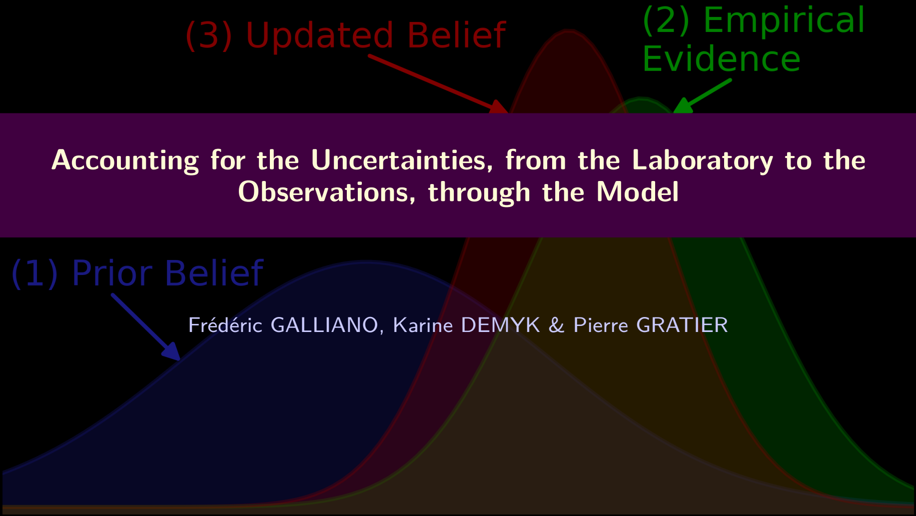
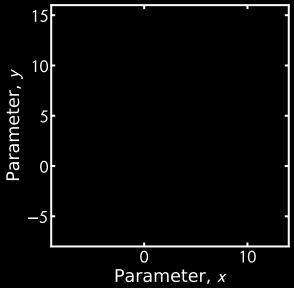


Accounting for the Uncertainties, from the Laboratory to the Observations, through the Model

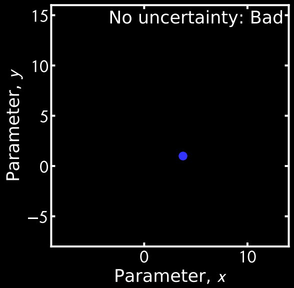
Frédéric GALLIANO, Karine DEMYK & Pierre GRATIER

Motivations: Why Uncertainties are Instrumental

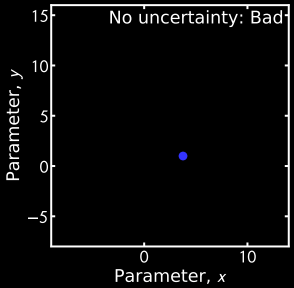
Motivations: Why Uncertainties are Instrumental



Motivations: Why Uncertainties are Instrumental

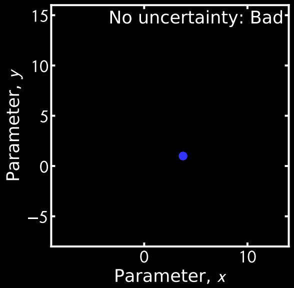


Motivations: Why Uncertainties are Instrumental



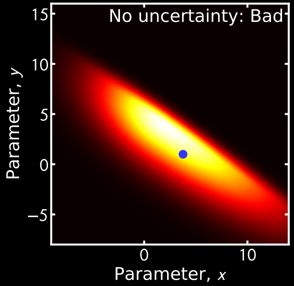
- A measure w/o uncertainties is meaningless.

Motivations: Why Uncertainties are Instrumental



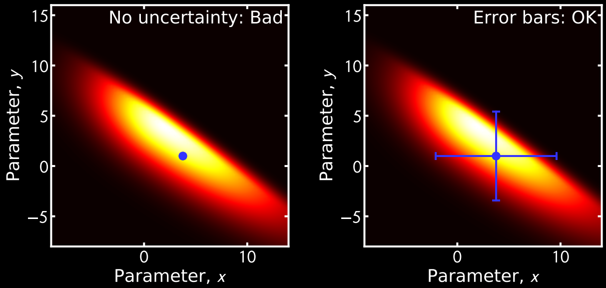
- A measure w/o uncertainties is meaningless.
- Uncertainties provide a metric to test models & theories.

Motivations: Why Uncertainties are Instrumental



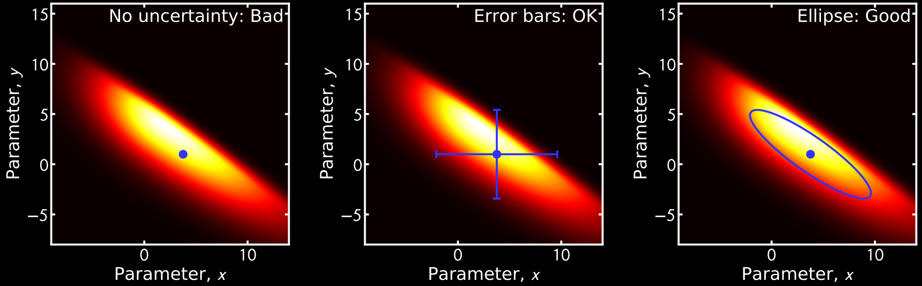
- A measure w/o uncertainties is meaningless.
- Uncertainties provide a metric to test models & theories.

Motivations: Why Uncertainties are Instrumental



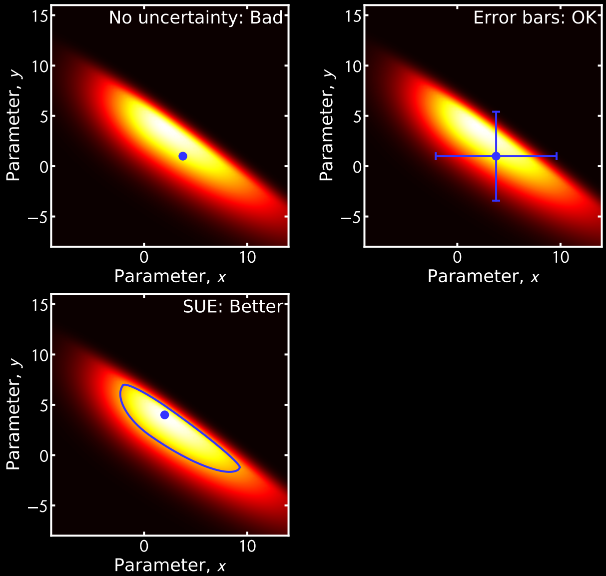
- A measure w/o uncertainties is meaningless.
- Uncertainties provide a metric to test models & theories.

Motivations: Why Uncertainties are Instrumental

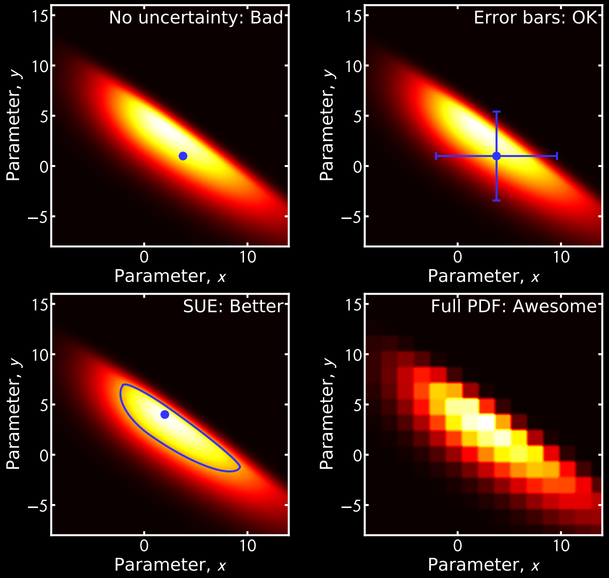


- A measure w/o uncertainties is meaningless.
- Uncertainties provide a metric to test models & theories.

Motivations: Why Uncertainties are Instrumental

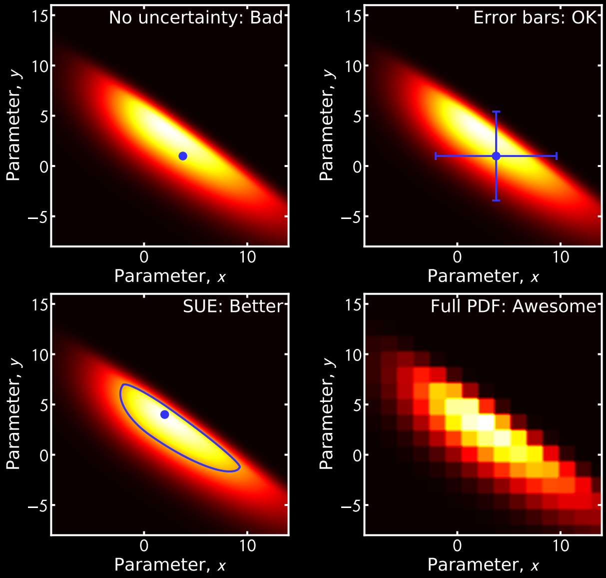


- A measure w/o uncertainties is meaningless.
- Uncertainties provide a metric to test models & theories.



- A measure w/o uncertainties is meaningless.
- Uncertainties provide a metric to test models & theories.

Motivations: Why Uncertainties are Instrumental



- A measure w/o uncertainties is meaningless.
- Uncertainties provide a metric to test models & theories.
- The uncertainty is not necessarily a single number.

The Particularity of ISM Studies

The Particularity of ISM Studies

ISM studies often rely on long chains of heterogeneous data

The Particularity of ISM Studies

ISM studies often rely on long chains of heterogeneous data

Experiments

The Particularity of ISM Studies

ISM studies often rely on long chains of heterogeneous data

Experiments

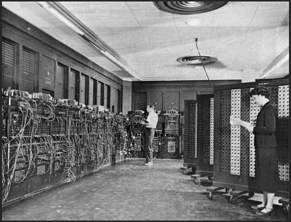
Observations

The Particularity of ISM Studies

ISM studies often rely on long chains of heterogeneous data

Experiments

Observations



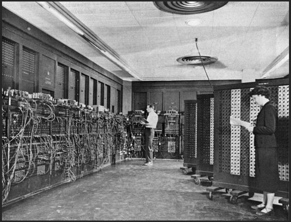
Models/Simulations

The Particularity of ISM Studies

ISM studies often rely on long chains of heterogeneous data


Experiments


Observations


Models/Simulations

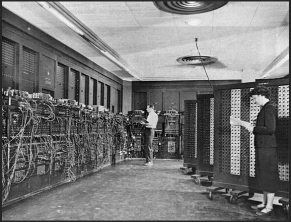
Our object of study impacts the way we work

The Particularity of ISM Studies

ISM studies often rely on long chains of heterogeneous data

Experiments

Observations



Models/Simulations

Our object of study impacts the way we work

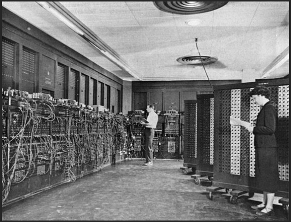
- Small samples.

The Particularity of ISM Studies

ISM studies often rely on long chains of heterogeneous data

Experiments

Observations



Models/Simulations

Our object of study impacts the way we work

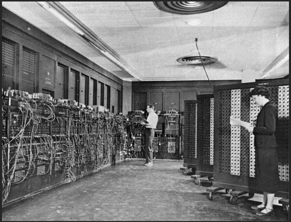
- Small samples.
- Rather small teams.

The Particularity of ISM Studies

ISM studies often rely on long chains of heterogeneous data

Experiments

Observations



Models/Simulations

Our object of study impacts the way we work

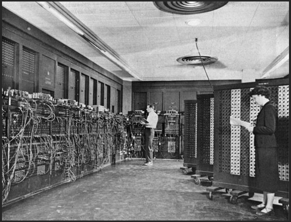
- Small samples.
- Rather small teams.
- Cosmology: large teams on the same data $\Rightarrow$ need standardized, reproducible uncertainties.

The Particularity of ISM Studies

ISM studies often rely on long chains of heterogeneous data

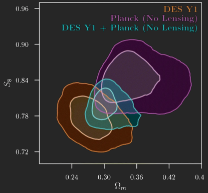

Experiments


Observations


Models/Simulations

Our object of study impacts the way we work

- Small samples.
- Rather small teams.
- Cosmology: large teams on the same data $\Rightarrow$ need standardized, reproducible uncertainties.

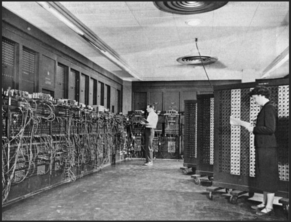


The Particularity of ISM Studies

ISM studies often rely on long chains of heterogeneous data

Experiments

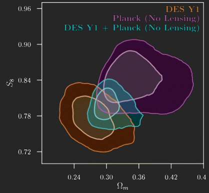
Observations

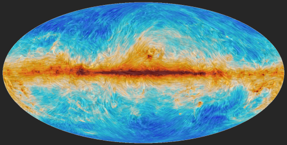


Models/Simulations

Our object of study impacts the way we work

- Small samples.
- Rather small teams.
- Cosmology: large teams on the same data $\Rightarrow$ need standardized, reproducible uncertainties.





(*Planck* collaboration)

An Example of a Nested Uncertainty Study

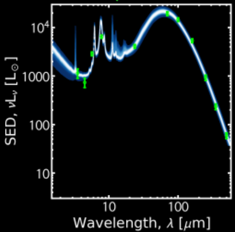
An Example of a Nested Uncertainty Study

Galaxies (multi-λ)

(Galliano *et al.*, 2021)

An Example of a Nested Uncertainty Study

Galaxies (multi-λ)

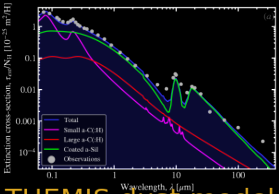


SED fit

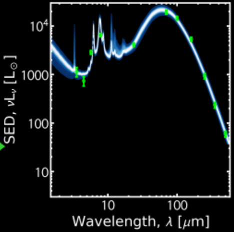
(Galliano *et al.*, 2021)

An Example of a Nested Uncertainty Study

Galaxies (multi-λ)



THEMIS dust model

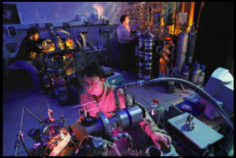


SED fit

(Galliano *et al.*, 2021)

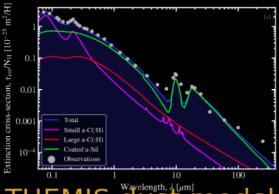
An Example of a Nested Uncertainty Study

Galaxies (multi-λ)

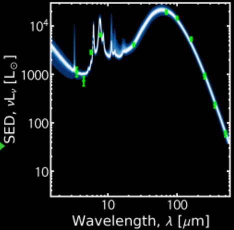


Laboratory data

+



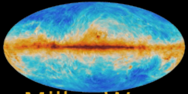
THEMIS dust model



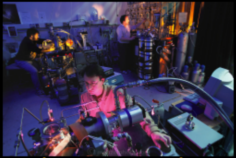
SED fit

(Galliano *et al.*, 2021)

An Example of a Nested Uncertainty Study

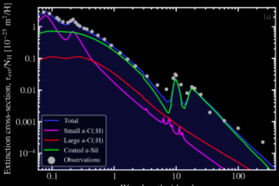


Milky Way



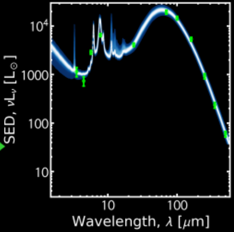
Laboratory data

+



THEMIS dust model

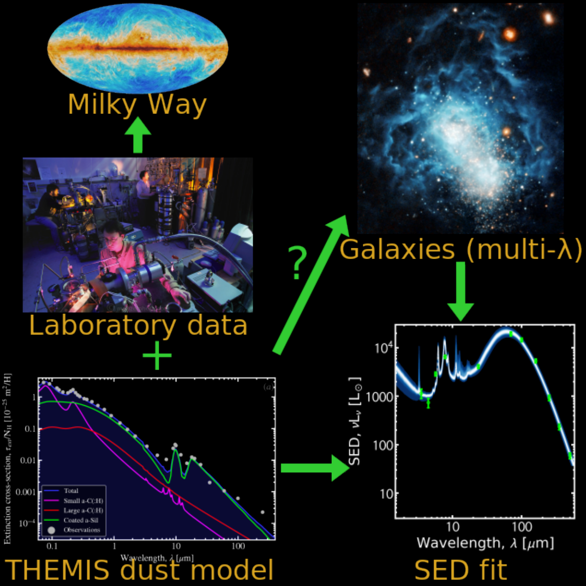
Galaxies (multi-λ)



SED fit

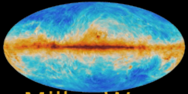
(Galliano *et al.*, 2021)

An Example of a Nested Uncertainty Study

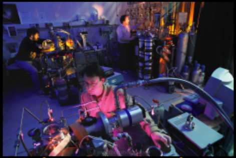


(Galliano *et al.*, 2021)

An Example of a Nested Uncertainty Study

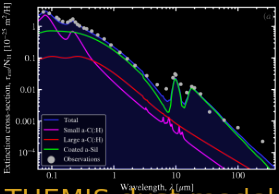


Milky Way



Laboratory data

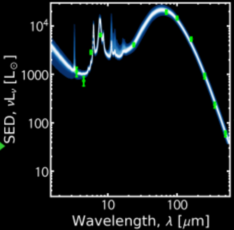
+



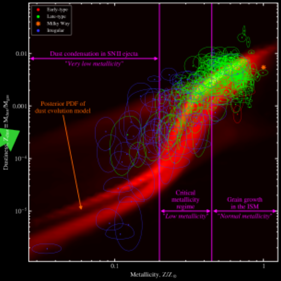
THEMIS dust model

?

Galaxies (multi-λ)



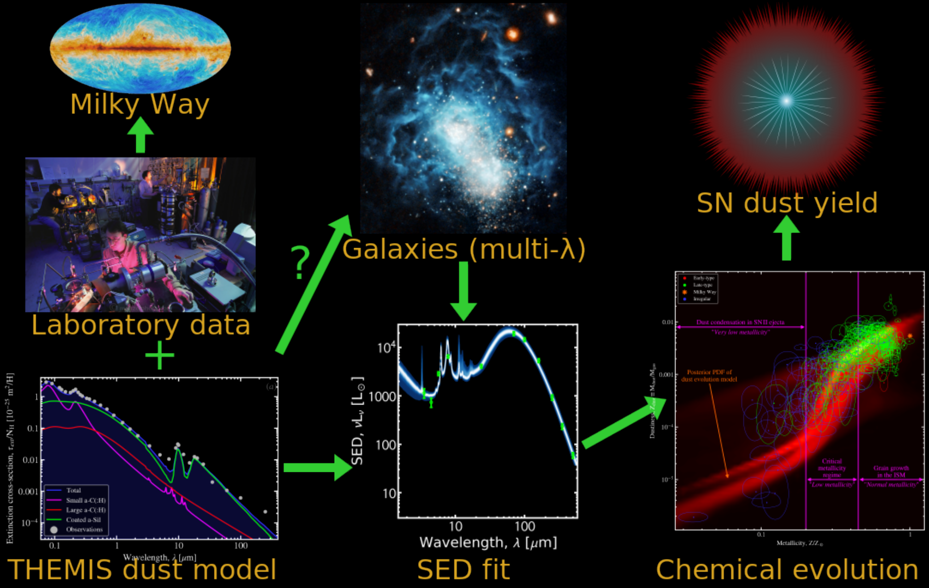
SED fit



Chemical evolution

(Galliano *et al.*, 2021)

An Example of a Nested Uncertainty Study



(Galliano *et al.*, 2021)

Objectives of the Workshop

Objectives of the Workshop

1. Have a general discussion about uncertainties & their associated methods.

Objectives of the Workshop

1. Have a general discussion about uncertainties & their associated methods.

2. Provide an overview of the way they are taken into account in the different fields of ISMology $\Rightarrow$ give grounded examples.

Objectives of the Workshop

1. Have a general discussion about uncertainties & their associated methods.
2. Provide an overview of the way they are taken into account in the different fields of ISMology ⇒ give grounded examples.
3. Give momentum to initiatives that could lead to a standardization of the way they are taken into account and published.

The Program of the Workshop

The Program of the Workshop

08:30–08:55	INTRODUCTION: WHY UNCERTAINTIES ARE INSTRUMENTAL	
08:30–08:40	Frédéric GALLIANO	Motivations & objectives of the workshop - An example of a nested uncertainty problem
08:40–08:50	Marie GUEGUEN	A philosopher's viewpoint on identifying, quantifying & communicating uncertainties
08:50–08:55	Everyone	Discussion

The Program of the Workshop

08:30–08:55	INTRODUCTION: WHY UNCERTAINTIES ARE INSTRUMENTAL	
08:30–08:40	Frédéric GALLIANO	Motivations & objectives of the workshop - An example of a nested uncertainty problem
08:40–08:50	Marie GUEGUEN	A philosopher's viewpoint on identifying, quantifying & communicating uncertainties
08:50–08:55	Everyone	Discussion

08:55–10:00	QUANTIFYING EXPERIMENTAL & OBSERVATIONAL UNCERTAINTIES	
08:55–09:05	Karine DEMYK	An overview of the challenges of estimating experimental uncertainties
09:05–09:15	Marco MINISSALE	Uncertainties in ice laboratory experiments
09:15–09:20	Everyone	Discussion
09:20–09:30	Lucas EINIG	Signal extraction from noisy line cubes: the problem of applying hyperspectral imaging methods
09:30–09:35	Everyone	Discussion
09:35–09:45	Nathalie YSARD	Uncertainties in dust models
09:45–10:00	Everyone	Discussion

The Program of the Workshop

08:30–08:55	INTRODUCTION: WHY UNCERTAINTIES ARE INSTRUMENTAL	
08:30–08:40	Frédéric GALLIANO	Motivations & objectives of the workshop - An example of a nested uncertainty problem
08:40–08:50	Marie GUEGUEN	A philosopher's viewpoint on identifying, quantifying & communicating uncertainties
08:50–08:55	Everyone	Discussion
08:55–10:00	QUANTIFYING EXPERIMENTAL & OBSERVATIONAL UNCERTAINTIES	
08:55–09:05	Karine DEMYK	An overview of the challenges of estimating experimental uncertainties
09:05–09:15	Marco MINISSALE	Uncertainties in ice laboratory experiments
09:15–09:20	Everyone	Discussion
09:20–09:30	Lucas EINIG	Signal extraction from noisy line cubes: the problem of applying hyperspectral imaging methods
09:30–09:35	Everyone	Discussion
09:35–09:45	Nathalie YSARD	Uncertainties in dust models
09:45–10:00	Everyone	Discussion
10:00–10:30	COFFEE BREAK	

The Program of the Workshop

08:30–08:55	INTRODUCTION: WHY UNCERTAINTIES ARE INSTRUMENTAL	
08:30–08:40	Frédéric GALLIANO	Motivations & objectives of the workshop - An example of a nested uncertainty problem
08:40–08:50	Marie GUEGUEN	A philosopher's viewpoint on identifying, quantifying & communicating uncertainties
08:50–08:55	Everyone	Discussion
08:55–10:00	QUANTIFYING EXPERIMENTAL & OBSERVATIONAL UNCERTAINTIES	
08:55–09:05	Karine DEMYK	An overview of the challenges of estimating experimental uncertainties
09:05–09:15	Marco MINISSALE	Uncertainties in ice laboratory experiments
09:15–09:20	Everyone	Discussion
09:20–09:30	Lucas EINIG	Signal extraction from noisy line cubes: the problem of applying hyperspectral imaging methods
09:30–09:35	Everyone	Discussion
09:35–09:45	Nathalie YSARD	Uncertainties in dust models
09:45–10:00	Everyone	Discussion
10:00–10:30	COFFEE BREAK	
10:30–11:20	PROPAGATING UNCERTAINTIES THROUGH DATA PROCESSING & MODELING	
10:30–10:40	Frédéric GALLIANO	Techniques to propagate uncertainties through data processing
10:40–10:45	Everyone	Discussion
10:45–10:55	Lise RAMAMBASON	Challenges for topological models of the interstellar medium
10:55–11:00	Everyone	Discussion
11:00–11:10	Erwan ALLYS	Evaluating uncertainties for components separation from observational data
11:10–11:20	Everyone	Discussion

The Program of the Workshop

08:30–08:55	INTRODUCTION: WHY UNCERTAINTIES ARE INSTRUMENTAL	
08:30–08:40	Frédéric GALLIANO	Motivations & objectives of the workshop - An example of a nested uncertainty problem
08:40–08:50	Marie GUEGUEN	A philosopher's viewpoint on identifying, quantifying & communicating uncertainties
08:50–08:55	Everyone	Discussion
08:55–10:00	QUANTIFYING EXPERIMENTAL & OBSERVATIONAL UNCERTAINTIES	
08:55–09:05	Karine DEMYK	An overview of the challenges of estimating experimental uncertainties
09:05–09:15	Marco MINISSALE	Uncertainties in ice laboratory experiments
09:15–09:20	Everyone	Discussion
09:20–09:30	Lucas EINIG	Signal extraction from noisy line cubes: the problem of applying hyperspectral imaging methods
09:30–09:35	Everyone	Discussion
09:35–09:45	Nathalie YSARD	Uncertainties in dust models
09:45–10:00	Everyone	Discussion
10:00–10:30	COFFEE BREAK	
10:30–11:20	PROPAGATING UNCERTAINTIES THROUGH DATA PROCESSING & MODELING	
10:30–10:40	Frédéric GALLIANO	Techniques to propagate uncertainties through data processing
10:40–10:45	Everyone	Discussion
10:45–10:55	Lise RAMAMBASON	Challenges for topological models of the interstellar medium
10:55–11:00	Everyone	Discussion
11:00–11:10	Erwan ALLYS	Evaluating uncertainties for components separation from observational data
11:10–11:20	Everyone	Discussion
11:20–12:00	HOW TO PUBLISH UNCERTAINTIES & ALLOW FUTURE STUDIES TO USE THEM CONSISTENTLY	
11:20–11:25	Pierre GRATIER	Quoting and plotting errors, and accounting for their correlation with ancillary phenomena
11:25–11:35	Everyone	Discussion
11:35–11:40	Pierre GRATIER	How to store and distribute this information
11:40–11:50	Everyone	Discussion
11:50–12:00	Everyone	Conclusion: what to do next?

@nwo.nl

Why are uncertainties so instrumental?

A philosopher’s viewpoint

Marie Gueguen, Marie Słodowska Curie fellow
Institut de Physique de Rennes 1
PCMI, 26 Octobre 2022

Astrochemistry

- Astrochemistry is a young interdisciplinary field, that started with the detection of CH, CH^+, CN in the 1940's

(McKellar, *PASP*, 52, 187 ,1940; Adams, *Astrophysics J.*,93,11,1941; Douglas & Herzberg ,93, 11,1941, Douglas & Herzberg 94, 381, 94, 381, 1941)

@nwo.nl

@nwo.nl

Astrochemistry

- What characterizes young IDF:
  - Rapid collection of observational data that requires interpretation
  - Theoretical and experimental progresses not always able to keep up with this rapid pace

  => Non-predictive models.

@nwo.nl

What is a model?

- *Partial representation*
- *Minimal modelling* principle
- *Construction*: idealizations, approximations and simplifications

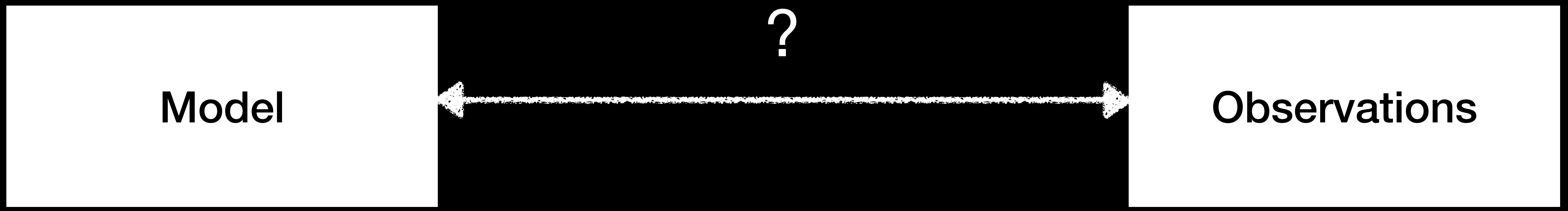
What is a model?

@nwo.nl

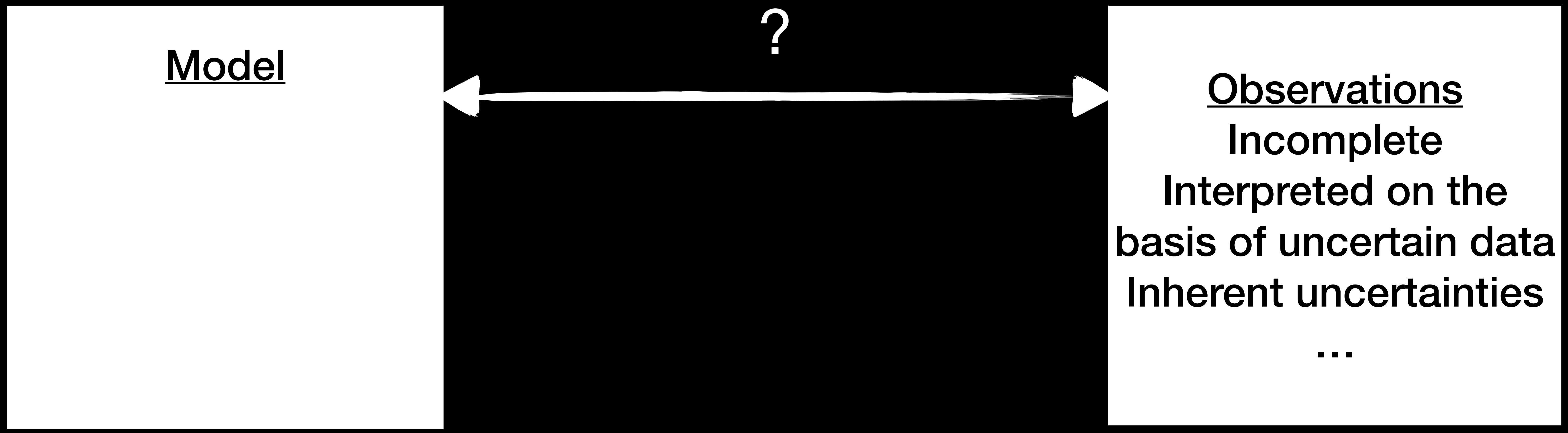
- *Partial representation*
- *Minimal modelling* principle
- *Construction*: idealizations, approximations and simplifications
- *Computational model:* high epistemic opacity (= not an easy task to contribute which input data contribute the most to the model's output)

Model development in context of high uncertainties

@nwo.nl



Model development in context of high uncertainties



@nwo.nl


@nwo.nl

Model development in context of high uncertainties

<u>Model</u>
Partial representation
Idealizations (chemical networks, type of chemistry, astrophysical conditions, etc..)
Uncertainties in the input
….

? ⟷

<u>Observations</u>
Incomplete
Interpreted on uncertain data
Inherent uncertainties
…

Astrochemistry

@nwo.nl

- Non-predictive models:
  - Uncertainties higher on the theoretical side than on the observational side
  - (Dis)agreement with observations not interpretable

Astrochemistry

- Non-predictive models:
    - But: identifying and reducing uncertainties is not only the path to predictivity, but it is also the only tool you have to break the epistemic opacity of your model and get a better understanding of what your model is sensitive to!
    - It also allows you to target where experimental and theoretical progresses are the most needed.

But which uncertainties for which task?

- Parametric uncertainties
- Model uncertainties
- Unknown unknowns

Example 1: *Ab initio* calculations & rate coefficients

Uncertainties → PES

Uncertainties → Dynamical calculations

Uncertainties → Approximations

Example 1: *Ab initio* calculations & rate coefficients

Uncertainties →

- PES
  - *Size of the grid, basis, fit*
- Dynamical calculations
- Approximations

Example 1: *Ab initio* calculations & rate coefficients

Uncertainties

- PES
- Dynamical calculations
  - *Cross-sections, Collisional data*
- Approximations

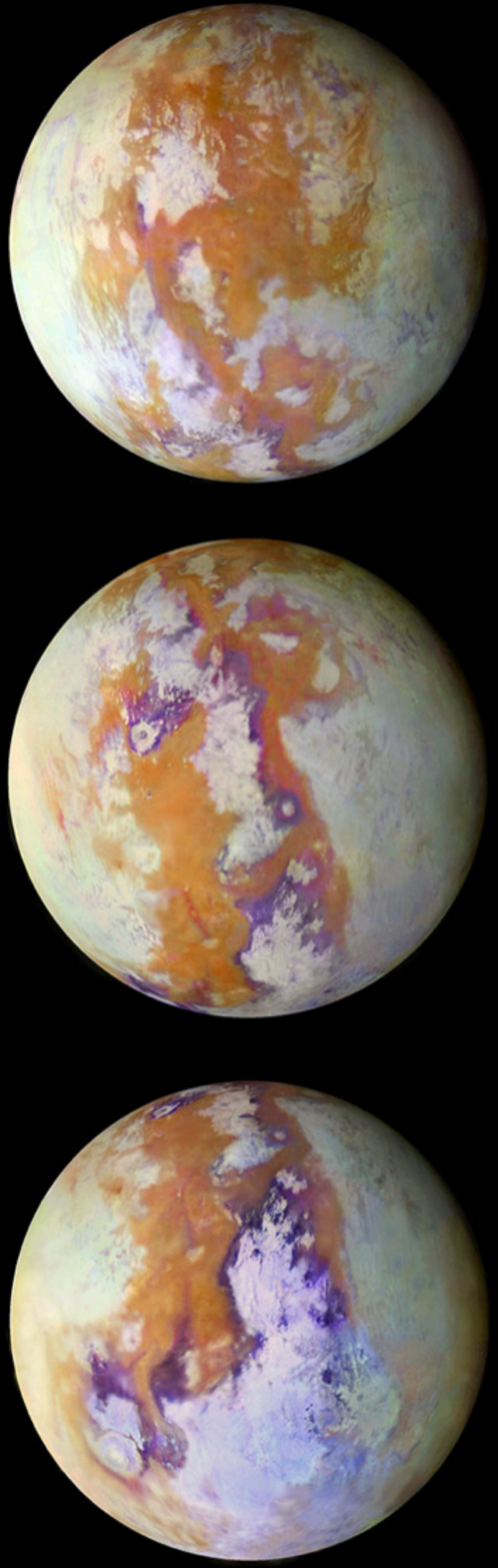
Example 1: *Ab initio* calculations & rate coefficients

Uncertainties → PES

Uncertainties → Dynamical calculations

Uncertainties → Approximations

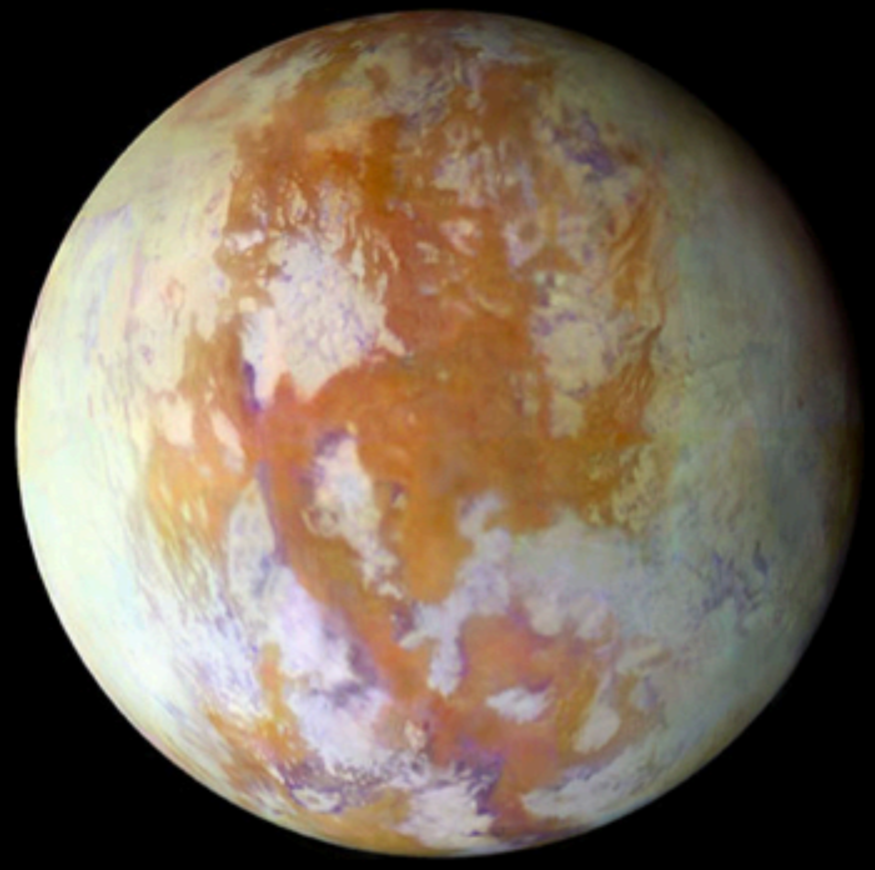
Born-Oppenheimer ?

