of chemical engineering

Mines de Saint-Étienne (France)

1 Introduction

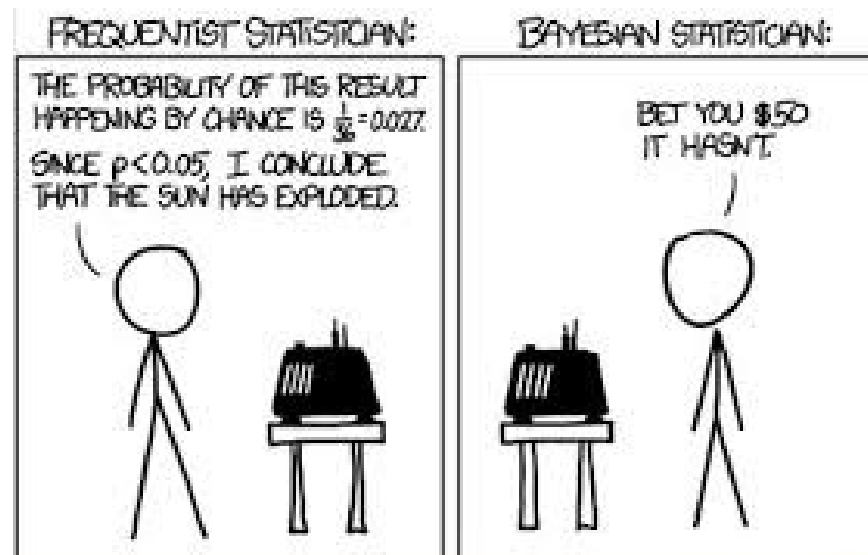
- Chemical kinetics: study of the speed of chemical reactions and the factors bearing on it.
- Application to a wide range of fields: combustion, catalysis, atmospheric chemistry, water pollution...
- The ultimate goal is to derive predictive models.
- Both frequentist and Bayesian approaches are used to describe and propagate the parameter uncertainty (Wang, 2015).

Hai Wang and David A Sheen. Combustion kinetic model uncertainty quantification, propagation and minimization. Progress in Energy and Combustion Science, 47:1–31, 2015.

The goals of this presentation are twofold:

1) To give an overview of the field which respect to the frequentist and Bayesian approaches to chemical kinetic parameter uncertainty.

2) To show how classic **precise** Bayesian methods can be very misleading in cases where there aren't enough experimental data.



2 Presentation of the main approaches

2.1 Basic notions

If a chemical system undergoes only the elementary reaction

$v_A A + v_B B \rightarrow v_C C + v_D D$, the reaction rate r is given by

$$r = \frac{1}{v_A} \frac{d[A]}{dt} = \frac{1}{v_B} \frac{d[B]}{dt} = \frac{1}{v_C} \frac{d[C]}{dt} = \frac{1}{v_{[D]}} \frac{dD}{dt} = k[A]^{v_A}[B]^{v_B}$$

v_j is always positive for products and negative for reactants.

If there are N elementary reactions, the concentrations are governed by an ordinary differential equation system (ODE) which is such that $\frac{1}{v_A} \frac{d[A]}{dt} = \sum_{i=1}^{i=N} v_{i,A} r_i$. It includes all reactions where A is involved as a product or a reactant.

2.2

Frequentist approach

- Feasible set defined as $F = \{k \mid p(|X_i(m_i(k)) - m_i(k)| \geq |e_i - m_i(k)|) \geq \alpha, i \in [1;n]\}$.
- Simplification: $F_\varepsilon = \{k \mid |m_i(k) - e_i| \leq \varepsilon \sigma_i\}$.
- The predictions must be close enough to **each** measurements.
- The more relevant data come in, the narrower the hyperspace becomes

→ parameter uncertainty reduction (Frenklach,2007).

