A&A, 698, A311 (2025)
https://doi.org/10.1051/0004-6361/202554266

Astronomy & Astrophysics

BEETROOTS: Spatially regularized Bayesian inference of physical parameter maps. Application to Orion

Pierre Palud[1,2,3,⋆], Emeric Bron[1], Pierre Chainais[2], Franck Le Petit[1], Pierre-Antoine Thouvenin[2], Miriam G. Santa-Maria[4,5], Javier R. Goicoechea[4], David Languignon[1], Maryvonne Gerin[6], Jérôme Pety[6,7], Ivana Bešlić[6], Simon Coudé[10,11], Lucas Einig[7,8], Helena Mazurek[6], Jan H. Orkisz[7], Léontine Ségal[7,9], Antoine Zakardjian[12], Sébastien Bardeau[7], Karine Demyk[12], Victor de Souza Magalhães[13], Pierre Gratier[14], Viviana V. Guzmán[15], Annie Hughes[12], François Levrier[16], Jacques Le Bourlot[1], Dariusz C. Lis[17], Harvey S. Liszt[13], Nicolas Peretto[18], Antoine Roueff[9], Evelyne Roueff[1], and Albrecht Sievers[7]

(Affiliations can be found after the references)

Received 25 February 2025 / Accepted 15 April 2025

ABSTRACT

Context. The current generation of millimeter (mm) receivers is capable of producing cubes of 800 000 pixels over 200 000 frequency channels to cover a number of square degrees over the 3 mm atmospheric window. Estimating the physical conditions of the interstellar medium (ISM) with an astrophysical model on the basis of such large datasets is challenging. Common approaches tend to converge to local minima and end up poorly reconstructing regions with a low signal-to-noise ratio (S/N) in most cases. This instrumental revolution thus calls for new scalable data analysis techniques with more advanced approaches to statistical modeling and methods.

Aims. Our aim is to design a general method to reconstruct large maps of physical conditions from the rich datasets produced by new and future instruments. The requirements of the method include the ability to scale to very large maps, to be robust to varying S/N, and to escape from the local minima. In addition, we want to quantify the uncertainties associated with our reconstructions to produce reliable analyses.

Methods. We present BEETROOTS, a PYTHON software that performs Bayesian reconstructions of maps of physical conditions based on observation maps and an astrophysical model. It relies on an accurate statistical model, exploits spatial regularization to guide estimations, and uses state-of-the-art algorithms. It can also assess the ability of the astrophysical model to explain the observations, providing feedback to improve ISM models. In this work, we demonstrate the power of BEETROOTS with the Meudon PDR code on synthetic data. We then apply it to estimate physical condition maps in the full Orion molecular cloud 1 (OMC-1) star-forming region based on *Herschel* molecular line emission maps.

Results. The application to the synthetic case shows that BEETROOTS can currently analyze maps with up to ten thousand pixels, addressing large variations among the S/N values within the observations while escaping from local minima and providing consistent uncertainty quantifications. On a personal laptop, the inference runtime ranges from a few minutes for maps of 100 pixels to 28 hours for maps of 8100 pixels. Regarding OMC-1, our reconstructions of the incident UV radiation field intensity, G_0, are consistent with those obtained from FIR luminosities. This demonstrates that the considered molecular tracers are able to constrain G_0 over a wide range of environments. In addition, the obtained thermal pressures are high in all dense regions of OMC-1 and positively correlated with G_0. Finally, the Meudon PDR code successfully explains the observations and the obtained G_0 values are reasonable, which indicates that UV photons control the gas physics and chemistry across the rim of OMC-1.

Conclusions. This work paves the way toward systematic and rigorous analyses of observations produced by current and future instruments. Subsequent efforts still need to be made in parallelizing the algorithm and thereby gaining two orders of magnitude for the map sizes.

Key words. methods: data analysis – methods: numerical – methods: statistical – ISM: clouds – ISM: lines and bands – photon-dominated region (PDR)

1. Introduction

Many aspects of star and planet formation are only partially understood. Major questions remain with respect to the impact of turbulence and feedback mechanisms on star formation efficiency, as well as the development of chemical complexity alongside the evolution of interstellar matter from diffuse clouds to planet-forming disks. Recent large hyper-spectral surveys of giant molecular clouds and star forming regions are key to tackle these questions. For instance, the "Orion B" IRAM-30 m Large Program (Pety et al. 2017) covers the full scale of a giant molecular cloud (∼10 pc size) at a dense core resolution (<0.1 pc). It produces a hyper-spectral image with a million pixels and 200 000 spectral channels, making it possible to map the emission of dozens of molecules over the whole cloud. More generally, recent instruments such as the IRAM-30 m, ALMA, NOEMA, and the James Webb space telescope, have provided observation maps with thousands of pixels in multiple emission lines. For more details, we refer to the work of Habart et al. (2024), for instance.

Astrophysical codes for ISM environments can model observed regions and link observables (e.g., line intensities) to a few local physical parameters (e.g., the gas density). For

⋆ Corresponding author: palud@apc.in2p3.fr

instance, radiative transfer codes can be used to relate gas density, temperature and column densities of detected species to observable line intensities. Such codes include RADEX (van der Tak et al. 2007), RADMC-3D (Dullemond et al. 2012), LIME (Brinch & Hogerheijde 2010), MCFOST (Pinte et al. 2022), MOLPOP-CEP (Asensio Ramos & Elitzur 2018), and LOC (Juvela 2020). It is also possible to adopt a more holistic approach to model the ISM by taking multiple physical phenomena into account (e.g., large chemical networks, thermal balance, and radiative transfer), as well as their coupling. These more comprehensive codes are usually dedicated to a specific type of environment, described with specific parameters. Shock models, such as the Paris-Durham code (Godard et al. 2019) and the MAPPINGS code (Sutherland et al. 2018), compute observables in shock-dominated environments from shock speeds, pre-shock densities or magnetic field. Photodissociation regions (PDR) models such as the Meudon PDR code (Le Petit et al. 2006) and HII region models such as CLOUDY (Ferland et al. 2017) describe ultraviolet (UV) irradiated regions of the ISM and predict observables from the incident UV flux properties and gas properties (e.g., density or metallicity). For each model, small changes in the physical parameters can cause large changes in the predicted observables (Röllig et al. 2007).

The process of inference consists of finding the physical parameters that lead to predicted observations that are consistent with the actual observations. The most widespread approach for this adjustment is to solve a chi-square minimization problem, that is, a least-squares problem (see e.g., Wu et al. 2018). However, this approach relies on the assumption of a Gaussian additive noise, which is applicable to many ISM observations. In addition, it produces physical parameter maps that may not be physically consistent due to the existence of multiple solutions reconstructing the observations equally well. These issues are particularly present in low signal-to-noise ratio (S/N) regions, where observations mostly provide upper bounds on the true observable. Neighboring pixels in reconstructed maps thus might contain large physically-inconsistent differences. However, a chi-squared minimization approach does not ensure a satisfactory reconstruction even in high-S/N regimes, due to local minima in the cost function. These specificities become more problematic as the size of the observation map increases. Most of the estimation methods currently in use in ISM studies do not scale to observation maps of more than a few hundreds pixels.

In this work, we introduce BEETROOTS (BayEsian infErence with spaTial Regularization of nOisy multi-line ObservaTion mapS). This PYTHON software performs Bayesian reconstruction of maps of physical parameters from a set of observation maps and an astrophysical model. Our approach is based on a finer statistical modeling of the noise affecting the observations. Identifiability issues are mitigated using a spatial regularization that enables pixels to exploit the information contained in their neighboring pixels. This regularization favors smooth maps that are physically consistent and is particularly helpful for large diffuse regions where the parameters are assumed to vary smoothly. BEETROOTS runs the sampling algorithm introduced in Palud et al. (2023) and can thus provide much more information than optimization-based approaches, such as credibility intervals, covariance estimations, and detections of multiple solutions or degeneracies. Finally, it automatically quantifies and assesses the ability of the astrophysical model to explain the observations, using the Bayesian hypothesis testing approach introduced in Palud et al. (2023a). This model checking enables users to determine whether reconstructions are trustworthy. BEETROOTS is publicly available as a PYTHON package on PyPi[1].

In this paper, Section 2 reviews the general framework that we adopt for statistical modeling. Section 3 describes the inference and model checking methods implemented in BEETROOTS. The contributions in BEETROOTS are highlighted and compared with common choices in ISM studies. Section 4 describes the specificities considered to illustrate the proposed package, such as the astrophysical model and noise properties. Section 5 applies the method to a synthetic yet realistic case at different spatial resolutions. Section 6 applies the method to Orion molecular cloud 1 observation maps. Section 7 provides concluding remarks. Appendix A explains how to use BEETROOTS in practice.

2. Statistical modeling for studies of the ISM

This work considers observations, $\mathbf{Y}$, which correspond to a set of either the integrated intensities of emission lines, their ratios, or the full line profiles. An astrophysical model $\mathbf{M}$ simulates the physics of the observed region. It is assumed to predict the observables, $\mathbf{Y}$, from a set of physical parameters, $\boldsymbol{\Theta}$, including for instance the volume density, the thermal pressure, the visual extinction, or the radiative field intensity. The goal of BEETROOTS is to infer the physical parameters, $\boldsymbol{\Theta}$, and the associated uncertainties from the observation, $\mathbf{Y}$, and the astrophysical model, $\mathbf{M}$, and to assess whether the reconstruction actually explains the observation. To do so, we resorted to a Bayesian framework.

This section presents the key components of statistical modeling and standard choices in ISM studies. First, we introduce fundamental notions of Bayesian statistics and draw their relative links – see Robert & Casella (2004, chapter 1) for a more complete introduction. Then, we discuss the main advantages and drawbacks of the standard choices in ISM studies for the astrophysical model, the uncertainties on observations, and the prior distribution. For more details on the statistical model adopted in BEETROOTS, we refer to Palud (2023).

2.1. Bayesian modeling

The Bayesian framework is a statistical approach that encodes uncertainty using random variables and probability distributions. It models the physical parameters to infer, $\boldsymbol{\Theta}$, and the observation, $\mathbf{Y}$, as random variables. In a nutshell, Bayesian statistical modeling is aimed at building a posterior distribution, $\pi(\boldsymbol{\Theta}|\mathbf{Y})$. It involves the definition of a likelihood function $\pi(\mathbf{Y}|\boldsymbol{\Theta})$ – from a deterministic astrophysical model, $\mathbf{M}$, and a deterioration, $\mathcal{D}$, which includes the noise, for instance; in addition, it includes the choice of a prior distribution, $\pi(\boldsymbol{\Theta})$.

The likelihood function, $\pi(\mathbf{Y}|\boldsymbol{\Theta})$, evaluates how well the parameter, $\boldsymbol{\Theta}$, reconstructs the observations, $\mathbf{Y}$. It is the probability density function (PDF) of the observations, $\mathbf{Y}$, given a fixed value for the parameters $\boldsymbol{\Theta}$, according to an observation model:

$$\mathbf{Y} = \mathcal{D}\left[\mathbf{M}(\boldsymbol{\Theta})\right], \tag{1}$$

where $\mathbf{M}$ is the deterministic astrophysical model and $\mathcal{D}$ is a deterioration that can include measurement noise, calibration errors, modeling error from the choice of astrophysical model, or upper limits on observables.

[1] https://pypi.org/project/beetroots/. All results in this work are obtained with version 1.1.0.

The prior distribution, $\pi(\boldsymbol{\Theta})$, encodes prior information on the physical parameters of interest, $\boldsymbol{\Theta}$. It may be non-informative (e.g., a uniform distribution) or informative (e.g., to include the results of other trusted studies or to constrain acceptable solutions to be physically meaningful). For informative priors, the negative log-prior, that is, the function $\boldsymbol{\Theta} \mapsto -\log \pi(\boldsymbol{\Theta})$, is called the regularization function.

In contrast, the distribution of interest for inference is the posterior distribution, that is, the probability of $\boldsymbol{\Theta}$ knowing the observation, $\mathbf{Y}$. It combines the likelihood function and the prior distribution following the Bayes theorem:

$$\pi(\boldsymbol{\Theta}|\mathbf{Y}) = \frac{\pi(\mathbf{Y}|\boldsymbol{\Theta})\,\pi(\boldsymbol{\Theta})}{\pi(\mathbf{Y})}. \tag{2}$$

The Bayesian evidence, $\pi(\mathbf{Y})$, is a normalization constant that does not depend on $\boldsymbol{\Theta}$. In inference, Eq. (2) is thus simplified to

$$\pi(\boldsymbol{\Theta}|\mathbf{Y}) \propto \pi(\mathbf{Y}|\boldsymbol{\Theta})\ \pi(\boldsymbol{\Theta}). \tag{3}$$

Estimations of the physical parameters, $\boldsymbol{\Theta}$, are derived from the likelihood function, $\pi(\mathbf{Y}|\boldsymbol{\Theta})$, or from the posterior, $\pi(\boldsymbol{\Theta}|\mathbf{Y})$. Uncertainty quantifications are extracted from the posterior. Before describing how in Sect. 3, we discuss the choices of astrophysical model, $\mathbf{M}$, random deterioration, $\mathcal{D}$, and prior, $\pi(\boldsymbol{\Theta})$.