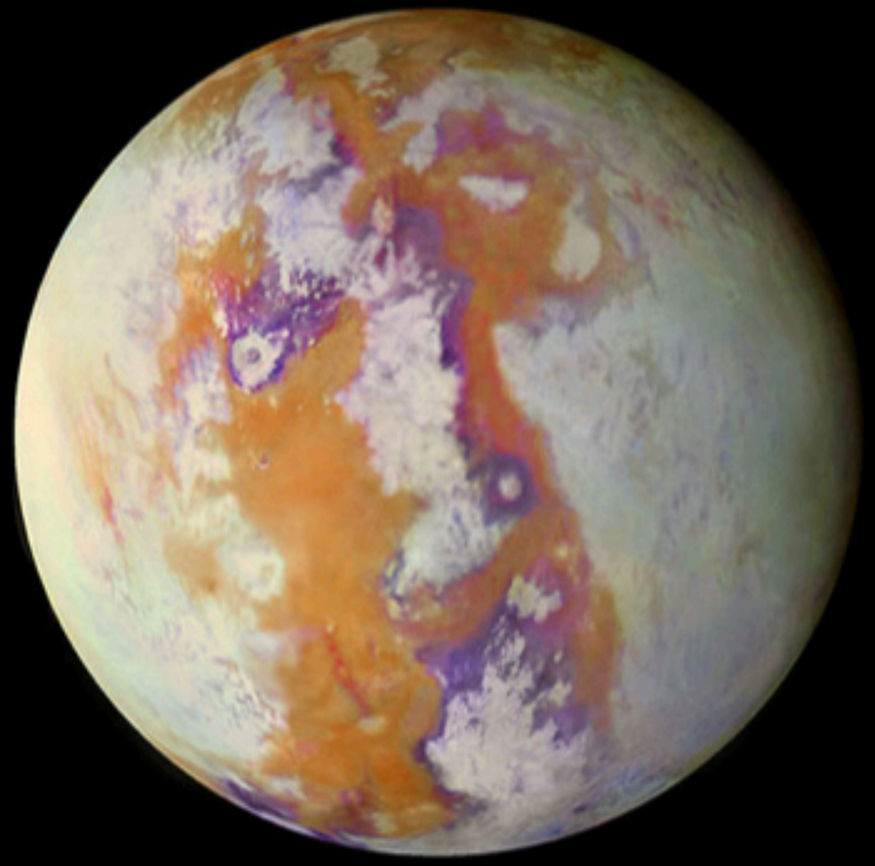


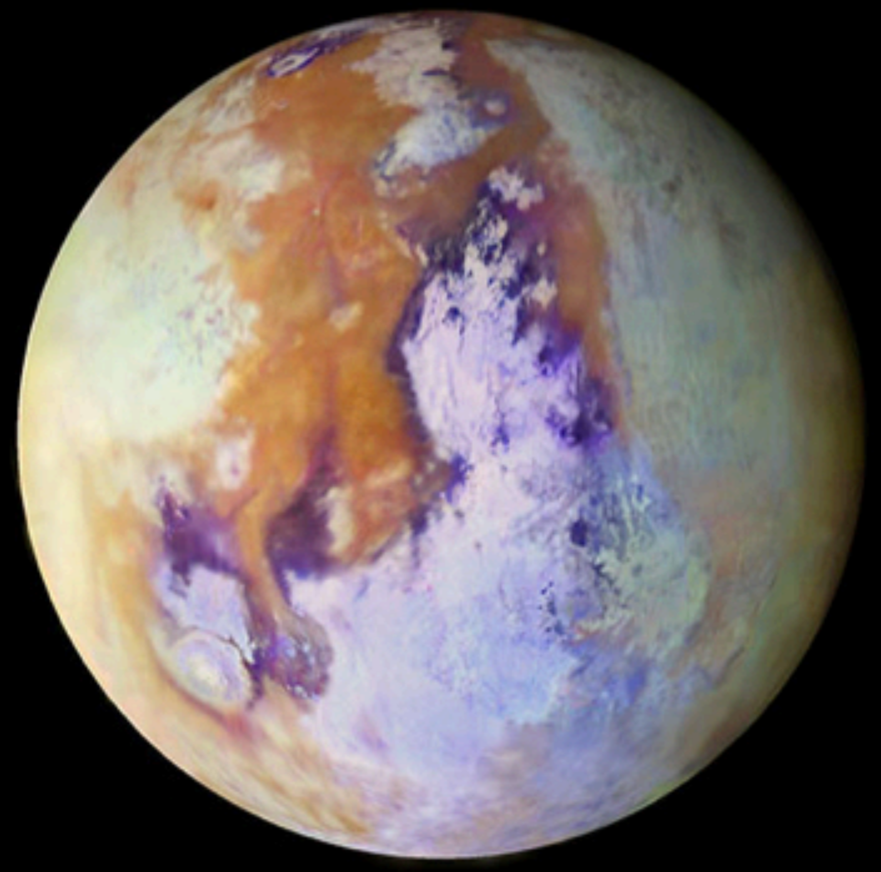
Example 2: Low-T reaction rate constants

Uncertainties

- Parametric
- Extrapolation
- Model
- Unknown unknowns







Example 2: Low-T reaction rate constants

Uncertainties

- Parametric

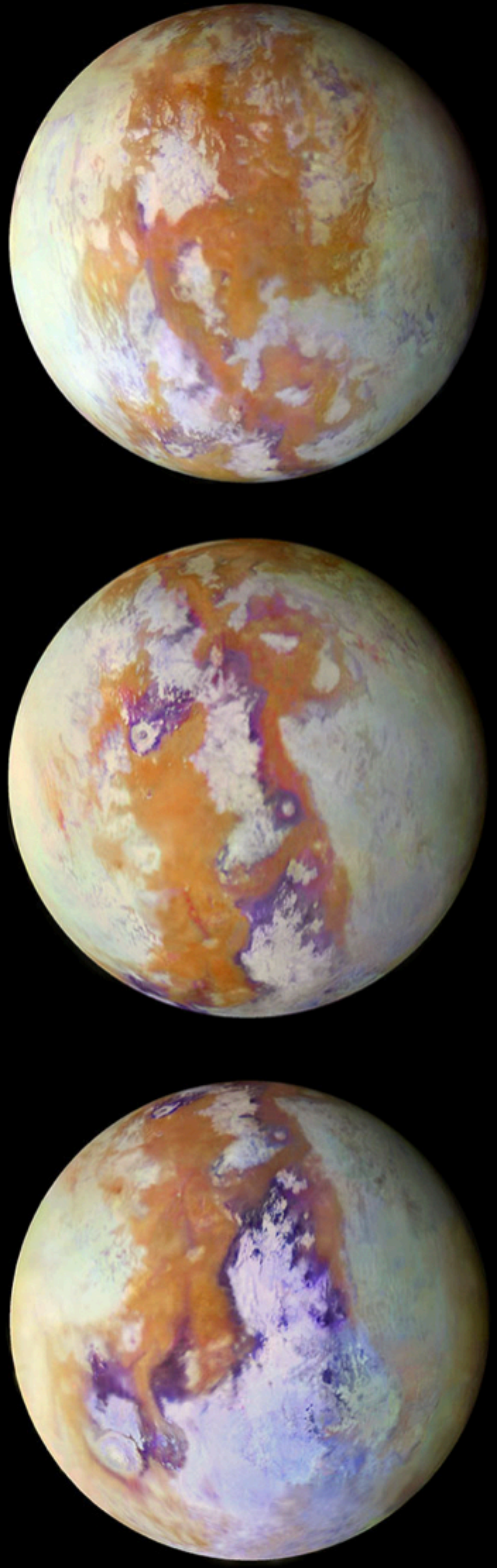
  $\ln k(T) = \alpha + \beta x(T)$

  -> $\alpha, \beta, \sigma_\alpha, \sigma_\beta$ et $\varrho_{\alpha\beta}$

- Extrapolation

- Model

- Unknown unknowns

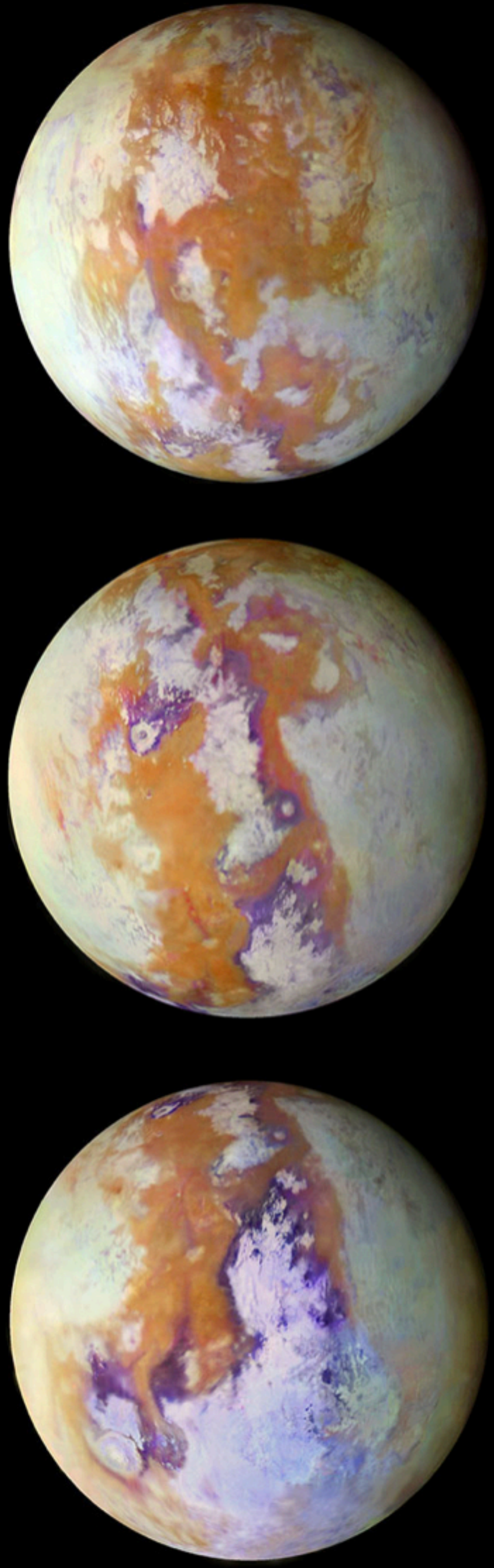


Example 2: Low-T reaction rate constants

Uncertainties →

- Parametric
- Extrapolation

  T range?
- Model
- Unknown unknowns

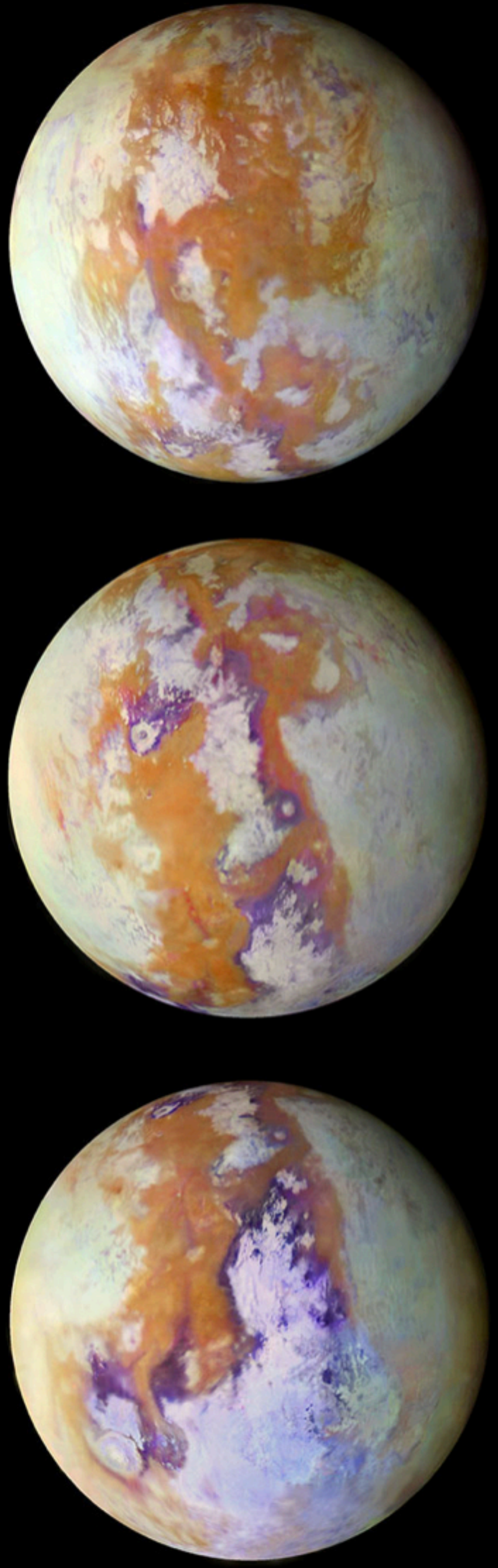


Example 2: Low-T reaction rate constants

Uncertainties →

- Parametric
- Extrapolation
- Model

  Arrhenius laws?
- Unknown unknowns



Example 2: Low-T reaction rate constants

Uncertainties

- Parametric
- Extrapolation
- Model
- Unknown unknowns

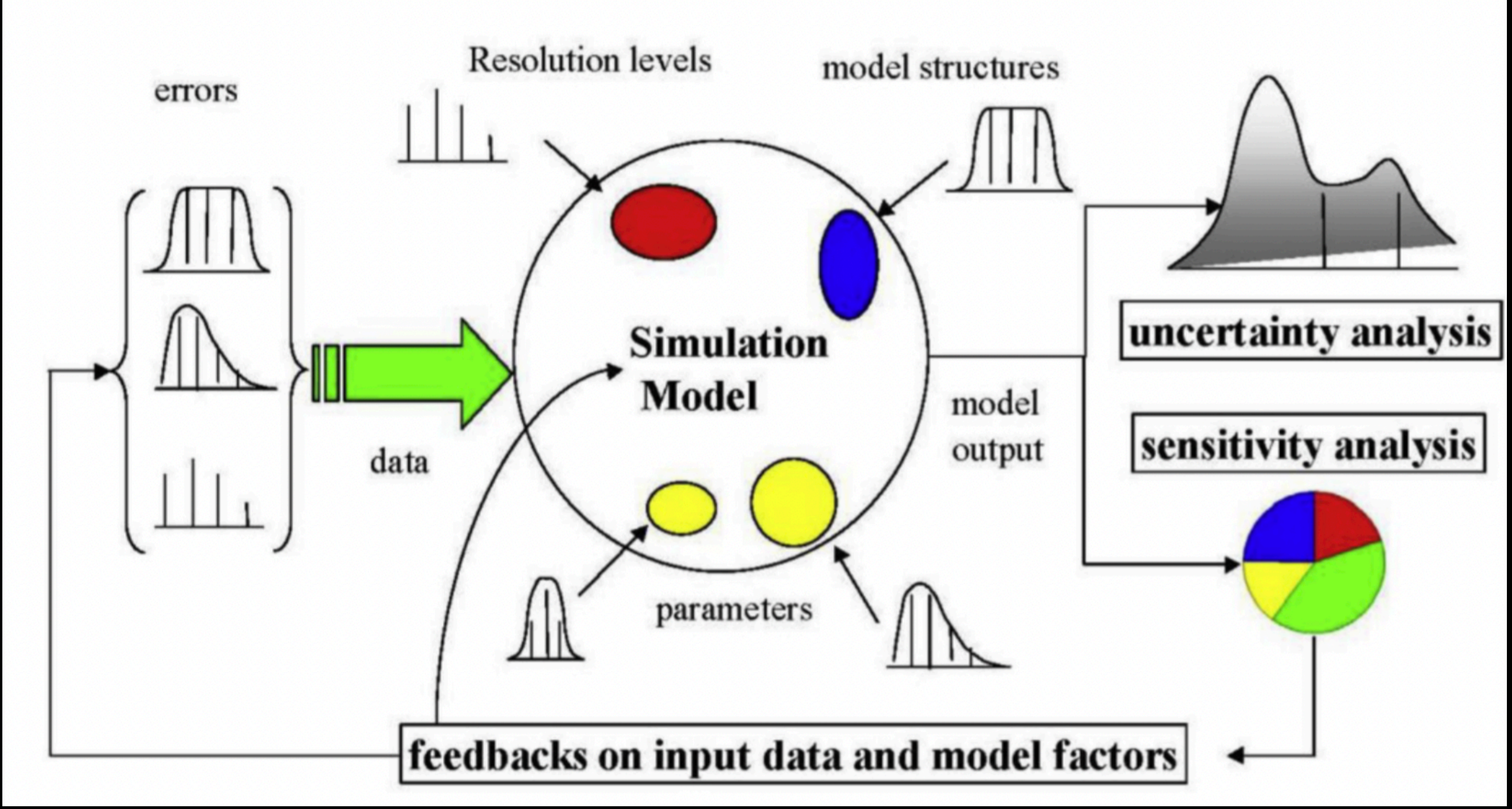
Missing chemistry? Missing physics?

Example 1: Low-T reaction rate constants

- **Uncertainty Propagation**: each reaction rate constant k_i / all parameters are randomly perturbed a large number of times according to a pre-definite probability distribution

- **Sensitivity analysis** (Saltelli, 2020): study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input
    - identify correlations between inputs and outputs

Saltelli et al., 2019, *Env. Mod. &Softw 114*



Uncertainties coming from heterogenous sources are propagated in the model

Generates an empirical distribution of the output of interest

UC in the output decomposed according to source

Saltelli et al., 2019, *Env. Mod. &Softw 114*

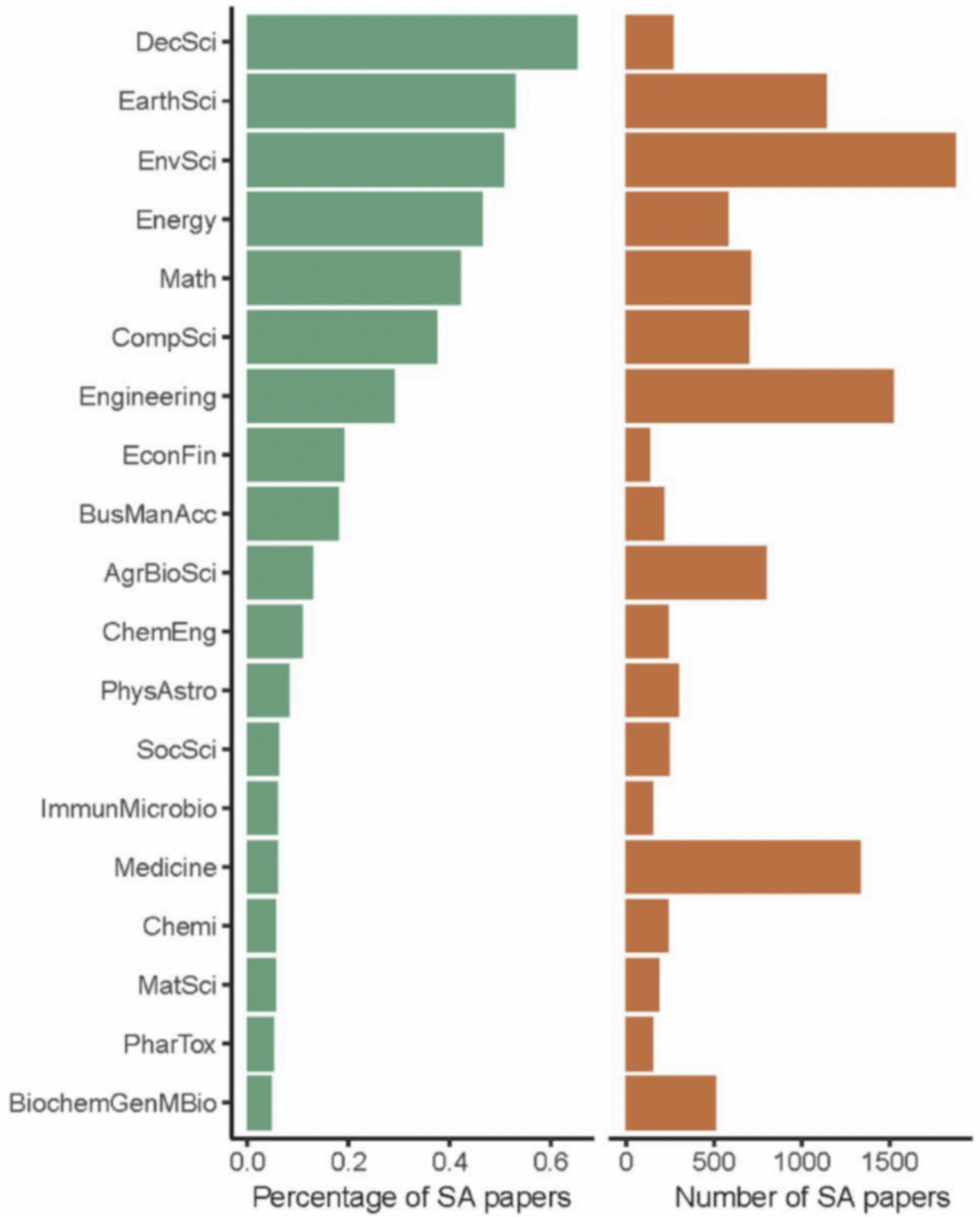
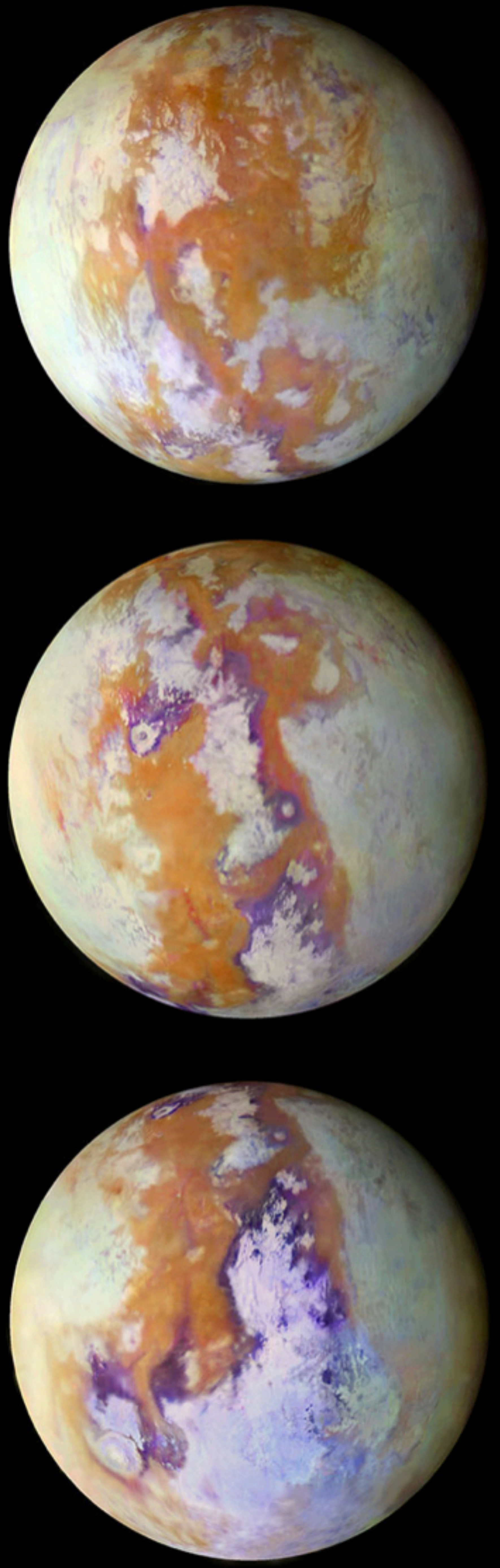


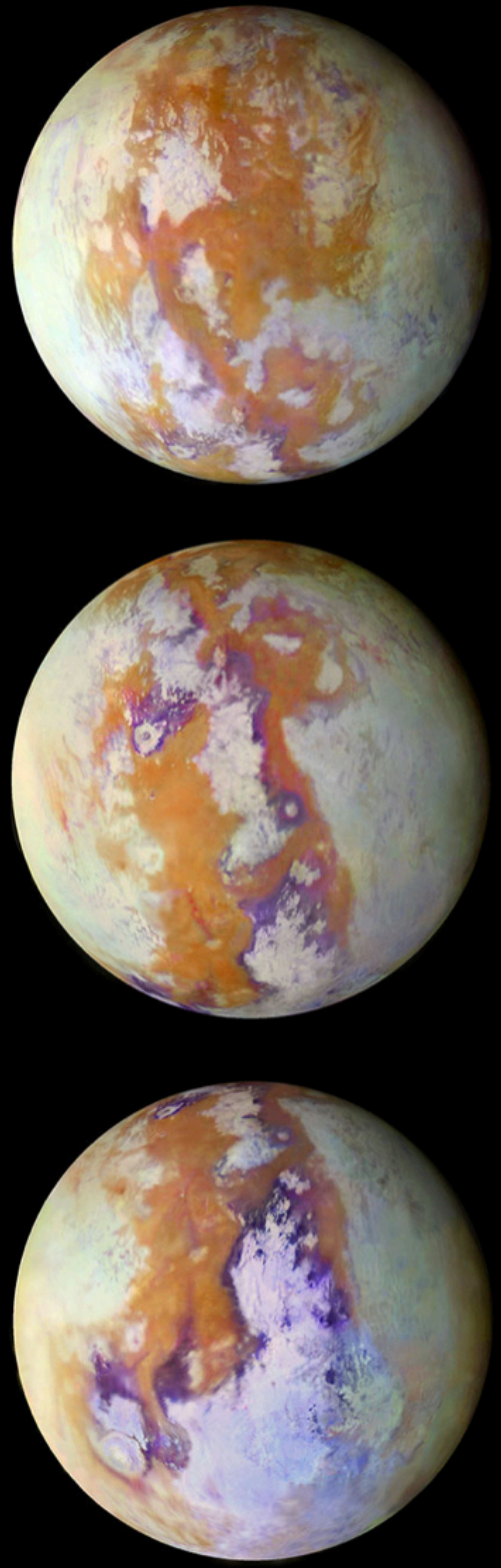
Fig. 4. Density and number of sensitivity analysis articles returned by search criteria, by subject.



Example 1: Low-T reaction rate constants

Uncertainties

- Parametric ✔
- Extrapolation ✔
- Model
- Unknown unknowns



Example 1: Low-T reaction rate constants

Uncertainties →

- Parametric ✔
- Extrapolation ✔
- ~~Model~~
- ~~Unknown unknowns~~

Loison, Hickson, Hebrard and Dobrijevic, 2014

Methodology

The methodology we use to improve our knowledge of the photochemistry of Titan's atmosphere is the following:

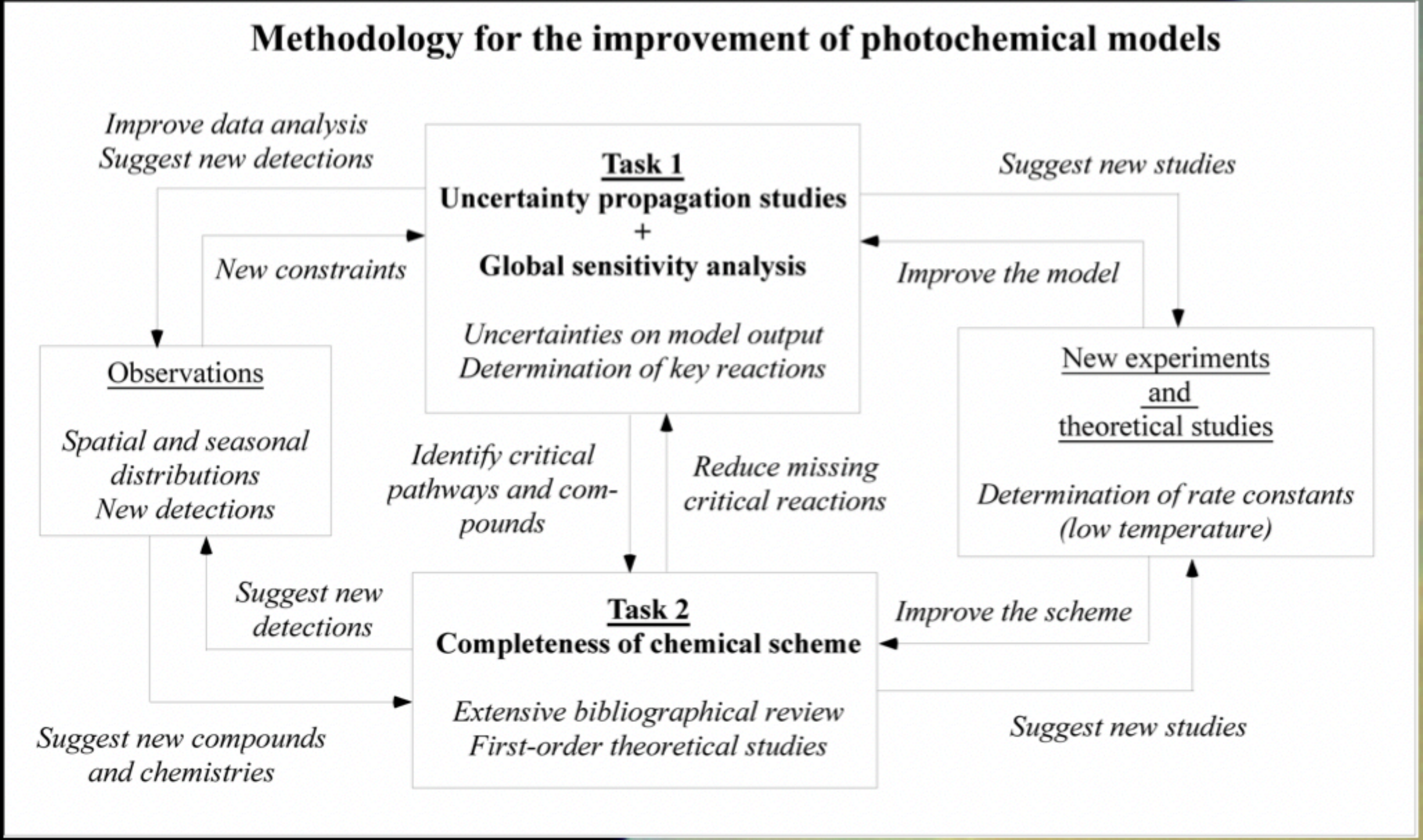


Figure 1. Methodology based on two tasks to improve the chemical networks of photochemical models. These two tasks serve as an efficient basis for new studies focused on selected reactions, which in return can improve significantly the chemical scheme used in models. Improvement of models favour new detection attempts and put better constraints on physical parameters.

Conclusion

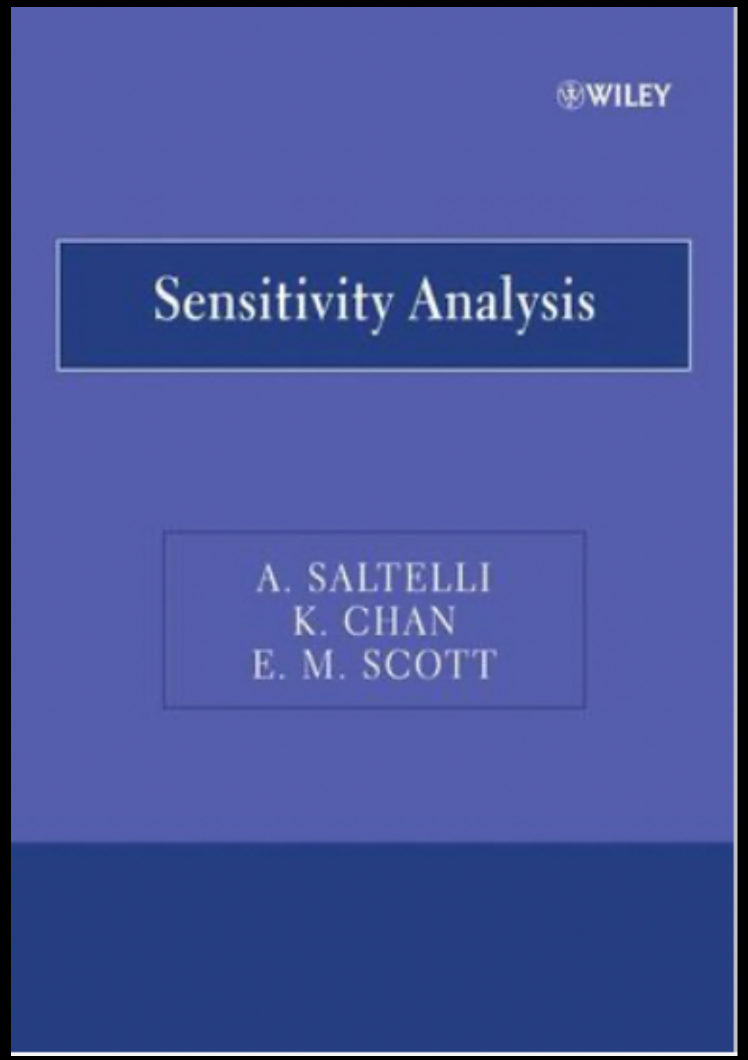
- Why are uncertainties instrumental? Because exploiting uncertainties helps to better understand a model that would otherwise remain a black box - especially what the model is sensitive too
  - Target theoretical and experimental progresses thanks to uncertainty propagation methods and sensitivity analysis.

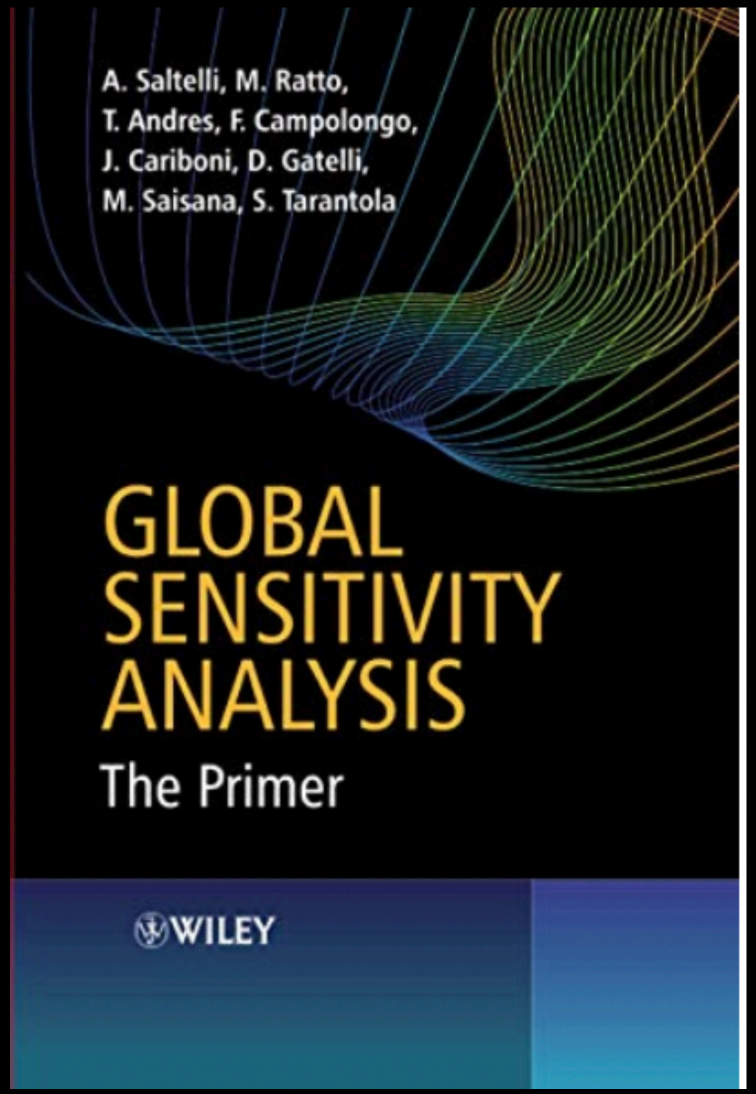
and what is negligible

  - Optimization of your model in terms of computational cost.

- Forcing a non-predictive model to match observations bu tuning parameters not empirically constrained amounts to increasing its epistemic opacity and to loose your two main sources of information!
- UQ methods and sensitivity analysis: important interdisciplinary facilitators both in terms of uncertainty communication and of targetting where experimental and theoretical progresses will pay off the most. (Saltelli, 2020,

Conclusion



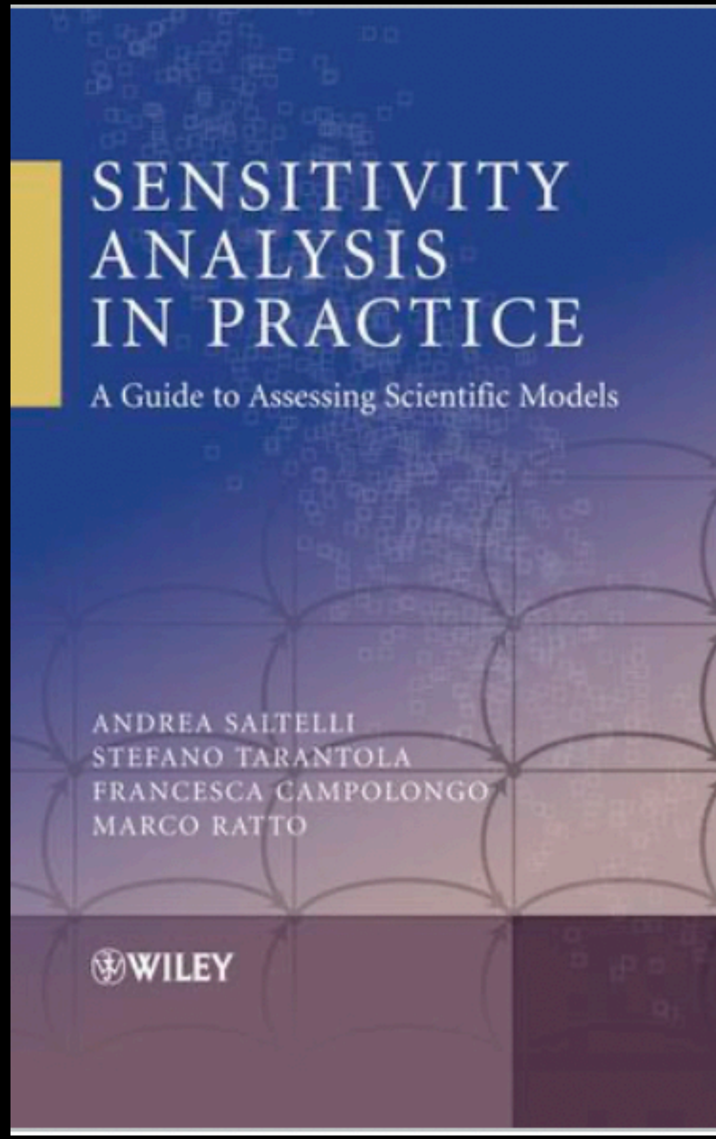


- Why are uncertainties instrumental? Because exploiting uncertainties helps to better understand a model that would otherwise remain a black box - especially what the model is sensitive too
    - Target theoretical and experimental progresses thanks to uncertainty propagation methods and sensitivity analysis.

and what is negligible

- Optimization of your model in terms of computational cost.