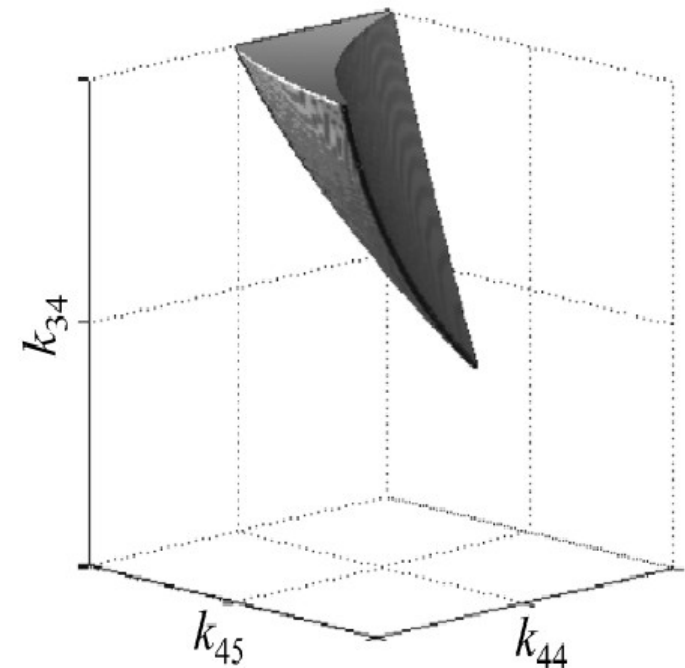


Figure 1: Example of feasible set

M. Frenklach. Transforming data into knowledge process informatics for combustion chemistry. Proc. Combust. Inst., 31(1): 125, 2007b.

2.3 Bayesian approach

- The posterior probability distribution of k is computed as

$$f(k | e) = \frac{L(e | k) f_0(k)}{\int_{k \in K} L(e | k') f_0(k') dk'}$$

- $L(e|k)$: likelihood function. $f_0(k)$: prior distribution.
- Example: posterior probability distribution of the reaction $\text{C}_{12}\text{H}_{26} + \frac{25}{2}\text{O}_2 \rightarrow 12\text{CO} + 13\text{H}_2\text{O}$
Kinetic parameters A and E_a .

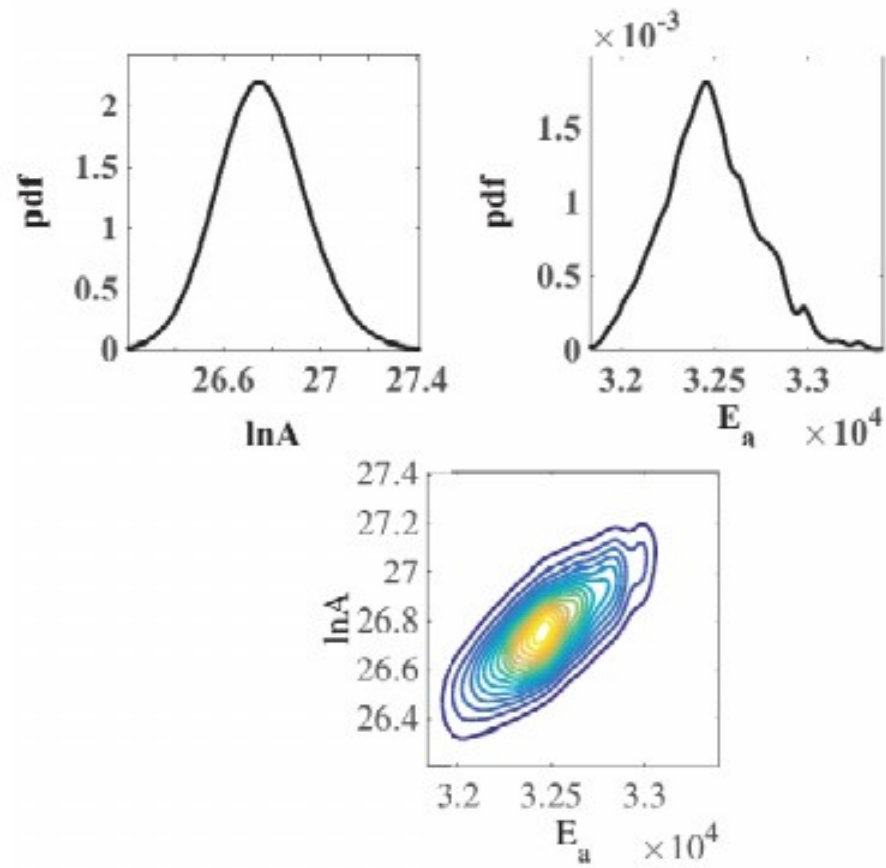


Figure 2: Posterior predictive marginal and joint distributions of Arrhenius rate parameters

- An increasing number of researchers in chemical kinetics are adopting a Bayesian approach to parameter uncertainty (see references in my paper).
- Most of them do so uncritically and don't explain why they prefer Bayesianism to frequentism.
- The justifications given by others mostly boil down to the following reasons:
 - higher simplicity, intuitiveness and appealing nature
 - frequentism ignores mistakes in the model assumption
 - dividing a complex problem into a series of simpler ones
 - avoiding overfitting through the use of Bayes factors and uniform priors.
- However, the overwhelming majority of authors only utilise one single (mostly) uniform prior.