2.2. The astrophysical model

The astrophysical model, $\mathbf{M}$, predicts observables from physical parameters. In ISM studies, some relatively simple models such as RADEX (van der Tak et al. 2007) and UCLCHEM (Holdship et al. 2017) are sometimes used directly in inference (Behrens et al. 2022; Keil et al. 2022). More comprehensive codes provide more reliable predictions at the cost of longer evaluation times. Such codes are usually dedicated to a specific physical regime, such as photodissociation regions (PDRs) for Kosma-τ (Röllig & Ossenkopf-Okada 2022) or the Meudon PDR code (Le Petit et al. 2006), HII regions for CLOUDY (Ferland et al. 2017), or shock-dominated regions for the Paris-Durham code (Godard et al. 2019). These codes usually take as the input some parameters at the cloud scale and return emissions integrated over the full cloud. It is common in ISM studies to associate these inputs and outputs with pixels in an observation map (see, e.g., Wu et al. 2018). That is to say, for a map of N pixels, the full set of physical parameters, $\boldsymbol{\Theta}$, can be broken down into N vectors $\boldsymbol{\theta}_n$ (each corresponding to a pixel) of D physical parameters (e.g., gas density). Similarly, the full set of observations, $\mathbf{Y}$, can be broken down into N vectors $\mathbf{y}_n$ of L observables. For each pixel, the astrophysical model outputs $\mathbf{M}(\boldsymbol{\theta}_n) = (M_\ell(\boldsymbol{\theta}_n))_{\ell=1}^{L}$, which is to be compared with the observation of the corresponding pixel $\mathbf{y}_n = (y_{n\ell})_{\ell=1}^{L}$. Therefore, for a map with 100 pixels, comparing the physical parameters and observations requires 100 evaluations of the astrophysical model; namely, one evaluation per pixel. In addition, as shown in Sect. 3, extracting estimators from the posterior distribution requires many evaluations of the likelihood function and, thus, of the astrophysical model for each pixel. The astrophysical model thus needs to be fast while providing accurate predictions.

BEETROOTS was implemented assuming that an evaluation of the astrophysical model is fast (less than 0.1 second) and can be processed in batches. These specifications are nearly impossible to satisfy with a comprehensive numerical code and require the use of a fast and accurate emulator. Using emulators built from precomputed model grids is common practice in ISM studies, using interpolation methods (Galliano 2018; Wu et al. 2018; Ramambason et al. 2022), artificial neural networks (ANN, Grassi et al. 2011; de Mijolla et al. 2019; Holdship et al. 2021; Grassi et al. 2022; Palud et al. 2023b), or other machine learning approaches (Smirnov-Pinchukov et al. 2022; Bron et al. 2021).

2.3. The random deterioration and noise model

To compare the observations, $\mathbf{Y}$, with the predictions of the astrophysical model, $\mathbf{M}(\boldsymbol{\Theta})$, it is necessary to describe the random deteriorations (i.e., noise) that affect the observations. This section reviews common noise models, $\mathcal{D}$, in the ISM community and those implemented in BEETROOTS.

The most widespread random deterioration in ISM studies is Gaussian and additive (Galliano et al. 2003; Chevallard et al. 2013; Pérez-Montero 2014; Chevance et al. 2016; Wu et al. 2018; Lee et al. 2019; Roueff et al. 2021; Keil et al. 2022; Behrens et al. 2022). For a pixel $1 \leq n \leq N$ and an observable $1 \leq \ell \leq L$, it is expressed as:

$$y_{n\ell} = M_\ell(\boldsymbol{\theta}_n) + \varepsilon_{n\ell}^{(a)}, \quad \boldsymbol{\varepsilon}^{(a)} \sim \mathcal{N}(0, \boldsymbol{\Sigma}^{(a)}), \tag{4}$$

where $\boldsymbol{\varepsilon}^{(a)}$ is a measurement noise (typically thermal) with zero mean and covariance matrix $\boldsymbol{\Sigma}^{(a)}$. The Gaussian noise is sometimes considered on the log of the observations, $\log y_{n\ell}$ (Vale Asari et al. 2016), which is equivalent to a multiplicative error, $\boldsymbol{\varepsilon}^{(m)}$, following a lognormal distribution. In all these works, the covariance matrix, $\boldsymbol{\Sigma}^{(a)}$, is assumed to be diagonal, that is, all the noise components are independent. An additive Gaussian model is often a good approximation thanks to the central limit theorem. Besides, it is simple to manipulate. However, it fails to accurately capture all uncertainties associated with ISM observations.

To consider a more realistic random deterioration, some ISM studies combine a Gaussian additive noise, $\boldsymbol{\varepsilon}^{(a)}$, and a Gaussian multiplicative noise, $\boldsymbol{\varepsilon}^{(m)}$. This multiplicative noise represents calibration errors (Gordon et al. 2014; Ciurlo et al. 2016; Galliano 2018; Galliano et al. 2021) or modeling errors in the astrophysical model, $\mathbf{M}$ (Blanc et al. 2015; Vale Asari et al. 2016; Jóhannesson et al. 2016). The observation model then reads:

$$y_{n\ell} = \varepsilon_{n\ell}^{(m)} M_\ell(\boldsymbol{\theta}_n) + \varepsilon_{n\ell}^{(a)},\ \boldsymbol{\varepsilon}^{(m)} \sim \mathcal{N}(1, \boldsymbol{\Sigma}^{(m)}), \boldsymbol{\varepsilon}^{(a)} \sim \mathcal{N}(0, \boldsymbol{\Sigma}^{(a)}). \tag{5}$$

Using a Gaussian model for multiplicative noise simplifies computations, as the overall uncertainty model is Gaussian. Although practical, a Gaussian model may not accurately account for multiplicative noise depending on the standard deviation (STD) of the multiplicative error. In particular, a Gaussian distribution allows for unphysical negative multiplicative factors to be included. Figure 1 illustrates that the Gaussian approximation for multiplicative errors is relevant for small errors, but inappropriate for large ones. In statistics, a typical model for multiplicative noise is the lognormal distribution, that is, a Gaussian distribution for the multiplying factor on the log scale. A deterioration combining an additive Gaussian noise and a multiplicative lognormal noise has already been used in ISM (Kelly et al. 2012). We refer to Palud (2023, chapter 3) for more details on how this deterioration arises in the context of integrated intensities.

In BEETROOTS, the user can choose a deterioration involving an additive Gaussian uncertainty only (Eq. (4)) or along with multiplicative noise, Gaussian (Eq. (5)), or lognormal (Eq. (10)). It can also handle upper limits on observations, as described in Appendix B.1 (Eq. (B.2)). The current implementation assumes that additive and multiplicative errors are independent.

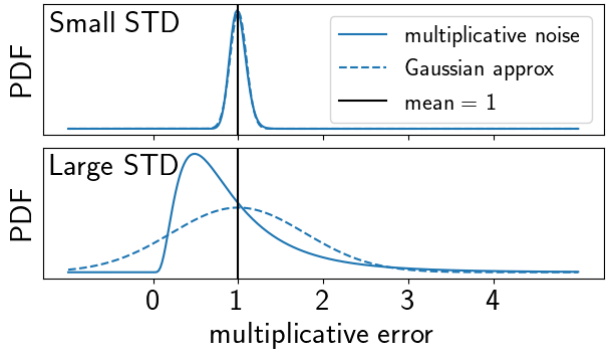


Fig. 1. Quality of a Gaussian approximation to a lognormal distribution for low and high STD. Top: multiplicative noise associated with an average error of 10%. The lognormal distribution is well approximated by a normal distribution. Bottom: multiplicative noise associated with an average error of a factor of two. The PDF of the lognormal distribution is significantly asymmetric and thus poorly approximated by a normal distribution.

2.4. The prior distribution

The prior distribution, $\pi(\boldsymbol{\Theta})$, encodes a priori information on the maps of physical parameters that we wish to reconstruct. In BEETROOTS, we combine a uniform prior on validity intervals with a spatial regularization prior. The first guarantees that the estimated values lie in a meaningful range. The second guides reconstructions towards spatially smooth maps and removes spurious local minima.

Using validity intervals on the components of $\boldsymbol{\Theta}$ is common in ISM studies (see, e.g., Behrens et al. 2022; Blanc et al. 2015; Thomas et al. 2018; Holdship et al. 2018; Wu et al. 2018). Validity intervals are also implicitly involved when working on a lattice dataset. In a Bayesian framework, wide validity intervals are usually used to define a weakly informative uniform prior that does not bias estimations. In BEETROOTS, a uniform prior can be set on the linear or log scale, depending on the physical parameter and its dynamic range.

Spatial regularizations have been used in ISM studies to fit Gaussian line profiles on hyperspectral observations (Paumard et al. 2014; Ciurlo et al. 2016; Marchal et al. 2019; Paumard et al. 2022). Assuming that spatial variations of physical parameters in low-S/N regions are smooth, we used an L_2 penalty on the discrete gradient of each parameter map. For each physical parameter map, this spatial regularization is weighted by a hyperparameter, $\tau_d > 0$. We consider equal values τ_d for all d, since they all describe the same cloud and, thus, correspond to the same typical size. Such a regularization improves reconstructions particularly in the low-S/N regions, where the observations mostly contain noise. Concretely, spatial regularization enables pixels to exploit the information contained in their neighbors, which improves estimations. In low-S/N regions, which typically contain more diffuse gas with larger spatial scales, the regularization orients the likelihood and can remove unphysical solutions. In high-S/N pixels, the likelihood overcomes the regularization and dominates the posterior.

Informative priors such as spatial regularization require careful hyperparameters tuning, as different values yield different trade-offs between prior and likelihood. For instance, Ciurlo et al. (2016) presents a manual setting of the six hyperparameters of its spatial regularization. Hierarchical Bayesian models infer the prior parameters from the data along with the physical parameters of interest, $\boldsymbol{\Theta}$. In ISM studies, hierarchical models are mostly used in dust studies (Kelly et al. 2012; Juvela et al. 2013; Veneziani et al. 2013; Galliano 2018; Galliano et al. 2021). In this work, we estimate the regularization weights, τ_d, along with physical parameters, $\boldsymbol{\Theta}$, with sampling methods (see Sect. 3.2) by maximizing the Bayesian evidence, $\pi(\mathbf{Y})$ (Eq. (3)), as proposed in Vidal et al. (2020). For the optimization methods (Sect. 3.1), we set them manually. In practice, additive error bars in integrated intensity observations sometimes contain multiplicative noise, resulting in overestimated additive standard deviations, $\sigma_{a,n\ell}$. For instance, the STDs $\sigma_{a,n\ell}$ considered in Joblin et al. (2018) or Wu et al. (2018) are correlated with the intensities, which indicates that these STDs include multiplicative errors. In such conditions, adjusting automatically τ_d can lead to over-smoothed maps.

3. Beetroots: Performing estimations and model checks

Building on the described statistical model, this section details some methods to infer physical parameters, $\boldsymbol{\Theta}$, from observations, $\mathbf{Y}$, with a focus on tendencies in ISM studies. First, we depict optimization methods that are usually fast. Then, we present sampling methods that are slower but more informative, as they natively provide uncertainty quantifications together with the estimates. Both approaches can be used with BEETROOTS. For more details on the estimation method adopted in BEETROOTS, we refer to Palud et al. (2023). Finally, we describe the model checking approach implemented in BEETROOTS, first presented in Palud et al. (2023a).

3.1. Optimization-based estimation

Physical parameter estimations (i.e., reconstructions) can be extracted either from the likelihood or the posterior distribution. As their name suggests, the maximum likelihood estimator (MLE), $\widehat{\boldsymbol{\Theta}}_{\mathrm{MLE}}$, and the maximum a posteriori (MAP), $\widehat{\boldsymbol{\Theta}}_{\mathrm{MAP}}$, are the values of $\boldsymbol{\Theta}$ that achieve the maximum value for the likelihood and for the posterior probability density function, respectively. These estimators are often evaluated by minimizing the negative logarithm of these functions:

$$\widehat{\boldsymbol{\Theta}}_{\mathrm{MLE}} = \arg\min_{\boldsymbol{\Theta}} \left[-\log \pi(\mathbf{Y}|\boldsymbol{\Theta}) \right], \tag{6}$$

$$\widehat{\boldsymbol{\Theta}}_{\mathrm{MAP}} = \arg\min_{\boldsymbol{\Theta}} \left[-\log \pi(\boldsymbol{\Theta}|\mathbf{Y}) \right], \tag{7}$$

with $-\log \pi(\boldsymbol{\Theta}|\mathbf{Y}) = -\log \pi(\mathbf{Y}|\boldsymbol{\Theta}) - \log \pi(\boldsymbol{\Theta})$ up to an additive constant (Eq. (3)). In this context, the negative log-likelihood and the negative log-posterior are often called "loss function" or "cost function". The MLE is, by definition, the value of $\boldsymbol{\Theta}$ that leads to the predictions, $\mathbf{M}(\boldsymbol{\Theta})$, that are most compatible with the observations, $\mathbf{Y}$. It is sensitive to noise and hence unstable in low-S/N regions or when the observables are not good tracers of the physical parameters of interest. The MAP defines a trade-off between fitting the observation and satisfying prior knowledge. It leads to more robust results. Both the MLE and MAP are widespread estimators in ISM studies.

Solving the optimization problems in Eqs. (6) and (7) is challenging due to the potential existence of multiple local minima in the cost function. Three main strategies prevail in ISM studies: a discrete search within a grid of models (Sheffer et al. 2011; Sheffer & Wolfire 2013; Joblin et al. 2018; Lee et al. 2019), meta-heuristics (Möller et al. 2013), and gradient descent algorithms (Schilke et al. 2010; Galliano et al. 2003;

Paumard et al. 2022; Wu et al. 2018), such as Levenberg-Marquardt. Gradient descent algorithms fail to escape local minima and may return unphysical results. Discrete searches and meta-heuristics approaches may escape from local minima, but their computational cost becomes prohibitive when the dimension of $\boldsymbol{\Theta}$ exceeds 10. When used for optimization, BEETROOTS combines a gradient descent for fast convergence with a global exploration step to escape from local minima (Sect. 3.3).