- Forcing a non-predictive model to match observations bu tuning parameters not empirically constrained amounts to increasing its epistemic opacity and to loose your two main sources of information!
- UQ methods and sensitivity analysis: important interdisciplinary facilitators both in terms of uncertainty communication and of targetting where experimental and theoretical progresses will pay off the most. (Saltelli, 2020,



Acknowledgments

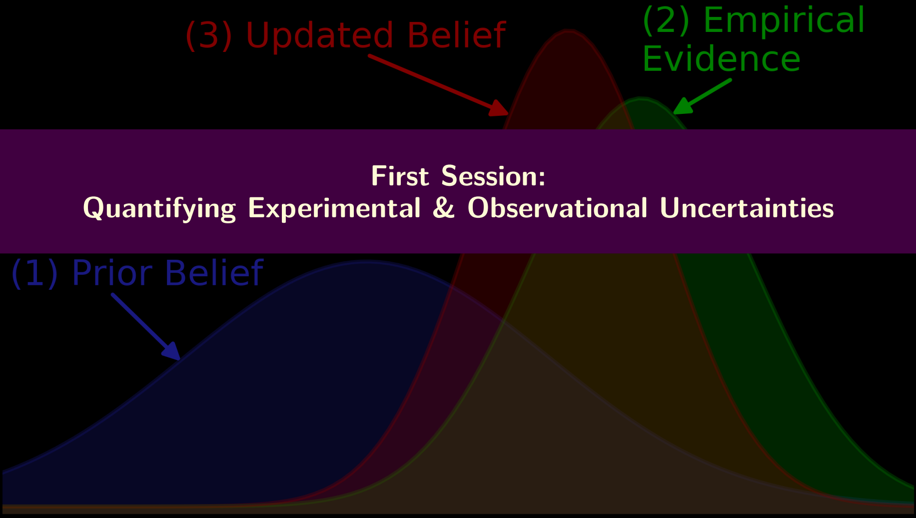
COLLEXISM

Institut de Physique de Rennes 1

Thank you!

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No.101026214.

First Session:
Quantifying Experimental & Observational Uncertainties



Overview of the challenges of estimating experimental uncertainties

Karine Demyk

How do we / should we do?

- list all sources of uncertainties and errors
- calibrate the experiments
- estimate the uncertainties/errors coming from
  - measurements
  - data reduction
  - data modelling / fitting needed to extract the studied quantity
- explain all choices made, keep track of all steps
- not always easy to quantify!
- highly dependent on the experiment

Example 1: measurement of opacities (MAC)

- From spectroscopic measurements on a population of grains embedded in a matrix

$$MAC = -\frac{S}{m} \times ln(T)$$

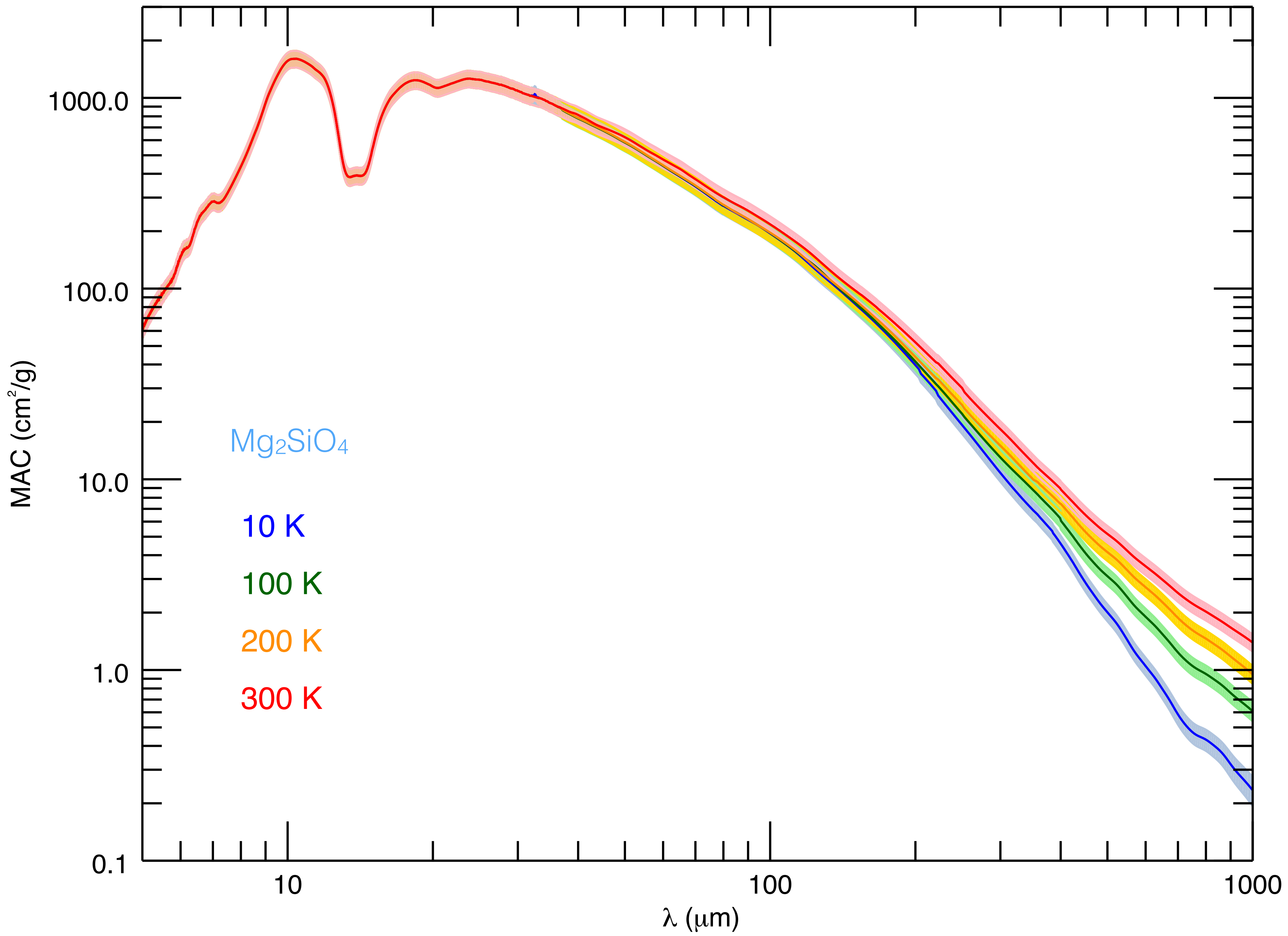
T = transmittance
S = sample surface
m = sample mass

- spectrometer stability

$$\Delta MAC = \frac{\delta T}{T} \times \frac{S}{m}$$

- uncertainty on the sample mass

$$\Delta MAC_m = MAC(\lambda) \times \frac{\delta m}{m}$$



Example 1: calculation of optical constants

A number of assumptions have to be made :

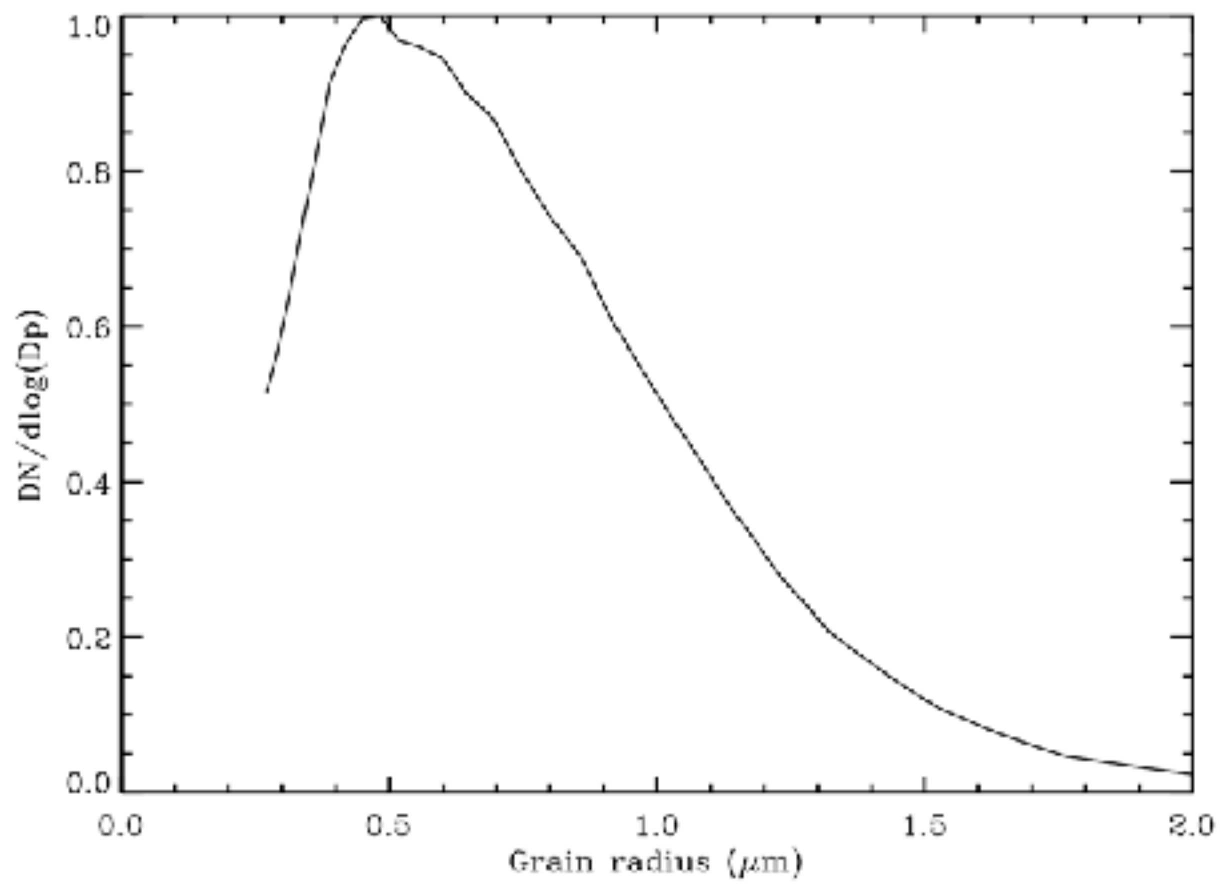
➡ To relate the MAC to (n,k) :

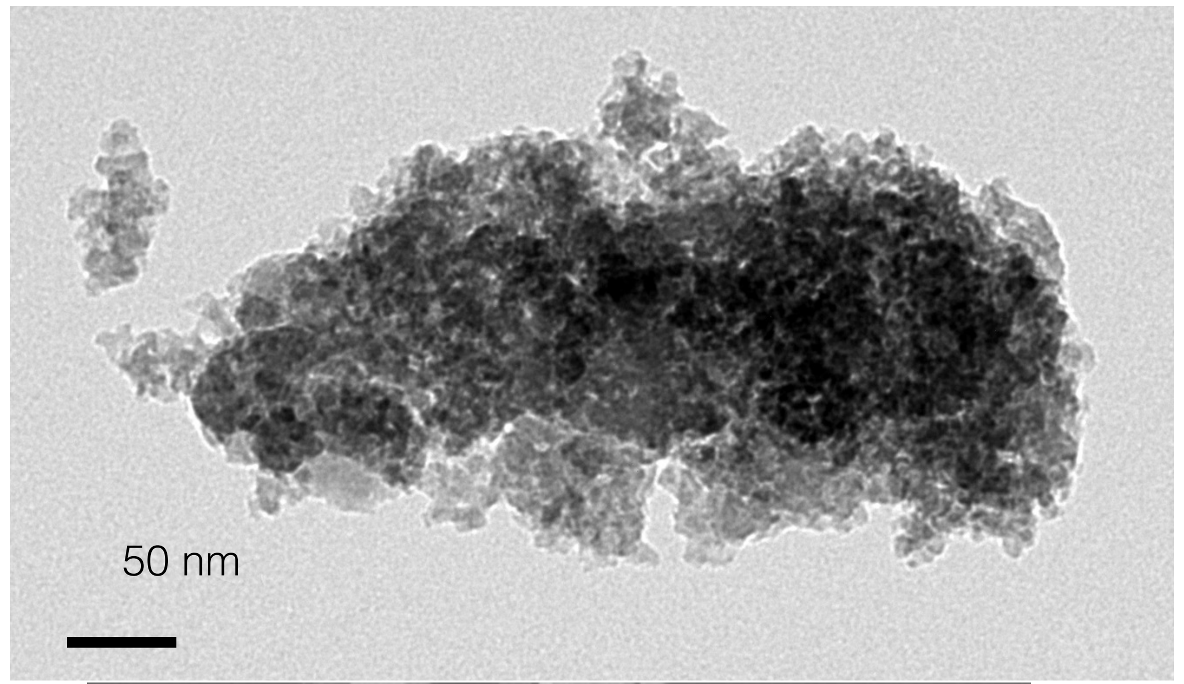
- grain size
- grain shape

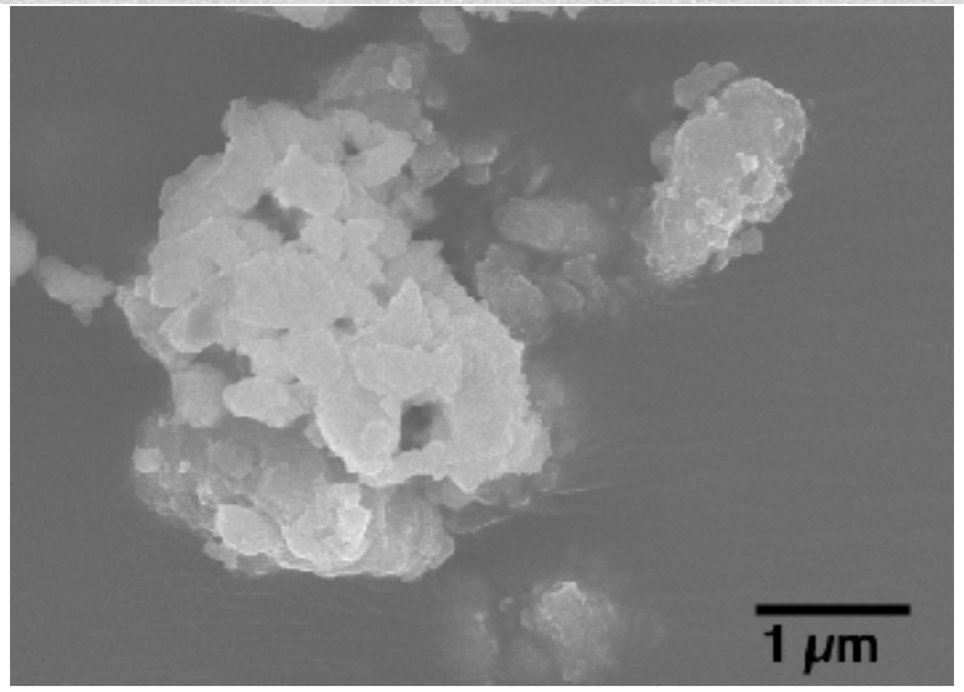
➡ To calculate (n,k) knowing the MAC

- value of the refractive index in the visible

➡ Effect of a possible agglomeration in the pellet ?



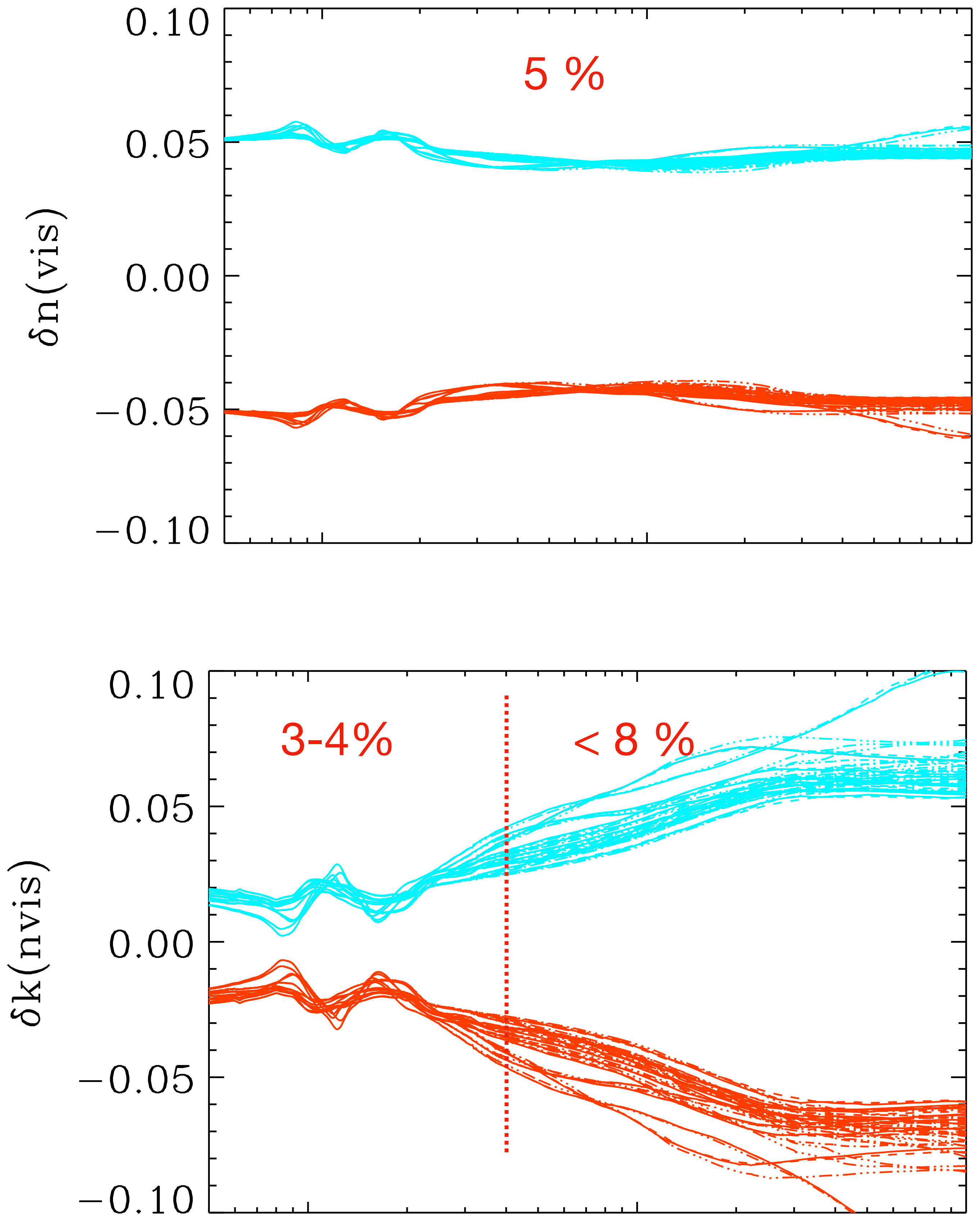




Example 1: estimation of the error on the optical constants

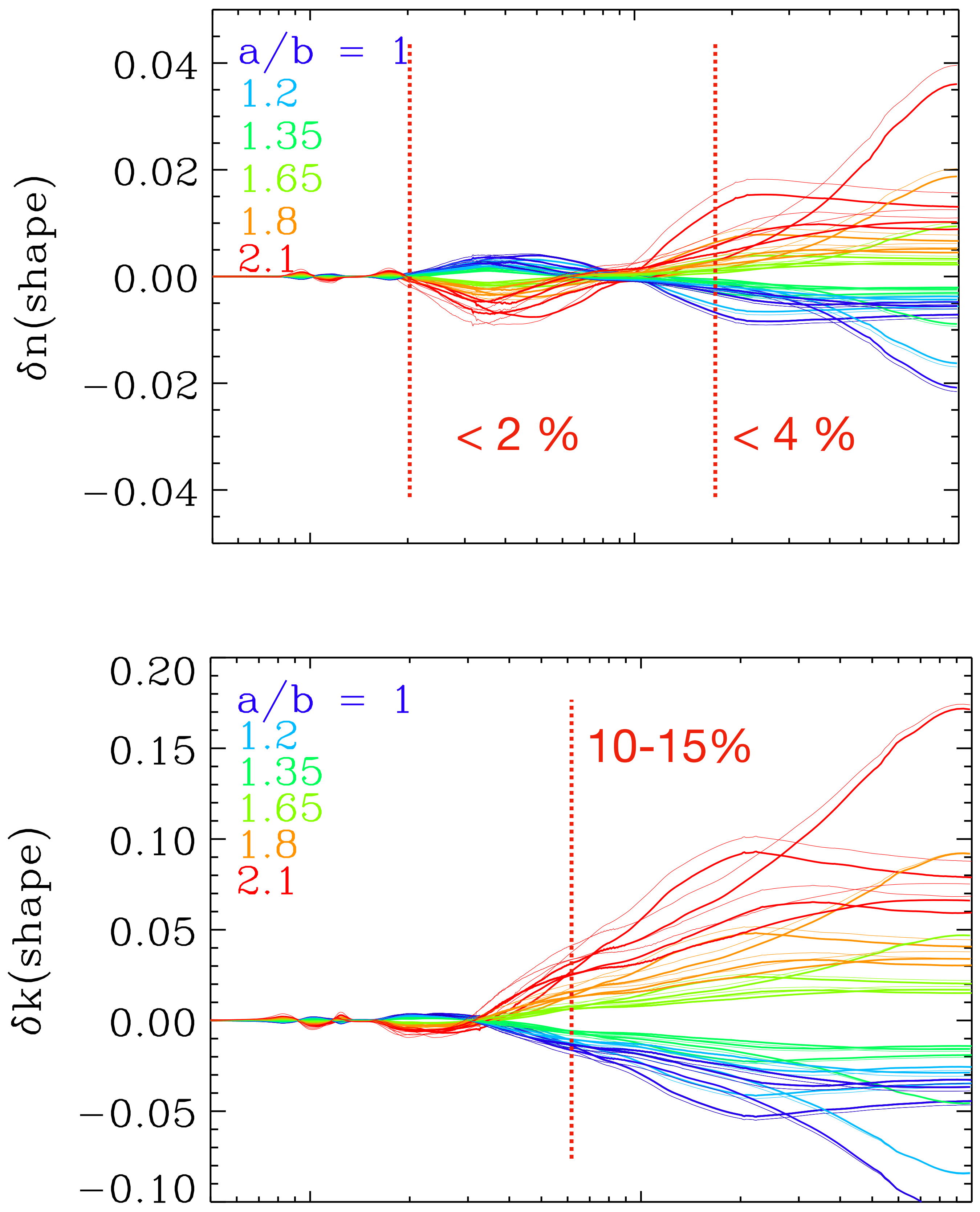
Error on n_{vis} :

- n_{vis} varies by ± 5%



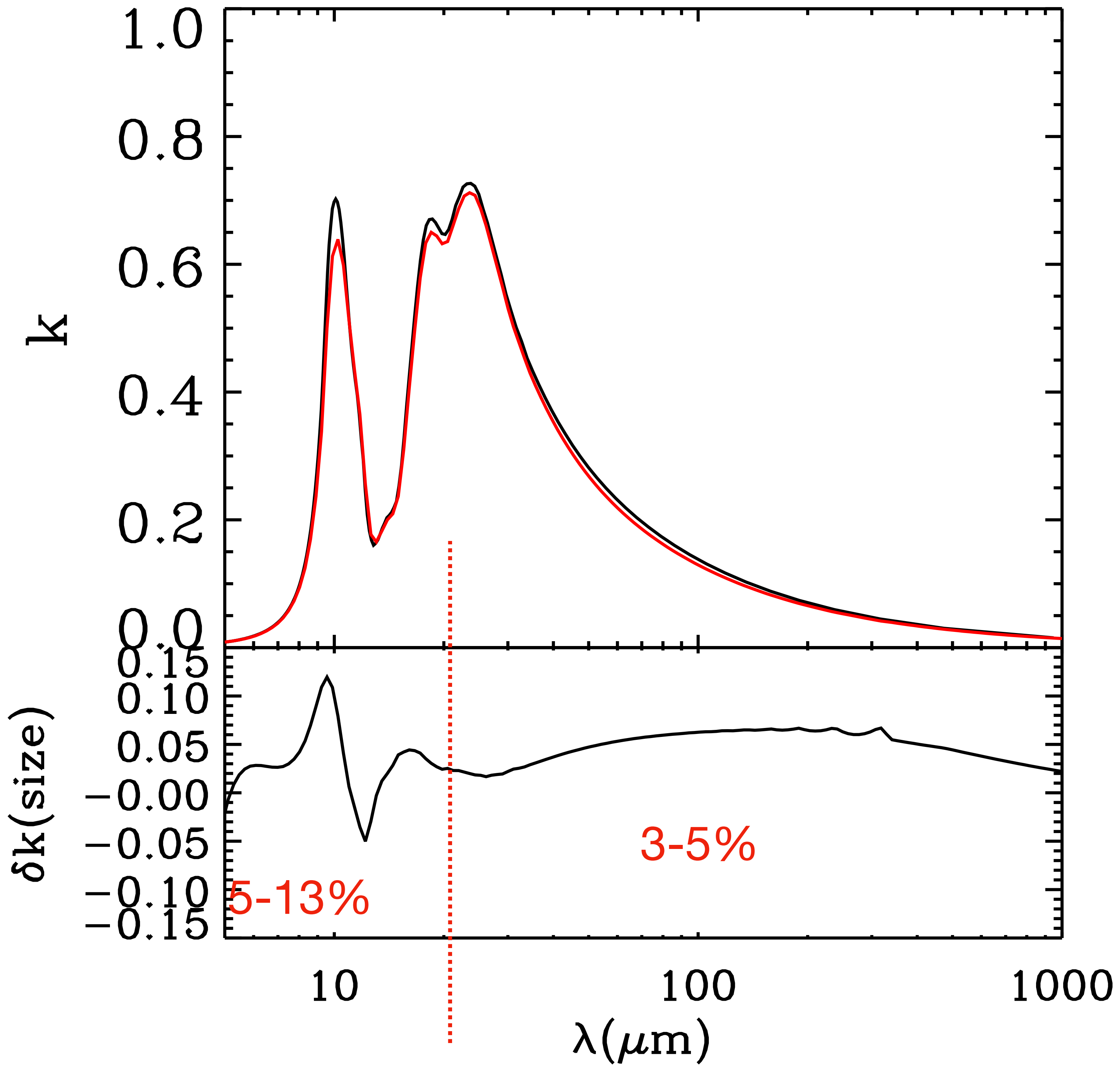
Error on grain shape :

- prolate vs oblate
- a/b varies by ± 40%



Error on grains size (Rayleigh limit assumption):

- DDA calculations
- prolate grains
- measured size distribution
- Mg_2SiO_4 silicate from Jäger+2003



[Demyk+*2022*]

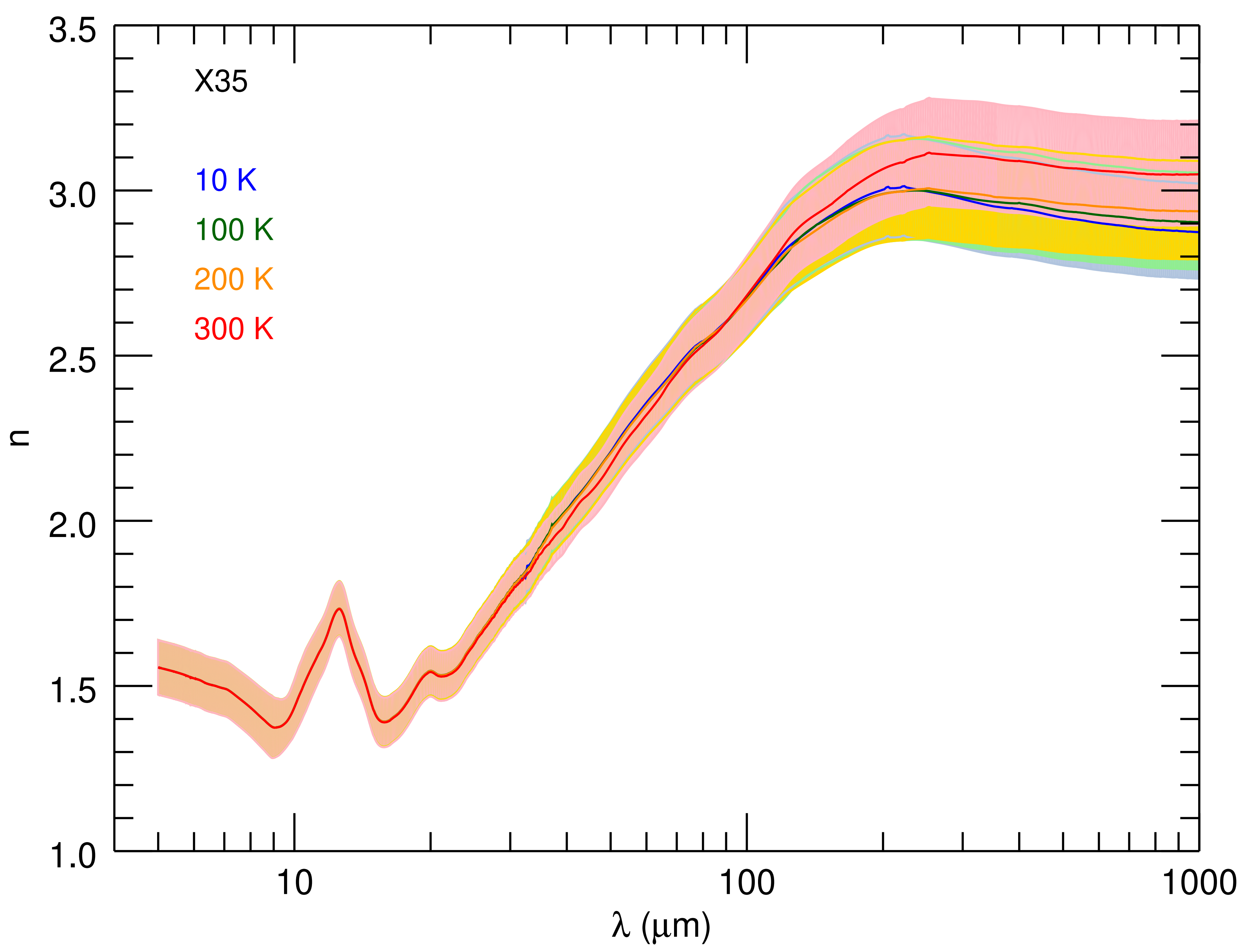
Example 1 : Error estimation on Mg-rich silicate optical constants:

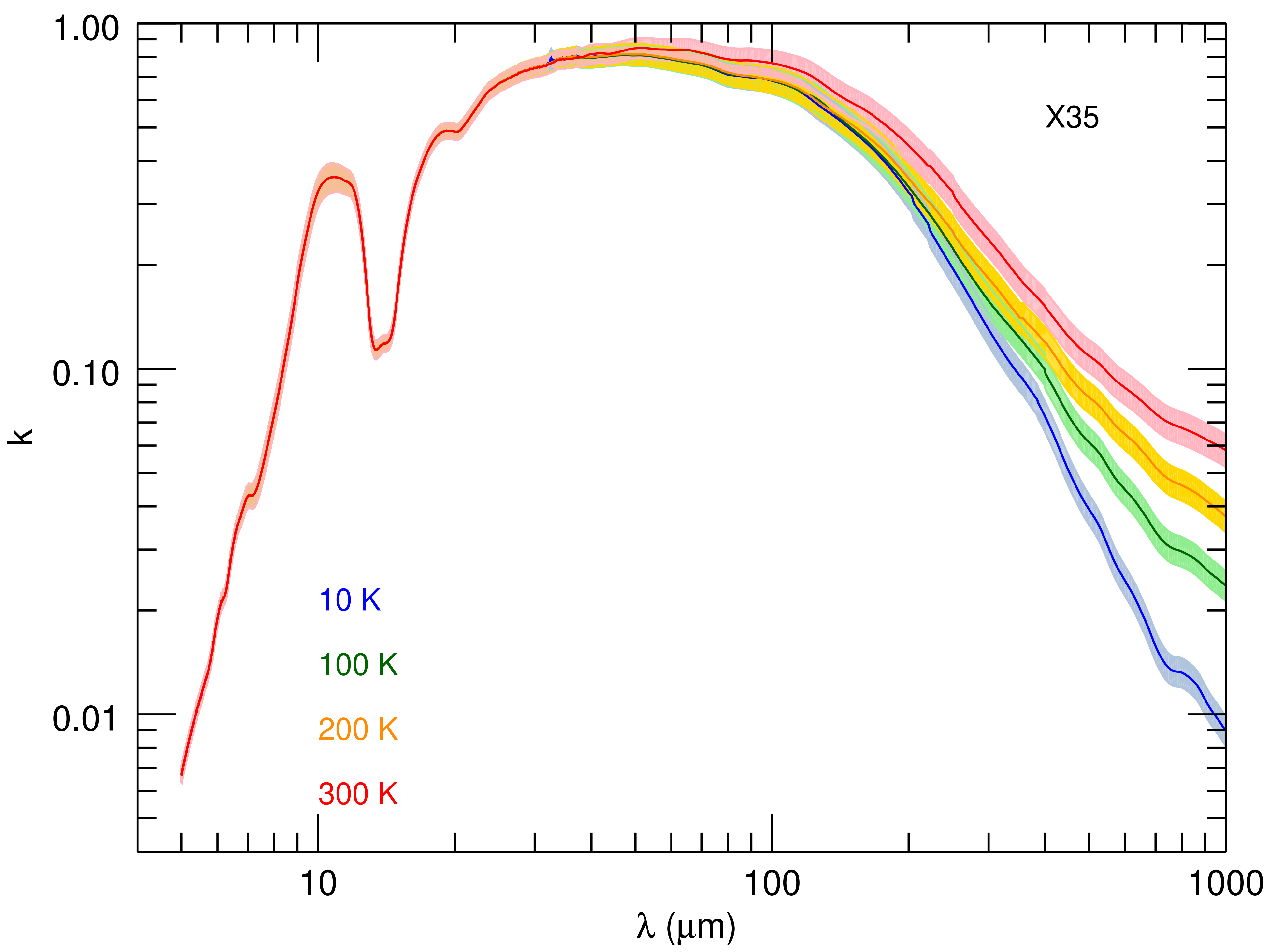
Quadratic sum of the errors ⇒ total uncertainty on n and k

δn_{tot} ~ 4 - 6% , dominated by uncertainty on nvis

δk_{tot}: λ < 30 µm: ~5% (up to ~13% in the silicate feature , dominated by grain size uncertainty