- They aren't concerned about the effect of the choice of the prior on the posterior distribution.
- Actually, many seem even unaware that such a choice was made at all and believe that a single uniform prior is a perfect representation of our ignorance.
- In my contribution, I show how this can lead to very misleading results in cases when there are only few experimental data to “wash out” the prior.

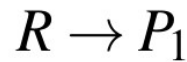
3

How not to mix knowledge and ignorance

Simple problem of chemical kinetics: 2 reaction models possible, one experimental profile to be reproduced.

Initial conditions: $[R]_0 = 1200 \text{ mol/m}^3$

M_1 : An isomerisation towards product P_1 : M_2 : A recombination towards product P_2 :



$$\frac{d[R]}{dt} = -k_1[R]$$

$$k_1 \in [2; 2E + 3]$$

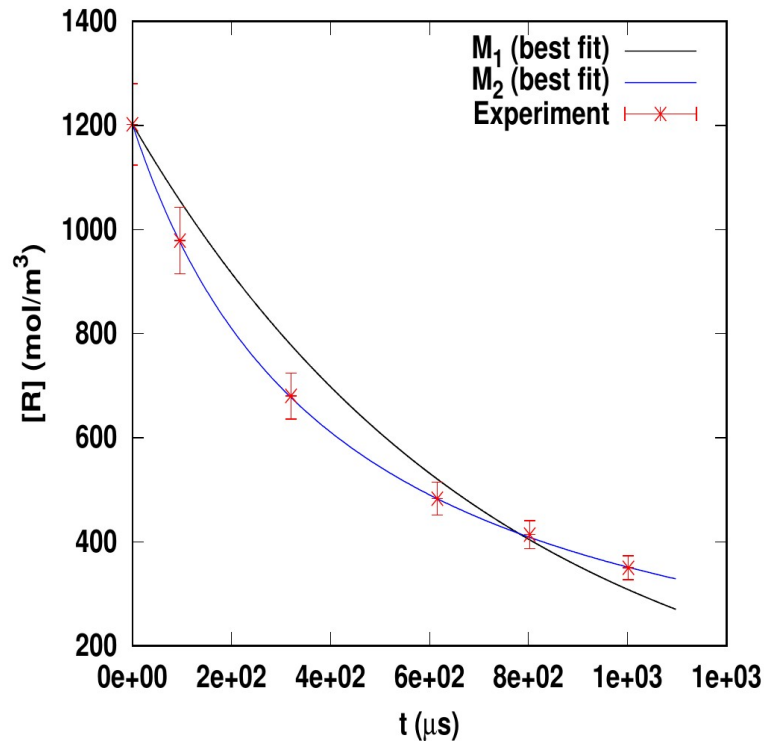


$$\frac{d[R]}{dt} = -2k_2[R]^2$$

$$k_2 \in [5E + 04; 5E + 09]$$

3.1 Feasible set approach

For $\varepsilon = 2$ and $F_\varepsilon = \{k \mid |m_i(k) - e_i| \leq \varepsilon \sigma_i\}, i \in [|1;n|]$



$M_1: F_\varepsilon = \{\emptyset\}$. Incompatible with the data.

$M_2: F_\varepsilon = [8.51E+05; 1.22E+06] mol.cm^{-3}.s^{-1}$
Very good agreement with the shape.

3.2

Precise Bayesian approach

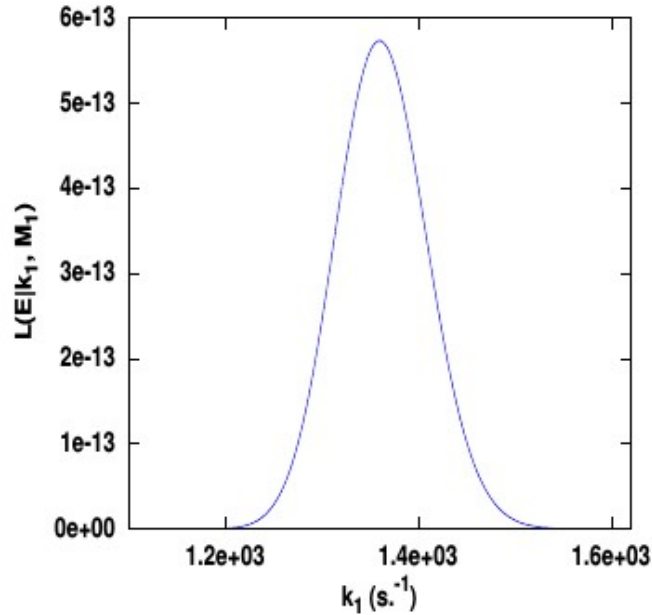


Figure 4: Likelihood (M_1)