The main limitation of optimization methods is that they do not natively provide uncertainty quantifications on $\boldsymbol{\Theta}$ associated with estimations. In this case, producing uncertainty quantifications requires the use of additional methods. The Cramér–Rao bound is sometimes used (Roueff et al. 2021), but it is only relevant when the posterior is well approximated by a Gaussian at its mode, which is often not the case in astrophysics, as noted by Panter et al. (2003). In contrast, sampling-based approaches natively provide uncertainty quantifications along with point estimates for general posterior distributions.

3.2. Sampling-based estimation

A Bayesian framework usually exploits the full posterior distribution to define estimators. It is the case of the posterior mean, also called minimum mean squared error (MMSE), $\widehat{\boldsymbol{\Theta}}_{\mathrm{MMSE}} = \mathbb{E}[\boldsymbol{\Theta}|\mathbf{Y}]$. Uncertainties on physical parameters (which can help to identify the existence of multiple solutions or degeneracies among physical parameters) can be quantified in multiple ways. The covariance matrix provides a typical error for each parameter and the correlations between pairs of parameters. However, it is mostly relevant for nearly Gaussian distributions. Credibility intervals yield lower and upper limits on each physical parameter, without any assumption on their probability distribution. Both quantities require evaluating integrals over multiple variables. In ISM studies, these integrals are sometimes evaluated with an integration on a discrete grid (Da Cunha et al. 2008; Pacifici et al. 2012; Pérez-Montero 2014; Blanc et al. 2015; Vale Asari et al. 2016; Thomas et al. 2018; Villa-Vélez et al. 2021), which does not scale to high dimensions. An alternative approach resorts to Monte Carlo (MC) estimators, computed from samples of the posterior distribution. Noting T_{MC} the number of samples $\boldsymbol{\Theta}^{(t)} \sim \pi(\boldsymbol{\Theta}|\mathbf{Y})$, the MMSE is estimated as

$$\widehat{\boldsymbol{\Theta}}_{\mathrm{MMSE}} \simeq \frac{1}{T_{\mathrm{MC}}} \sum_{t=1}^{T_{\mathrm{MC}}} \boldsymbol{\Theta}^{(t)}. \tag{8}$$

The generation of these samples is often performed iteratively with a Markov chain Monte Carlo (MCMC) algorithm (Robert & Casella 2004). At each iteration, a transition kernel first generates a candidate from a proposal distribution, and then accepts or rejects this candidate with a certain probability that involves the ratio of the posterior PDF of the candidate and of the previous iterate. Candidates with low posterior PDF are thus likely to be rejected. Therefore, the proposal distribution needs to generate candidates with high posterior PDF, while sufficiently exploring the parameter space to be able to escape from a local mode.

MCMC algorithms were first popularized in astronomy through cosmology in Christensen et al. (2001). The first public MCMC code, COSMOMC, was published in Lewis & Bridle (2002). Both articles used random walk Metropolis-Hastings (RWMH) to generate posterior samples, arguably the most widespread transition kernel. The RWMH kernel was also applied in ISM studies (Makrymallis & Viti 2014; Paradis et al. 2010). Although it can be applied to maps of physical parameters, this transition kernel becomes quickly inefficient in dimensions higher than about 20. In BEETROOTS, we have implemented a new and efficient MCMC algorithm that scales to large maps. Other sampling methods are also already popular in the ISM community. Appendix B.2 lists such algorithms and their applications in ISM studies and discusses their limitations in comparison with the sampler implemented in BEETROOTS (see Sect. 3.3).

Exhaustively describing a quantification of uncertainties can be quite tedious. For instance, for D maps of physical parameters with N pixels, a complete covariance matrix would contain $ND(ND+1)/2$ different terms. We favor credibility intervals (CIs) that allow us to reduce the number of uncertainty quantification terms to $2ND$; namely, one map for the CI lower bound and one map for its upper bound. We further reduce the number of terms to ND (that is, one single map per physical parameter) by quantifying the CI size. For a 95% CI, $[q^{(\mathrm{l},95)}, q^{(\mathrm{u},95)}]$, a natural description of the CI size would involve the difference $0.5(q^{(\mathrm{u},95)} - q^{(\mathrm{l},95)})$, where the 0.5 factor returns the distance from the interval center. However, in ISM studies, the range of physical parameters such as the thermal pressure, P_{th}, covers numerous orders of magnitude, such that the difference $0.5(q^{(\mathrm{u},95)} - q^{(\mathrm{l},95)}) \simeq 0.5q^{(\mathrm{u},95)}$ is not informative. Therefore, for strictly positive physical parameters, using this difference in logarithmic scale is more relevant. As we describe all our estimators in linear scale in the following applications, we summarize the CI size using a quantity we call "uncertainty factor" (UF) defined as the exponential of this difference in logarithmic scale

$$\mathrm{UF}\ 95\% = \exp\left\{\frac{1}{2}\left(\log q^{(\mathrm{u},95)} - \log q^{(\mathrm{l},95)}\right)\right\} = \sqrt{\frac{q^{(\mathrm{u},95)}}{q^{(\mathrm{l},95)}}}. \tag{9}$$

For instance, an uncertainty factor of two indicates that an estimation is associated with an uncertainty of a factor of two.

3.3. Estimation methods in BEETROOTS

In ISM studies, the numerical model is almost always non-linear, which can causes the negative log-likelihood (Eq. (6)) and negative log-posterior (Eq. (7)) to have many local minima. Common samplers, such as RWMH, usually fail to escape from a local minimum and should be avoided. Besides, with many pixels, the dimension of the parameter space becomes very large and usual quadrature-based numerical integration becomes intractable.

BEETROOTS implements an MCMC algorithm introduced in Palud et al. (2023), which specifically addresses these two difficulties. It relies on two transition kernels: multiple-try Metropolis kernel (MTM), which performs a "global exploration," and preconditioned Metropolis-adjusted Langevin algorithm (PMALA), which performs a "local exploration." At each step t, the transition kernel is randomly selected with probability p_{MTM} for MTM and $1 - p_{\mathrm{MTM}}$ for PMALA. Figure 2 illustrates the principle of these two kernels and applies them to a two-dimensional Gaussian mixture model restricted to a square.

The MTM kernel enables the escape from local minima. For each pixel, it randomly proposes $K \geq 1$ candidates, and selects one with a high probability density. The proposal distribution (used to generate candidates) is defined from the prior. When the prior includes spatial regularization, the proposal distribution for each pixel is a Gaussian mixture defined from the neighboring pixels. This choice guides the reconstruction towards smooth maps. When the prior is a uniform distribution on validity intervals, this proposal distribution is set to the same

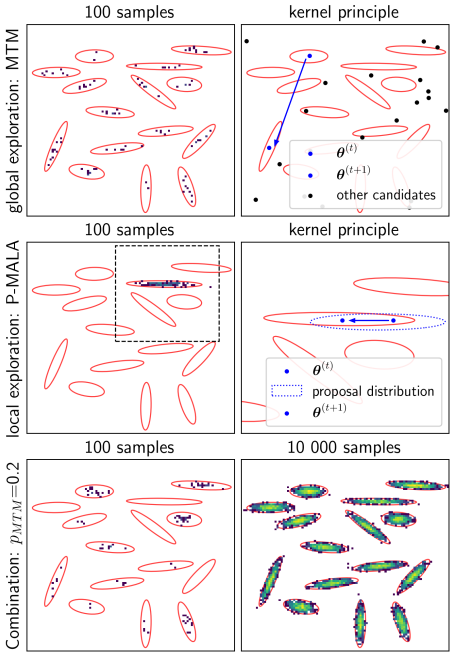


Fig. 2. Illustration of the two transition kernels PMALA and MTM on a two-dimensional probability distribution: a Gaussian mixture model (shown with the red ellipses) restricted to validity intervals (shown with the large black square). Top: MTM update rule (on the right), and a histogram after 100 sampling steps (on the left). With this update rule only, the sampling visits all the distribution modes at the cost of computationally heavier individual steps. Middle: PMALA update rule (on the right), and a histogram after 100 sampling steps (on the left). With this cheaper update rule only, the sampling fails to visit all the distribution modes. Bottom: Two histograms obtained using both kernels after 100 steps (left) and 10 000 steps (right). The histograms converge to the correct distribution.

uniform distribution. This is the case in single-pointing observations (i.e., one-pixel maps) or when no spatial regularization is considered. This is also the case on Fig. 2 where candidates are generated from a uniform distribution on the large black square. Even with 100 iterations, this kernel manages to visit all the modes. However, generating many candidates at each iteration is expensive.

The PMALA kernel efficiently explores the posterior distribution locally by exploiting its geometry. Similarly to a gradient descent algorithm, it relies on a step-size $\eta > 0$. It is faster than the MTM kernel, as it only generates one candidate. However, with local exploration only, the sampling fails to visit other modes, and stays trapped in the mode closest to the initialization.

Combining PMALA with MTM exploits the best of the two methods. As shown on Fig. 2, the combination leads to sample histograms converging to the correct distribution, at a cheaper computational cost than with MTM alone.

The implementation of BEETROOTS is general enough so that it can be used for optimization, to return either the MAP or the MLE. In this case, PMALA boils down to an efficient preconditioned gradient descent update step, and MTM, to keeping the candidate with the lowest cost only if this cost is lower than that of the current iterate. When $0 < p_{\mathrm{MTM}} < 1$, the optimization combines a fast convergence to a local minimum with the ability to escape from local minima. However, the parameter p_{MTM} can also be set to 0 or 1 to exploit only one of the two transition kernels. In practice, setting p_{MTM} to 0.1, 0.2 or 0.5 exploits both kernels well. For the MTM kernel, we recommend setting the number of candidates to a few tens if considering spatial regularization and to hundreds or thousands if not, as the spatial regularization guides the generation of candidates. For the PMALA kernel, the step size, η, should be set so that the sampling has an accept rate of about 60% ($\eta = 10^{-2}$–10^{-3} has been observed to yield such an accept rate in the experiments reported in Sect. 5).

3.4. Methods: Model checking

Once the inference is performed, the quality of the reconstruction needs to be assessed. Model selection is a quite common approach in astrophysics (Ramambason et al. 2022; Chevallard & Charlot 2016; Zucker et al. 2021; Kamenetzky et al. 2014). However, it only compares different models, assessing their relative success, as the absolute values of the associated criteria are not interpretable. Conversely, model checking approaches assess a model individually, and yield interpretable diagnoses on whether a model can explain the observations.

In ISM studies, the obtained loss function value in a non-linear least squares problem, often noted χ^2, is used in a three-case interpretation. When $\chi^2 > 1$, the estimated parameters are judged not able to reproduce observations. A $\chi^2 \ll 1$ suggests an overfit or an overestimation of uncertainties on observations. The ideal case, $\chi^2 \simeq 1$, indicates that the physical parameters reasonably reproduce the observations. This interpretation appears for instance in Chevance et al. (2016); Joblin et al. (2018); Villa-Vélez et al. (2021). However, it comes with limitations and was criticized in the astrophysics community (Andrae et al. 2010). In particular, the degree of freedom is challenging to estimate for non-linear astrophysical models and when a prior distribution is considered. Besides, this χ^2 rule only applies to Gaussian noise. Finally, when $\boldsymbol{\Theta}$ is described by a posterior distribution and not by a single estimation $\widehat{\boldsymbol{\Theta}}$, the obtained χ^2 value is approximated with an MC estimator. The error associated with this approximation is seldom accounted for in astrophysics.

Bayesian model checking (Gelman et al. 1996), also known as posterior predictive checking, is a statistically principled model checking technique that was also considered in ISM studies (Chevallard & Charlot 2016; Galliano et al. 2021; Lebouteiller & Ramambason 2022). It is a hypothesis testing method that tests the following hypothesis:

Hypothesis 1 (H1) The astrophysical model, $\mathbf{M}$, random deterioration, $\mathcal{D}$, and prior, $\pi(\boldsymbol{\Theta})$, can reproduce the observations, $\mathbf{Y}$.

To do so, reproduced observations are generated from the posterior samples, $\boldsymbol{\theta}^{(t)}$, using the observation model (Eq. (1)). The distribution of these noisy predicted observations is called the "posterior predictive distribution." The goal is then to compare the posterior predictive distribution and the true observations, $\mathbf{Y}$, with respect to some discrepancy measure, T, on $(\mathbf{Y}, \boldsymbol{\Theta})$. The results of this comparison are summarized in a p-value defined as the probability of the parameters $\boldsymbol{\Theta}$ yielding observations that are worse (in the sense of the chosen discrepancy measure) than the one we have observed. H1 is then rejected with a confidence level $1 - \alpha$ when $p \leq \alpha$. The posterior predictive distribution can also be visualized to get a more detailed diagnosis, for instance, to identify the observables that led to a model rejection.

In BEETROOTS, to ensure that the p-value is consistent with the likelihood function in general cases, the discrepancy measure is set to the negative log-likelihood: $T(\mathbf{Y}, \boldsymbol{\Theta}) = -\log \pi(\mathbf{Y}|\boldsymbol{\Theta})$. As all noise realizations are assumed independent, we can test

H1 for each pixel and provide maps of p-values. This approach helps to identify regions that are poorly modeled by the astrophysical model, $\mathbf{M}$. The implemented approach also accounts for the uncertainties on the p-value that come from the MC evaluation. This avoids rejections that are due to chance and caused by insufficient number of samples. More details on the implemented model checking approach can be found in Palud et al. (2023a).