λ > 30 µm: 5 to 25 % depending on sample, dominated by grain shape

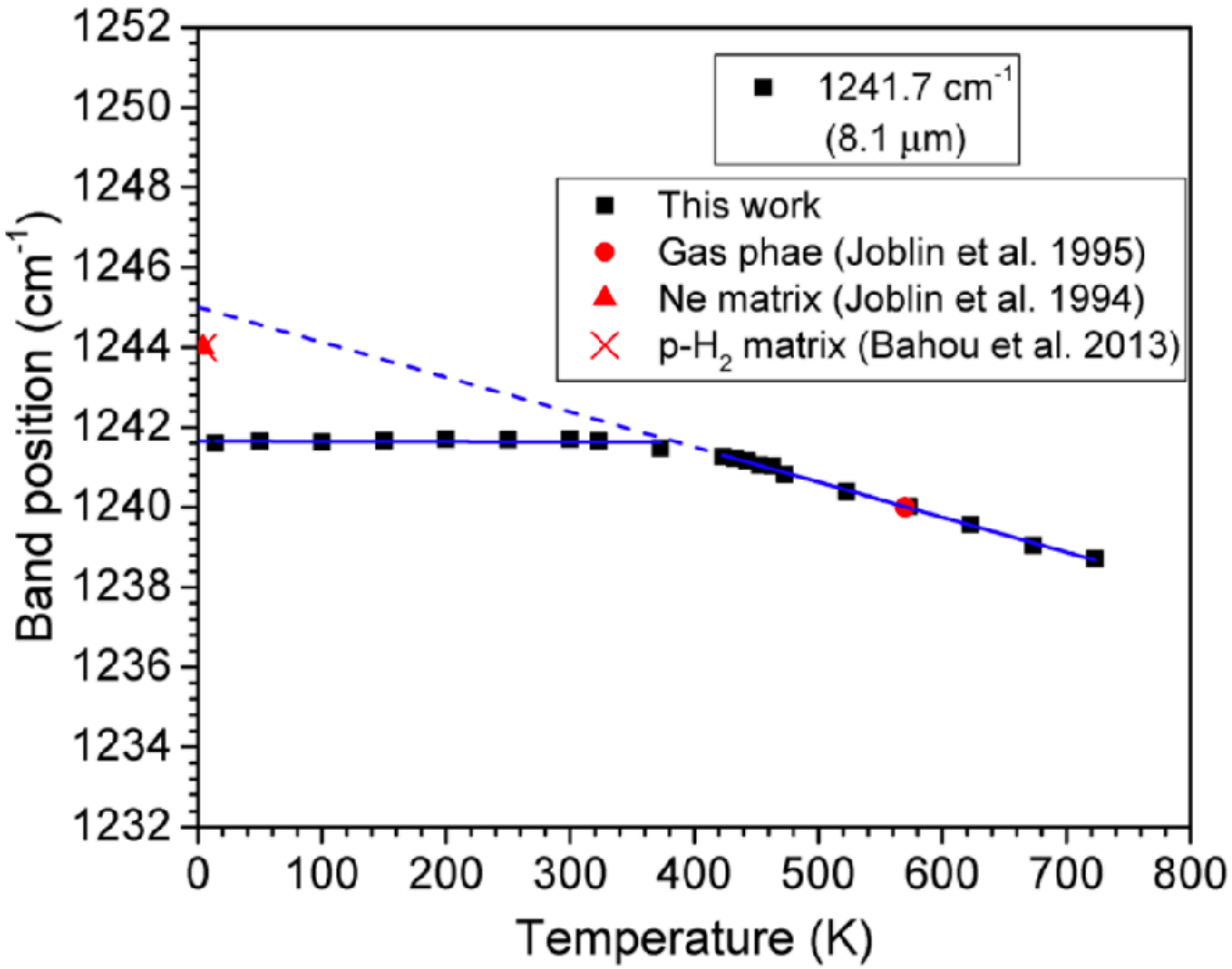




[Demyk+202

Example 2 : determination of anahormanicity factor

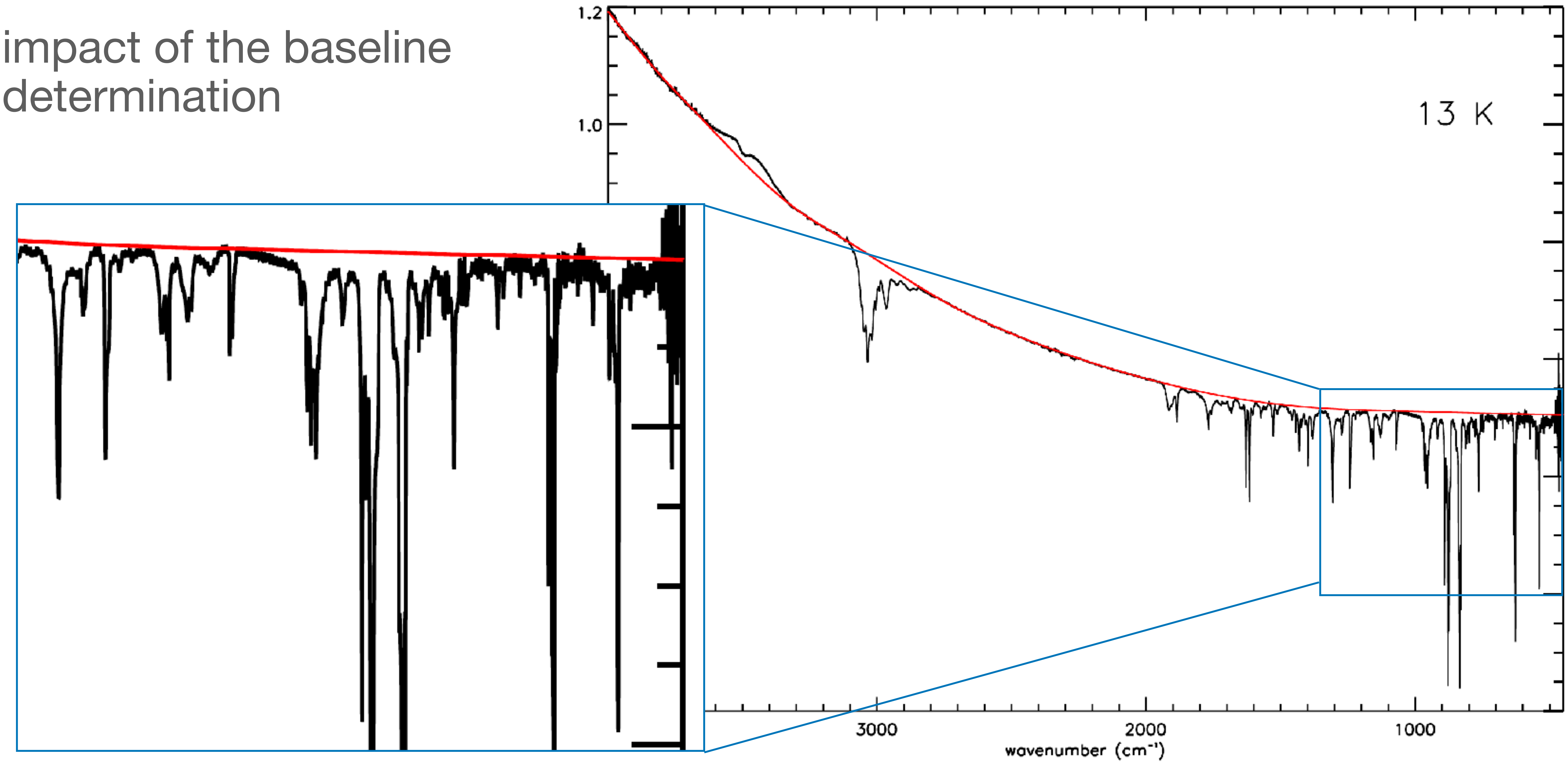
- From spectroscopic measurements on a population of grains embedded in a matrix at varying temperature: 13 - 723 K



- baseline subtraction
- peak position determination
- fit the λpeak - T relation

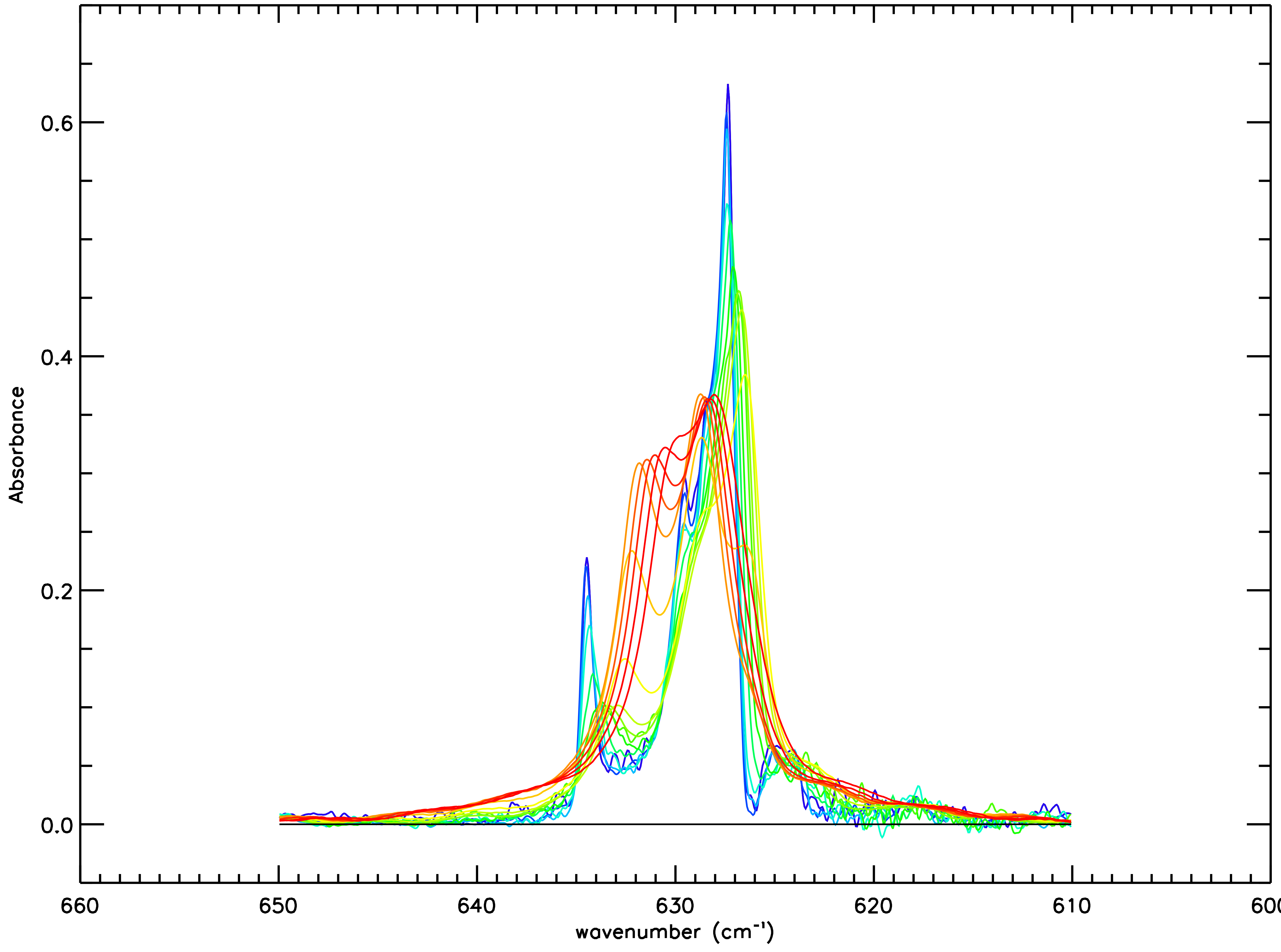
Example 2: determination of anahormanicity factor

- data reduction:
  - impact of the baseline determination

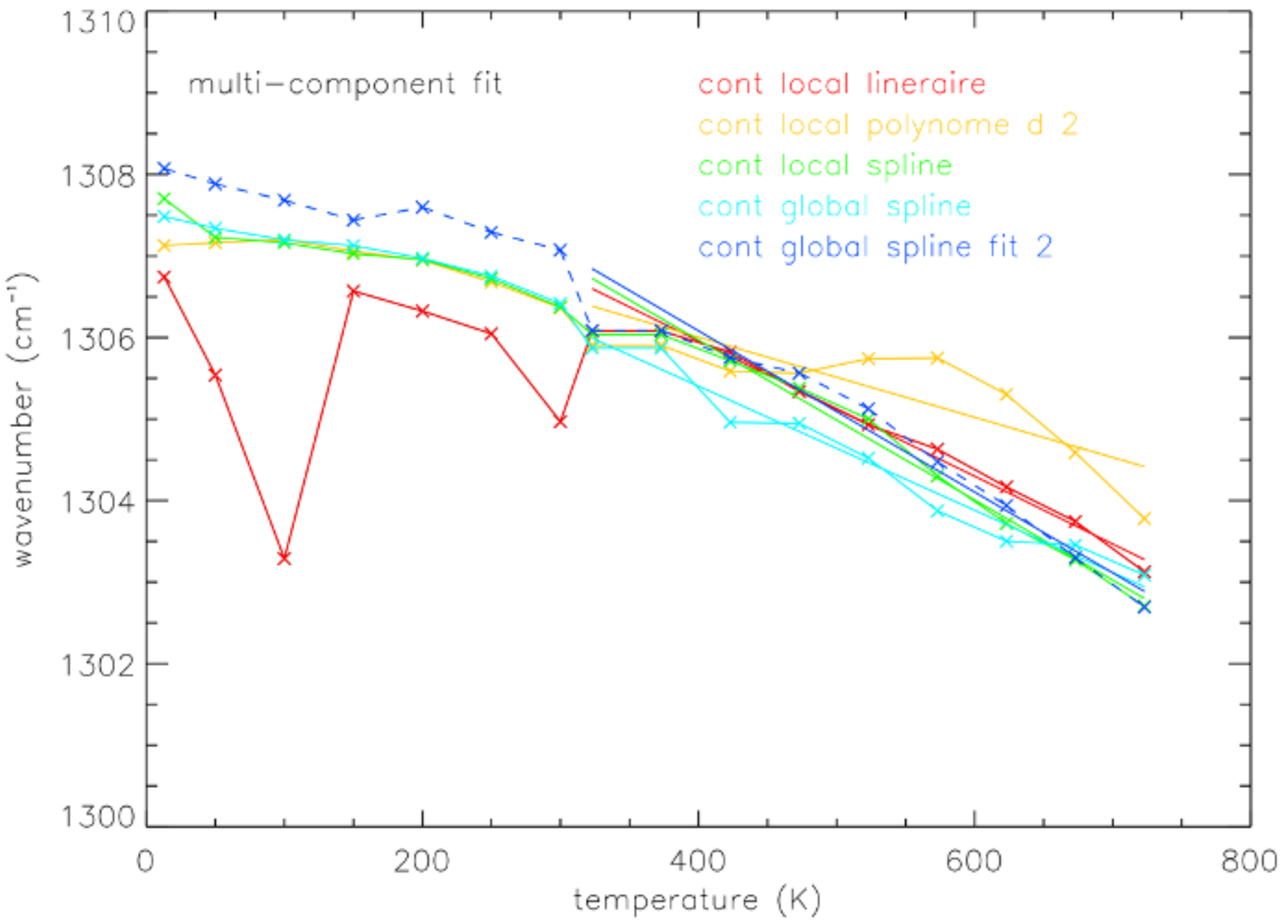


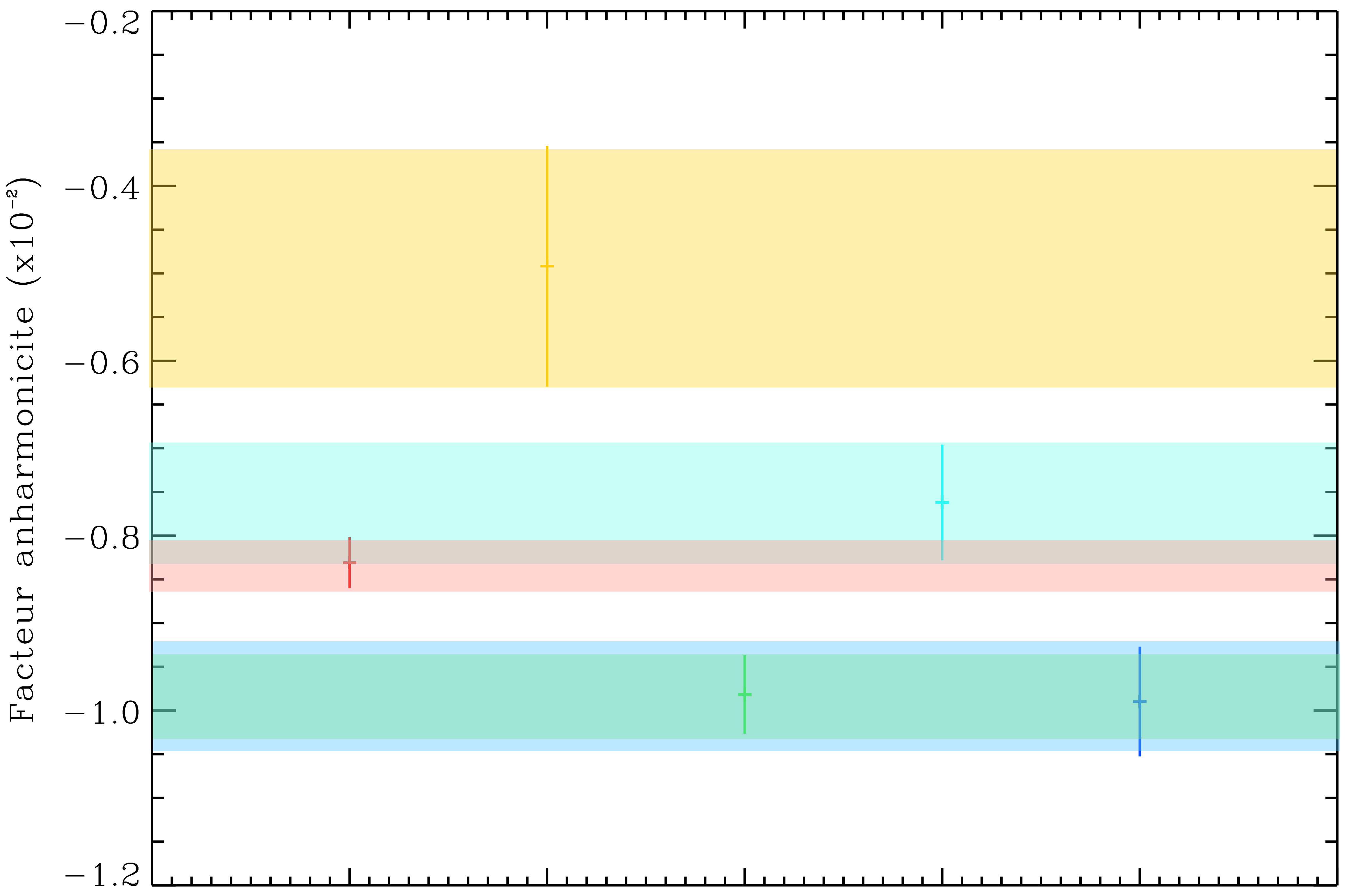
Example 2: determination of anahormanicity factor

- data modelling: method to determine the band position and width
  - peak maximum
  - area-weighed peak maximum
  - spectral decomposition with gaussians, lorentzians



Example 2: determination of anahormanicity factor





UNCERTAINTIES IN
ICE LABORATORY EXPERIMENTS

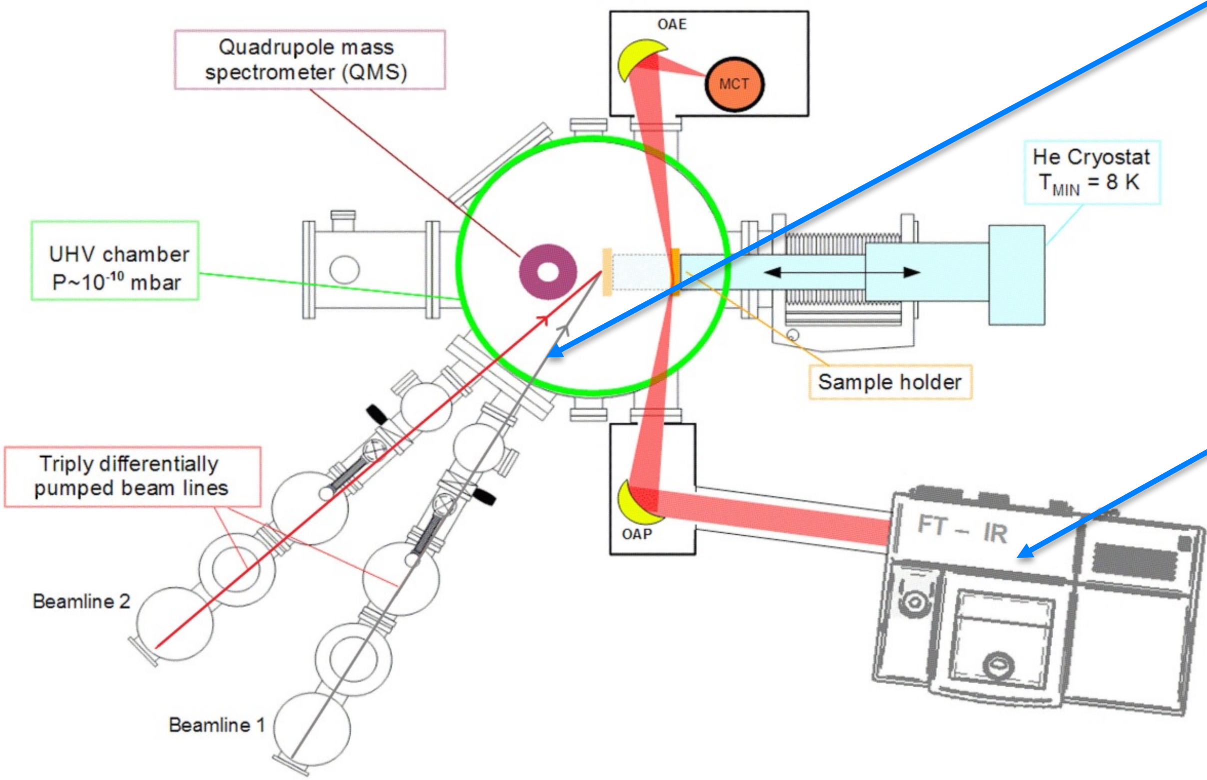
Marco Minissale

What we measure?	- Reaction rates - Thermal and non-thermal desorption rates - Diffusion constants - *	

*** Non-exhaustive list**

Ingredients: what we need?	- Flux of atom/molecule - Sample temperature - Flux of particles (photons, electrons, ions, …) - …*	Calibration and systematic errors
How we measure?	- Thermocouples - Mass and Infrared spectroscopy - …*	Instrumental errors (i.e. Accuracy and sensitivity)
What we measure?	- Reaction rates - Thermal and non-thermal desorption rates - Diffusion constants - *	Model uncertainty – Which parameters and physical-chemical processes?

* Non-exhaustive list



Calibration and systematic errors

Instrumental errors (i.e. Accuracy and sensitivity)

Model uncertainty – Which parameters and physical-chemical processes?

Experimental uncertainty

Analysis and interpretation of results

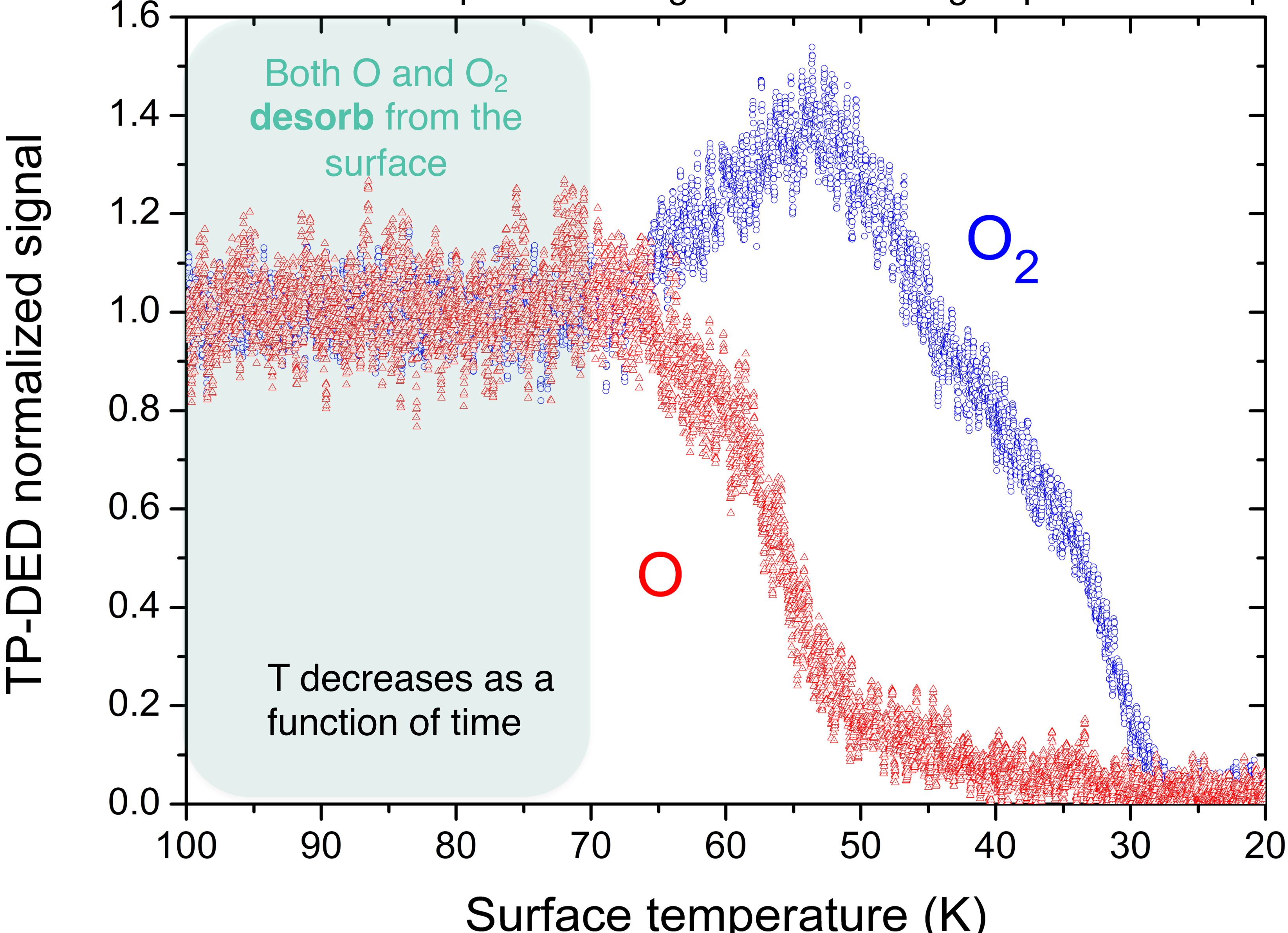
FORMOLISM setup, CY Cergy-Paris University

Two examples:

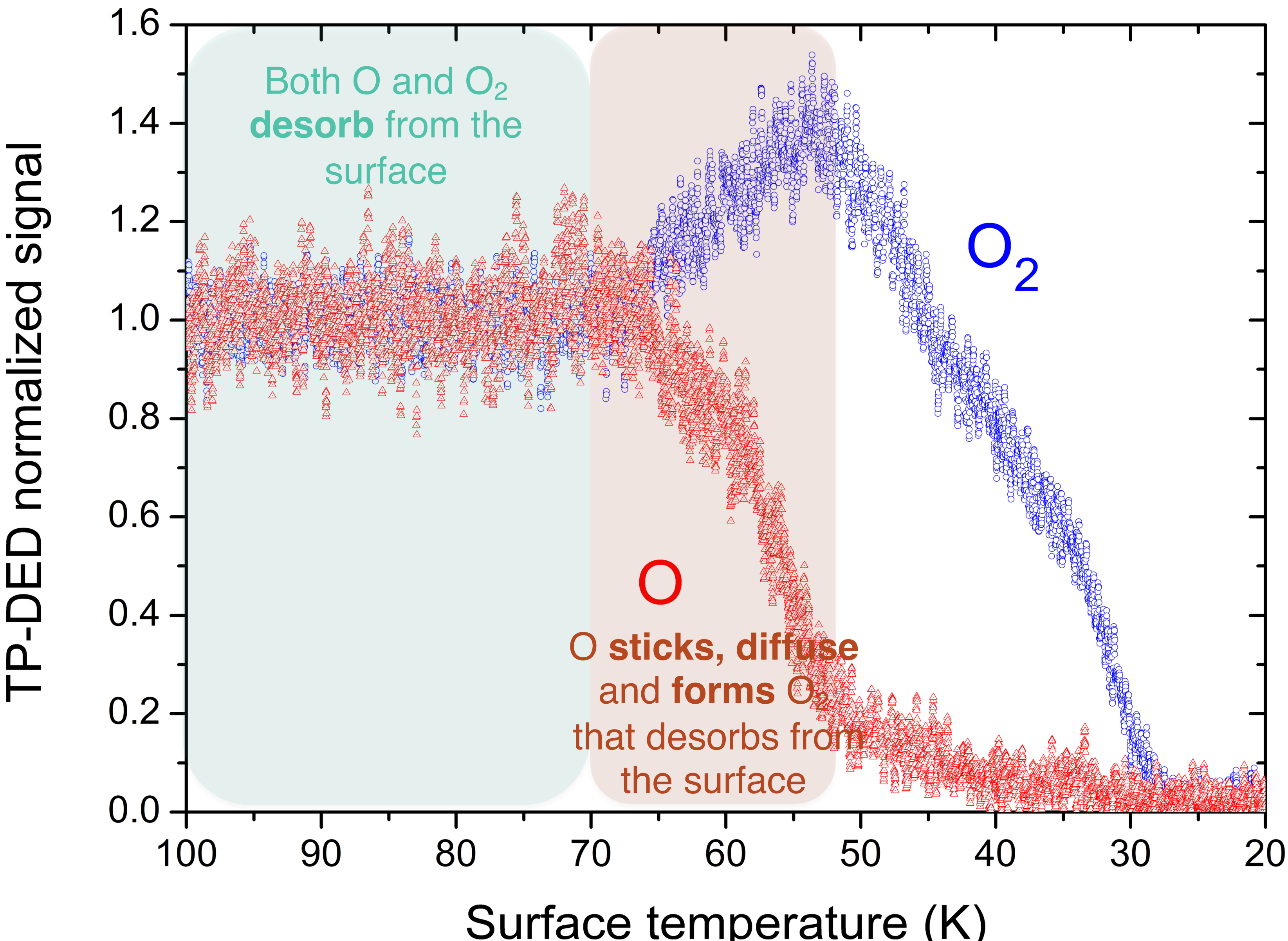
-Diffusion/desorption of oxygen atoms on cold surfaces

- Reaction on solid-phase: H_2CO+O

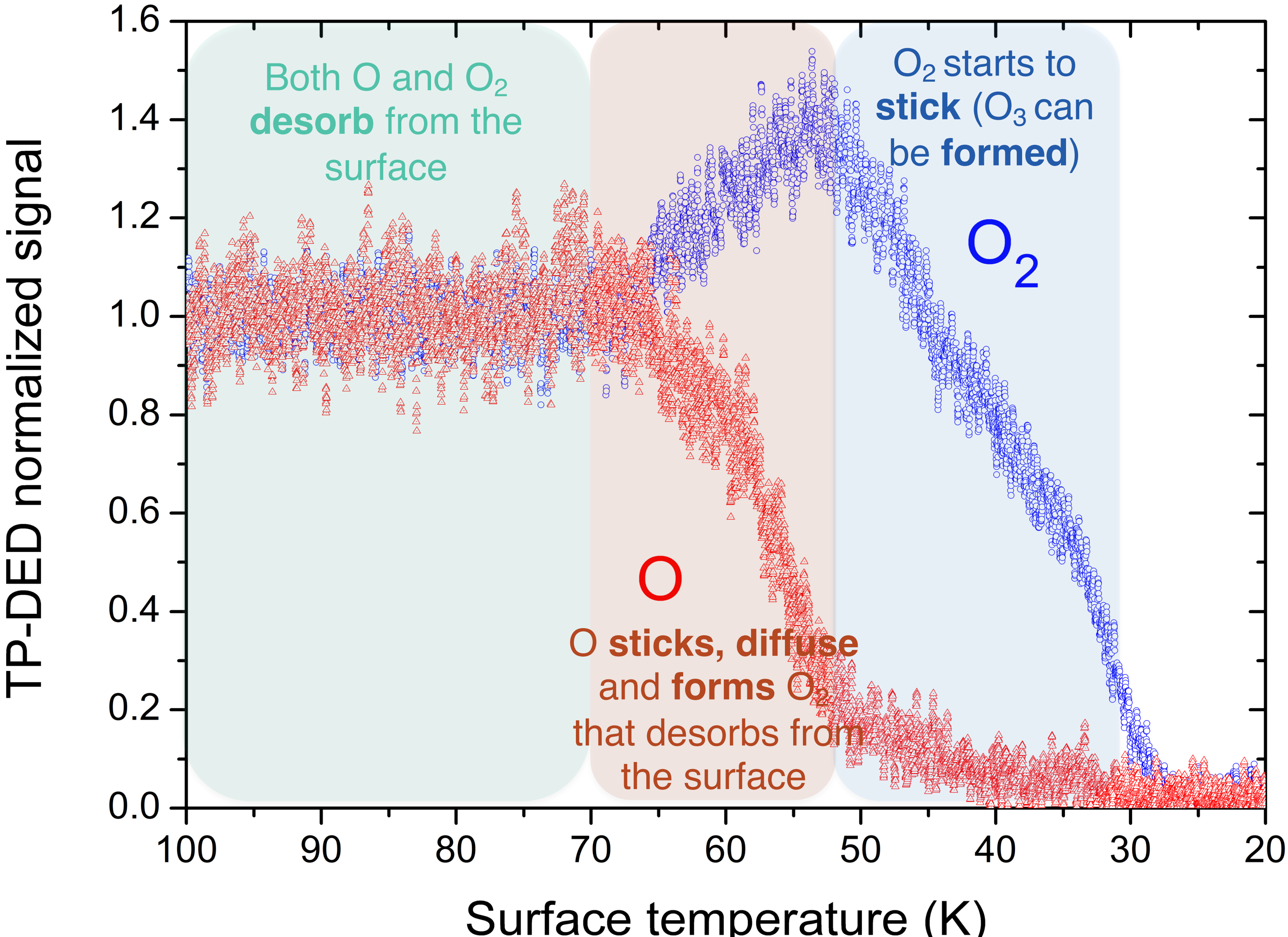
TP-DED: Temperature Programmed –During Exposure Desorption



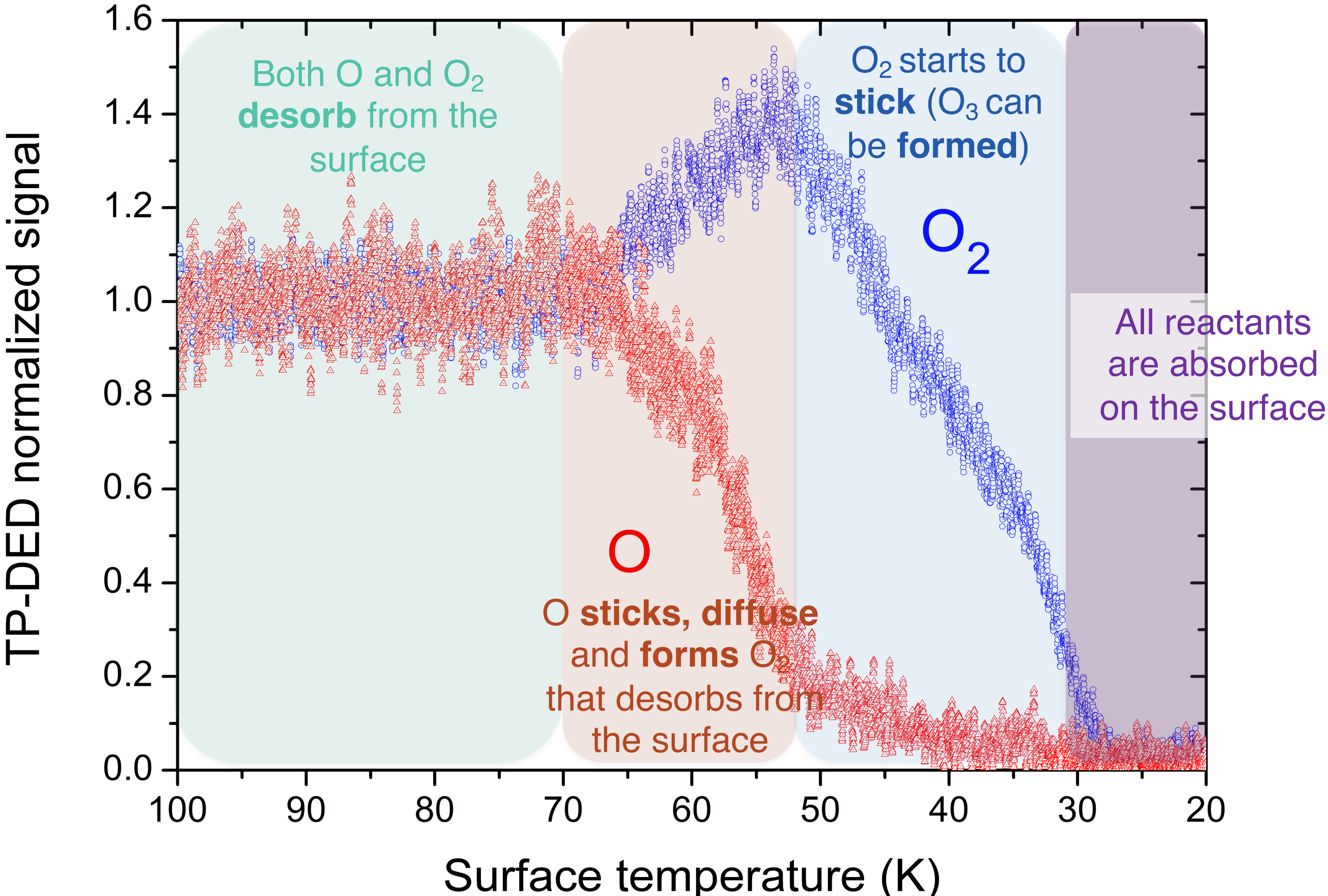
From *Minissale, Congiu & Dulieu A&A 2016*



From *Minissale, Congiu & Dulieu A&A 2016*

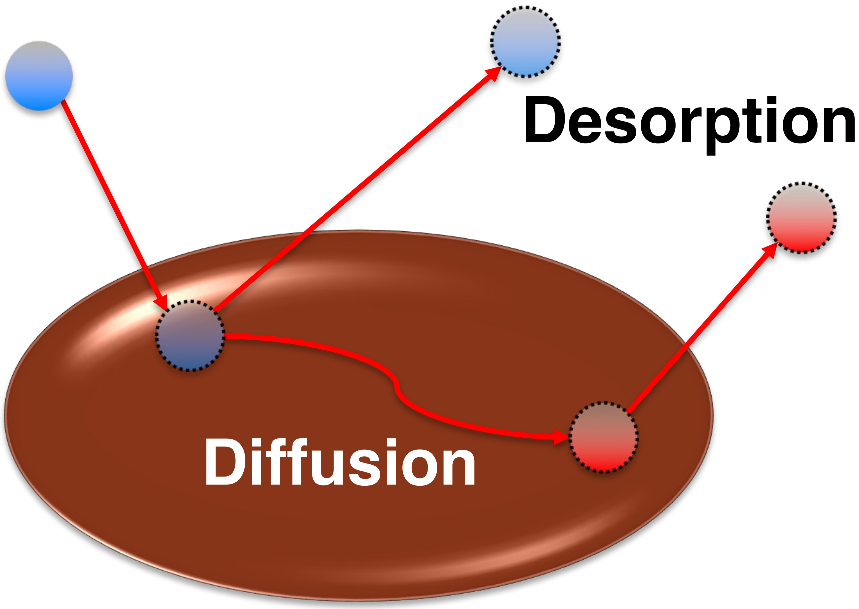


From *Minissale, Congiu & Dulieu A&A 2016*



From *Minissale, Congiu & Dulieu A&A 2016*

Modeling diffusion/desorption kinetics



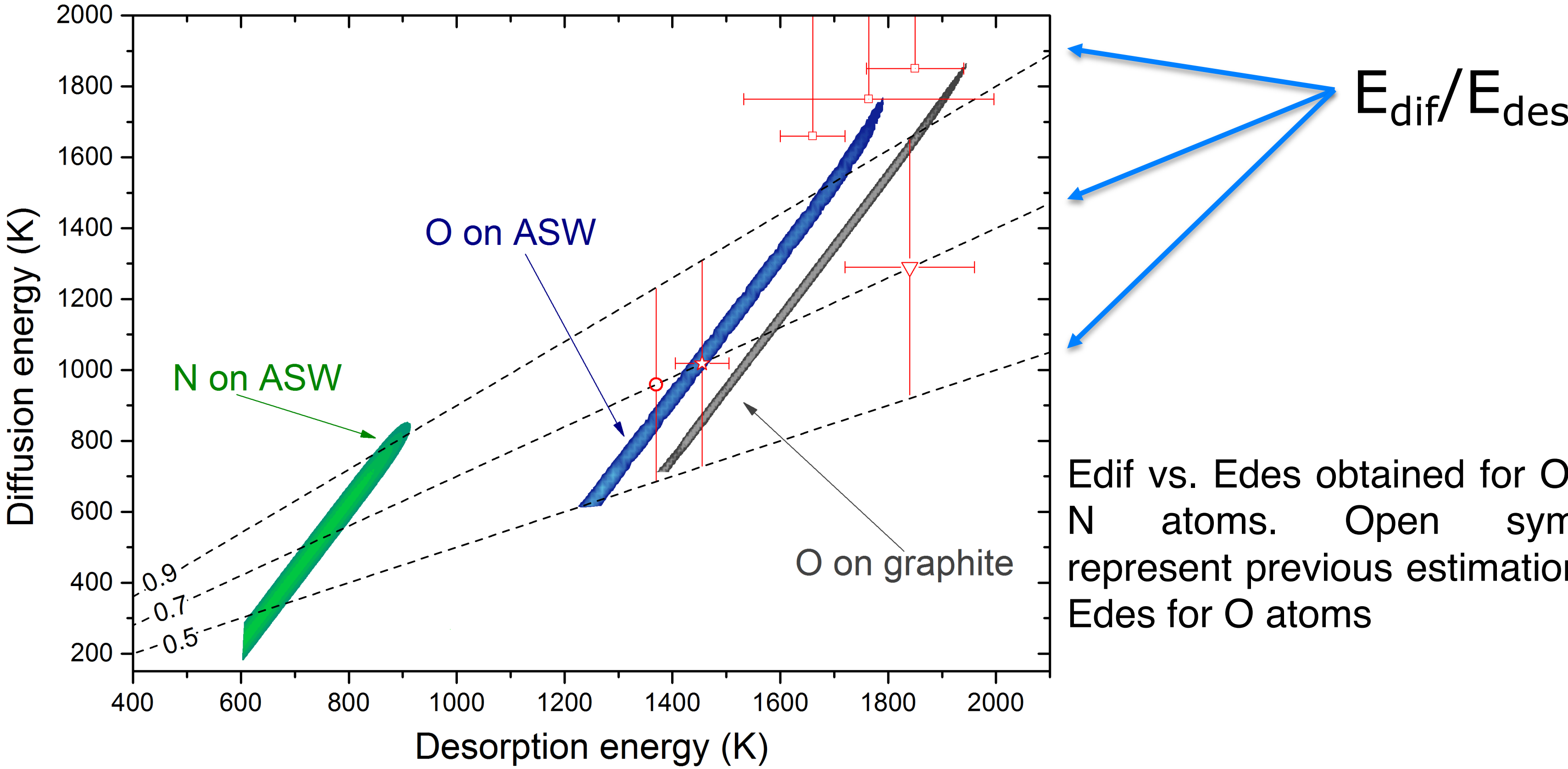
Experimental parameter

$$\frac{dX}{dt} = \phi_X - X k_{X-des} - R(X, k_{X-dif})$$

Flux | Desorption rate | Reaction rate

Fitting TP-DED experiments, we can find the couple E_{diff}-E_{des}

Modeling diffusion/desorption kinetics



Edif vs. Edes obtained for O and N atoms. Open symbols represent previous estimations of Edes for O atoms

Which ratio is the most appropriate?