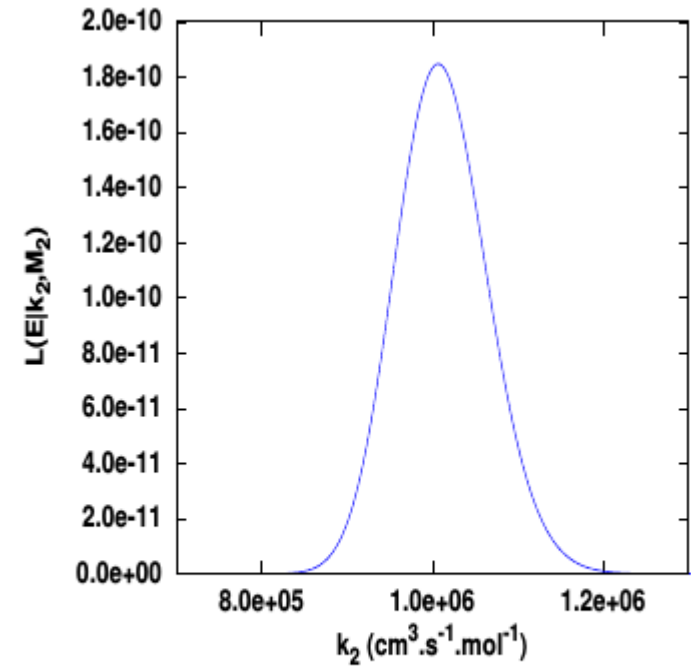


Figure 5: Likelihood (M_2)

Principle of indifference

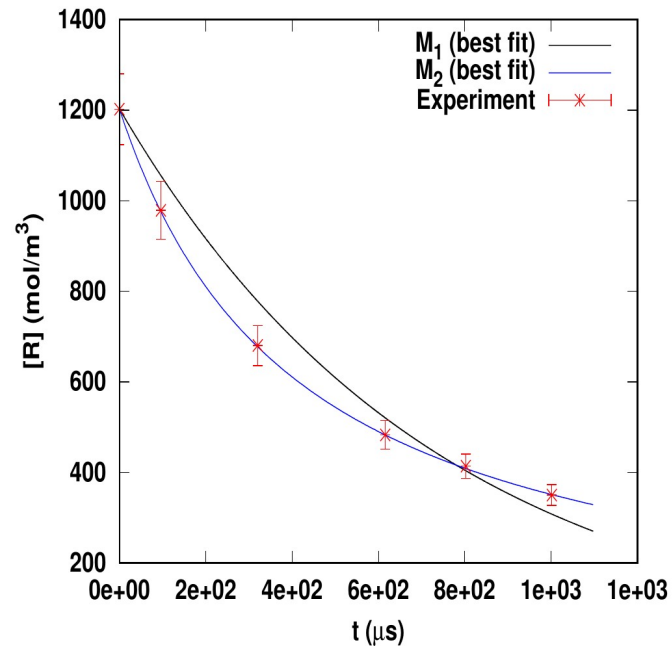
→ uniform distribution on k_i

$$P(k_1|M_1) = 1/(2^3 - 2) = 0.00050$$

$$P(k_2|M_2) = 1/(5^9 - 5^4) = 2^{-10}$$

- Bayes' factor: $B_k = \frac{L(e|M_2)}{L(e|M_1)} = 0.312$.

→ We should reject model M_2 in favour of model M_1 in spite of the poor agreement!



- Example of Lindley's paradox

3.3

Imprecise Bayesian approach

- We consider five additional priors.
- They are all uniform with respect to a reparametrisation of k_i .
- The Bayes' factors were computed.

Variable	B
k	3.12E-01
$x = 1/k$	1.67E+04
$l = \log_{10}(k)$	3.29E+02
$y = 1/\log_{10}(k)$	3.74E+03
$z = k^{0.25}$	7.08E+01
$w = k^2$	1.98E-04

Table 1: Dependency of B on the prior

3.4

Imprecise Bayesian approach

- Bayes' factor B Ockham's factor O strongly dependent on the prior.
- No single posterior probability distribution is enough to represent the situation.
- Using only the first uniform prior leads one to illegitimately mix up knowledge (about the measurements and their uncertainties) and ignorance (about the parameter spaces).
- Misguided mixture of epistemic and aleatory uncertainty.

4 Perspectives

Bayesianism brings lots of exciting possibilities in chemical kinetics.

However, the reliance on one single prior undermines the reliability of the results.

The large majority of chemical kineticists use one single uniform prior.

There is an urgent need for robust Bayesian analyses embedded in an imprecise framework.