4. Experimental setting

This section describes two applications: a synthetic case, where the ground truth is known and observations of the Orion molecular cloud 1 (OMC-1). In both applications, we consider observation maps of integrated intensities, $\mathbf{Y} \in \mathbb{R}^{N \times L}$, that contain N pixels and L channels, and analyze them with the Meudon PDR code, $\mathbf{M}$. For a pixel n, we assume the Meudon PDR code to be able to reproduce the line integrated intensities, $\mathbf{y}_n$, from a set of D physical parameters, $\boldsymbol{\theta}_n \in \mathbb{R}^D$, such as the thermal pressure. The goal of both applications is to reconstruct physical parameter maps, $\boldsymbol{\Theta} = (\boldsymbol{\theta}_n)_{n=1}^N$, and the associated uncertainties.

This section presents the likelihood elements that we use in the two applications, namely, the Meudon PDR code, its ANN emulator, and the random deterioration. It also introduces the estimators to be evaluated and compared.

4.1. The Meudon PDR code

The Meudon PDR code[2] (Le Petit et al. 2006) is a one-dimensional stationary code that simulates a PDR, that is, neutral interstellar gas illuminated with a UV radiation field. It allows for the investigation of the radiative feedback of a newborn star on its parent molecular cloud, for instance. Although primarily designed for PDRs, it can also simulate a wide variety of environments, such as diffuse clouds, nearby galaxies, damped Lyman alpha systems, and protoplanetary disks.

The user specifies physical conditions such as the intensity of the incident UV radiation field, G_0, the elemental abundances, and the depth of the slab of gas expressed as its total visual extinctions, A_V^{tot}. The user can also choose the thermal pressure in the cloud, P_{th}(for isobaric models), or the gas volume density (for constant density models). The code then solves multiphysics coupled balance equations of radiative transfer, thermal balance, and chemistry on an adaptive spatial grid over a one-dimensional slab of gas. For the radiative transfer equation, the code considers absorption in the continuum by dust and gas and in the lines of key atoms and molecules such as H and H_2 (Goicoechea & Le Bourlot 2007). For thermal balance, it computes the gas and grain temperatures from the specific intensity of the radiation field. The code accounts for many heating and cooling processes, in particular photoelectric and cosmic ray heating, and line cooling. For chemistry, the code provides the densities of about 200 species at each position. About 3000 reactions are considered, both in the gas phase and on the grains. The chemical reaction network was built combining sources such as the KIDA database (Wakelam et al. 2012), the UMIST database (McElroy et al. 2013), and articles. For key photoreactions, cross sections are taken from Heays et al. (2017) and from Ewine van Dishoeck's photodissociation and photoionization database[3]. The successive resolution of these three coupled aspects is iterated until reaching a global stationary state.

The code yields 1D-spatial profiles of the volume density of all chemical species and of temperature of both grains and gas as a function of depth in the PDR. From these spatial profiles, it also computes the line integrated intensities emerging from the cloud that can be compared to observations. As of version 7 (released in 2024), thousands line intensities are predicted from species such as H_2, HD, H_2O, C^+, C, CO, ^{13}CO, $C^{18}O$, $^{13}C^{18}O$, SO, HCO^+, OH, HCN, HNC, CH^+, CN, or CS.

4.2. The neural network emulator

A single full run of the Meudon PDR code typically lasts a few hours for one input vector, $\boldsymbol{\theta}_n$. Running a full reconstruction with the original code would thus be very slow, as it requires thousands of astrophysical model evaluations, $M_\ell(\boldsymbol{\theta}_n)$. In this work, we use the fast, light (memory-wise) and accurate ANN approximation of the Meudon PDR code proposed in Palud et al. (2023b). This ANN emulator was built with the NNBMA software[4] (neural-network-based model approximation).

This approximation was generated for isobaric models. It is valid for $\log_{10} P_{\mathrm{th}} \in [5, 9]$, $\log_{10} G_0 \in [0, 5]$, $\log_{10} A_V^{\mathrm{tot}} \in [0, \log_{10}(40)]$. As ANNs can process multiple inputs at once in batches, the evaluation of 10^3 input parameter vectors with this approximation lasts about 10 ms on a personal laptop. For the lines studied in this paper and chosen validity intervals, the emulator has an average error of 3.5%, which is three times lower than the typical calibration errors considered for our applications. In the remainder of this paper, to simplify notation, we denote this ANN approximation as $\mathbf{M}$.

Some secondary parameters were set to standard values in the Meudon PDR grid used to train our ANN emulator. For instance, we considered standard galactic grains and the average galactic dust extinction curve, and we set $R_V = 3.1$ (Fitzpatrick & Massa 2007), $N_H/E(B-V) = 5.8 \times 10^{21}$ cm^{-2} (Bohlin et al. 1978), and the cosmic ray ionization rate to 10^{-16} s^{-1} per H_2 (Le Petit et al. 2004; Indriolo et al. 2007). Finally, the interstellar standard radiation field was set to that of Mathis et al. (1983). Although these standard values might not be the most adequate for a specific environment, they represent a relevant starting point to infer the considered main parameters over wide maps that cover dense and diffuse gas regions[5]. This is the case, for instance, for the synthetic case (Sect. 5) and for the OMC-1 map (Sect. 6).

4.3. Noise model and characteristics

As indicated in Sect. 2.3, here we consider a combination of additive Gaussian and multiplicative lognormal uncertainties. The additive noise contains thermal noise. The multiplicative noise contains calibration error and model mis-specification. We assume no model mis-specification for synthetic observations,

[2] https://ism.obspm.fr/pdr.html
[3] https://home.strw.leidenuniv.nl/~ewine/photo/index.html
[4] https://github.com/einigl/ism-model-nn-approximation
[5] A better approach would include these secondary parameters as free parameters in the grid used for training, and then infer them with BEETROOTS. However, including many free parameters is expensive, as the number of Meudon PDR evaluations necessary to train a neural network emulator scales exponentially with the number of free parameters. Besides, the considered observations of OMC-1 (Sect. 6) only contain 5 lines per pixel, and we chose to keep the number of parameters to estimate per pixel lower than the number of lines. Therefore, we focused on three physical parameters from the Meudon PDR code, namely P_{th}, G_0, and A_V^{tot}, and set the others to standard values from the literature.

since the data are generated with the Meudon PDR code emulator. For real observations, the model mis-specification STD σ_{model} is set so that a 3σ error corresponds to an error of a factor of 3, that is, $\sigma_{\text{model}} = \frac{1}{3}\log 3$. All noise realizations are assumed to be independent.

To account for beam dilution and geometrical effects (viewing angle, irradiation angle, and multiple overlapping PDR fronts), we consider a last multiplicative parameter, $\boldsymbol{\kappa} = (\kappa_n)_{n=1}^N$, that affects all lines identically in a pixel. This scaling parameter, κ_n, is introduced in Sheffer & Wolfire (2013) as a product of four terms representing the four effects listed above, which can be smaller than 1 (e.g., if beam dilution dominates) or larger than 1 (e.g., if limb brightening dominates, due to a very oblique viewing angle). Joblin et al. (2018) considered a similar multiplicative parameter. As in Joblin et al. (2018), we inferred this parameter. We consider only values of κ of the order of unity to be physically reasonable and, thus, we assume that $\log_{10}\kappa_n \in [-1, 1]$. Using a single multiplicative parameter is only an approximation; for a limb brightening, it assumes that all lines are optically thin, while for beam dilution, it assumes that they are all emitted in a similar fraction of the beam. In the following, for any line ℓ and pixel n, the observation model from Eq. (1) can be written

$$y_{n\ell} = \varepsilon_{n\ell}^{(m)} \kappa_n M_\ell(\boldsymbol{\theta}_n) + \varepsilon_{n\ell}^{(a)}, \tag{10}$$

with $\varepsilon_{n\ell}^{(m)} \sim \log \mathcal{N}(-\sigma_m^2/2, \sigma_m^2)$ and $\varepsilon_{n\ell}^{(a)} \sim \mathcal{N}(0, \sigma_{a,n\ell}^2)$. To simplify notation, this parameter, κ_n, is added to the parameter vector, $\boldsymbol{\theta}_n$. In other words, in the following, each pixel is described with the physical parameter vector, $\boldsymbol{\theta}_n = (\kappa, P_{\text{th}}, G_0, A_V^{\text{tot}}) \in \mathbb{R}^4$.

Finally, in the synthetic observation test, we also included the case where only an upper limit on the observation was available, that is, when the line was undetected. More precisely, for a pixel n and a line ℓ, when the simulated intensity, $\varepsilon_{n\ell}^{(m)} \kappa_n M_\ell(\boldsymbol{\theta}_n) + \varepsilon_{n\ell}^{(a)}$, was smaller than $3\sigma_{a,n\ell}$, we only kept the upper limit $\omega_{nl} = 3\sigma_{a,n\ell}$ in our synthetic observations for this line in this pixel.

4.4. Considered reconstruction methods

In the following applications, we demonstrate the potential of BEETROOTS by computing multiple estimators. First, we consider the MMSE computed with the MCMC algorithm presented in Sect. 3.3. The algorithm is also used in optimization mode to return the MAP. In the remainder, we call these estimators "MMSE (ge)" and "MAP (ge)" to highlight the use of the MTM kernel (i.e., of global exploration), together with the PMALA kernel. For each application, the "MAP (ge)" is evaluated for multiple values of the regularization weight, τ_d. For the "MMSE (ge)," the regularization weight, τ_d, is estimated along with the physical parameters, $\boldsymbol{\Theta}$. This estimation relies on the approach from Vidal et al. (2020) that maximizes the Bayesian evidence, $\pi(\mathbf{Y})$, as expressed in Eq. (3).

In addition to the MMSE (ge) and MAP (ge), we considered two approaches to evaluate the maximum likelihood estimator (MLE) to highlight the critical roles of the spatial regularization and global exploration. The first exploits global exploration and, thus, is called "MLE (ge)." The second approach, that we call "MLE (le)," uses local exploration only by setting $p_{\text{MTM}} = 0$. This approach is the closest to standard gradient descent estimations used in ISM studies. We used it to demonstrate the importance of the MTM kernel to escape from local minima. To achieve better results with local exploration only, a common strategy consists in running the optimization procedure multiple times with different initial values (e.g., 100 times for each pixel in Wu et al. 2018). This strategy assumes that at least one initialization is close to the global minimum, which is not guaranteed. For fair comparisons, for each application in the following, all estimators were evaluated only once with the same random initialization, using CPU only. All the experiments presented in this work were run on a personal laptop with an Apple M3 Pro chip with 11 cores for CPU and a 18 GB memory.

5. Application to synthetic observation maps

This section introduces a synthetic yet realistic reconstruction problem, and applies the estimation algorithm presented in Sect. 3.3. From synthetic true physical parameter maps, $\boldsymbol{\Theta}^*$, we generate synthetic observations, $\mathbf{Y}$. The goal in this experiment is to provide an estimator of the physical parameter maps, $\widehat{\boldsymbol{\Theta}}$, from $\mathbf{Y}$ as close as possible to the truth, $\boldsymbol{\Theta}^*$. To demonstrate the power of the presented algorithm, we consider three sets of observations of the same cloud with different spatial resolutions.

First, the synthetic molecular cloud structure and the observation generation method are presented. Then, the metrics used to assess the performance of the reconstruction are described. Finally, the reconstructions for the three cases are described and a comparison between the proposed methods is detailed.

5.1. Generation of synthetic observation maps

This section describes the maps of the true physical parameter, $\boldsymbol{\Theta}^* \in \mathbb{R}^{N\times D}$, and of simulated observations, $\mathbf{Y} \in \mathbb{R}^{N\times L}$. Figure 3a shows the structure of the considered fictitious cloud, representative of a plausible astrophysical scenario. It represents a roughly spherical dark cloud, illuminated from the right by nearby massive stars creating a high-UV, high pressure surface layer, as well as a buried source inside the cloud (on the left).

Figure 3b presents true physical parameters maps, $\boldsymbol{\Theta}^*$. The same scenario is considered at multiple spatial resolutions, with $N_{\text{side}} = 10$, 30 and 90. The corresponding total number of pixels for each case is $N = N_{\text{side}}^2$, that is, 100, 900 and 8 100, respectively. In this synthetic application, the line intensities in each pixel are taken as the face-on intensities from the model and we do not include any beam dilution, so that the true map of the scaling parameter, κ, is set to 1 everywhere.

From the true maps, $\boldsymbol{\Theta}^*$, the emulator of the Meudon PDR code, $\mathbf{M}$, generates observation maps of $L = 10$ ^{12}CO emission lines of mid-J rotational transitions, from $J = 4$–3 to $J = 13$–12. These lines are selected because their emission in PDRs observed with *Herschel* SPIRE FTS has been well studied with PDR models before (see, e.g., Wu et al. 2018). The physics governing their emission is thus relatively well understood. For each line ℓ, the noiseless integrated intensities, $M_\ell(\boldsymbol{\theta}_n)$, range from 10^{-18} to 10^{-2} erg cm^{-2} s^{-1} sr^{-1}. As detailed in Sect. 4.3, these maps are deteriorated with a Gaussian additive noise and a lognormal multiplicative noise, and can contain upper limits of the observations. The standard deviation of the multiplicative noise is set to $\sigma_m = \log(1.1)$, which roughly represents a 10% alteration in average due to calibration errors (model mis-specification noise is not considered in this example). The additive noise is set constant on the map for all lines to $\sigma_{a,n\ell} = \sigma_a = 1.39 \times 10^{-10}$ erg cm^{-2} s^{-1} sr^{-1}, so that the S/N varies between 10^{-8} and 10^{8}. This noise level is typical for CO observations with *Herschel* (Joblin et al. 2018; Wu et al. 2018).