We need **input from other experiments** to evaluate the ratio

From *Minissale, Congiu & Dulieu MNRAS 2016*

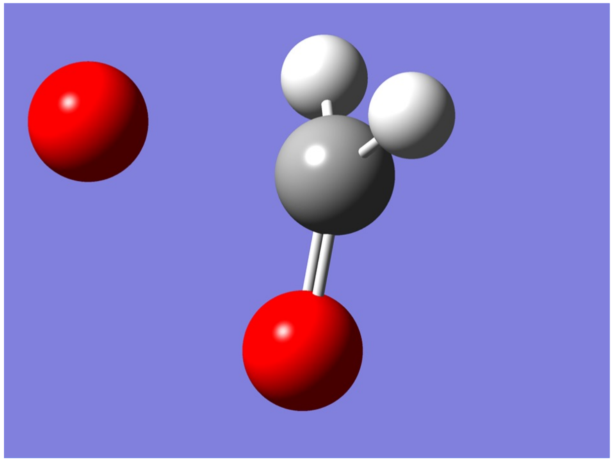
H_2CO+O reaction in solid-phase

In gas phase H_2CO+O → CO_2+H_2

In gas phase H_2CO+O → H_2O+CO

Chang & Barker (1979)
Wellman et al.(1991)

Activation barrier= 1560 K

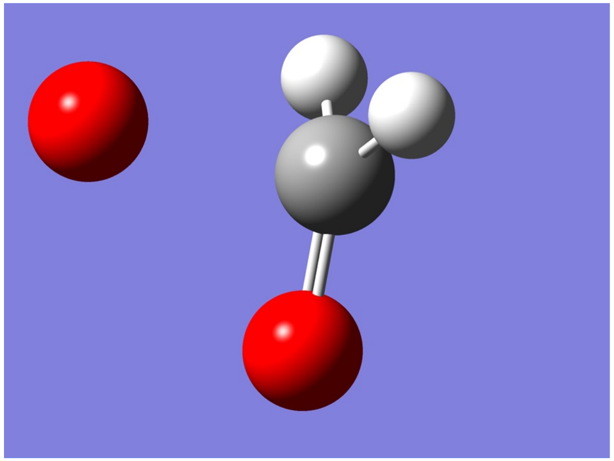


H_2CO+O reaction in solid-phase

In gas phase H_2CO+O → CO_2+H_2
→ H_2O+CO

Chang & Barker (1979)
Wellman et al.(1991)

Activation barrier= 1560 K

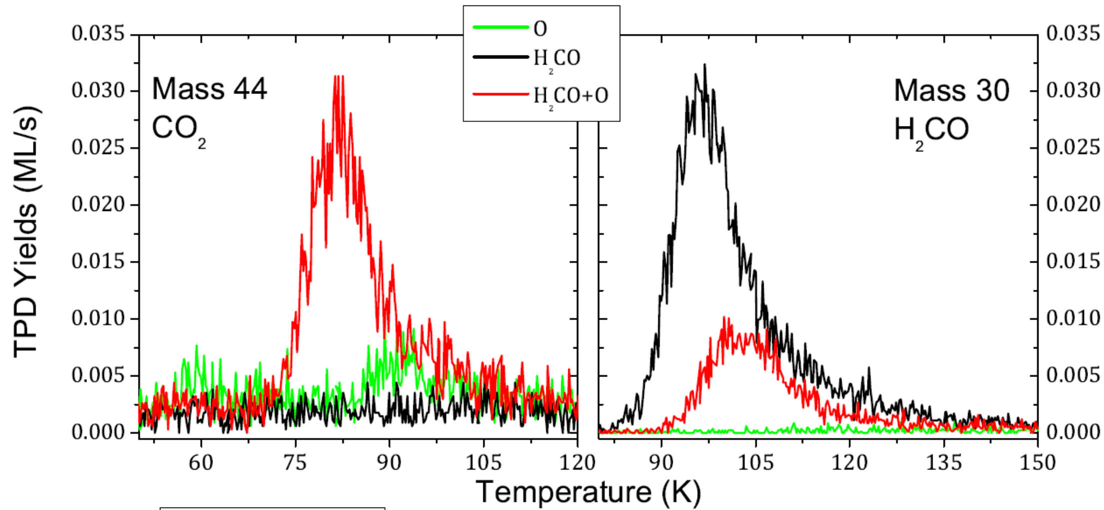


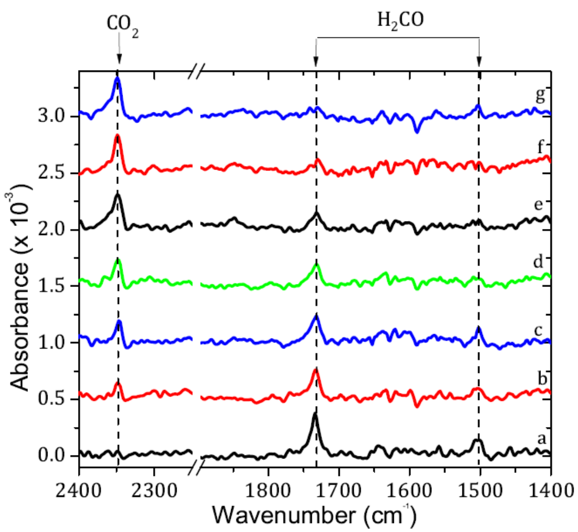
The presence of oxygen atoms makes harder the estimation of activation barrier in solid-phase →

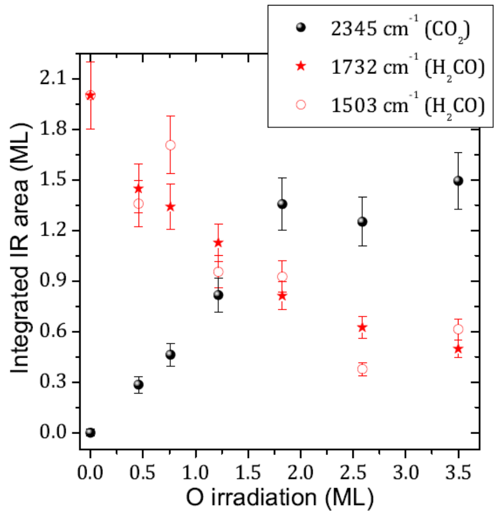
- The O atoms are mixed with O_2 molecules.
- O atoms diffuse quite fast on surface
- O atoms can react each other

H_2CO+O reaction in solid-phase

Temperature programmed desorption







Infrared spectoscopy results @55K

Minissale et al., A&A 2015

H_2CO+O reaction in solid-phase

$$O'(t) = 2\tau\, \phi_{O_2 off}\,(1 - 2O - O_2) - (1-\tau)\,\phi_{O_2 off}\, O - r_{aER} 2\tau \phi_{O_2 off} H_2CO - O\, r_{Deso_O}$$

$$O_2'(t) = (1-\tau)\,\phi_{O_2 off}\,(1 - O\,(1-\epsilon)) - 2\tau\,\phi_{O_2 off}\, O_2 + 2\tau\,(1-\epsilon)\,\phi_{O_2 off}\, O - O_2\, r_{Deso_{O_2}}$$

$$O_3'(t) = (1-\tau)\,\phi_{O_2 off}\, O + 2\tau\,\phi_{O_2 off}\, O_2$$

$$H_2CO'(t) = -r_{aER}\, 2\tau\,\phi_{O_2 off}\, H_2CO$$

$$CO_2'(t) = r_{aER}\, 2\tau\,\phi_{O_2 off}\, H_2CO.$$

$$r_{aER} = e^{-\frac{E_a}{T_{eff}}}$$

$$r_{Deso_O} = \nu\, e^{\frac{-E_{O des}}{T}}$$

$$r_{Deso_{O_2}} = \nu\, e^{\frac{-E_{O_2 des}}{T}},$$

Eley-Rideal

$$O'(t) = -4\, k_{Odiff}\, O\, O - k_{Odiff}\, O\, O_2 - r_{aLH}\, k_{Odiff}\, OH_2CO - O\, r_{Deso_O}$$

$$O_2'(t) = 2\, k_{Odiff}\, O\, O\, \epsilon - k_{Odiff}\, O\, O_2 - O_2\, r_{Deso_{O_2}}$$

$$O_3'(t) = k_{Odiff}\, O\, O_2$$

$$H_2CO'(t) = -r_{aLH}\, k_{Odiff}\, O\, H_2CO$$

$$CO_2'(t) = r_{aLH}\, k_{Odiff}\, O\, H_2CO,$$

where

$$k_{Odiff} = \nu\, e^{\frac{-E_{Odiff}}{T}}$$

$$r_{aLH} = e^{\frac{-E_a}{T}}$$

Langmuir-Hinshelwood

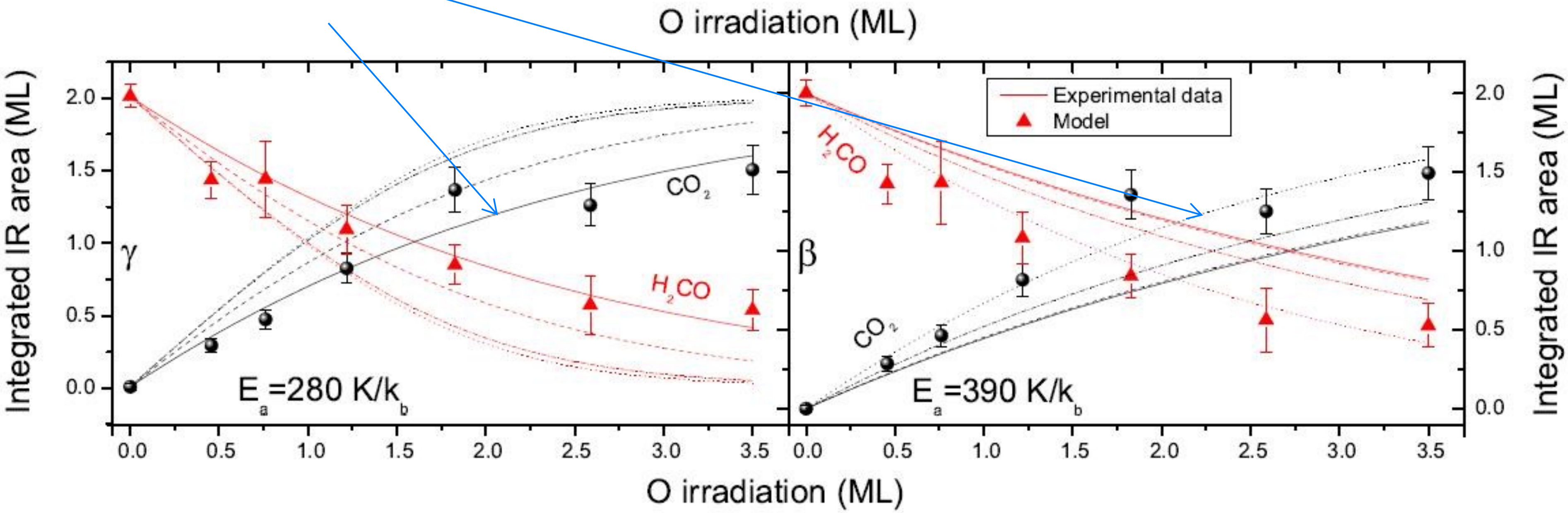
Source of uncertainty: fluxes of O and O_2 (ϕ), H_2CO initial coverage, chemical desorption (ε), desorption parameters (ν, E_{des}), diffusion constant (E_{diff})

→ Activation energy

Simulation of experimental results

H_2CO+O reaction in solid-phase

- Two coupled parameters: H_2CO+O barrier and O diffusion
- The pure thermal O diffusion estimated to be between 900 and 600 K

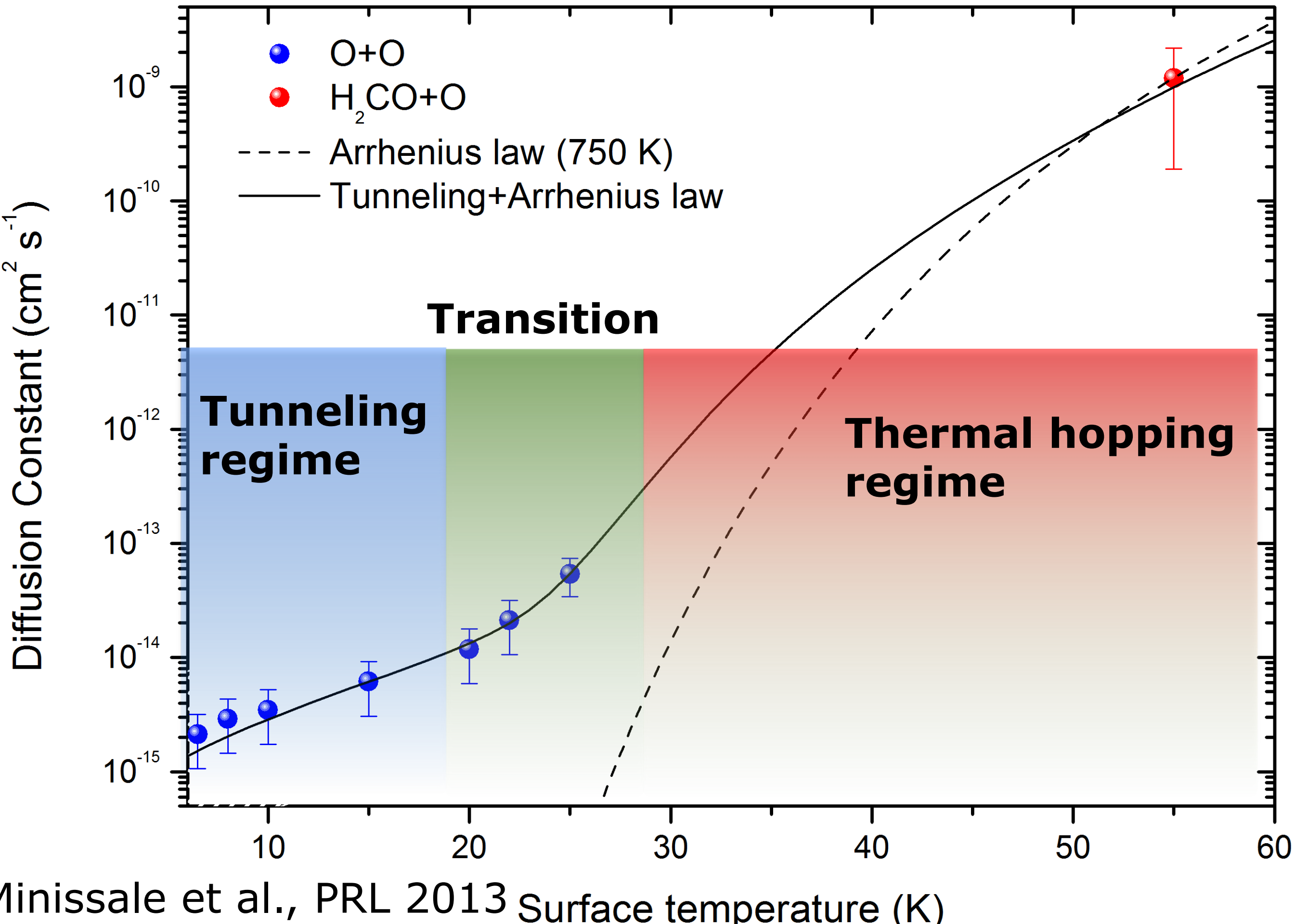


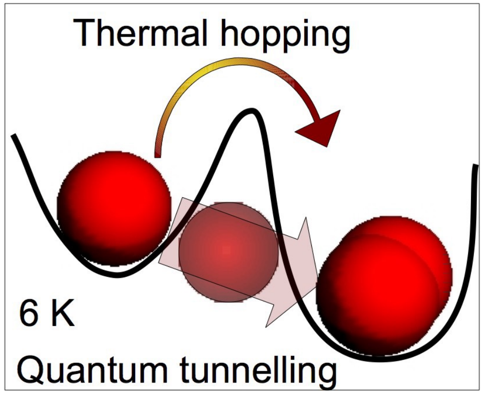
If O diffusion 750 K,
H_2CO+O barrier=335 K

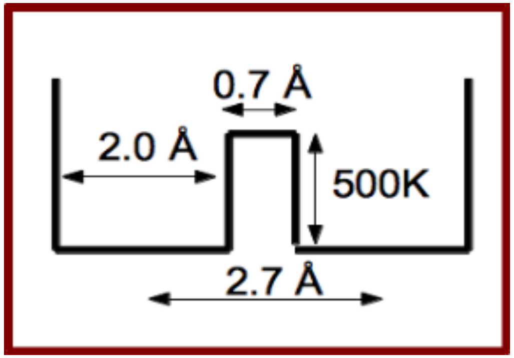
Exp uncertainty ~ 10%
"Model" uncertainty ~ 30%

Minissale et al., A&A 2015

Diffusion: oxygen atoms







O atoms start to desorb

Minissale et al., PRL 2013
Minissale et al., A&A 2015

Ingredients: what we need?	**Experimental and systematic errors**
How we measure?	**Instrumental errors**
What we measure?	**Model uncertainty – which physical-chemical processes?**

Experimental uncertainty can be “**easily**” taken into account and **estimated or reduced**

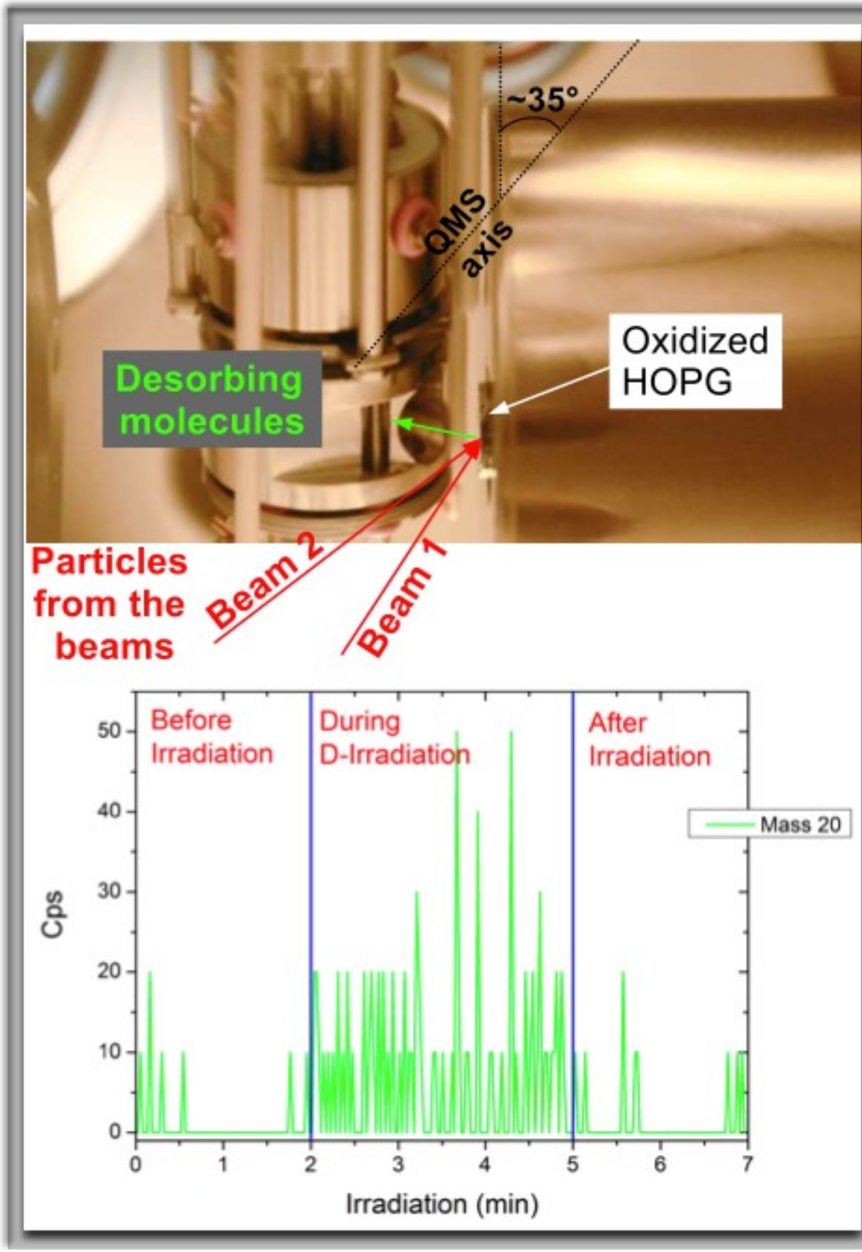
Often the main source of uncertainty comes from **coupled parameters / coupled processes**

*** Non-exhaustive list**

Desorption induced by chemistry

DED
(During Exposure Desorption)

O_2 ices exposed to D atoms

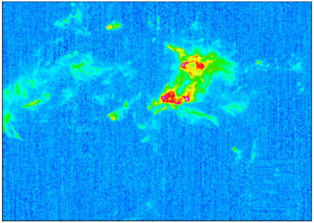


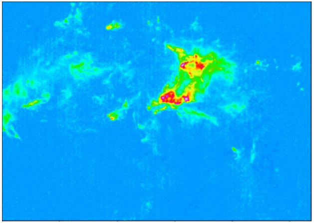
Thank you for your attention

Signal extraction from noisy line cubes

The problem of applying hyperspectral imaging methods

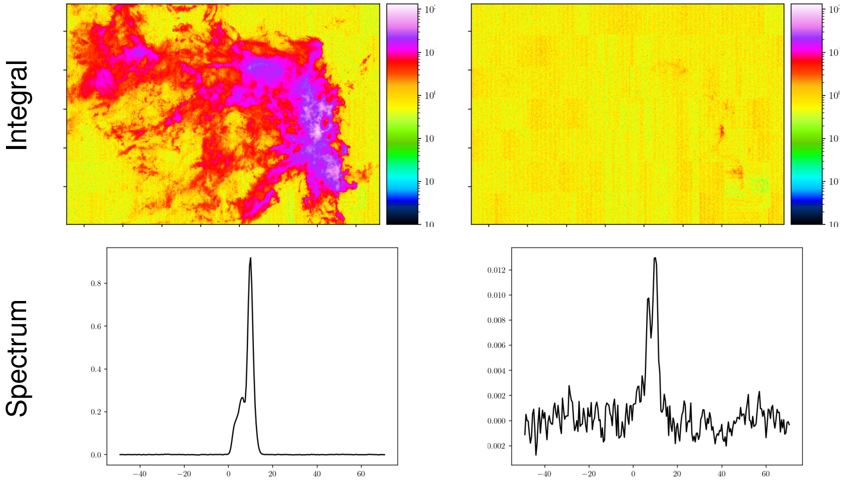
Lucas Einig



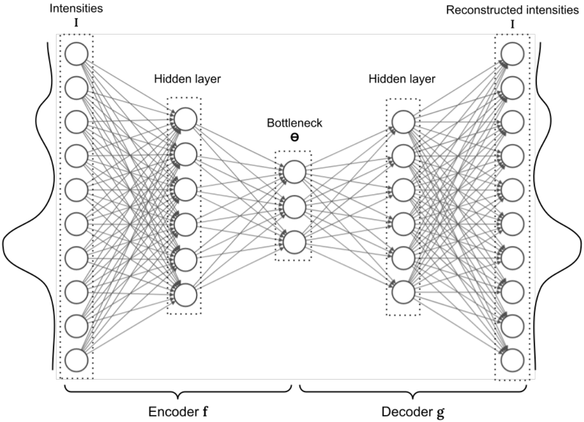


ORION-B dataset

^{13}CO (1-0) line | $C^{17}O$ (1-0)



Low rank asumption based methods

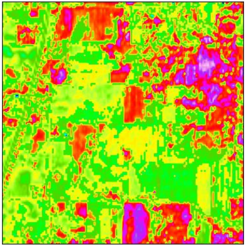


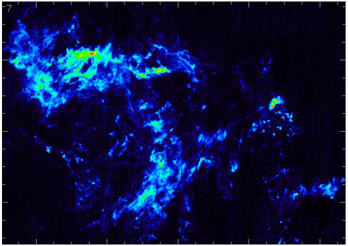
$$\hat{f},\ \hat{g} = \arg\min_{f,g} \|\mathrm{I} - g(\Theta)\|_2^2 \quad \text{s.t.} \quad \Theta = f(\mathrm{I}) \text{ and } \dim\Theta \ll \dim\mathrm{I}$$

Redundancy between channels

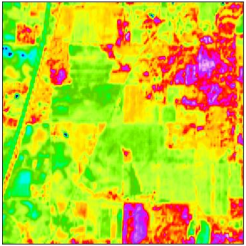
Hyperspec. data Indian Pines | ^{13}CO (1-0) line

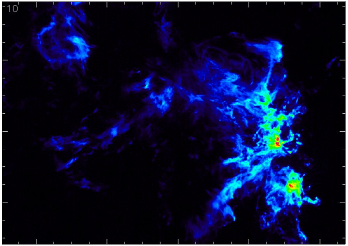
Example channel 1





Example channel 2





Intrinsic dimension estimation

We propose to estimate the intrinsic dimension of a dataset using the well known “elbow method” in a non-linear framework.

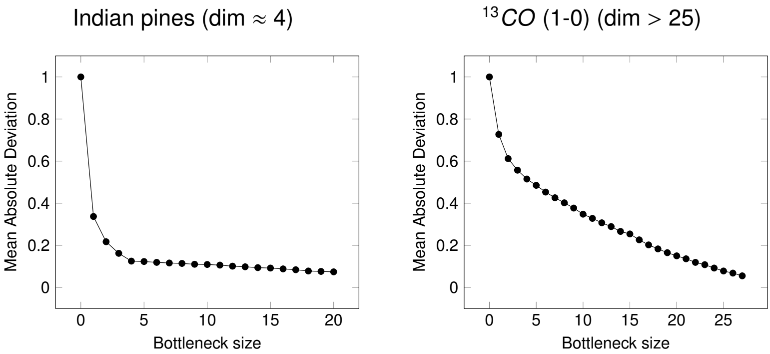


Figure: Mean absolute deviation between input and reconstructed data.

Limitations of low rank methods

Limitations

- The methods based on a low rank assumption are very suitable for continuum cubes but more limited for line cubes.
- The higher the intrinsic dimension, the lower the redundancy and the more complex the signal extraction.

Concerned methods

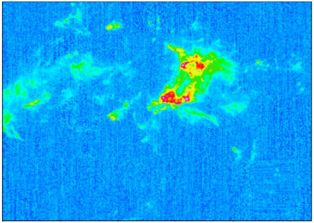
These conclusions apply to any method based on a low rank assumption, including

- Principal Component Analysis (PCA).
- Autoencoder neural network (AE).
- Low rank tensor decomposition.

Improved neural network for molecular line cubes

Developed solutions

- Adapt the network architecture to the data
- Use prior knowledge



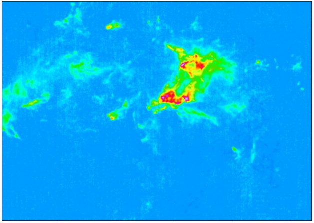
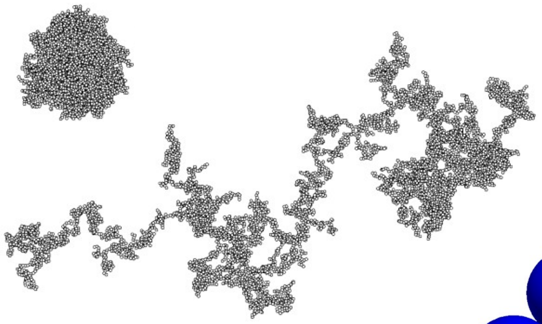
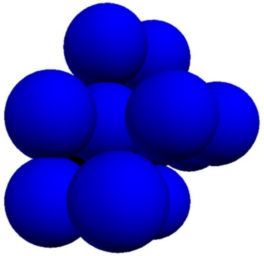


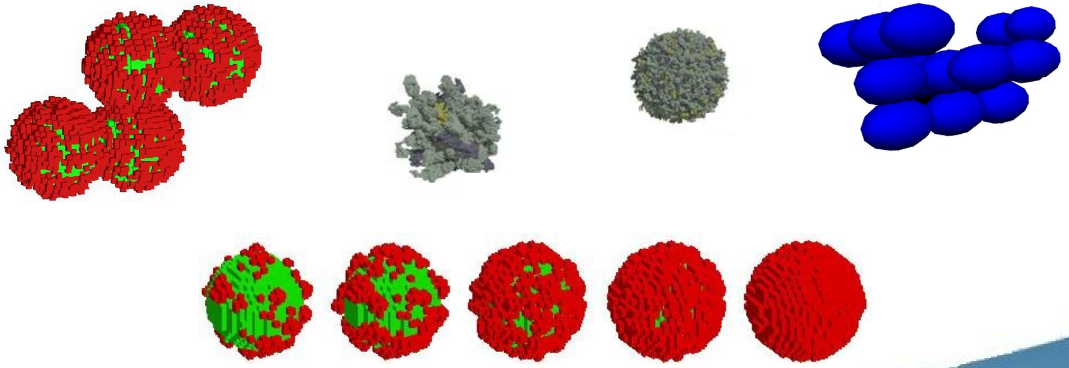
Figure: Example of noisy and denoised data with the **Local autoencoder with prior knowledge.**

(A very small part of the)
Uncertainties in the ISM grain models

N. Ysard (IAS, Orsay)







Mukai et al. (1992)
Koehler et al. (2011, 2012)
Min et al. (2016)
Ysard et al. (2018)

Basics of all dust models

- Chemical composition
  - → $m = n + ik$: from the lab ? Empirical ?
  - → composite grains ?
  - → inclusions, ice mantle ?
- Structure
  - → compact vs. porous
  - → core/mantle
  - → single grains vs. aggregates
  - → spheres vs. spheroids

Absorption efficiency $Q_{abs}(a,\lambda,T?)$
Scattering efficiency $Q_{sca}(a,\lambda)$
Scattering phase function $G(a,\lambda)$
Heat capacity $C(a, T)$

non-trivial step

- Size distribution
  - → a_{min}, a_{max}
  - → log-normal, power law, MRN, weird ?

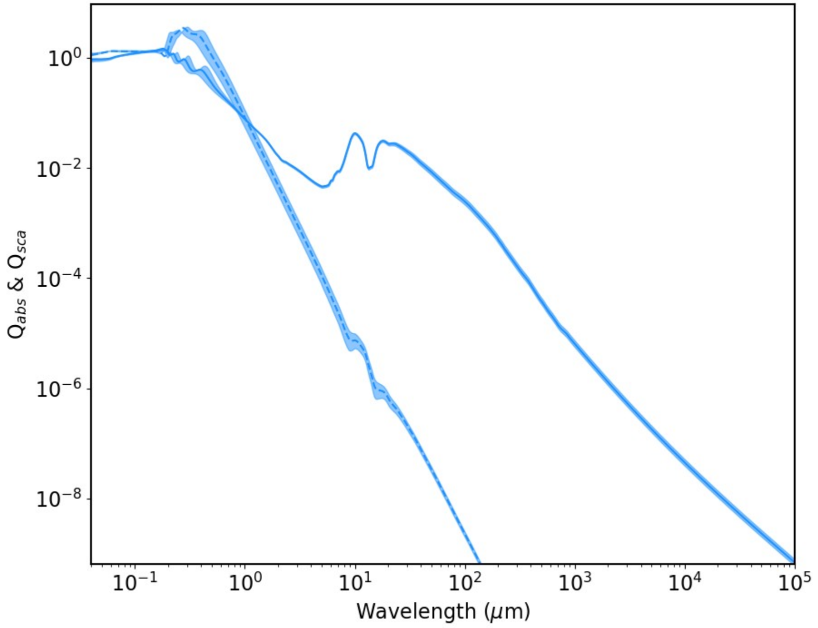
Calculations of the optical properties
Which model to choose ?