Figure 3c shows the observation maps of three of the ten considered lines and the maps of the number of upper limits considered in each pixel. For instance, in the harsh conditions (high

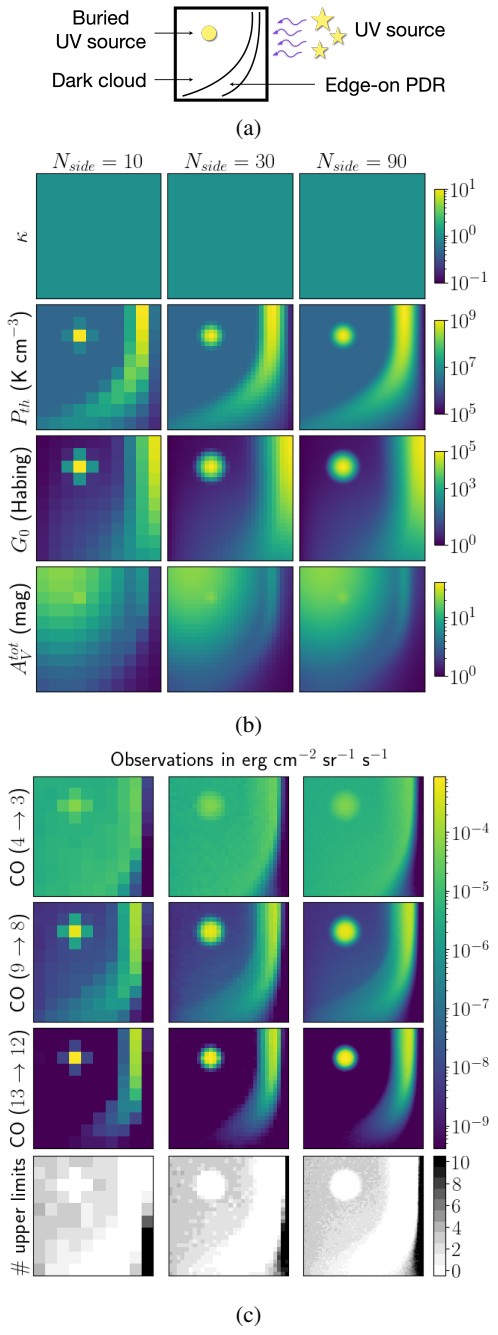


Fig. 3. Synthetic case set up. (a) Structure of the synthetic cloud. (b) True physical parameter maps at resolution $N_{\text{side}} = 10$, 30, and 90 (from left to right). The rows show the maps of the scaling factor κ, the thermal pressure P_{th}, the intensity of the radiation field G_0, and the visual extinction A_V^{tot}. (c) Simulated observation maps of three out of the ten considered lines, and maps of number of lines with upper limits per pixel.

G_0, low A_V^{tot}) of the rightmost region of the map, no molecule can survive: all ten lines are thus undetected and only provided as upper limits. Besides, to be bright, the $J = 13 \to 12$ transition requires both a high pressure and a high UV irradiation for ^{12}CO to form in a layer warm enough to populate such an excited rotation level. This condition is only met in the edge-on PDR and around the buried source. In the rest of the map, this line only has an upper limit. Similarly, in the deep cloud region and far from the buried source, P_{th} and G_0 are low ($P_{\text{th}} \sim 10^7$ K cm^{-3}, $G_0 \lesssim 10^1$), causing the two or three line intensities with the highest energies to be upper limits in this region.

5.2. Comparison metrics

We assess the performance of the reconstruction methods in terms of speed and quality of the resulting estimate. The reconstruction quality is evaluated with the error factor (EF), introduced in Palud et al. (2023b), computed for each estimated physical parameter on each pixel $\widehat{\theta}_{nd}$ as

$$\text{EF}\left(\theta^*_{nd}, \widehat{\theta}_{nd}\right) = 10^{\left|\log_{10}\theta^*_{nd} - \log_{10}\widehat{\theta}_{nd}\right|} = \max\left\{\frac{\theta^*_{nd}}{\widehat{\theta}_{nd}}, \frac{\widehat{\theta}_{nd}}{\theta^*_{nd}}\right\}. \tag{11}$$

Similarly to the UF for uncertainty quantification, the EF provides a multiplicative description of the error in linear scale. This is more relevant than using a distance in linear scale, as the range of a physical parameter typically covers many orders of magnitude. We also find it more intuitive than an error in logarithmic scale, although equivalent. For instance, EF = 2 indicates that the reconstruction, $\widehat{\theta}_{nd}$, has an error of a factor of two with the truth, θ^*_{nd}. A perfect reconstruction corresponds to EF = 1. We quantitatively compare estimators with the mean EF, averaged over each physical parameter map.

The inference speed is quantified with the complete runtime of each inference procedure and with the mean runtime per iteration. The complete runtime is the time required to run an estimation with BEETROOTS. The mean runtime per iteration provides a fair comparison of speed.

5.3. Discussion of experimental results

Optimization algorithms are run for $T_{\text{MC}} = 5000$ steps, and the sampling method for $T_{\text{MC}} = 10\,000$ iterations. Since the goal of sampling – for the evaluation of the MMSE (ge) – is to approximate the posterior distribution instead of only reaching a mode, significantly more iterations are required. In all cases involving the MTM kernel, its selection probability is set to $p_{\text{MTM}} = 20\%$ and the number of candidates, K, to 25.

Table 1 compares the performance of the considered estimators for all spatial resolutions. Figure 4 shows the estimation results on the $N_{\text{side}} = 90$ case. The reconstructed maps for the $N_{\text{side}} = 10$ and 30 cases are displayed in Appendix C.1.

5.3.1. Maximum likelihood: Effect of global exploration

This section analyzes the two presented evaluations of the MLE, shown in the second and third rows in Fig. 4. The first evaluation, "MLE (le)," only exploits the PMALA kernel, that is, local exploration. The second evaluation, "MLE (ge)," uses both PMALA and MTM kernels, that is, local and global exploration. The "MLE (le)" leads to a very poor reconstruction for all spatial resolutions. It yields mean EFs of up to about 10^2 for the thermal pressure and about 10^4 for the intensity of the radiation field, that is, the estimation misses the true values in average by a factor of 10^2 for P_{th} and of 10^4 for G_0[6]. Although they display the main regions of the synthetic cloud, the reconstructed maps contain

[6] The validity interval of G_0, $[10^0, 10^5]$, covers five orders of magnitude. In contrast, the validity interval of P_{th}, $[10^5, 10^9]$ K cm^{-3}, only covers four, that of κ, $[10^{-1}, 10^1]$, only two and that of A_V^{tot}, $[1, 40]$ mag, less than two. In addition, G_0 has the largest spatial variations in the true maps – see Fig. 3b. The mean EF of G_0 is therefore expected to be the largest for most estimators.

Table 1. Performance of inference methods on a synthetic case.

	Estimator	τ_d	Mean error factor				Runtime	
			κ	$P_{\rm th}$	G_0	$A_V^{\rm tot}$	Total	Per iter.
$N_{\rm side}=10$	MLE (le)	–	1.9	31.1	56.9	3.1	4m 09s	0.06s
	MLE (ge)	–	1.6	2.3	15.6	2.8	5m 11s	0.06s
	MAP (ge)	10^{-1}	1.4	1.8	11.8	2.1	5m 24s	0.07s
	MAP (ge)	10^{0}	**1.1**	1.4	**1.4**	**1.3**	5m 27s	0.07s
	MAP (ge)	10^{1}	**1.1**	**1.3**	**1.4**	1.5	5m 21s	0.06s
	MAP (ge)	10^{2}	1.5	4.7	2.2	4.8	4m 50s	0.06s
	MMSE (ge)	1.1	**1.1**	1.5	1.9	**1.3**	27m 49s	0.17s
$N_{\rm side}=30$	MLE (le)	–	2.3	20.2	1 231	2.9	24m 17s	0.29s
	MLE (ge)	–	1.7	2.4	31.4	2.9	29m 48s	0.36s
	MAP (ge)	10^{-1}	1.4	2.0	9.5	2.1	31m 45s	0.38s
	MAP (ge)	10^{0}	**1.2**	1.7	3.2	1.6	31m 21s	0.38s
	MAP (ge)	10^{1}	**1.2**	**1.3**	**1.3**	**1.3**	34m 14s	0.41s
	MAP (ge)	10^{2}	**1.2**	1.5	1.5	1.5	32m 14s	0.39s
	MMSE (ge)	4.7	**1.2**	**1.3**	1.4	**1.3**	2h 09m 44s	0.78s
$N_{\rm side}=90$	MLE (le)	–	2.3	21.4	1 223	3.0	3h 01m 23s	2.18s
	MLE (ge)	–	1.7	2.4	22.8	2.8	4h 53m 46s	3.53s
	MAP (ge)	10^{-1}	1.5	2.6	32.8	2.1	6h 10m 57s	4.45s
	MAP (ge)	10^{0}	1.3	1.6	5.5	1.6	7h 13m 01s	5.20s
	MAP (ge)	10^{1}	1.2	**1.3**	1.5	1.4	7h 30m 36s	5.41s
	MAP (ge)	10^{2}	1.2	**1.3**	**1.3**	**1.2**	6h 15m 13s	4.50s
	MMSE (ge)	31.1	**1.1**	**1.3**	**1.3**	1.3	28h 43m 02s	10.34s

Notes. le: local exploration, i.e., with MTM kernel disabled ($p_{\rm MTM}=0$). ge: global exploration, i.e., with MTM kernel enabled ($p_{\rm MTM}>0$). In each case, the number of pixels in each map is $N=N_{\rm side}^2$. The three considered cases thus contain $N=100$, 900 and 8100 pixels, respectively. For the MMSE (ge), the regularization weight, τ_d, is estimated along with the physical parameters, $\boldsymbol{\Theta}$. For each spatial resolution and physical parameter, the lowest mean error factor is highlighted in bold.

many unphysical large variations and are thus not exploitable for physical interpretation. With these reconstructions and only one initialization, it is difficult to assess whether poor reconstructions missed the global minimum or whether the model itself defines an unphysical global minimum.

Using global exploration answers this question: with only one initialization, the maps obtained with "MLE (ge)" miss the true values in average by a factor of up to 2.4 for $P_{\rm th}$ and of up to 31 for G_0, which represents a significant improvement. It is safe to assume that this reconstruction procedure attains the global mode of the likelihood function. However, some pixels still have unphysical values for $P_{\rm th}$ and G_0 because of an identifiability issue. Besides, as we only consider rotationally excited ^{12}CO transitions that originate from a surface layer of the cloud, both MLE evaluations poorly reconstruct $A_V^{\rm tot}$. Additional information is thus required to further improve the reconstructions.

5.3.2. Effect of spatial regularization

Spatial regularization can provide this additional information. With a large enough regularization weight τ_d, the "MAP (ge)" estimator systematically produces lower mean EF than the "MLE (ge)." For $P_{\rm th}$ and G_0, for each pixel, it selects the mode that is consistent with the pixel observations and with those in its neighbors. In addition, it compensates for the likelihood inability to reconstruct $A_V^{\rm tot}$ by returning smooth maps that are physically exploitable. However, the regularization weight needs careful tuning. On the one hand, using too low regularization weight can yield artifacts in the reconstructions, due to a local minimum. This happened for instance in the $N_{\rm side}=90$ case with $\tau_d=1$, with two sets of neighboring pixels staying in a local minimum (visible in the bottom middle part of the map as a small contiguous group of high G_0, low $P_{\rm th}$ pixels contrasting with the surrounding pixels). On the other hand, in the $N_{\rm side}=10$ case where true maps are not smooth, using a too large regularization weight τ_d biases the estimations. It erases the buried source, and under-estimates the spatial variations of $P_{\rm th}$ and G_0 in the transition from the edge-on PDR to the deep cloud. Appendix C.2 provides more quantitative results on the regularization weight influence on the mean EF achieved with MAP (ge).

5.3.3. Sampling-based approach

A sampling-based approach provides more information and can further improve estimation. We find that the MMSE (ge) can be a better estimator than MAP (ge), that is, with lower mean EF. This is particularly true for loosely constrained observations, such as the rightmost region of the map that contains mostly upper limits of the considered ^{12}CO transitions. It is also true for loosely constrained physical parameters such as the visual extinction, $A_V^{\rm tot}$, in the deep molecular cloud, or the scaling parameter, κ, in the rightmost part of the maps. As the MMSE (ge) is the mean of the posterior samples (i.e., that explain the observations while respecting the prior), it can be closer to the true value. On the $N_{\rm side}=90$ use case, the MMSE (ge) yields a mean EF of only 1.1 for κ, 1.3 for $P_{\rm th}$, 1.3 for G_0 and 1.3 for $A_V^{\rm tot}$, that is, mean errors of at most 30%. Achieving such low errors is remarkable, especially considering the variety of environments and of S/N in the considered scenario.

In addition to the MMSE, the sampling approach quantifies the uncertainty associated with the estimation. The obtained CIs and UFs are consistent with the physical intuition. In all three cases, the UFs of κ, $P_{\rm th}$ and G_0 are highest in the rightmost part of the maps. In particular, it only yields upper bounds on κ and $A_V^{\rm tot}$, and a lower bound on G_0. Beyond these obtained bounds, most molecules are photodissociated, which explains the observation upper limits. For the deep molecular cloud (on the left), the reconstruction only yields lower bounds on $A_V^{\rm tot}$ as the considered lines only trace the surface of the cloud. For the PDR (on the right) and the buried source (on the left), the observations are in high S/N and do not include upper limits on observations. The reconstructions of these two regions are thus associated with the smallest UFs for $P_{\rm th}$ and G_0. Finally, κ is well reconstructed in all three cases when at least one transition is in a high-S/N regime. The estimated values are indeed close to the true value (i.e., no bias), and the UFs are close to one (i.e., very small uncertainty) everywhere but in the rightmost part of the maps.

The sampling-based approach allows for a model-checking that accounts for as many uncertainty sources as possible. Appendix C.3 provides the results of the model checking implemented in BEETROOTS applied to these three synthetic cases. Here, this assessment confirms that the model is fully compatible with the observations, as expected for these synthetic observations.

This application shows that BEETROOTS produces significantly more accurate estimators compared to the usual MLE approach, both in optimization (with MAP (ge) estimator) and in sampling (with MMSE estimator). The automatic tuning of the regularization weight proves to yield relevant values. The uncertainty quantification results are interpretable and consistent with the physics of the considered ^{12}CO rotational emission lines.

5.3.4. Comparison of inference runtimes

Table 1 shows the total inference runtime for each estimation. It shows that for small maps (with up to 10^3 pixels), running an

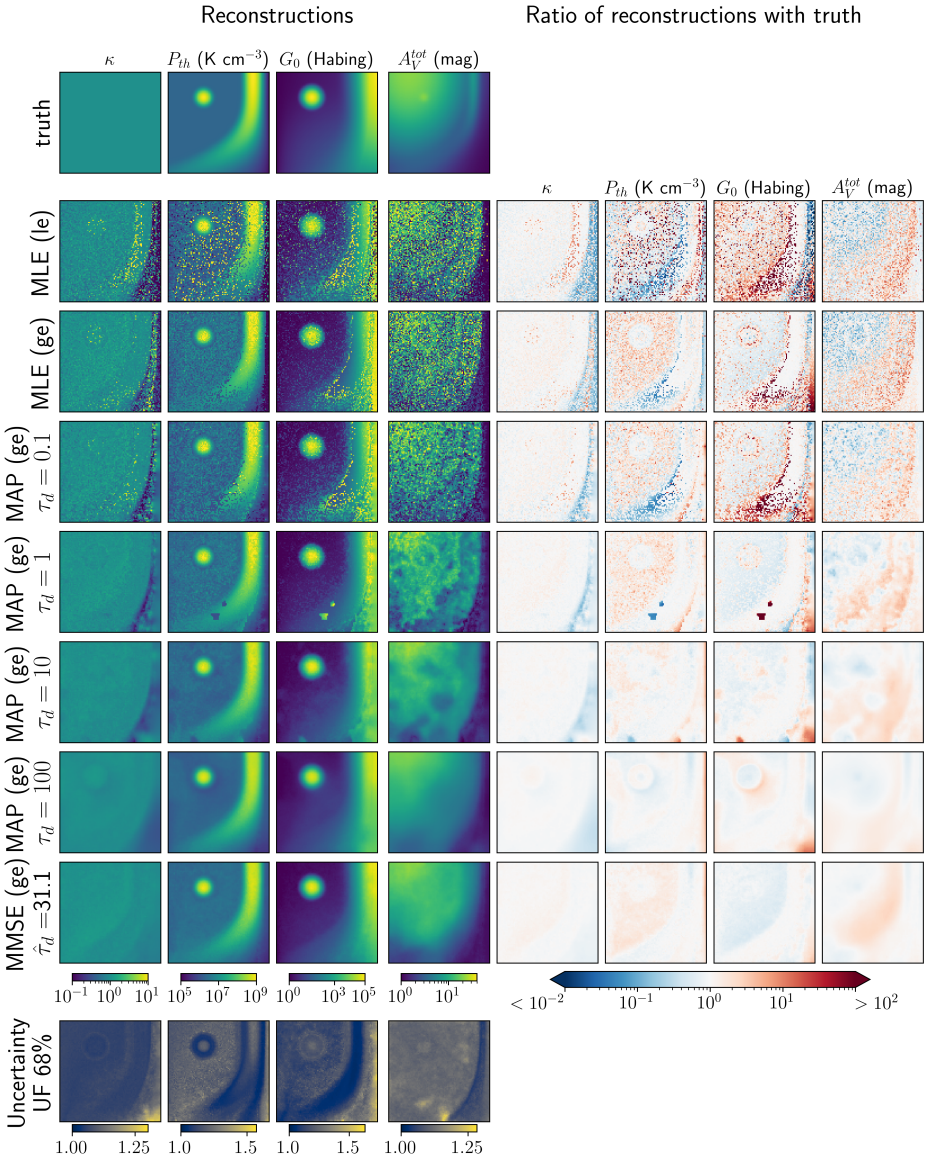


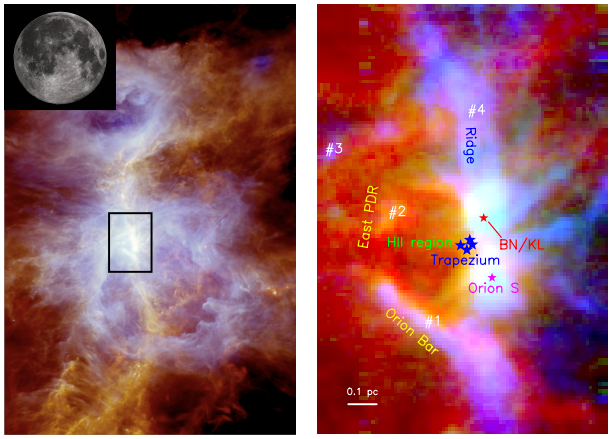
Fig. 4. Reconstruction results for the $N_{\text{side}} = 90$ spatial resolution. The first row shows the $D = 4$ true maps that each estimator tries to reconstruct. The obtained reconstructions are shown below the true maps. The first estimator is "MLE (le)": the MLE evaluated using the PMALA kernel only, that is, only local exploration. The second estimator is "MLE (ge)": the MLE evaluated using both the PMALA and MTM kernels, that is, a combination of local and global exploration. The third to fifth estimators are MAP estimators (i.e., including spatial regularization) with different values for the regularization weight, τ_d. The sixth estimator is the MMSE, obtained with sampling. The associated regularization weight, τ_d, is inferred along with the physical parameters, $\boldsymbol{\Theta}$. The last row, "UF 68%", quantifies the uncertainties associated with the MMSE by indicating the size of the 68% CI. The maps on the right display the ratios between the estimated and true maps, to better assess the quality of each estimation.

optimization (MLE or MAP) or a sampling procedure takes less than two hours. It shows that BEETROOTS can analyze most past observations similar to this synthetic case. For larger maps with about 10^4 pixels, reconstructions are longer: from 3 to 8 hours for optimization, and about 28 hours for sampling. As inference runtime does not scale linearly, we believe that BEETROOTS currently can manage maps of up to $N \simeq 30\,000$ pixels.

The runtime per iteration provides a more fair comparison of the actual speed of each estimation, as sampling was run for twice as many iterations as optimization. As PMALA only considers one candidate at each iteration, a local exploration only (le, with only PMALA) is faster than a global exploration (ge, combining MTM and PMALA). Because of the way the spatial regularization is used to generate candidates in MTM, larger values of τ_d lead to faster iterations (this effect should be removed in a future version). Finally, a sampling iteration is slower than an optimization iteration, as it requires more intermediate steps.

6. Application to the Orion molecular cloud 1

In the previous section, we demonstrate the potential of BEETROOTS on three synthetic cases. In this section, we apply our inference procedure on ∼7.5′ × ∼11.5′ far-infrared and (sub)millimeter line emission maps of the Orion molecular cloud 1 core (OMC-1 core) described in Goicoechea et al. (2019). To the best of our knowledge, this inference is the first performed on a large map (that is, with 1000 pixels or more) for multiple physical parameters from molecular tracers. It is also the first time that one quantifies the uncertainties in an inference of maps of physical parameters from molecular tracers. Unlike the synthetic



(a) (b)

Fig. 5. Structure of OMC-1. (a) The Orion A star-forming cloud seen by *Herschel* with dust thermal emission. The black rectangle outlines the observed region of OMC-1, centered on the Trapezium star cluster. The image is a composite of 70 µm (blue), 160 µm (green) and 250 µm (red) emission. It spans about $1.3 \times 1.6\,\mathrm{deg}^2$. The moon is shown for scale. (b) OMC-1 composite image covering ~85 arcmin^2 at ~12″ with three emission lines: the [CII] 158 µm line which traces the FUV-illuminated surface of the molecular cloud (red), the $C^{18}O$, $J = 2 \rightarrow 1$ line which traces cold dense gas (blue) and the HCO^+ $J = 3 \rightarrow 2$ line (green). Reconstruction results will be detailed for positions 1 to 4. Adapted from Goicoechea et al. (2019).

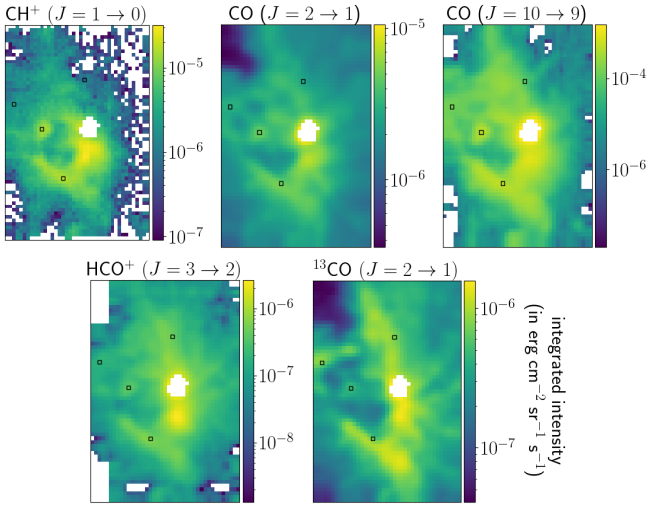


Fig. 6. OMC-1 core observed integrated intensity maps of the lines used for inversion. The white region in the middle of the maps is the BN/KL region, dominated by shocks, and not considered in the inversion. The remaining white pixels correspond to negative integrated intensities due to the additive Gaussian noise. The inference results for the four pixels highlighted with a black square are detailed in Sect. 6.3.3.

applications, it is here not clear a priori whether the observed lines are sufficient to constrain the physical parameters.

First, we describe the region and the observation map. Then, we set up the inversion procedure. The results are discussed and compared with the literature, including the analysis of the Orion Bar from Joblin et al. (2018). Third, we compare our estimates of G_0 with those derived by Goicoechea et al. (2015) from total far infrared (FIR) luminosities. Finally, we explore the relationship between G_0 and P_{th} that appears in our results.

6.1. Cloud, observation, and noise description

OMC-1 is a bright region of the Orion A cloud. It is located at about 414 pc from the Sun (Menten et al. 2007), making it the closest region of intermediate- and high-mass star formation. Figure 5 shows its position in Orion A and its general structure. The surface of OMC-1 is illuminated by the Trapezium star cluster, which contains a few massive stars. The FUV radiation from these massive stars heats and photodissociates the molecular surface of OMC-1, at the border of the central HII region. The Orion Bar and the eastern PDR, first reported by Goicoechea et al. (2015), are prominent examples of dense PDRs located at this interface. The two major star-forming sites are the Becklin-Neugebauer/Kleinmann-Low region (BN/KL) and Orion South (Orion S). Sect. 6.3.3 will detail reconstruction results for the positions highlighted in Fig. 5b.

The observation $\mathbf{Y}$ contains $N = 2\,475$ pixels and $L = 5$ molecular emission lines: CH^+ ($J = 1 \rightarrow 0$), CO ($J = 2 \rightarrow 1$), CO ($J = 10 \rightarrow 9$), ^{13}CO ($J = 2 \rightarrow 1$), and HCO^+ ($J = 3 \rightarrow 2$). A [CII] 158 µm emission map was also presented in Goicoechea et al. (2019) but was left aside in this study. Indeed, our attempts at simultaneously fitting [CII] with the molecular lines resulted in an incompatibility between the models and observations – identified with our model checking procedure. This incompatibility when trying to simultaneously fit [CII] emission and warm molecular emission with the Meudon PDR code was also observed in Joblin et al. (2018) in the Orion Bar, with a systematic underestimation of the [CII] intensity, which was proposed to be caused by the larger beam filling factor of [CII] emission in Herschel observations compared to warm molecular tracers that originate in denser filamentary structures.

Figure 6 shows the observation maps associated with the five lines used in the inversion. The associated STD maps are displayed in Appendix D.1. Four pixels are highlighted in this figure, corresponding to the Orion Bar PDR (lowest square, labeled #1 in Fig. 5b), the eastern PDR (mid-height square, labeled #2), the north-east border of the map (top left square, labeled #3) and the north-west ridge (top right square, labeled #4). Results for these pixels are detailed in Sect. 6.3.3. The BN/KL region is masked (in white, in the center of the maps). Since its emission contains a major contribution from shocks, and since such environments are not described by PDR models, reconstruction results on this region would not be relevant. Dedicated models such as the Paris-Durham code (Godard et al. 2019) would better simulate this region. Removing this region avoids biasing the neighboring pixels with the spatial regularization.

6.2. Inversion setup

The integrated intensity maps, $\mathbf{Y}$, and the associated additive noise STDs, $(\sigma_{a,n\ell})_{n\ell}$, are computed directly from the hyperspectral cubes. The multiplicative noise is modeled as a lognormal distribution that combines two terms. As indicated in Sect. 4.3, the model mis-specification STD, σ_{model}, is set so that a 3σ error corresponds to an error of a factor of three, that is, $\sigma_{\mathrm{model}} = \frac{1}{3}\log 3$. An estimated 8% calibration error is considered for the whole map[7], that is, $\sigma_c = \log 1.08$. The total multiplicative STD is $\sigma_m = (\sigma^2_{\mathrm{model}} + \sigma^2_c)^{0.5} \simeq \log 1.452$. The angle between the cloud surface and the line of sight, α, varies in the observed map.