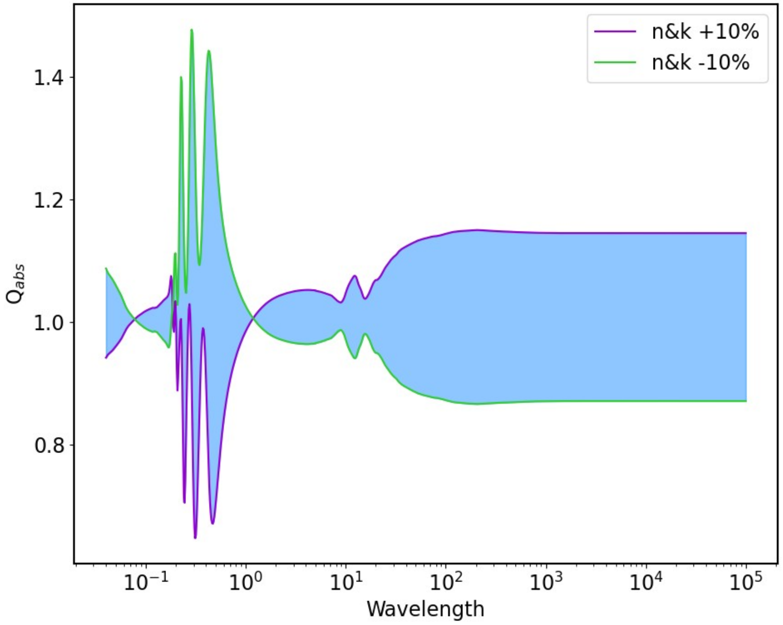
▪ Compact spherical grains Compact spherical grains with mantles	Mie: BHMIE BHCOAT *Bohren & Huffman (1983)*
▪ Porous grains Composite grains → random distribution	Effective Medium Theory EMT Maxwell Garnett or Bruggeman *Bohren & Huffman (1983)*
▪ Aggregates with one-point contact	T-MATRIX *Mischchenko (2000)*
▪ Aggregates with contact surface area Grains of any shape Composite grains → controlled distribution	Discrete Dipole Approximation DDA *Draine & Flatau (1994)*
▪ Spheroidal grains with or without mantles	DDA, T-MATRIX Analytic function in the Rayleigh limit Geometric limit in the UV *Bohren & Huffman (1983)*

Uncertainties in the optical constants
→ translation in the Q_{abs}

Let's assume that both n & k varies by +10 % or -10 %

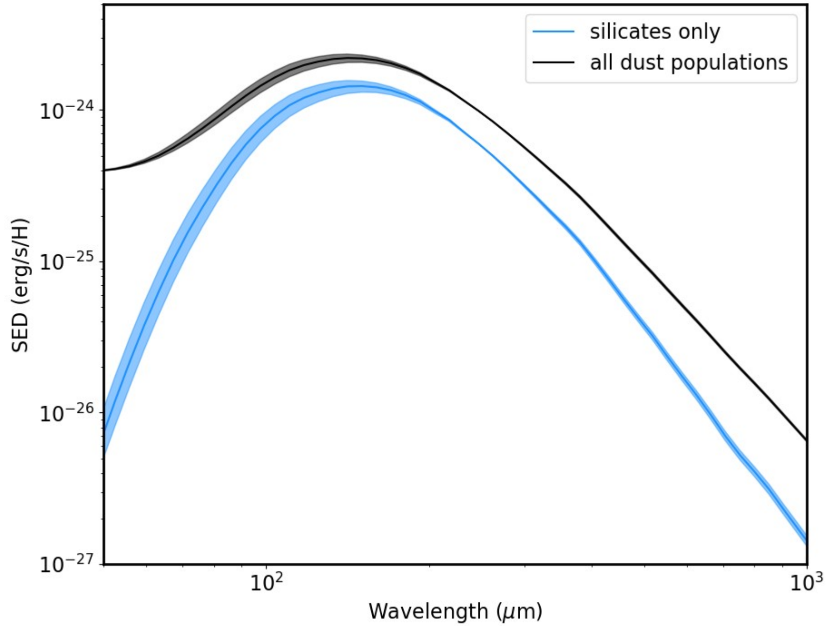
a = 0.1 µm

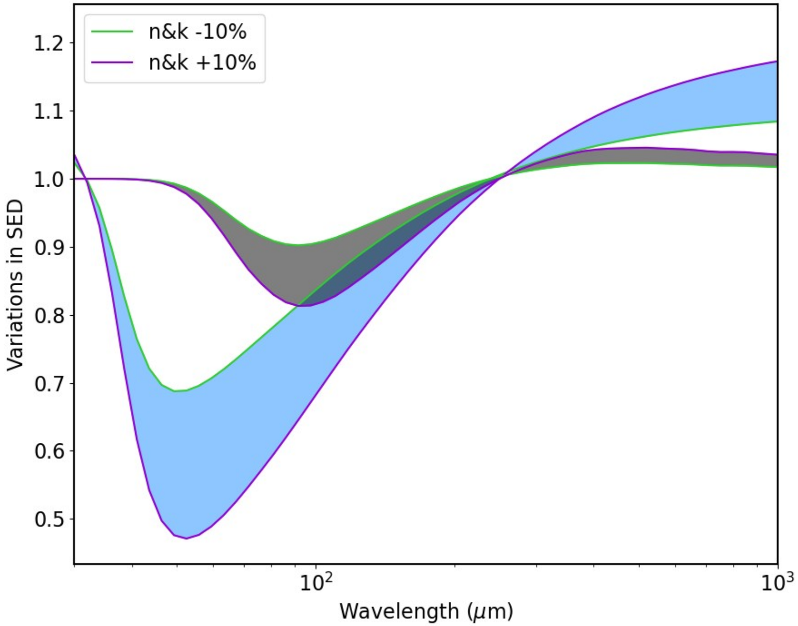




Uncertainties in the optical constants
→ translation in the SED

Let's assume that both n & k varies by +10 % or -10 %
→ silicates with a log-normal size distribution



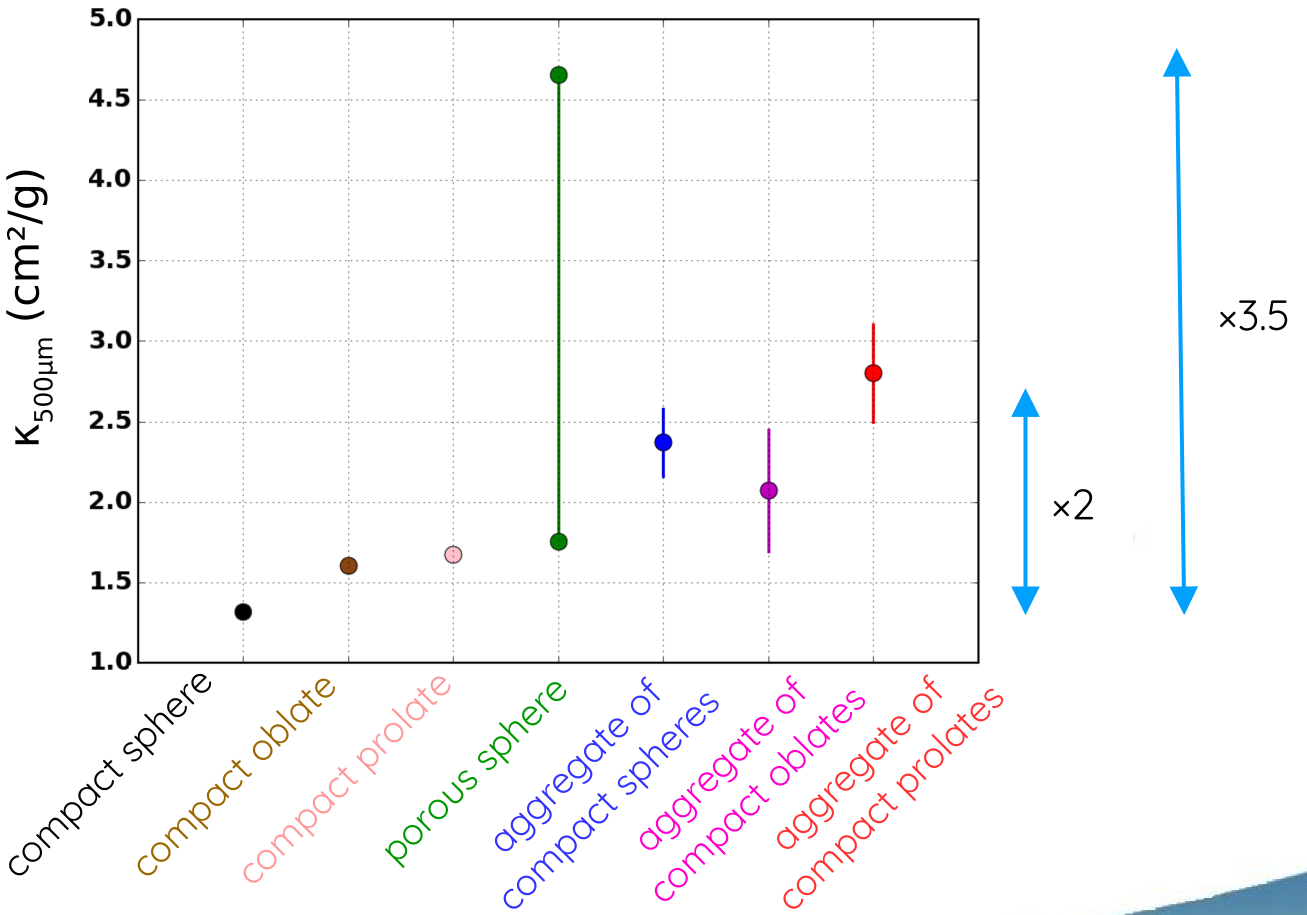


→ 10 % at long wavelength
→ more around the peak of the SED due to ≠ temperatures

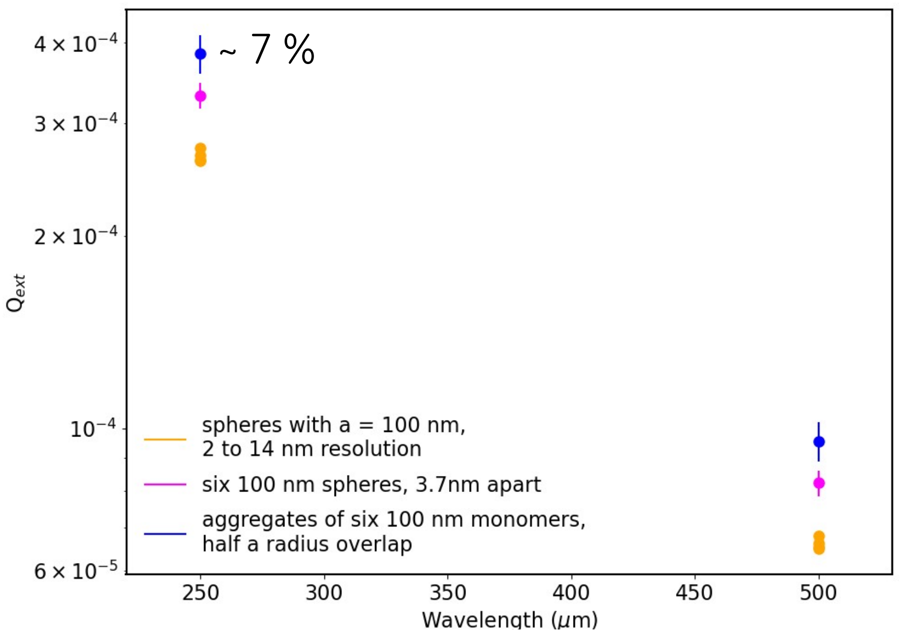
Structure of the grain

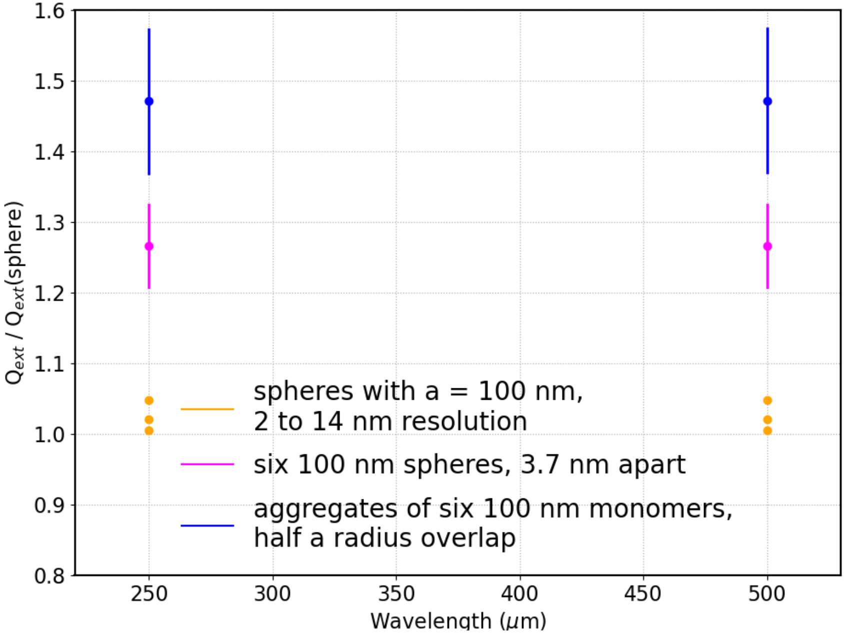
$$\kappa^{abs}\,[cm^2/g] = \frac{3}{4\rho}\frac{Q_{abs}}{a}$$

100% silicate
mass equivalent size of 0.1 µm
[no size distribution here]



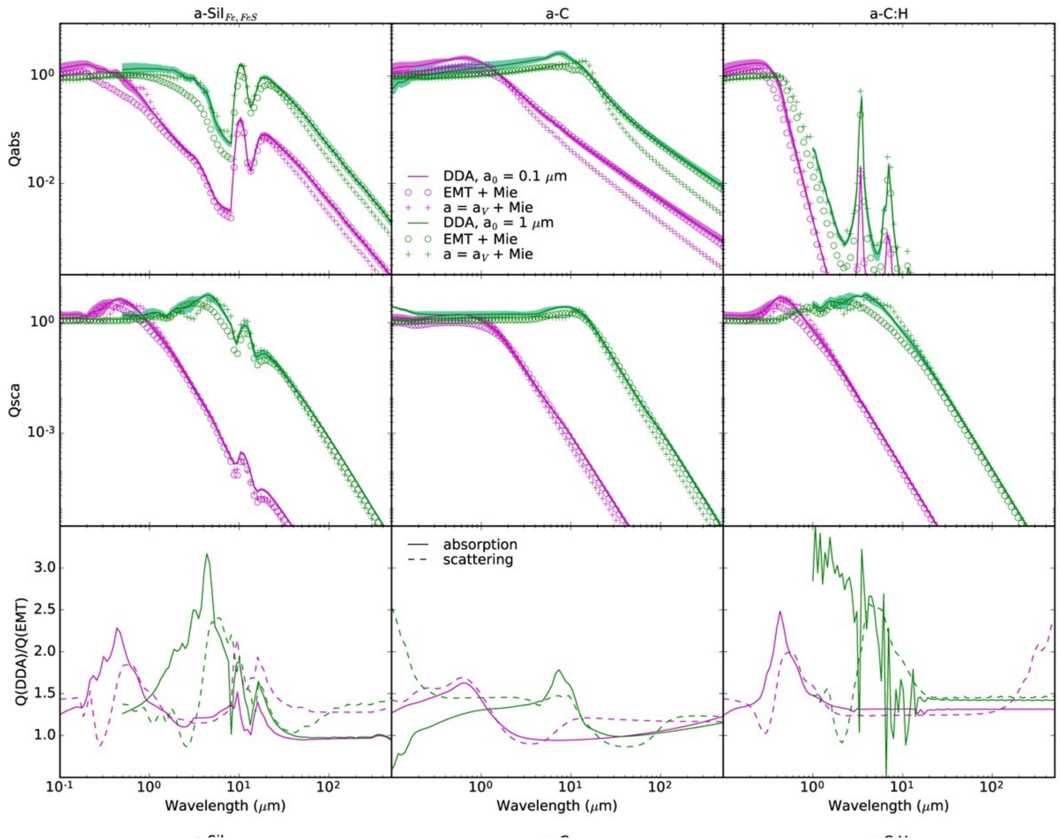
Description of the grain surface
→ completely smooth vs. irregular





→ single grains : increase by ~ 5 % for highly irregular surface
→ aggregates : increase by ~ 20 % for large contact area

Calculations of the optical properties

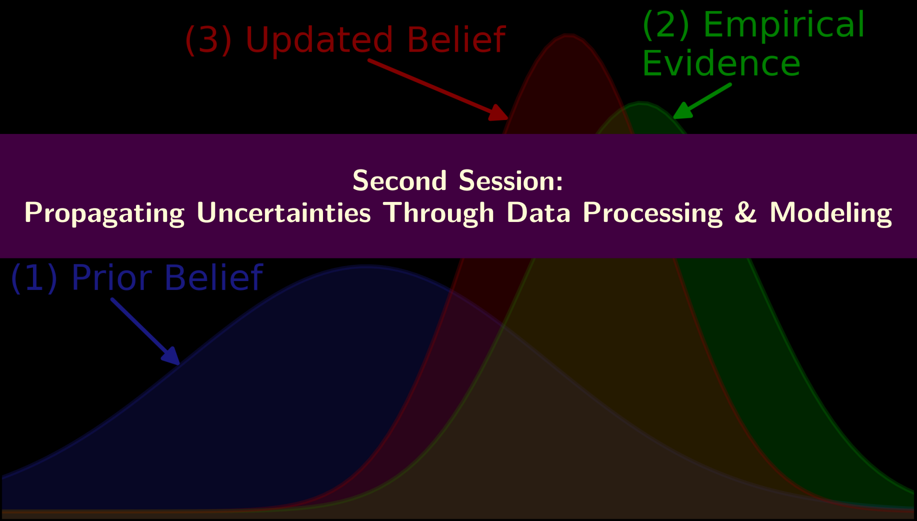


- Aggregates of 8 monomers
  monomer → 0.1 and 1 μm compact sphere
- Three types of calculations
  DDA → « exact »
  Mie for a sphere of equivalent mass
  EMT+Mie with a = $\mathcal{R}_g$ and $\mathcal{P}_{equivalent}$
- Significant differences
  → different grain temperatures
  → shifted SEDs
  → mid-IR silicate features ≠ size estimates

First Session Wrap-Up: What Can We Be Certain About?

- Should we advocate that the ability to precisely estimate the uncertainties must be taken into account in the design of new experiments & new telescopes?
- Can we use the scatter resulting from comparing different models as a way to quantify the absolute uncertainty on our hypotheses?
- Could machines learn estimating uncertainties?

Second Session:
Propagating Uncertainties Through Data Processing & Modeling

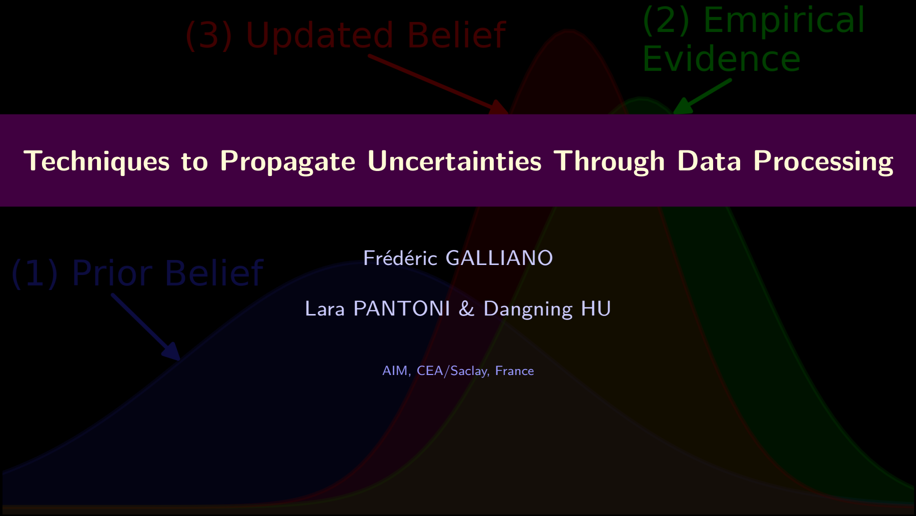


Techniques to Propagate Uncertainties Through Data Processing

Frédéric GALLIANO

Lara PANTONI & Dangning HU

AIM, CEA/Saclay, France



Propagating Uncertainties Through Complex, Non-Linear Operations

Propagating Uncertainties Through Complex, Non-Linear Operations

Multi-wavelength image homogenization

Given a collection of multi-λ images $\Rightarrow$ propagate the original uncertainties through:

Propagating Uncertainties Through Complex, Non-Linear Operations

Multi-wavelength image homogenization

Given a collection of multi-λ images $\Rightarrow$ propagate the original uncertainties through:

1. Complex background subtraction $\Rightarrow$ stars, cirrus, CIB, *etc.*;

Multi-wavelength image homogenization

Given a collection of multi-λ images $\Rightarrow$ propagate the original uncertainties through:

1. Complex background subtraction $\Rightarrow$ stars, cirrus, CIB, *etc.*;
2. Degradation to a common resolution $\Rightarrow$ kernel convolution;

Propagating Uncertainties Through Complex, Non-Linear Operations

Multi-wavelength image homogenization

Given a collection of multi-λ images $\Rightarrow$ propagate the original uncertainties through:

1. Complex background subtraction $\Rightarrow$ stars, cirrus, CIB, *etc.*;
2. Degradation to a common resolution $\Rightarrow$ kernel convolution;
3. Reprojection on a common grid $\Rightarrow$ resampling.

Propagating Uncertainties Through Complex, Non-Linear Operations

Multi-wavelength image homogenization

Given a collection of multi-λ images $\Rightarrow$ propagate the original uncertainties through:

1. Complex background subtraction $\Rightarrow$ stars, cirrus, CIB, *etc.*;
2. Degradation to a common resolution $\Rightarrow$ kernel convolution;
3. Reprojection on a common grid $\Rightarrow$ resampling.

Spectral analysis

Spectral cube analysis $\Rightarrow$ propagate the original uncertainties through:

Propagating Uncertainties Through Complex, Non-Linear Operations

Multi-wavelength image homogenization

Given a collection of multi-λ images $\Rightarrow$ propagate the original uncertainties through:

1. Complex background subtraction $\Rightarrow$ stars, cirrus, CIB, *etc.*;
2. Degradation to a common resolution $\Rightarrow$ kernel convolution;
3. Reprojection on a common grid $\Rightarrow$ resampling.

Spectral analysis

Spectral cube analysis $\Rightarrow$ propagate the original uncertainties through:

1. Degradation to a common resolution $\Rightarrow$ kernel convolution;

Propagating Uncertainties Through Complex, Non-Linear Operations

Multi-wavelength image homogenization

Given a collection of multi-λ images $\Rightarrow$ propagate the original uncertainties through:

1. Complex background subtraction $\Rightarrow$ stars, cirrus, CIB, *etc.*;
2. Degradation to a common resolution $\Rightarrow$ kernel convolution;
3. Reprojection on a common grid $\Rightarrow$ resampling.

Spectral analysis

Spectral cube analysis $\Rightarrow$ propagate the original uncertainties through:

1. Degradation to a common resolution $\Rightarrow$ kernel convolution;
2. Reprojection on a common grid $\Rightarrow$ resampling;

Propagating Uncertainties Through Complex, Non-Linear Operations

Multi-wavelength image homogenization

Given a collection of multi-λ images $\Rightarrow$ propagate the original uncertainties through:

1. Complex background subtraction $\Rightarrow$ stars, cirrus, CIB, *etc.*;
2. Degradation to a common resolution $\Rightarrow$ kernel convolution;
3. Reprojection on a common grid $\Rightarrow$ resampling.

Spectral analysis

Spectral cube analysis $\Rightarrow$ propagate the original uncertainties through:

1. Degradation to a common resolution $\Rightarrow$ kernel convolution;
2. Reprojection on a common grid $\Rightarrow$ resampling;
3. Line fitting $\Rightarrow$ flux extraction.

Propagating Uncertainties Through Complex, Non-Linear Operations

Multi-wavelength image homogenization

Given a collection of multi-λ images $\Rightarrow$ propagate the original uncertainties through:

1. Complex background subtraction $\Rightarrow$ stars, cirrus, CIB, *etc.*;
2. Degradation to a common resolution $\Rightarrow$ kernel convolution;
3. Reprojection on a common grid $\Rightarrow$ resampling.

Spectral analysis

Spectral cube analysis $\Rightarrow$ propagate the original uncertainties through:

1. Degradation to a common resolution $\Rightarrow$ kernel convolution;
2. Reprojection on a common grid $\Rightarrow$ resampling;
3. Line fitting $\Rightarrow$ flux extraction.

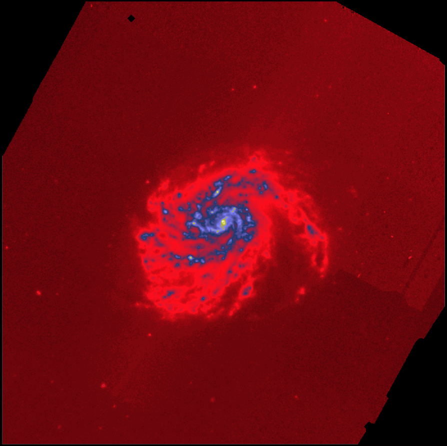
Consequences

These steps are necessary before modeling $\Rightarrow$ they change:

- noise level;
- its distribution;
- its correlation.

The Example of Multi-Wavelength Image Homogenization

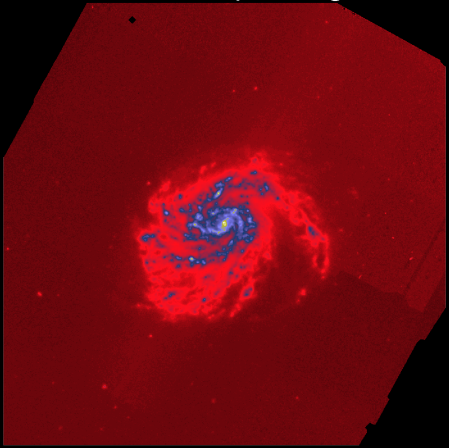
M 99 - IRAC 8 μm (original)

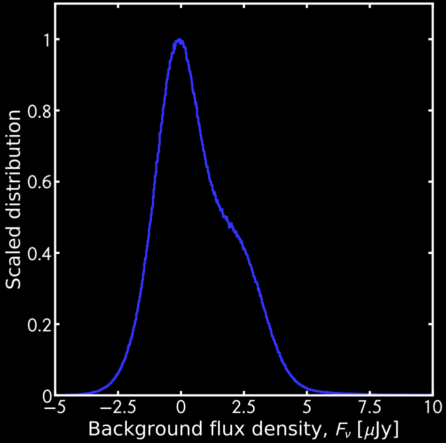


(Pantoni *et al.*, *in prep.*)

The Example of Multi-Wavelength Image Homogenization

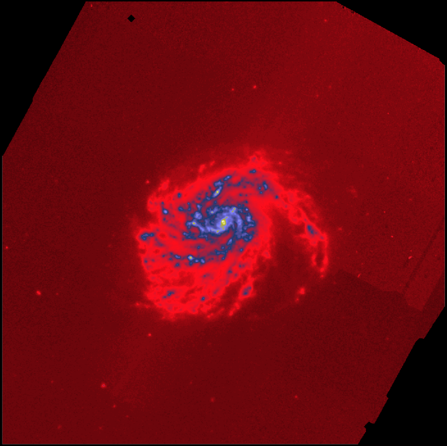
M 99 - IRAC 8 μm (original)



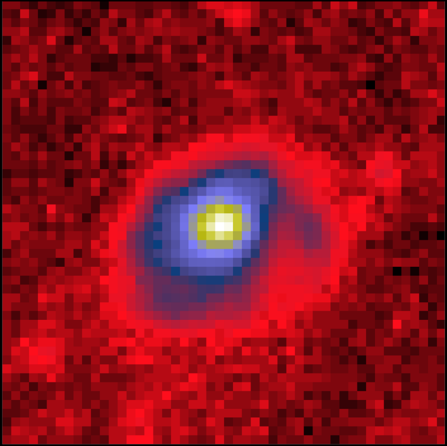


(Pantoni *et al.*, *in prep.*)

M 99 - IRAC 8 μm (original)

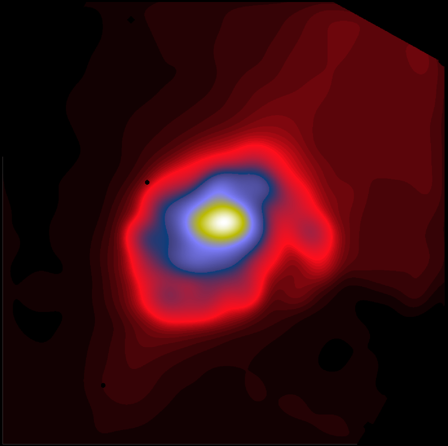


M 99 - SPIRE 500 μm (original)

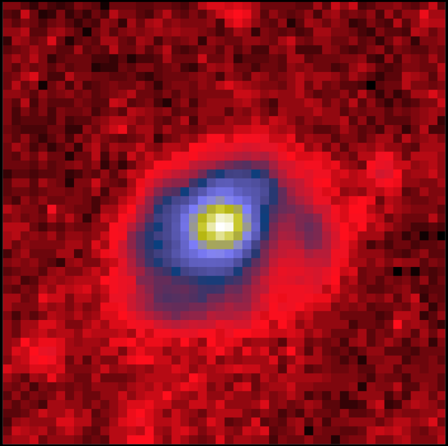


(Pantoni *et al.*, *in prep.*)

M 99 - IRAC 8 μm (convolved)



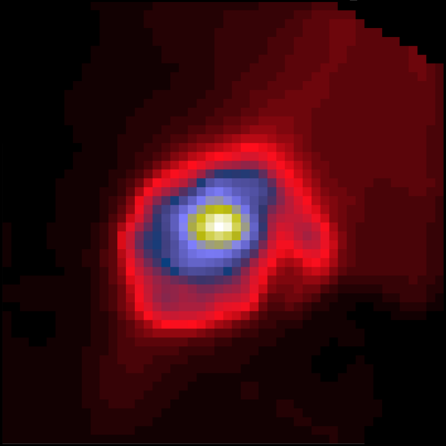
M 99 - SPIRE 500 μm (original)



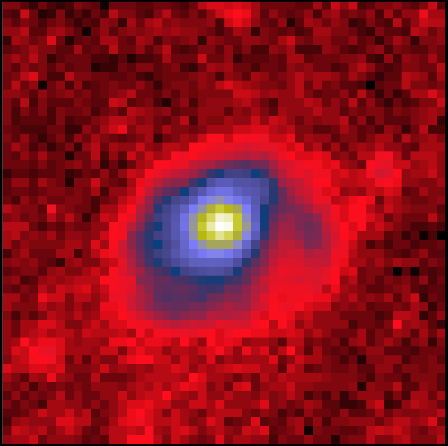
(Pantoni *et al.*, *in prep.*)

The Example of Multi-Wavelength Image Homogenization

M 99 - IRAC 8 μm (resampled)



M 99 - SPIRE 500 μm (original)



(Pantoni *et al.*, *in prep.*)

Methods to Propagate Uncertainties

Methods to Propagate Uncertainties

Analytic Work-Out

Possible to derive analytic expressions (*e.g.* Klein, 2021).

Analytic Work-Out

Possible to derive analytic expressions (*e.g.* Klein, 2021).

PRO:
- Quick to run;
- Elegant.

Methods to Propagate Uncertainties

Analytic Work-Out

Possible to derive analytic expressions (*e.g.* Klein, 2021).

PRO:
- Quick to run;
- Elegant.

CON:
- Requires some approximations: normal noise, Gaussian kernels, *etc.*;
- Can become excessively complex $\Rightarrow$ can not account for every effect.

Methods to Propagate Uncertainties

Analytic Work-Out

Possible to derive analytic expressions (*e.g.* Klein, 2021).

PRO:
- Quick to run;
- Elegant.

CON:
- Requires some approximations: normal noise, Gaussian kernels, *etc.*;
- Can become excessively complex $\Rightarrow$ can not account for every effect.

Monte-Carlo Bootstrapping (frequentist approach)

Consists in adding random perturbations to the data & looping over every process.

Methods to Propagate Uncertainties

Analytic Work-Out

Possible to derive analytic expressions (*e.g.* Klein, 2021).

PRO:
- Quick to run;
- Elegant.

CON:
- Requires some approximations: normal noise, Gaussian kernels, *etc.*;
- Can become excessively complex $\Rightarrow$ can not account for every effect.

Monte-Carlo Bootstrapping (frequentist approach)

Consists in adding random perturbations to the data & looping over every process.

PRO:
- Easy to implement;
- Can account for any effect;
- Accounts for the complexity of the noise: non-gaussianity, correlations, *etc.*

Methods to Propagate Uncertainties

Analytic Work-Out

Possible to derive analytic expressions (*e.g.* Klein, 2021).

PRO:
- Quick to run;
- Elegant.

CON:
- Requires some approximations: normal noise, Gaussian kernels, *etc.*;
- Can become excessively complex $\Rightarrow$ can not account for every effect.

Monte-Carlo Bootstrapping (frequentist approach)

Consists in adding random perturbations to the data & looping over every process.

PRO:
- Easy to implement;
- Can account for any effect;
- Accounts for the complexity of the noise: non-gaussianity, correlations, *etc.*

CON:
- Can be long to run ($\simeq 100\times$ the processing time);
- Not completely rigorous (not centered).

Methods to Propagate Uncertainties

Analytic Work-Out

Possible to derive analytic expressions (*e.g.* Klein, 2021).

PRO:
- Quick to run;
- Elegant.

CON:
- Requires some approximations: normal noise, Gaussian kernels, *etc.*;
- Can become excessively complex $\Rightarrow$ can not account for every effect.

Monte-Carlo Bootstrapping (frequentist approach)

Consists in adding random perturbations to the data & looping over every process.

PRO:
- Easy to implement;
- Can account for any effect;
- Accounts for the complexity of the noise: non-gaussianity, correlations, *etc.*

CON:
- Can be long to run ($\simeq 100\times$ the processing time);
- Not completely rigorous (not centered).

Bayesian Modeling

Methods to Propagate Uncertainties

Analytic Work-Out

Possible to derive analytic expressions (*e.g.* Klein, 2021).

PRO:
- Quick to run;
- Elegant.

CON:
- Requires some approximations: normal noise, Gaussian kernels, *etc.*;
- Can become excessively complex ⇒ can not account for every effect.

Monte-Carlo Bootstrapping (frequentist approach)

Consists in adding random perturbations to the data & looping over every process.

PRO:
- Easy to implement;
- Can account for any effect;
- Accounts for the complexity of the noise: non-gaussianity, correlations, *etc.*

CON:
- Can be long to run ($\simeq 100\times$ the processing time);
- Not completely rigorous (not centered).

Bayesian Modeling

PRO: The most rigorous method.

Methods to Propagate Uncertainties

Analytic Work-Out

Possible to derive analytic expressions (*e.g.* Klein, 2021).

PRO:
- Quick to run;
- Elegant.

CON:
- Requires some approximations: normal noise, Gaussian kernels, *etc.*;
- Can become excessively complex ⇒ can not account for every effect.

Monte-Carlo Bootstrapping (frequentist approach)

Consists in adding random perturbations to the data & looping over every process.

PRO:
- Easy to implement;
- Can account for any effect;
- Accounts for the complexity of the noise: non-gaussianity, correlations, *etc.*

CON:
- Can be long to run ($\simeq 100\times$ the processing time);
- Not completely rigorous (not centered).

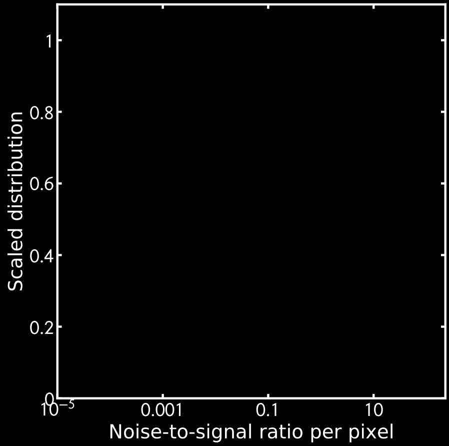
Bayesian Modeling

PRO: The most rigorous method.

CON: Requires a parametric model of the source morphology.