[7] See Roelfsema et al. (2012) for the calibration error of *Herschel*-HIFI and https://www.iram.fr/GENERAL/calls/s23/30mCapabilities.pdf for that of the IRAM-30 m.

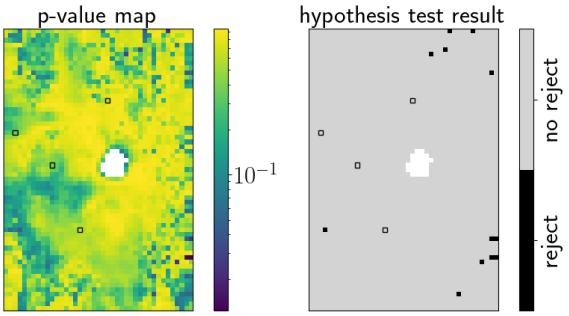


Fig. 7. Model checking for OMC-1. Left: Estimated p-value map. Right: p-value-based decision for a 95% confidence level. Only twelve pixels (i.e., 0.5%) are rejected and are not in the regions of interest. Inference results are thus trustworthy.

For instance, the Orion S region is considered to be mostly face-on ($\alpha \simeq 0$ deg) while the Orion Bar is edge-on ($\alpha \simeq 90$ deg). To avoid inferring more parameters, we set $\alpha = 0$ deg and assume the scaling factor κ to capture inclination effects.

We compare two estimators, namely "MLE (ge)" and "MMSE (ge)," both evaluated with BEETROOTS. The MLE (ge) (Sect. 3.1) is obtained with the optimization algorithm described in Sect. 3.3, run for 5000 iterations. The preconditioned gradient descent update (i.e., local exploration) step size is set to $\eta = 0.01$. The MMSE (ge) is evaluated from a Markov chain of $T_{\text{MC}} = 10^4$ iterates, including $T_{\text{BI}} = 2000$ of burn-in. The PMALA step size is set to $\eta = 0.1$. For both estimators and at each iteration, MTM (i.e., global exploration) has a probability $p_{\text{MTM}} = 0.2$ of being used, and generates $K = 25$ candidates per pixel.

6.3. Inference results

This section presents the inference results. After confirming that the Meudon PDR code successfully reproduces the observations (and, thus, the reconstructed physical parameter maps are trustworthy), we describe the reconstructed maps and explore the results more in detail for four selected pixels.

6.3.1. Model checking and Bayesian p-value

Before presenting the reconstructions themselves, we assess their relevance with the model checking approach presented in Sect. 3.4. As a reminder, this approach computes a p-value for each pixel to test the hypothesis H1, that is "the astrophysical model, the random deterioration and the prior can reproduce the observations, $\mathbf{Y}$." Figure 7 shows the map of estimated p-values. We find that the estimated p-values are larger than 0.5 in all the regions of interest of the maps. As a result, only 12 pixels out of 2475 – about 0.5% – lead to a model rejection. In particular, H1 is not rejected for any of the four highlighted positions. This indicates that the reconstructions to be presented next are trustworthy. One might expect the proportion of rejected pixels to be closer to 5% when H1 is true. This low rejection rate might be due to the fact that only $L = 5$ lines are used to reconstruct $D = 4$ physical parameters, or to the large considered model mis-specification STD, σ_{model}.

As mentioned in Sect. 3.4, this model validation is based on the posterior predictive distribution, that is, the distribution of predicted integrated intensities $\mathbf{M}(\boldsymbol{\theta})$ affected by noise (Eq. (10)). Visualizing this distribution is therefore a powerful diagnosis tool to identify lines that are not well predicted by the astrophysical model. Appendix D.2 shows this distribution for the four highlighted positions.

6.3.2. Inferred parameter maps

Figure 8 shows inference results for both likelihood minimization (top row) and posterior sampling (middle row). The spatial structures visible in the observations are recovered with the MMSE when including the spatial smoothness prior, but hardly visible with the MLE. As there are few lines and thus few constraints for the astrophysical model, the likelihood is very sensitive to noise and the MLE presents unphysical variations. The spatial regularization thus plays a key role in recovering the correct spatial structure.

We go on to focus on the MMSE and UF maps. The Orion Bar, the HII region, the Orion S region, the eastern PDR and the north-west ridge are visible in each of the four reconstructed maps. The scaling factor, κ, is close to 1 in the bright regions. In the Orion S region, in the eastern PDR and in the Orion Bar, the thermal pressure, P_{th}, is estimated at around 3×10^8 K cm^{-3} and the incident UV radiation field intensity, G_0, at around 4×10^4. In these regions, the recovered visual extinction, A_V^{tot}, is roughly 10 mag. These values are consistent with the literature. In particular, our results robustly confirm that the high thermal pressures found in the Orion Bar are widespread in the PDRs of OMC-1 (Goicoechea et al. 2019). Moreover, we will show in Sect. 6.4 that the full G_0 MMSE map shows very good agreement, both in terms of structure and values, with the map derived by Goicoechea et al. (2015) from FIR luminosities.

Over the whole map, the UF is lower than 1.2 for the scaling factor, κ, and the visual extinction, A_V^{tot}. This means that the typical 1σ error is at most 20% for κ and A_V^{tot}. Similarly, over most of the map, the UF is lower than 1.5 for the thermal pressure, P_{th}, and lower than 2 for the incident UV radiation field intensity, G_0. These four parameters are thus relatively well constrained considering their large validity intervals and the fact that we use only $L = 5$ lines to infer $D = 4$ parameters for each pixel.

6.3.3. Analysis of the highlighted positions

This subsection presents a discussion of the detailed reconstruction results on the four highlighted positions. For each of these pixels, Fig. 9 shows the posterior PDF for pairs of parameters, that is, pairwise histograms of samples.

Figure 9a show the results for the Orion Bar pixel (lowest of the four squares) It shows that the marginal posterior distributions for this pixel are simple, as they could be well approximated by a Gaussian distribution with a diagonal covariance matrix. The uncertainties on κ and A_V^{tot} are small and a bit larger on P_{th} and G_0, which is consistent with the UF maps in Fig. 8. These histograms also show that there is no degeneracy between parameters. The figure compares our estimation results with those from Joblin et al. (2018) (red star in the figure), which analyzed the Orion Bar with an observation of $L = 24$ lines in $N = 1$ pixel. The observed lines they used for inversion include ^{12}CO lines (from $J = 11 \rightarrow 10$ to $J = 23 \rightarrow 22$), rotational H_2 lines (from S(0) to S(5)) and low level CH^+ rotational lines (from $J = 1 \rightarrow 0$ to $J = 6 \rightarrow 5$). The noise on the observation was assumed additive, Gaussian and uncorrelated with known STDs, $(\sigma_{a,\ell})_{\ell=1}^{L}$. The authors inferred $\boldsymbol{\theta} = (\kappa, P_{\text{th}})$, while fixing the observation angle to $\alpha = 60$ deg and the total extinction to $A_V^{\text{tot}} = 10$ mag – as the observed lines did not provide sufficient constraint on A_V^{tot}. Based on Tielens & Hollenbach (1985)

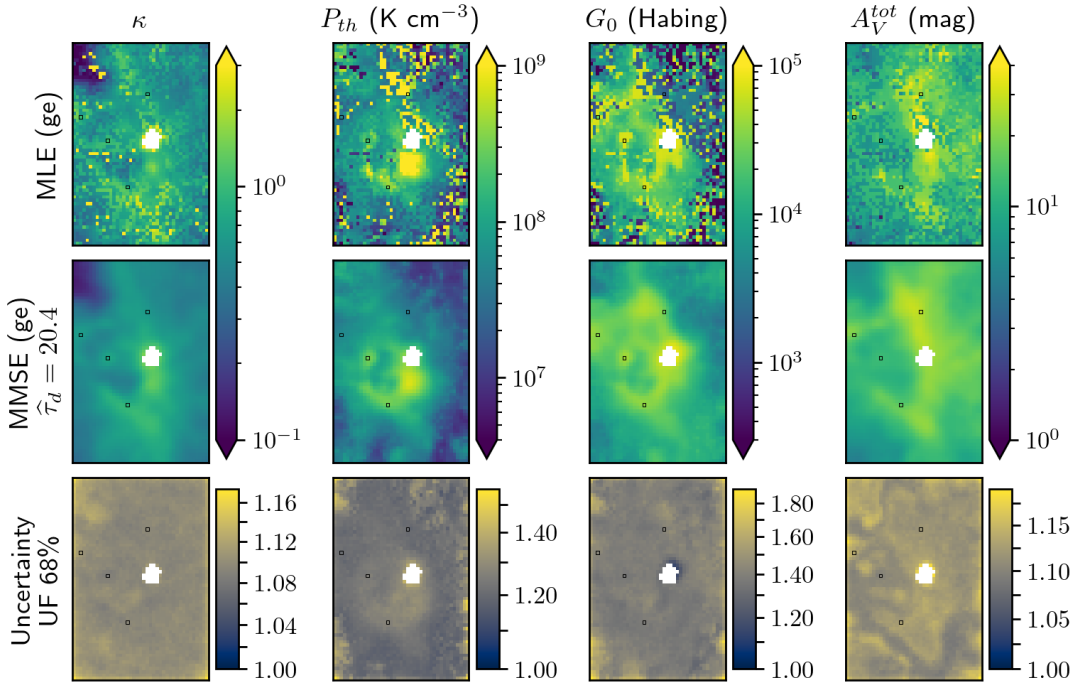


Fig. 8. Inference results for OMC-1. From top to bottom, the rows show the maximum likelihood estimator (MLE), the MMSE estimator, and then the maps of uncertainty factor (UF) for the 68% credibility interval (CI). From left to right, the rows show results on the scaling factor, κ, thermal pressure, $P_{\rm th}$, intensity of the radiation field, G_0, and the visual extinction, $A_V^{\rm tot}$.

and Marconi et al. (1998), they set $G_0 = 2 \times 10^4$. The fit was performed with a grid search on $P_{\rm th}$ and a continuous optimization on κ. They obtained $P_{\rm th} = 2.8 \times 10^8$ K cm^{-3} and $\kappa = 1.3$. Overall, Fig. 9a shows compatible results between the two estimations. It also displays the G_0 value estimated in Goicoechea et al. (2015) from FIR luminosities (black line in the figure). Again, the two G_0 reconstructions are compatible. These agreements between our estimations and these two independent reconstructions from the literature are remarkable given that they all resort to very different tracers.

Figure 9b shows the results for the eastern PDR pixel (mid-height square). These estimations are very similar to those of the Orion Bar, which is remarkable as the two pixels are distant on the map. It shows that these two pixels, located on the cavity edge, share very similar physical conditions. This figure also shows that the estimated G_0 value is still compatible with the FIR estimation, although less than for the Orion Bar.

Figures 9c and 9d show the north-east pixel (top left square) and the north-west ridge edge pixel (top right square), respectively. These two pixels and their surrounding regions yield a more moderate $P_{\rm th}$ ($\simeq$3–6 $\times 10^7$ K cm^{-3}) while G_0 remains similar to what is found in the Orion Bar ($\simeq 3 \times 10^4$), which seems in contradiction with the positive correlation between $P_{\rm th}$ and G_0 described in Joblin et al. (2018) and Wu et al. (2018) (see discussion in Sect. 6.5). We note that the histograms show a slight negative correlation between $P_{\rm th}$ and G_0, which indicates that the actual G_0 for these pixels might be lower and the actual $P_{\rm th}$ might be higher. In addition, our G_0 estimation in these two pixels appears to be only marginally compatible (or even clearly incompatible for the north-east edge pixel) with the FIR estimation. This might indicate that additional processes (not included in the models) are increasing the molecular emission in these pixels; this could, for instance, be the case if additional energy injection processes besides FUV irradiation are contributing to the warm molecular emission. Further investigation is required to better understand these regions.

6.4. Comparison with estimation of G_0 from far infrared luminosities

This section compares the G_0 reconstructions from Goicoechea et al. (2015) with our G_0 estimations (MMSE and CIs) from Sect. 6.3.2. Figure 10a shows an overall good agreement between these independent reconstructions, although the MMSE estimate tends to be larger than the estimations from FIR luminosities. For most pixels with high FIR G_0, this difference is lower than a factor of three. However, for pixels with lower G_0 from FIR luminosities, our reconstructions have a systematic bias much larger than the 68% CIs.

To better understand this discrepancy, we show in Figure 10b the G_0 map from Goicoechea et al. (2019) estimated using FIR luminosities, and in Figure 10c the map of the ratio between our MMSE estimate of G_0 and this FIR luminosity-based estimation. In both maps, the black contour identifies the regions where G_0 (MMSE) $\leq 3\,G_0$ (FIR). It shows that over all the central PDR regions bordering the walls of the cavity around the Trapezium stars, the two estimations are compatible within a factor of 3. In contrast, the regions where the discrepancy is larger than a factor of 3 are all located in the outer regions, where the physics and chemistry might not be dominated by the FUV illumination from the Trapezium cluster. In addition, these outer regions of the map also present lower S/N in the observations. Interestingly, the only region where our G_0 are lower than those based on FIR luminosities are in Orion S, where the G_0 is highest. This might be due to the upper bound of the validity interval on G_0, at 10^5 Habing units, constraining our MCMC posterior samples to remain below this threshold, and thus possibly biasing the MMSE towards lower values.