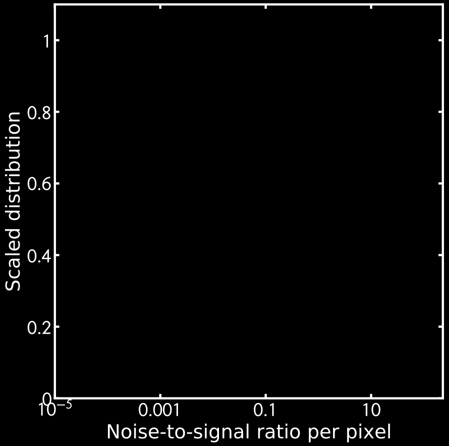
M 99 Noise Propagation: Pixel Statistics

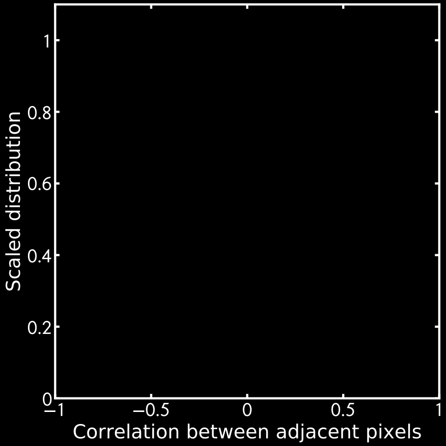
M 99 Noise Propagation: Pixel Statistics



(Pantoni *et al.*, *in prep.*)

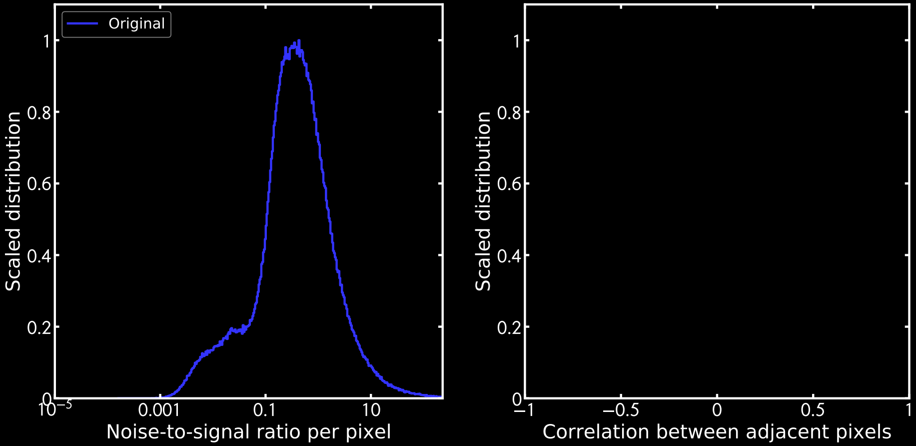
M 99 Noise Propagation: Pixel Statistics





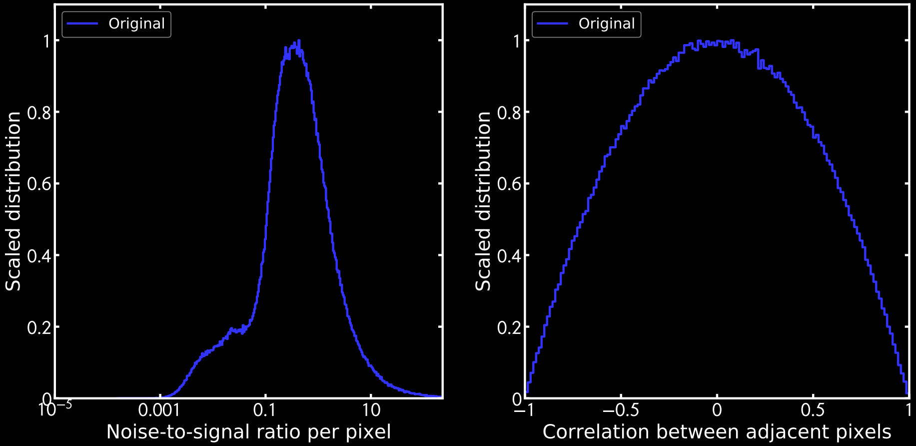
(Pantoni *et al.*, *in prep.*)

M 99 Noise Propagation: Pixel Statistics



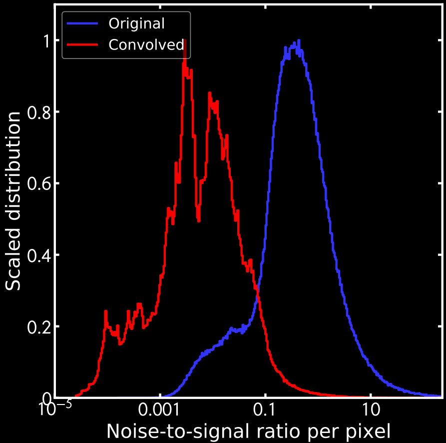
(Pantoni *et al.*, *in prep.*)

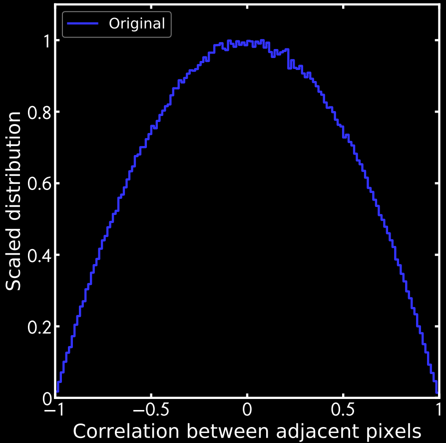
M 99 Noise Propagation: Pixel Statistics



(Pantoni *et al.*, *in prep.*)

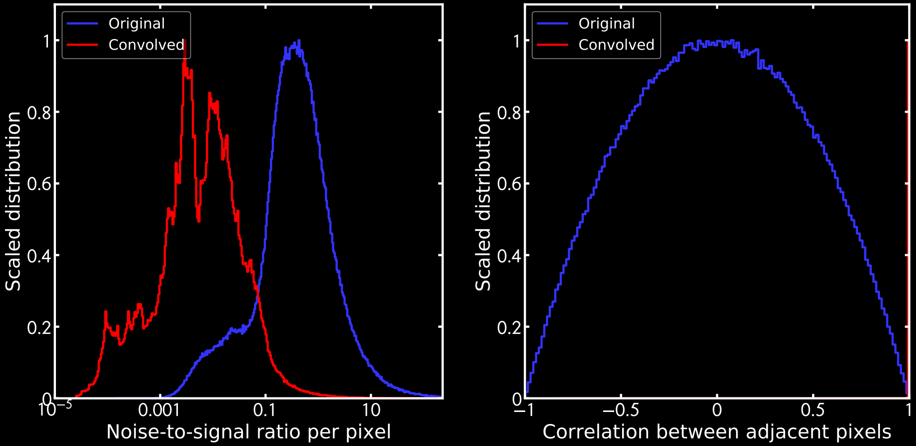
M 99 Noise Propagation: Pixel Statistics





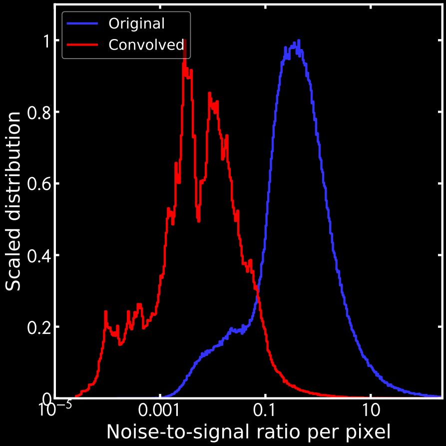
(Pantoni *et al.*, *in prep.*)

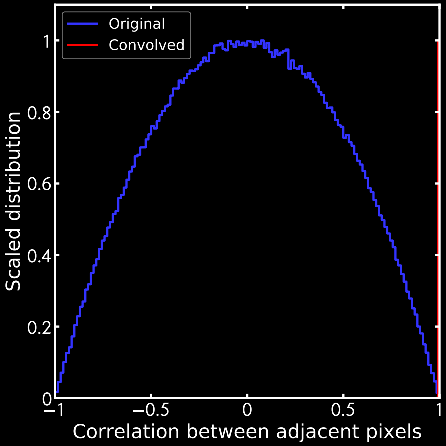
M 99 Noise Propagation: Pixel Statistics



(Pantoni *et al.*, *in prep.*)

M 99 Noise Propagation: Pixel Statistics

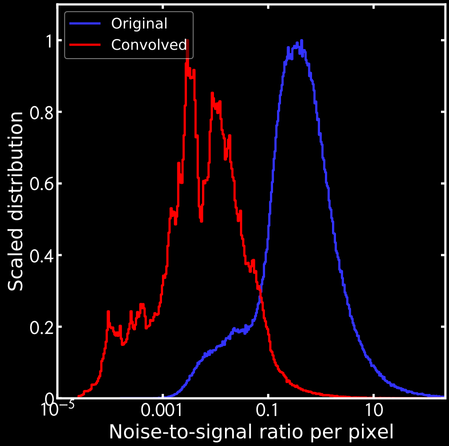


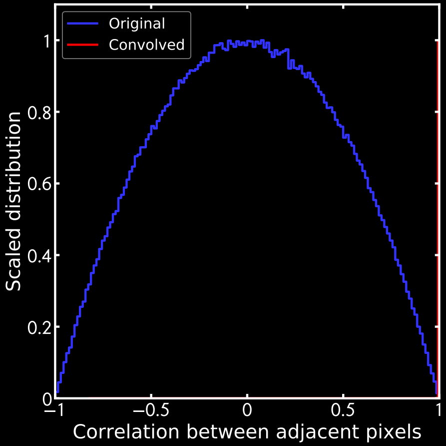


(Pantoni *et al.*, *in prep.*)

Independent ($\rho = 0$): $\sigma = \sqrt{\sigma_a^2 + \sigma_b^2 + 2\rho\sigma_a\sigma_b} = \sqrt{\sigma_a^2 + \sigma_b^2}$.

M 99 Noise Propagation: Pixel Statistics



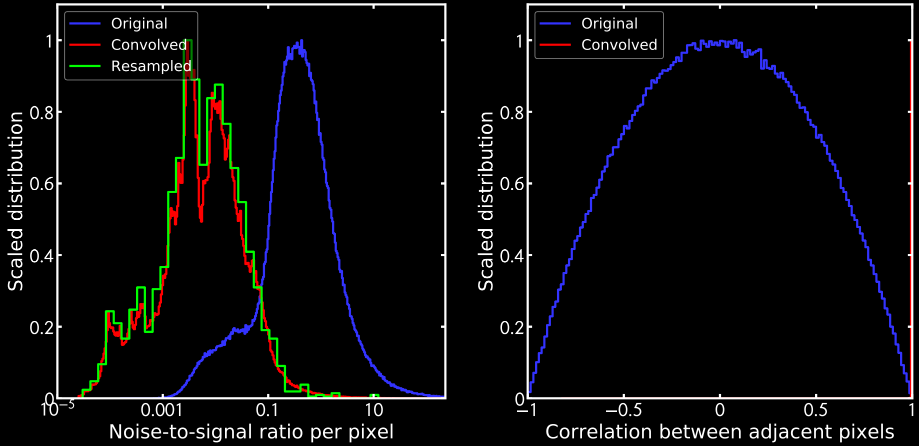


(Pantoni *et al.*, *in prep.*)

Independent ($\rho = 0$): $\sigma = \sqrt{\sigma_a^2 + \sigma_b^2 + 2\rho\sigma_a\sigma_b} = \sqrt{\sigma_a^2 + \sigma_b^2}.$

Correlated ($\rho = 1$): $\sigma = \sqrt{\sigma_a^2 + \sigma_b^2 + 2\rho\sigma_a\sigma_b} = \sigma_a + \sigma_b.$

M 99 Noise Propagation: Pixel Statistics

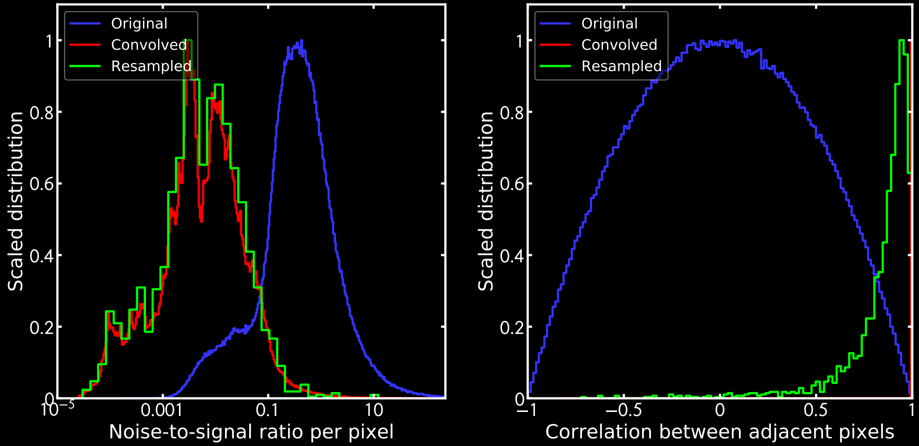


(Pantoni *et al.*, *in prep.*)

Independent ($\rho = 0$): $\sigma = \sqrt{\sigma_a^2 + \sigma_b^2 + 2\rho\sigma_a\sigma_b} = \sqrt{\sigma_a^2 + \sigma_b^2}$.

Correlated ($\rho = 1$): $\sigma = \sqrt{\sigma_a^2 + \sigma_b^2 + 2\rho\sigma_a\sigma_b} = \sigma_a + \sigma_b$.

M 99 Noise Propagation: Pixel Statistics

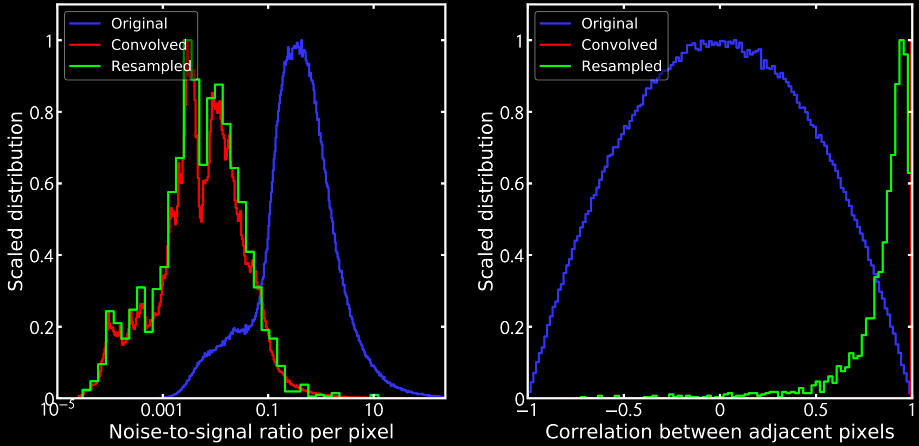


(Pantoni *et al.*, *in prep.*)

Independent ($\rho = 0$): $\sigma = \sqrt{\sigma_a^2 + \sigma_b^2 + 2\rho\sigma_a\sigma_b} = \sqrt{\sigma_a^2 + \sigma_b^2}$.

Correlated ($\rho = 1$): $\sigma = \sqrt{\sigma_a^2 + \sigma_b^2 + 2\rho\sigma_a\sigma_b} = \sigma_a + \sigma_b$.

M 99 Noise Propagation: Pixel Statistics



(Pantoni *et al.*, *in prep.*)

Independent ($\rho = 0$): $\sigma = \sqrt{\sigma_a^2 + \sigma_b^2 + 2\rho\sigma_a\sigma_b} = \sqrt{\sigma_a^2 + \sigma_b^2}$.

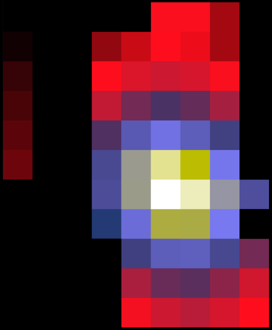
Correlated ($\rho = 1$): $\sigma = \sqrt{\sigma_a^2 + \sigma_b^2 + 2\rho\sigma_a\sigma_b} = \sigma_a + \sigma_b$.

$\Rightarrow$ can feed these errors to a Bayesian SED model.

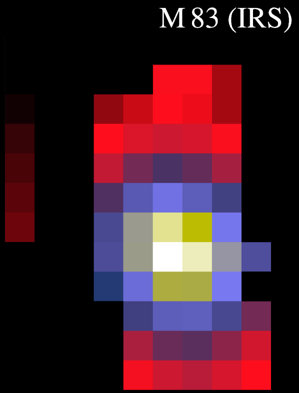
Accounting for Observational Biases: Pointing Errors

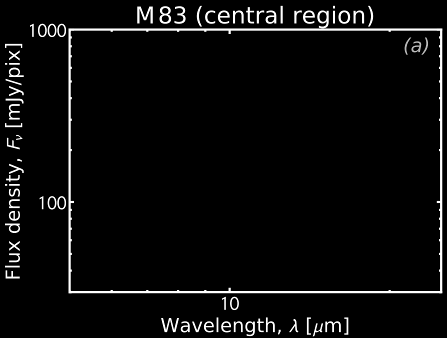
Accounting for Observational Biases: Pointing Errors

M 83 (IRS)



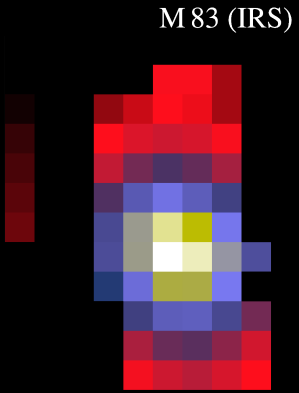
(Hu *et al.*, *in prep.*)

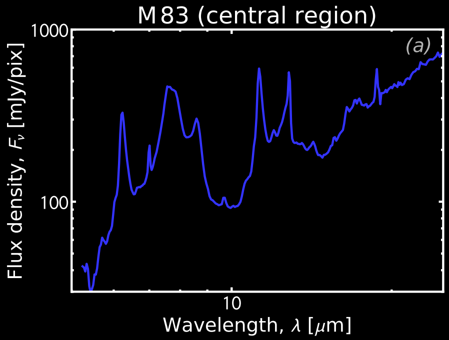




(Hu *et al.*, *in prep.*)

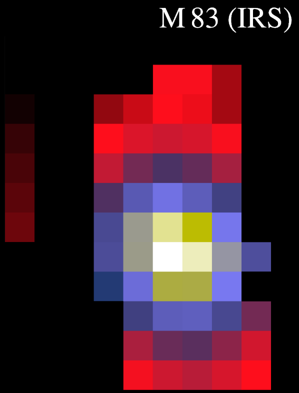
Accounting for Observational Biases: Pointing Errors



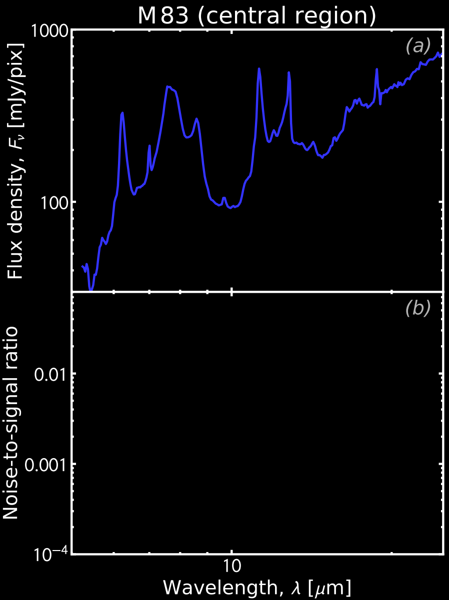


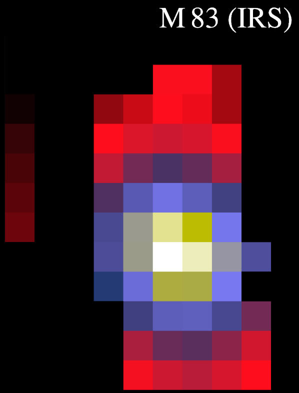
(Hu *et al.*, *in prep.*)

Accounting for Observational Biases: Pointing Errors

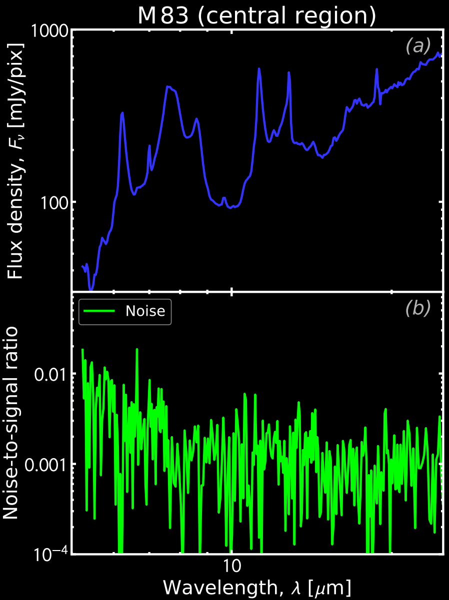


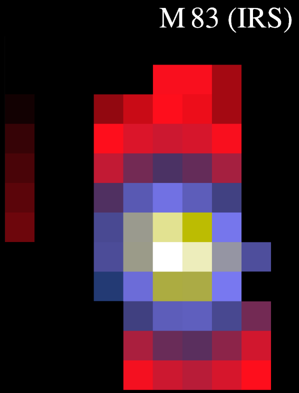
(Hu *et al.*, *in prep.*)

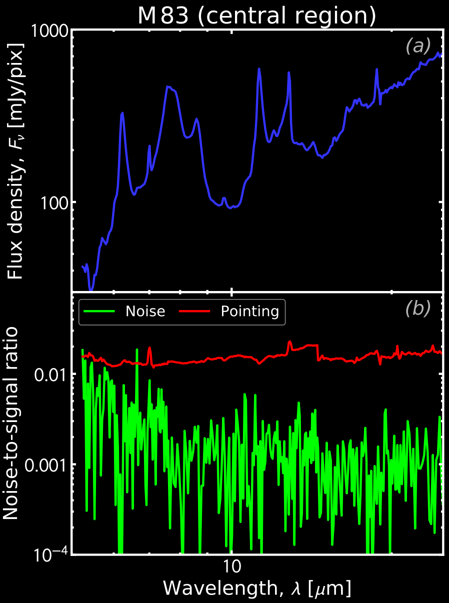




(Hu *et al.*, *in prep.*)







(Hu *et al.*, *in prep.*)

Take-Away

Take-Away

- Several applications require heavily processing original data before modeling.

Take-Away

- Several applications require heavily processing original data before modeling.
- Can account for any processing step and observing bias.

Take-Away

- Several applications require heavily processing original data before modeling.
- Can account for any processing step and observing bias.
- Bootstrapping is a very efficient & easy-to-implement method:
  1. Add random perturbations to the original data;
  2. Loop over the whole processing;
  3. Compress the distribution only at the end.

Take-Away

- Several applications require heavily processing original data before modeling.
- Can account for any processing step and observing bias.
- Bootstrapping is a very efficient & easy-to-implement method:
  1. Add random perturbations to the original data;
  2. Loop over the whole processing;
  3. Compress the distribution only at the end.

$\Rightarrow$ get a full uncertainty distribution w/ its correlations.

Challenges for topological models of the interstellar medium

PCMI - Uncertainty session
October 26th 2022

Lise RAMAMBASON
AIM, CEA Saclay
(*lise.ramambason@cea.fr*)

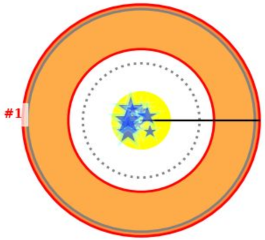
École Doctorale d'Astronomie & Astrophysique d'île-de-France

Topological models

From single-component models…

A **single set of parameters** to describe the ISM properties:

- photoionization models



parameters
-density
-ionization parameters
- stellar properties
-metallicity

but also:

- photodissociation models
- shock model
- dense gas models
- dust models
- …

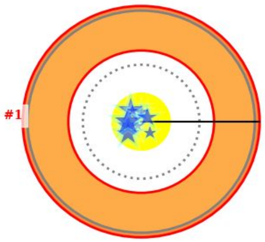
Topological models

(e.g., Lebouteiller+17, Cormier+19, Polles+19, Ramambason+22, Richardson+14,+16)

From single-component models…

A **single set of parameters** to describe the ISM properties:

- photoionization models



parameters
-density
-ionization parameters
- stellar properties
-metallicity

but also:

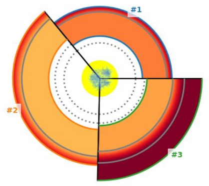
- photodissociation models
- shock model
- dense gas models
- dust models
- …

To "topological" models…

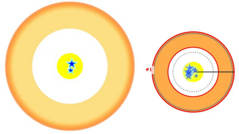
Increase the complexity by linearly **combining several components** under different configurations:

e.g.,

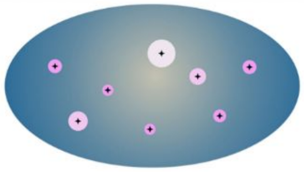
- **multisector** models (combining several sets of gas parameters)



- **multicluster** models (combining several sets of stellar parameters)



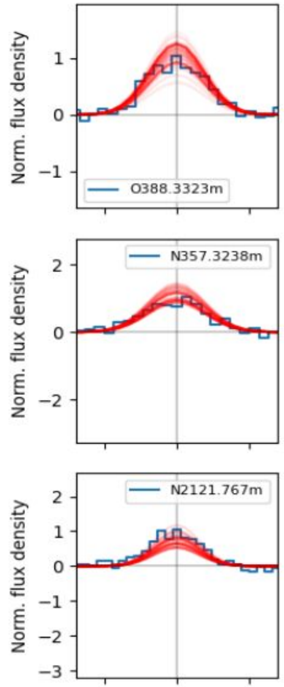
- **distributions** of parameters



$$\psi = U^{\alpha_U} n^{\alpha_n} \ldots$$

Comparing models and observations

PREDICTED LINE FLUXES



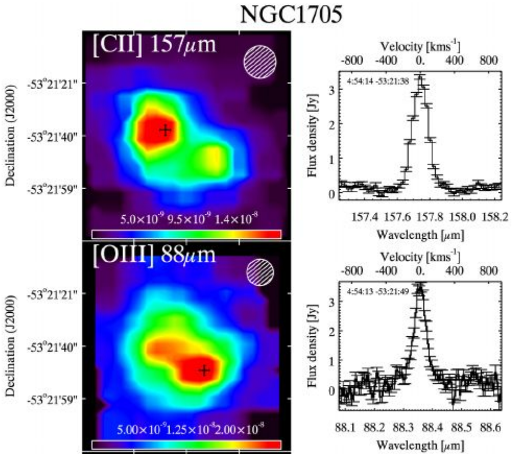
Predicted lines fluxes

How do we compare models to observations ?

OBSERVED LINE FLUXES

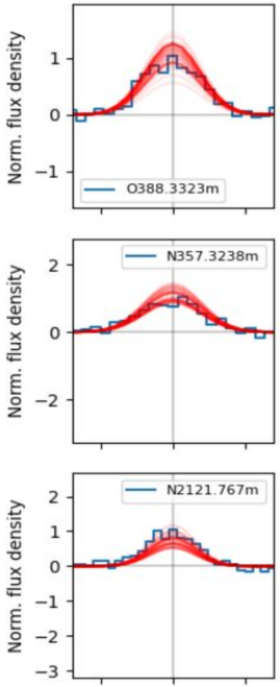
Integrated lines fluxes

Cormier et. al (2015)



Comparing models and observations

PREDICTED LINE FLUXES

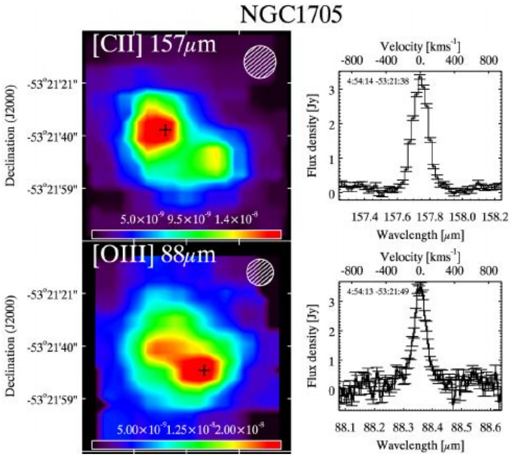


Predicted lines fluxes

OBSERVED LINE FLUXES

Integrated lines fluxes

Cormier et. al (2015)



How do we compare models to observations ?

Need for a **statistical framework**

- accounts for **upper limits**
- accounts for **priors**
- accounts for **systematic uncertainties**
- **asymmetric** uncertainties

MULTIGRIS: a Bayesian tool to automate multicomponent modeling

Lebouteiller & Ramambason, 2022
GitLab: https://gitlab.com/multigris/mgris

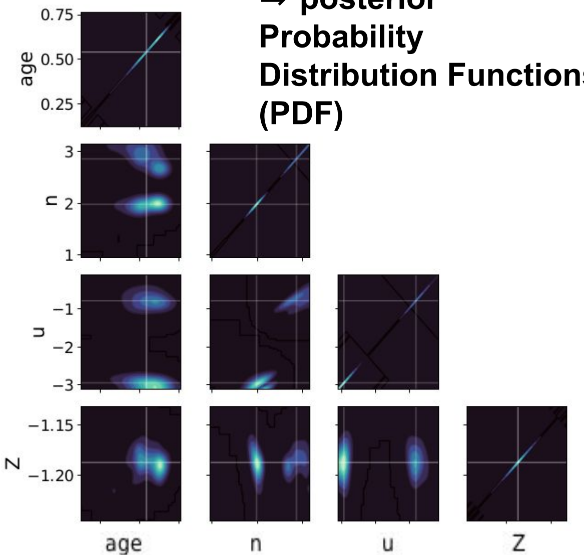
- **model M**= grid of predicted fluxes + interpolation function
- **Topological configuration** (number of sectors and parameters θ and priors P(θ))
- **data d** = observed emission lines + upper limits

PyMC3

$$\mathcal{L} = P(d|\theta) = \prod_{i=0}^{N} \mathcal{N}(\mu = O_i, \sigma^2 = U_i^2)$$

→ **SAMPLING**: draw from the likelihood with a given sampling algorithm (MCMC)

Use Bayes theorem: P(θ|d) ∝ P(d|θ)P(θ)
⇒ **posterior Probability Distribution Functions (PDF)**



MULTIGRIS: a Bayesian tool to automate multicomponent modeling

What is uncertain in our workflow?

Lebouteiller & Ramambason, 2022
GitLab: https://gitlab.com/multigris/mgris

- **model M**= grid of predicted fluxes + interpolation function
- **Topological configuration** (number of sectors and parameters θ and priors P(θ))
- **data d** = observed emission lines + upper limits

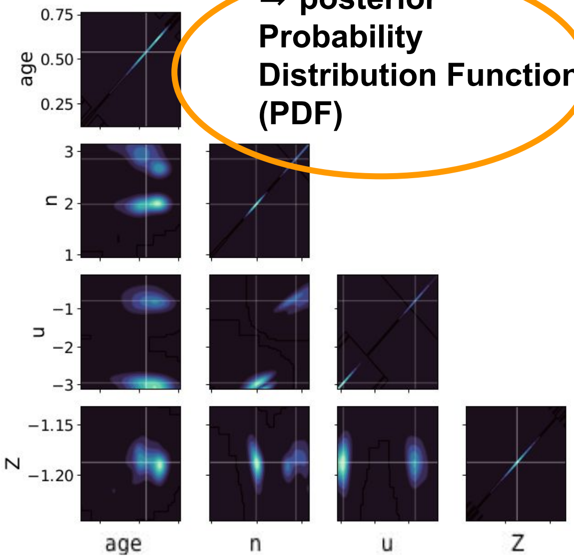
PyMC3

$$\mathcal{L} = P(d|\theta) = \prod_{i=0}^{N} \mathcal{N}(\mu = O_i, \sigma^2 = U_i^2)$$

→ **SAMPLING**: draw from the likelihood with a given sampling algorithm (MCMC)

Use Bayes theorem: P(θ|d) ∝ P(d|θ)P(θ)

⇒ **posterior Probability Distribution Functions (PDF)**



MULTIGRIS: a Bayesian tool to automate multicomponent modeling

What is uncertain in our workflow?

Lebouteiller & Ramambason, 2022
GitLab: https://gitlab.com/multigris/mgris

PyMC3

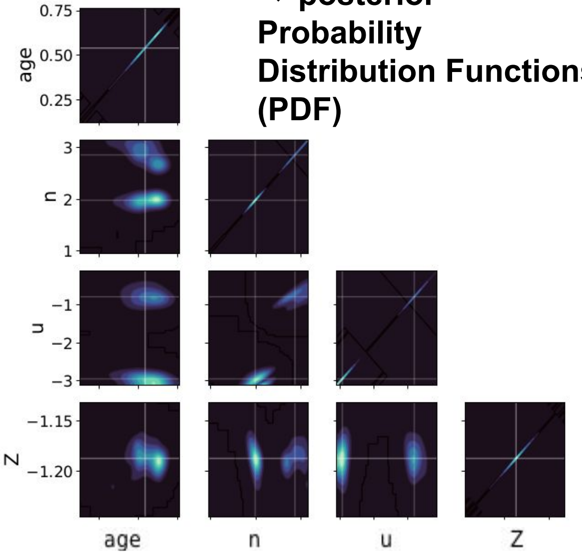
- **model M**= grid of predicted fluxes + interpolation function
- **Topological configuration** (number of sectors and parameters θ and priors P(θ))
- **data d** = observed emission lines + upper limits

$$\mathcal{L} = P(d|\theta) = \prod_{i=0}^{N} \mathcal{N}(\mu = O_i, \sigma^2 = U_i^2)$$

→ **SAMPLING**: draw from the likelihood with a given sampling algorithm (MCMC)

Use Bayes theorem: P(θ|d) ∝ P(d|θ)P(θ)

⇒ **posterior Probability Distribution Functions (PDF)**

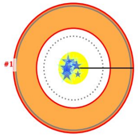


Uncertainties associated with the choice of the best configuration

<u>Different topological representations:</u>

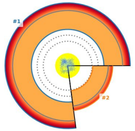
M1

1 sector



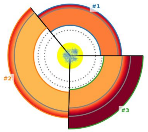
M2

2 sectors



M3

3 sectors



★ **Which topological <u>model is favored</u> by the knowledge from a given set of lines?**
⇒ Minimal level of model complexity

Need a **<u>metric</u>** to compare models!

PRIOR PROBABILITY OF THE MODEL

$$\begin{cases} R_{\mathrm{bayes}} = \dfrac{p(\mathcal{M}_1|\vec{O})}{p(\mathcal{M}_2|\vec{O})} \\ R_{\mathrm{bayes}} \propto \dfrac{p(\vec{O}|\mathcal{M}_1)\,p(\mathcal{M}_1)}{p(\vec{O}|\mathcal{M}_2)\,p(\mathcal{M}_2)} \end{cases}$$

MARGINAL LIKELIHOOD

Uncertainties associated with the sampling (MCMC)

Challenges with MCMC walkers:

★ known caveats of random walkers:
- can **get stuck in local maxima** ⇒ not well adapted to sample **multi-peaked distributions**
- stochasticity ⇒ solution may vary with **different starting points**

★ Do not sample the whole parameter space
⇒ the marginal likelihood is difficult to evaluate!

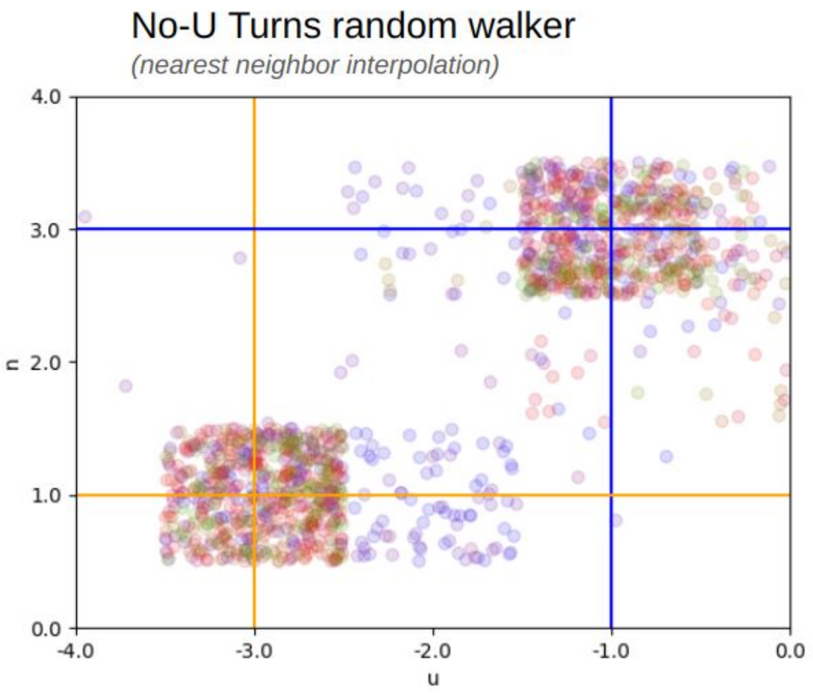
MARGINAL LIKELIHOOD

$$\boxed{p(\vec{O}|\mathcal{M})} = \int_\theta \underbrace{p(\vec{O}|\theta, \mathcal{M})}_{\text{likelihood}} \underbrace{p(\vec{\theta}|\mathcal{M})}_{\text{prior on } \theta} d\theta$$

integrate on the whole parameter space

likelihood

prior on θ



Uncertainties associated with the sampling (MCMC)

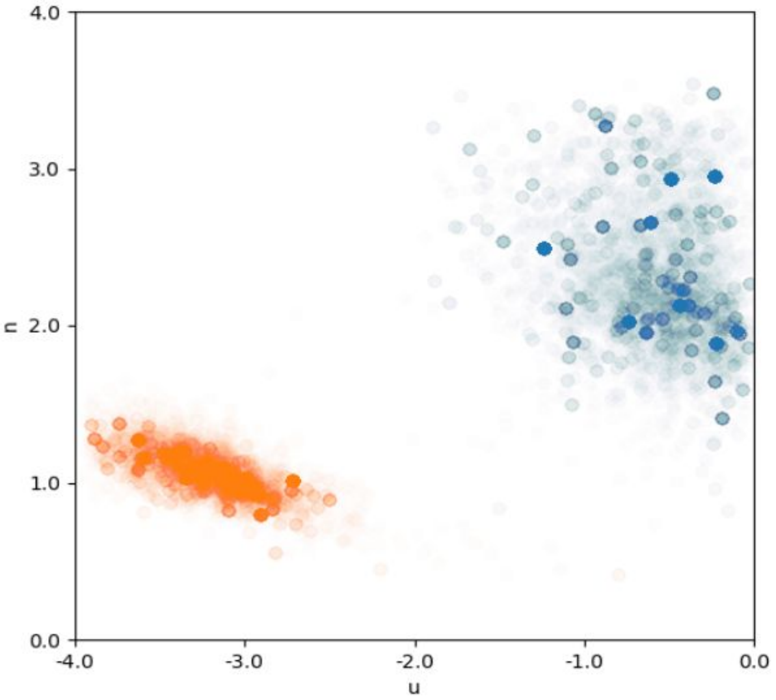
Particle filtering methods:

★ parallel MCMC chains that simultaneously sample the whole parameter space

★ less sensitive to starting values but requires large number of draw

★ marginal likelihood is easier to evaluate!
⇒ allows model comparison **assuming that prior probabilities are equal**

Sequential Monte Carlo (SMC)



MARGINAL LIKELIHOOD

$$\boxed{p(\overrightarrow{O}|\mathcal{M})} = \int_\theta p(\overrightarrow{O}|\theta, \mathcal{M})\, p(\overrightarrow{\theta}|\mathcal{M})\, d\theta$$

integrate on the whole parameter space

likelihood

prior on θ

Key points and some questions

★ Topological models add a layer of uncertainties with the **choice of the best configuration**

$$\begin{cases} R_{\text{bayes}} = \dfrac{p(\mathcal{M}_1|\vec{O})}{p(\mathcal{M}_2|\vec{O})} \\ R_{\text{bayes}} \propto \dfrac{p(\vec{O}|\mathcal{M}_1)}{p(\vec{O}|\mathcal{M}_2)} \dfrac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)} \end{cases}$$

assuming that the prior probabilities of all models are equal

#1; Is it ok to assume that *all models are a priori equivalent* ?

Should more complex models be favored as more likely to reproduce a complex ISM structure? (and how?)

Key points and some questions

★ Topological models add a layer of uncertainties with the **choice of the best configuration**

$$\begin{cases} R_{\text{bayes}} = \dfrac{p(\mathcal{M}_1|\vec{O})}{p(\mathcal{M}_2|\vec{O})} \\ R_{\text{bayes}} \propto \dfrac{p(\vec{O}|\mathcal{M}_1)}{p(\vec{O}|\mathcal{M}_2)} \boxed{\dfrac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)}} \end{cases}$$

assuming that the prior probabilities of all models are equal

★ **Particle filtering sampling** methods are well adapted to sample multi-peaked distributions and evaluate easily the marginal likelihood.

#1; Is it ok to assume that *all models are a priori equivalent* ?

Should more complex models be favored as more likely to reproduce a complex ISM structure? (and how?)

#2: How should we representent *multi-peaked distributions* in which the mean, median are not representative?

Smooth representations may be limited to interpret trends in samples, especially for incomplete samples

Key points and some questions

★ Topological models add a layer of uncertainties with the **choice of the best configuration**

$$\begin{cases} R_{\text{bayes}} = \dfrac{p(M_1|\vec{O})}{p(M_2|\vec{O})} \\ R_{\text{bayes}} \propto \dfrac{p(\vec{O}|M_1)}{p(\vec{O}|M_2)} \dfrac{p(M_1)}{p(M_2)} \end{cases}$$

assuming that the prior probabilities of all models are equal

★ **Particle filtering sampling** methods are well adapted to sample multi-peaked distributions and evaluate easily the marginal likelihood.

★ The posterior distribution reflects the knowledge associated with a given set of lines **and their associated (measured) uncertainties assumed to be gaussian.**

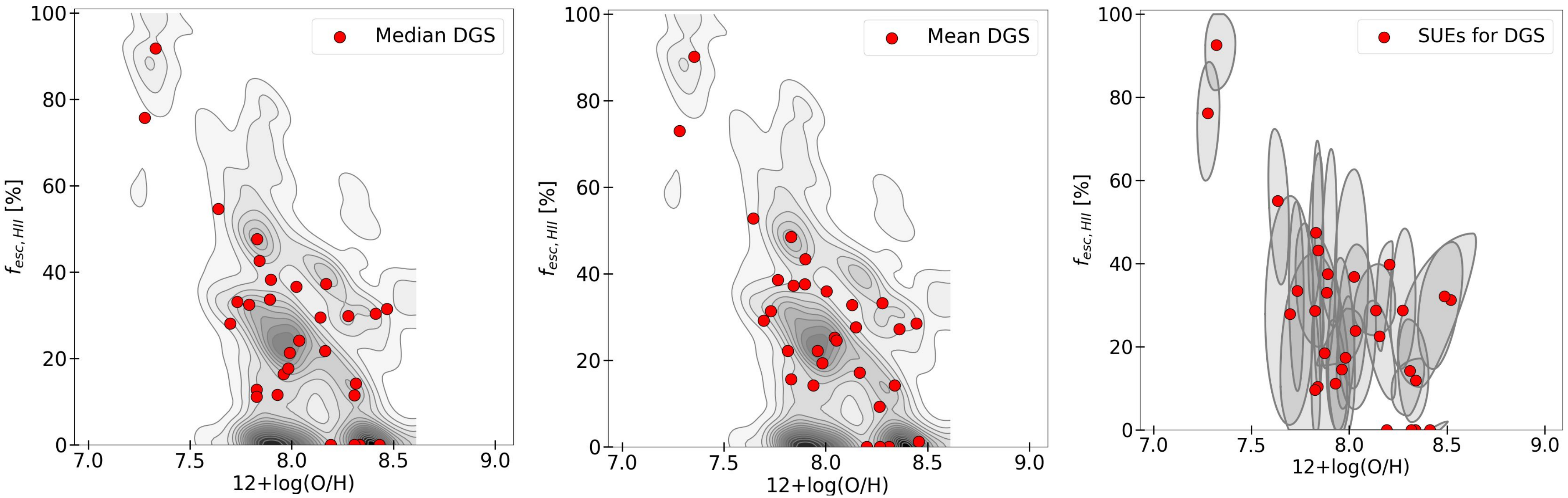
#1; Is it ok to assume that *all models are a priori equivalent* ?
Should more complex models be favored as more likely to reproduce a complex ISM structure? (and how?)

#2: How should we representent *multi-peaked distributions* in which the mean, median are not representative?
Smooth representations may be limited to interpret trends in samples, especially for incomplete samples

#3: What is the impact of the set of tracers used as input?
In theory better to have as many as possible but require more complex models. Tracers not well understood may bias the solution.

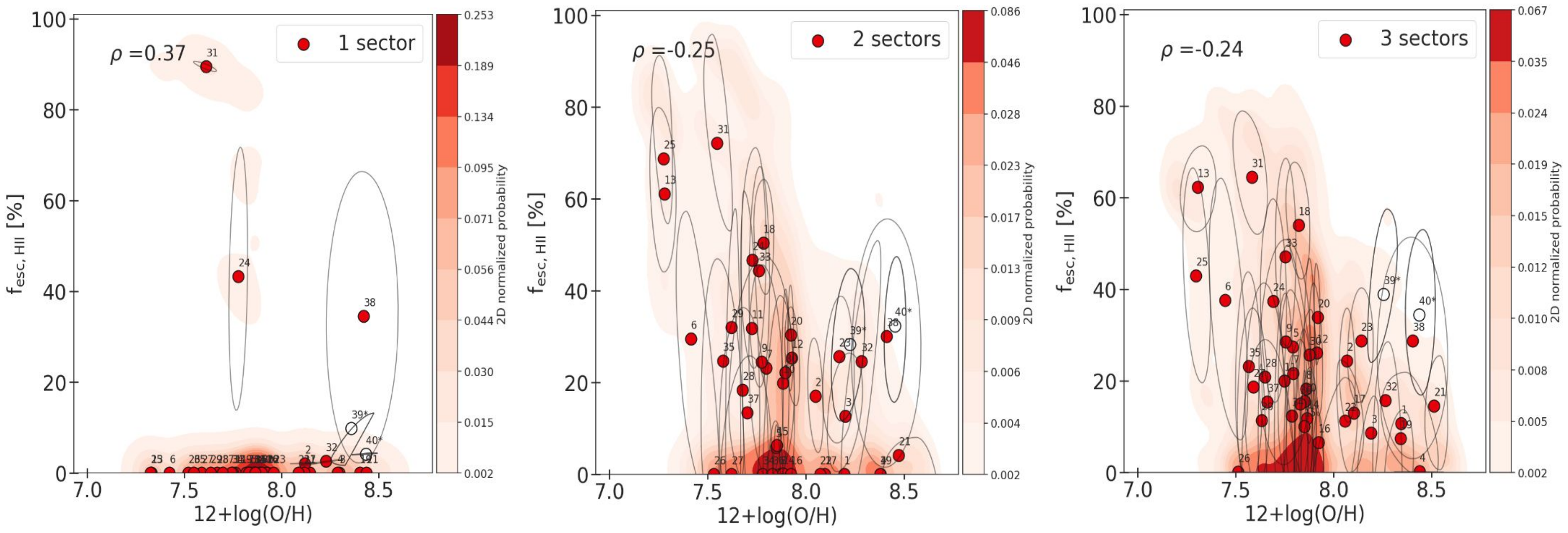
#4: What is the impact *assuming fixed gaussian uncertainties* for the input data?
Ideally, the fitting process should be included on-the-fly with a new fit at each draw in the MCMC (expensive)

Example: Representing the individual and global PDFs



Figures from Ramambason+22

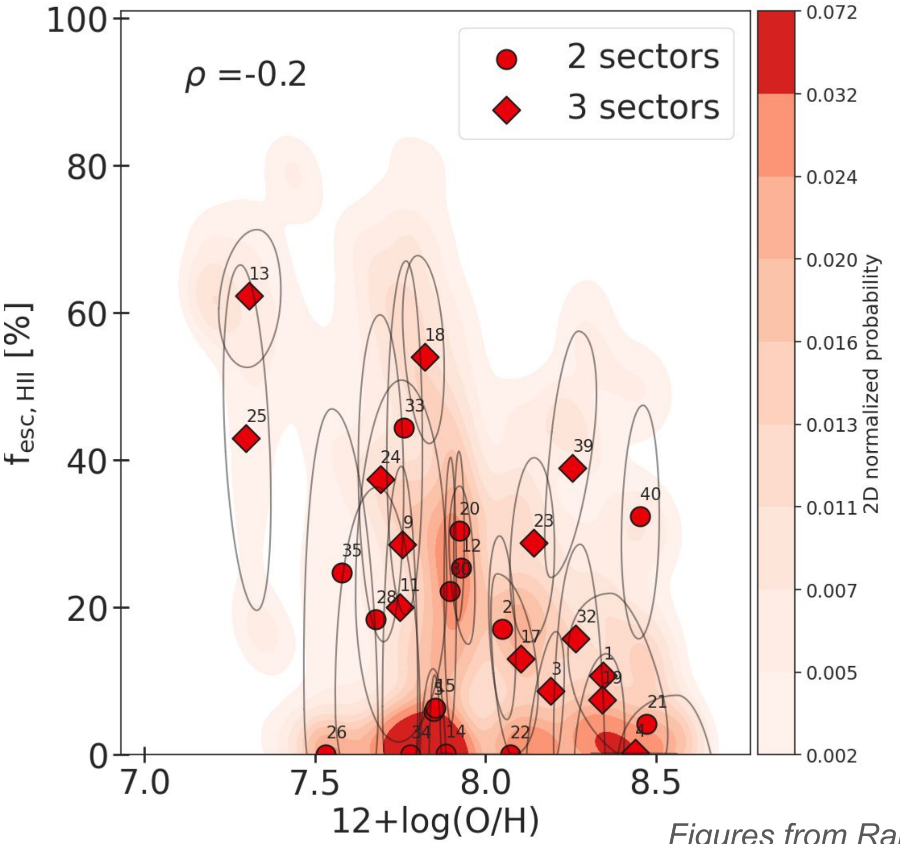
Example: Representing the uncertainty on topology



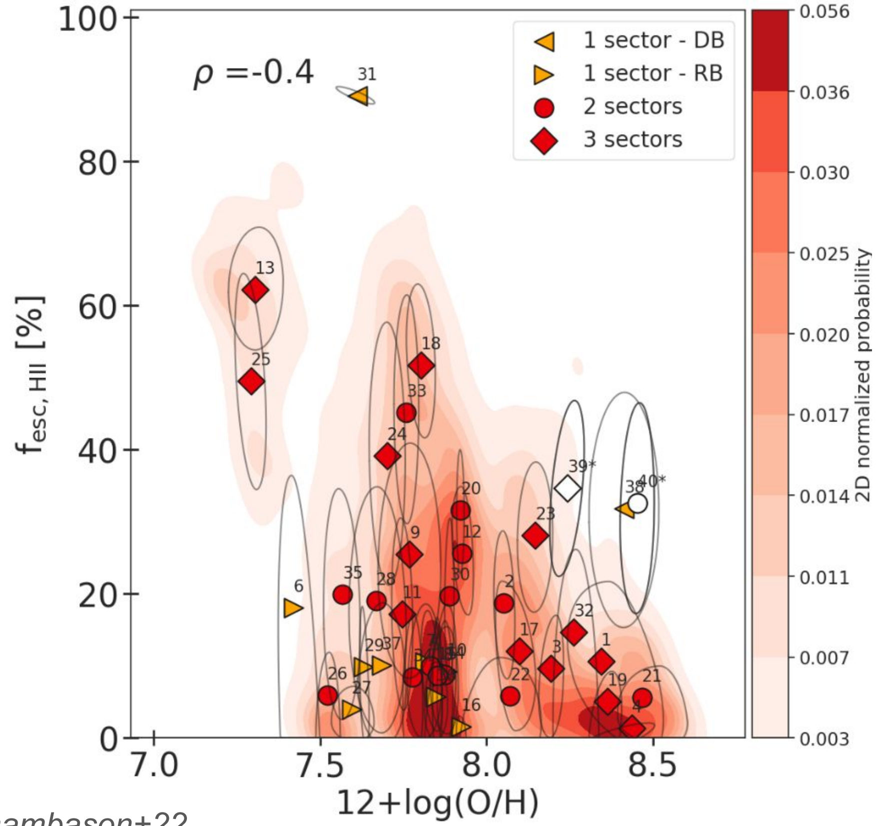
Figures from Ramambason+22

Example: Representing the uncertainty on topology

Choice #1: representing only the “best” models



Choice 2#: representing only the “best” models



Figures from Ramambason+22

Evaluating uncertainties for components separation

Erwan Allys - LPENS, Paris,
with C. Auclair, F. Boulanger, P. Richard

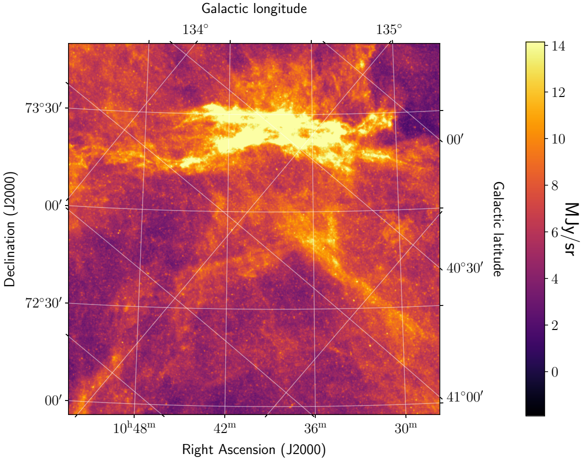
Colloque PCMI 2022
Paris, October 26$^{\text{th}}$ 2022

Outline

1. Dust/CIB components separation
2. Uncertainties in the components separation

Dust/CIB components separation

- **Herschel *spider* field at 250μm**



→ Thermal dust emission and Cosmic Infrared Background (CIB)

Dust/CIB components separation

- **Scientific objective**
  - Mixture of components in observations
    $\rightarrow d = s + k$ with d data, s dust, k CIB
  - Use non-Gaussian information to separate them
    $\rightarrow$ close SED of thermal dust and CIB
    $\rightarrow$ Work at a single frequency (to begin with)
  - Realistic simulations hard to find
    $\Rightarrow$ Work only from observational data

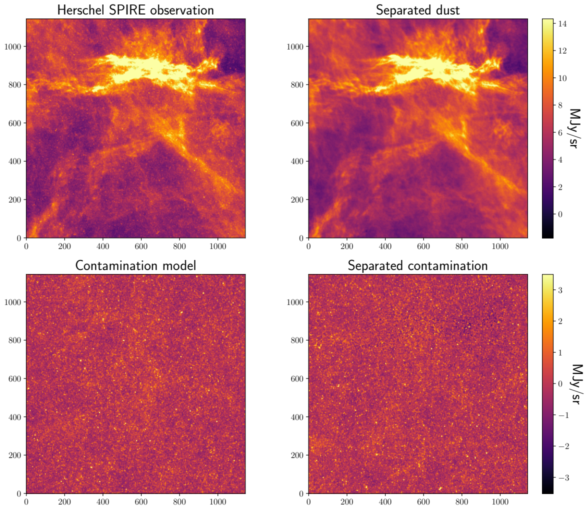
$\rightarrow$ **From observational statistics of k, recover statistics of s**
Auclair et al, in prep.

Components separation algorithm

- **Application on Herschel *spider* field**
  - Clean k_0 CIB observation from *Lockman hole* field
    $\rightarrow$ estimation of the statistics of k
  - Deformation of d to an estimate $\tilde{s}$ of s **(Regaldo+ 2021, Delouis+ 2022)**
    $\rightarrow$ gradient descent in pixel space
    $\rightarrow$ several constraints from scattering statistics
    $\rightarrow$ happy to discuss more :)
  - We obtain a $\tilde{s}$ map, on which we evaluate the statistics

$\rightarrow$ Focus on statistics of s (not deterministic at small scales)
$\rightarrow$ Only d and k_0 are used in the process !

- **Input data and separated components**



→ **Look nice, but what are the uncertainties ?**

Outline

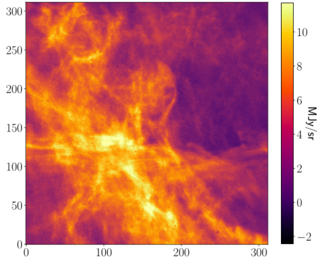
Evaluating uncertainties

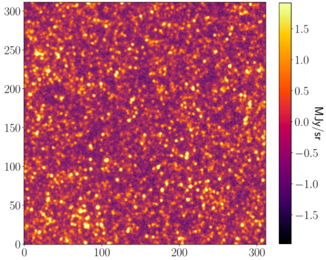
- **Difficulties for evaluating uncertainties...**
  - Only a few input maps
  - Highly non-linear separation

→ **First approach with a validation on mock data**

Validation pipeline on mock data

- **Constructing a set of mock data**
  - Perform a separation with known dust
  - Surrogate dust from observations
    → same field of view, gas from HI data
    → avoid CIB contamination
    → denoising and map construction with ROHSA **(Marchal+ 2019)**
    → lower resolution ⇒ smaller patch
  - CIB from *Lockman Hole* can be cut in several patches





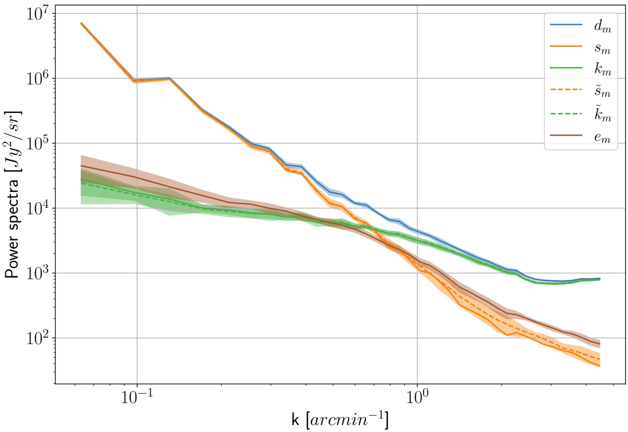
Validation pipeline on mock data

- **Different types of errors**
  - Same statistics for k and k_0
    $\rightarrow$ algorithm error
  - Statistical variance for k and k_0
    $\rightarrow$ model error

$\rightarrow$ **Model error dominates on all scales**

Results from mock data

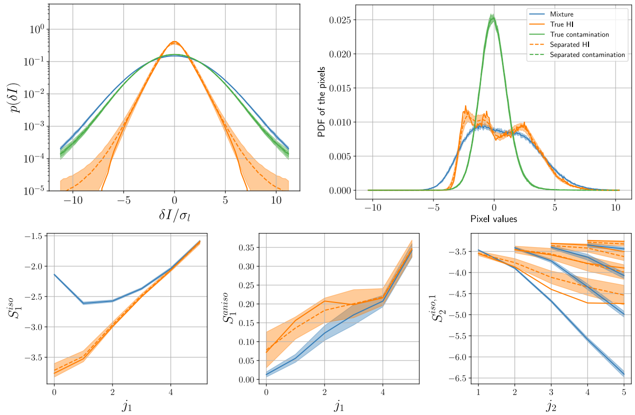
- **Power spectrum**



→ 1σ-variability + bias w.r.t. truth available.

Results from mock data

- **Beyond power spectrum (increments pdf, pixels pdf, RWST)**



→ Correct statistics reproduced on all but smallest scales

Estimating uncertainties with real data

- **Uncertainties: mock data are not real data...**
  - We have only one sample of the CIB
    → no direct variance assessment
  - Mock data are with smaller patches
    → how to extrapolate to larger patches ?
  - The HI map is not a dust map
    → how to extrapolate to an unknown component ?

- **We can extrapolate from the application on mock data**
  → What's the better way to do so ?

- **What other method could we use ?**

Estimating uncertainties with real data

- **Uncertainties: mock data are not real data...**
  - We have only one sample of the CIB
    → no direct variance assessment
  - Mock data are with smaller patches
    → how to extrapolate to larger patches ?
  - The HI map is not a dust map
    → how to extrapolate to an unknown component ?

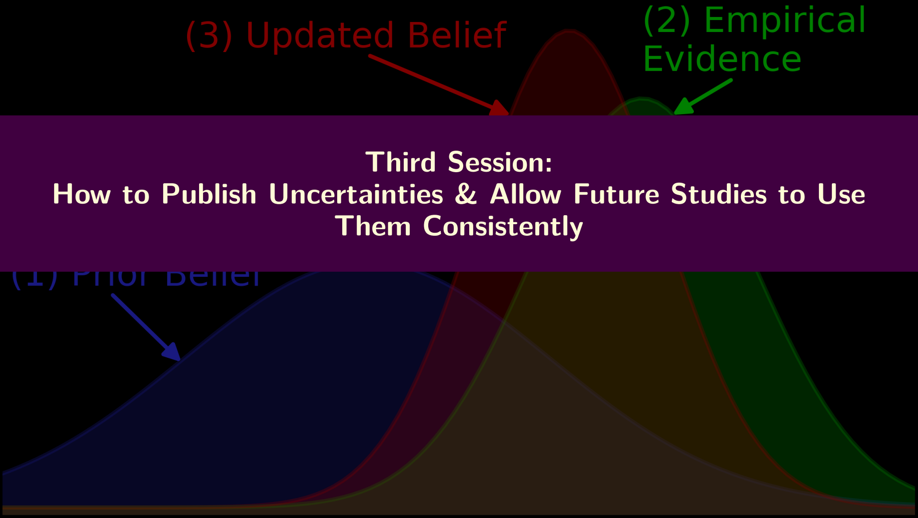
- **We can extrapolate from the application on mock data**
  → What's the better way to do so ?

- **What other method could we use ?**

Thanks for your attention !
– and happy to discuss components separation :) –

Second Session Wrap-Up: What Can We Be Certain About?

- Should we try to account for every effects, from the instrumental biases to the physics of our target, as well as the different contaminations, in one big single model?
- What is the meaning and potential usefulness of a good fit with a wrong model?
- How to validate a simulation that can not be fit to some observations?
- Have we been too pessimistic? Isn't there something called "the law of large numbers" that will guarantee that all our uncertainties average out?

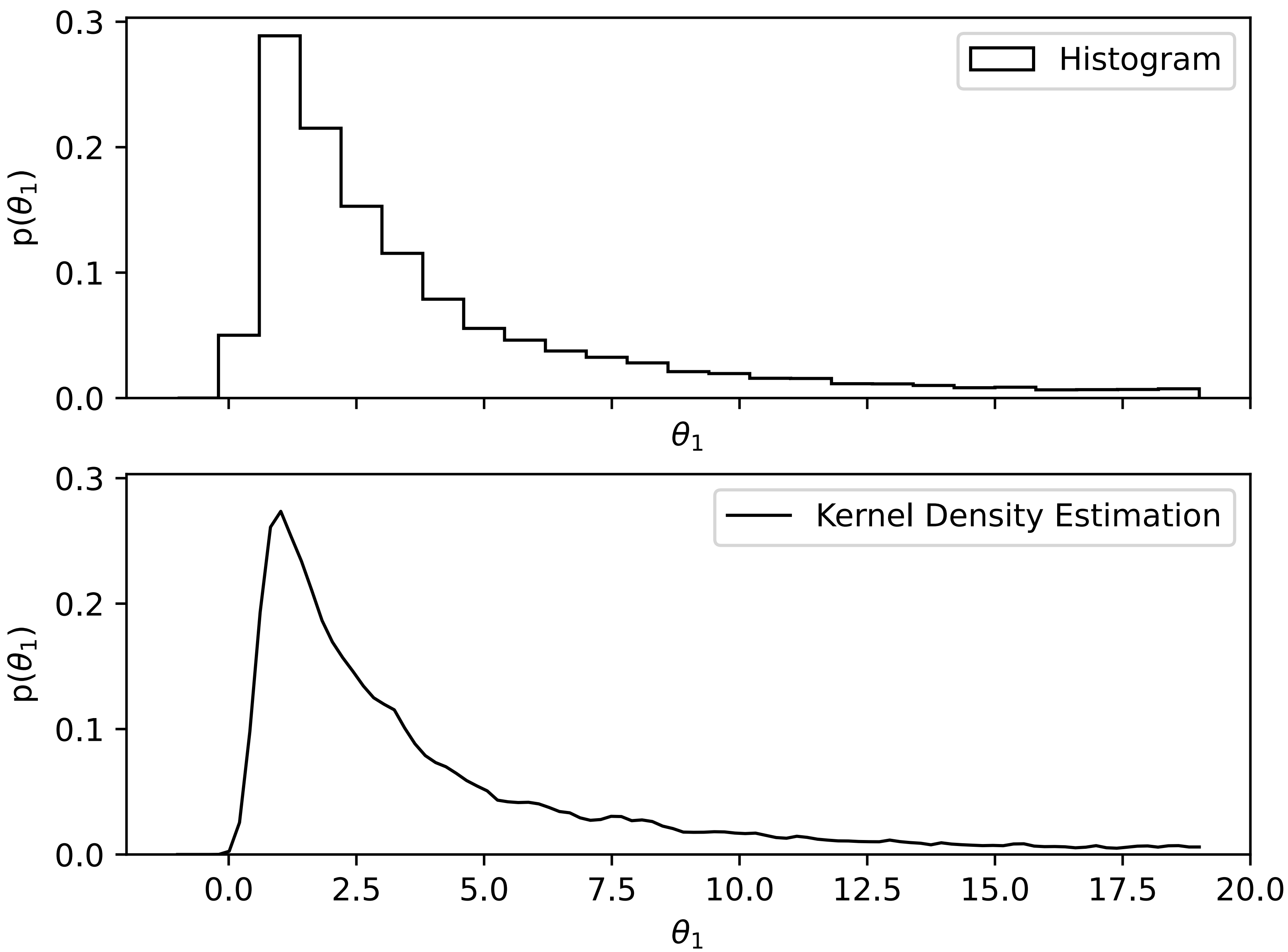


Third Session:
How to Publish Uncertainties & Allow Future Studies to Use Them Consistently

Plotting distributions

Most of the time we have access to a sample of points

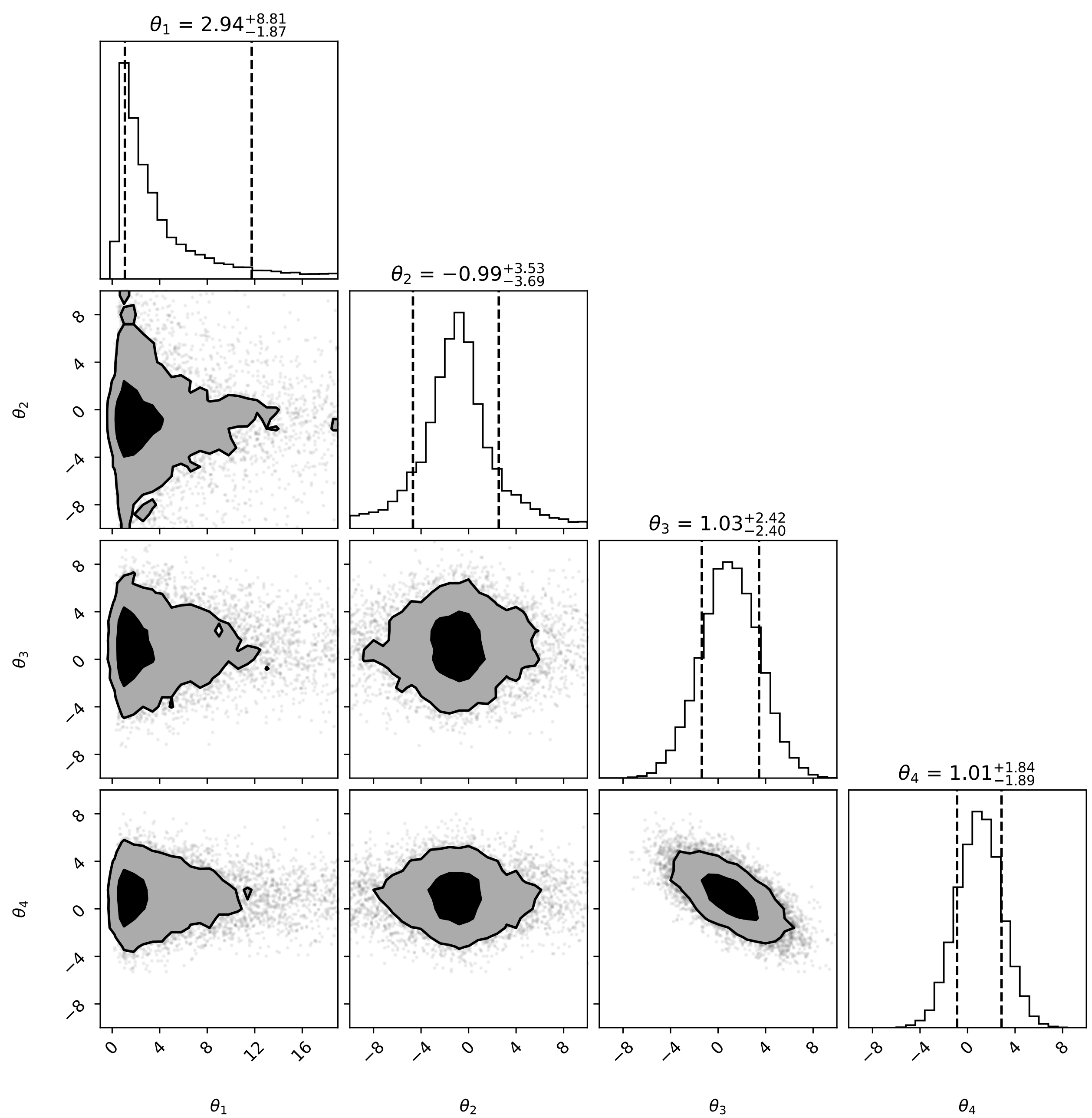
- Show approximations of the distributions (histograms, kde)



Plotting distributions

Most of the time we have access to a sample of points

- More difficult with increasing dimensions: corner (or triangle) plots



Quoting errors

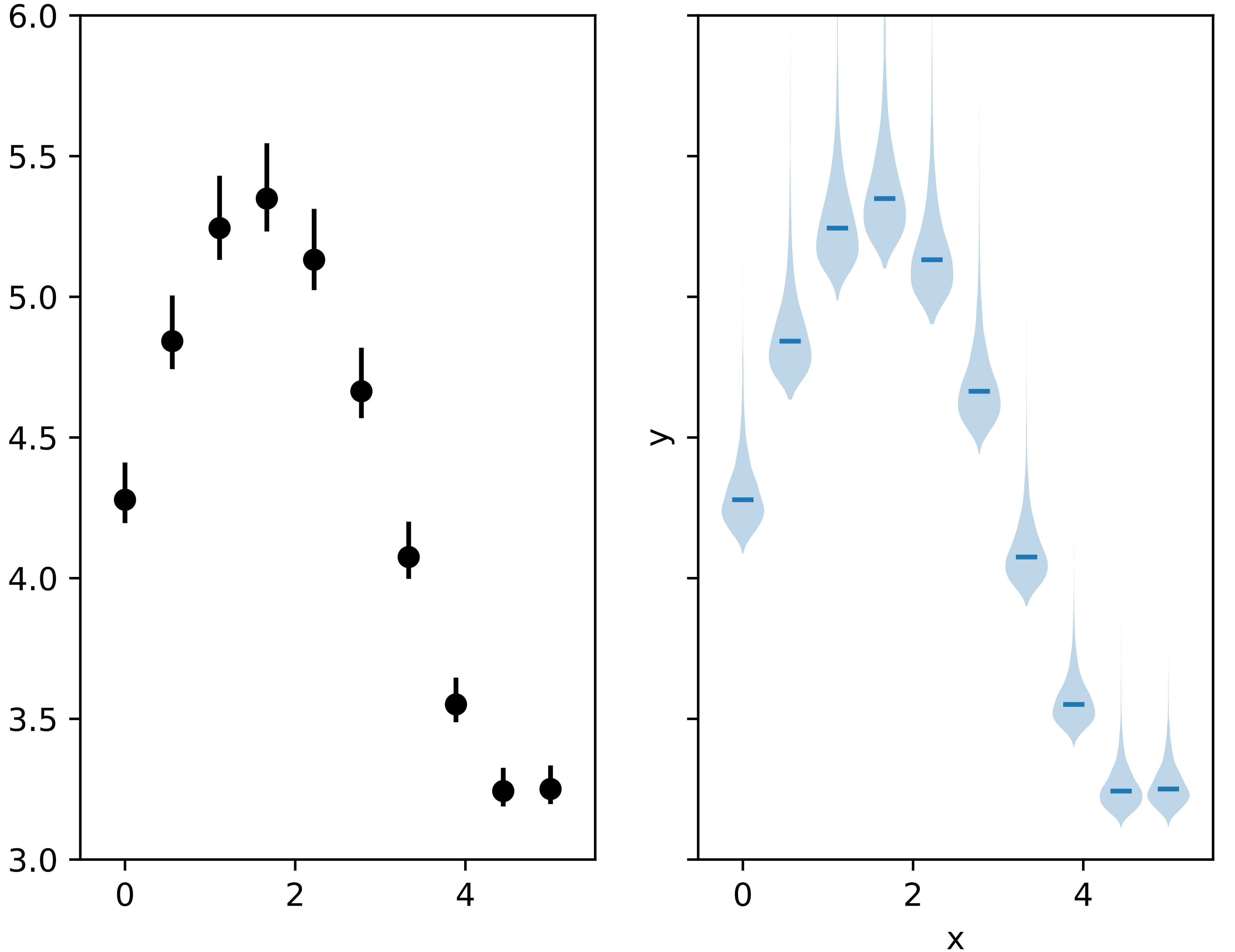
Most of the time we have access to a sample of points

Summarising the whole distribution with a “central value” and an “uncertainty” often called point estimates

- mean ± standard deviation
- median ± interpercentile range (often p16 and p84 to match the 1 sigma interval for a 1d gaussian distribution)
- For higher dimensions the uncertainty can be summarised in a covariance matrix (d x d but only need to store d(d-1)/2 values)

Plotting errors

Most of the time we have access to a sample of points



Sometimes we don't have a sample of points

- "Black box" optimisers
- If they give uncertainties usually computed from local curvature around "maximum likelihood" or "minimum chi2" and assume gaussian errors => covariance matrix.
- If no uncertainty given you can try bootstrapping
  - 1/ create many (10^{4-5}) new datasets by resampling with replacement
  - 2/ compute the value you want for each of these resampled dataset
  - 3/ you now have a sample of values you can deal with as in the previous slides

Transmitting this information

What to send ?

- send the samples themselves
- choose a family of analytic distributions, compute the associated parameters and send those
- send the parametric description (histogram or kde)
- send the “point estimates” (possibly with the covariance matrix)
- for bayesian inference: send the dataset + likelihood function + prior definitions and let the others resample as many points as they want

Transmitting this information

How to send it ?

- A table in your manuscript
- An ASCII table
- A structured format json, pandas, python pickle (beware of strange formats they don't live forever)
- for larger datasets: binary formats (hdf5, netCDF, fits, etc)
    - Some formats are trying to become standards eg arviz InferenceData structure for samples from a distribution
- Maybe one day: the python scripts that create the figures, tables of your manuscript from the data. Some editors already ask for the datasets

Third Session Wrap-Up: What Can We Be Certain About?

- Would it be profitable to the community to set a standard in the way uncertainties are estimated and published?
- Should we create a network of people interested in helping each other to achieve this task?
- Who should centralize the uncertainties of all published studies (A&A, the CDS, *etc.*)?
- Should we declare October 26 ± 1 "Uncertainty Day" at UNESCO?