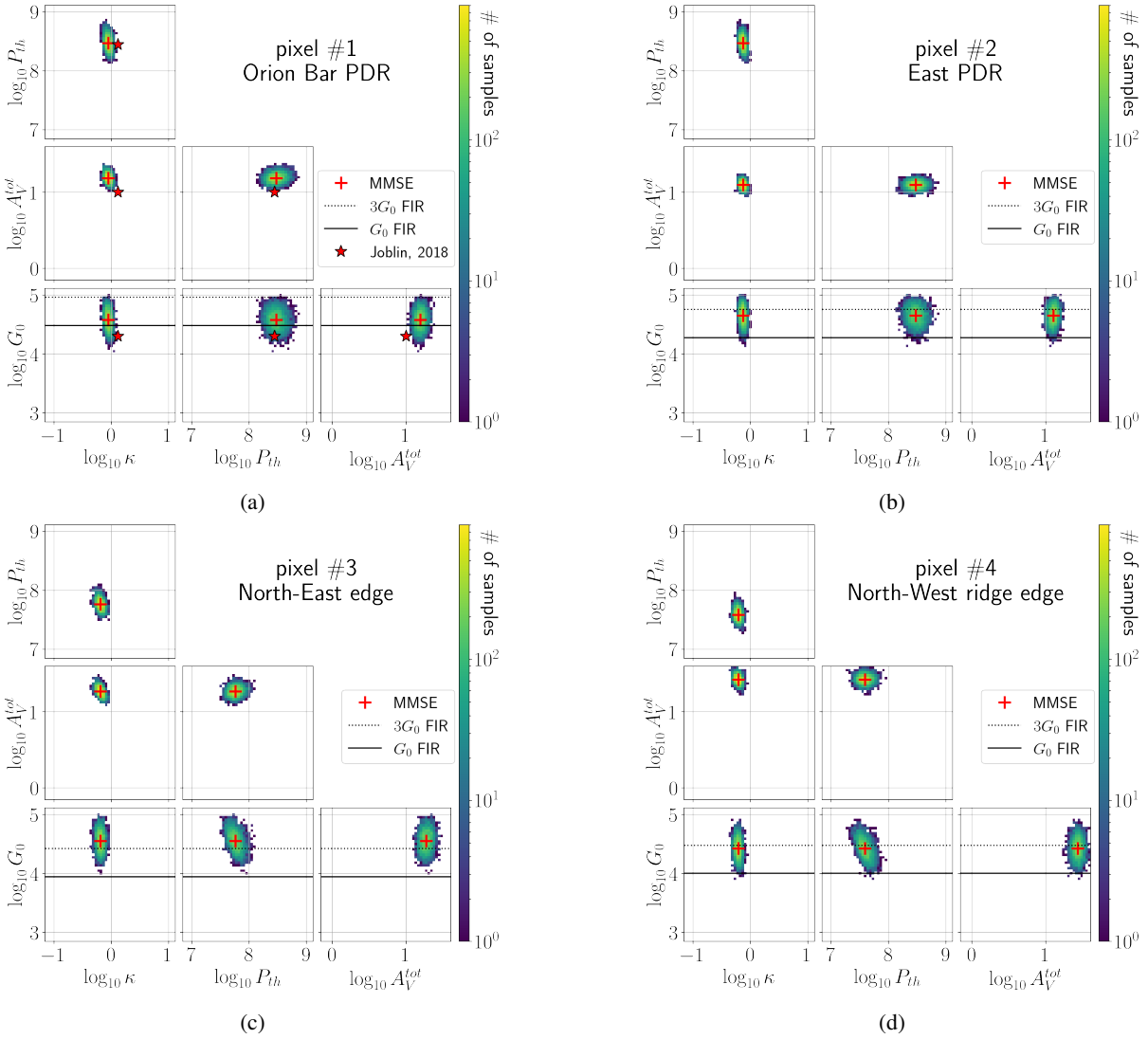


Fig. 9. Inference results for OMC-1 on four specific pixels, presented as 2D marginal histograms of the $\boldsymbol{\theta}_n$ samples. (a) Orion Bar PDR pixel (with the estimation from Joblin et al. 2018), (b) eastern PDR pixel, (c) North-east edge pixel, (d) North-west ridge edge pixel. The horizontal black line indicates the G_0 values estimated from FIR luminosities from Goicoechea et al. (2015).

This comparison to an independent measurement of G_0 indicates that despite inferring four parameters from five molecular lines only, our results are reliable over a large area of the OMC-1 map. Adding to that the fact that our PDR model succeeds in reproducing the observed line intensities over the full map, this provides evidence that FUV irradiation controls the physics and the chemistry of the warm molecular tracers over most of the central regions of OMC-1.

6.5. Relationship between G_0 and P_{th}

The existence of a positive correlation in PDRs between the incident UV radiation field intensity G_0 and the thermal pressure $P_{\rm th}$ has been discussed in multiple ISM studies (Joblin et al. 2018; Wu et al. 2018; Seo et al. 2019; Pabst et al. 2022; Wolfire et al. 2022). This relationship has been found to follow a linear relationship (with $P_{\rm th}/G_0 \in [1,4] \times 10^4$, Joblin et al. 2018) or near linear relationship (exponent of 0.9 in Wu et al. (2018) and of 0.75 in Seo et al. (2019)). A similar relationship has been suggested in Wu et al. (2015) from extragalactic observations in M83. This relationship suggests that the radiative feedback of massive stars can induce compression of the gas at the edge of PDRs (Bron et al. 2018). Whether this could trigger additional star formation remains an open question.

Figure 11a shows the MMSE pixel values along with the 68% CI (i.e., a 1σ error) for our OMC-1 estimation results. A similar positive correlation between G_0 and $P_{\rm th}$ can be seen. As the joint $(P_{\rm th}, G_0)$ histograms in Fig. 9 show null or negative correlations, we confirm that this positive correlation among multiple sources is not due to a degeneracy between the two parameter uncertainties and has a physical origin, as hypothesized in Joblin et al. (2018). In the following, we focus on the pixels that satisfy G_0 (MMSE) $\leq 3G_0$ (FIR) (in blue in Fig. 11a), as these are the most likely to be adequately described by our PDR models (see previous discussion in Sect. 6.4). A weighted least square fit on $\log_{10} P_{\rm th} = a \log_{10}(G_0) + b$ leads to $a = 0.89 \pm 0.02$ and $b = 4.2 \pm 0.1$ (results are shown with mean $\pm 1\sigma$), for a determination coefficient $R^2 = 0.62$ (in red in the figure). This fit is quite close to the relationship presented in Wu et al. (2018). The small set of points in blue lying above the rest at moderate G_0 correspond to the map's north-west corner. In this region, the S/N is much lower than on the rest of the map (see Fig. 6). Although

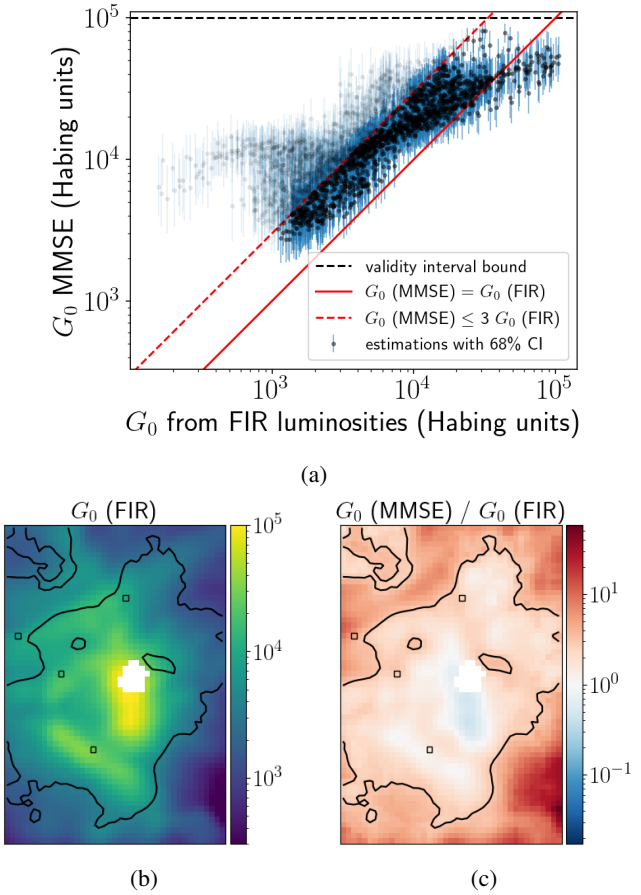


Fig. 10. Comparison of our estimations of G_0 with those of (Goicoechea et al. 2015) obtained from FIR luminosities. (a) Scatter plot of estimated G_0, from our reconstruction and from FIR luminosities. (b) Map of estimated G_0 with FIR luminosities, adapted from Goicoechea et al. (2019). (c) Map of the ratio between our G_0 reconstructions and those based on FIR luminosities. In both maps, the black contour identifies the regions where G_0 (MMSE) $\leq 3\,G_0$ (FIR).

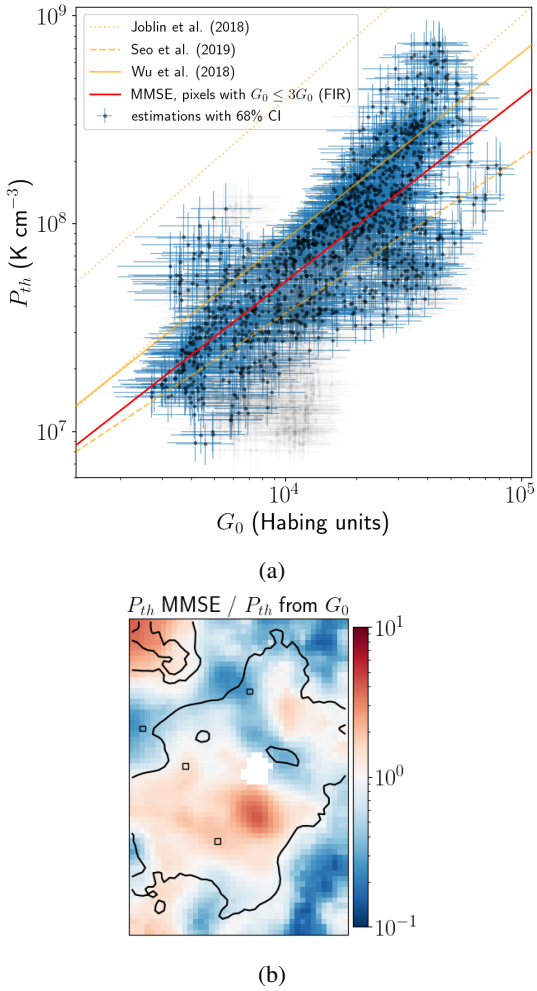


Fig. 11. Analysis of the relationship between G_0 and $P_{\rm th}$. (a) Scatter plot of the MMSE pixel values for G_0 and $P_{\rm th}$, that shows a strong positive correlation. The pixels such that G_0 (MMSE) $\leq 3\,G_0$ (FIR) are highlighted in blue, and the other pixels are in light grey. A power law fit on those highlighted pixels is shown in red, compared with previous power law relations proposed in the literature (in yellow). (b) Map of ratio of the MMSE for $P_{\rm th}$ and the $P_{\rm th}$ predicted from G_0 with our power law fit. In both maps, the black contour identifies the regions where G_0 (MMSE) $\leq 3\,G_0$ (FIR), as in Fig. 10.

the G_0 MMSE estimate is close to the FIR luminosity-based G_0 estimate in this region, the uncertainties on both $P_{\rm th}$ and G_0 are large, as the UF maps from Fig. 8 clearly show. These pixels thus have a small weight in our weighted least square fit.

We observe that the $P_{\rm th}$–G_0 relation found from warm molecular tracers (this work, Joblin et al. 2018, Wu et al. 2018) appears shifted towards higher thermal pressures compared to the theoretical relation of Seo et al. (2019). Moreover, Wolfire et al. (2022) suggests that magnetic and turbulent pressure contribution in the PDR should be accounted for, which, assuming equipartition, would result in thermal pressures a factor of 3 lower than the Seo et al. (2019) curve. This would further increase the discrepancy with our estimations. The theoretical law of Seo et al. (2019) is based on an argument of pressure equilibrium between the PDR and ionized gas at the Strömgren radius. In addition, it attempts to account for the photoevaporation flows present in blister-type HII regions by equating the PDR pressure to *twice* the ionized gas pressure (which corresponds to a D-critical ionization front, e.g., Draine 2011, Chap. 37). However, the ionized gas pressure in this D-critical relation should be that at the base of the so-called ionized boundary layer (Bertoldi 1989), which is higher than the global average thermal pressure in the HII region (especially near the densest parts of the surrounding neutral wall). Using a global Strömgren model is thus not justified for a blister-type HII region, and will tend to underestimate thermal pressure in the PDRs. The fact that the relationship observed from molecular tracers is higher than the Strömgren-based law of Seo et al. (2019) is thus compatible with a pressure unbalance, with the global thermal pressure in the HII region unable to contain the thermal pressure of the molecular PDR, and is thus consistent with the expected photoevaporation flows from dense PDRs bordering blister-type HII regions, as initially proposed in Joblin et al. (2018).

Figure 11b shows the map of the ratio between our pressure MMSE estimates and the pressures expected from our G_0 estimates according to the power law fit. This shows which regions of the map are adequately described by this power law relation. In most of the PDRs bordering the central cavity, including the Orion Bar and the eastern PDR, $P_{\rm th}$ from the MMSE is consistent within a factor of two with the power law fit. It reaches a factor of about three above the value expected from G_0 in the Orion S region. Larger discrepancies are found in the outer regions of the map, with the power law fit generally predicting

too high $P_{\rm th}$ values compared to the MMSE estimate. Values of $P_{\rm th}$ from the MMSE reach a factor of up to five lower than expected from G_0 in the south-west and north-east corners of the map. In particular, a ratio below one is found in the ridge that is north of the BN/KL region and the bar that goes east, each already mentioned in the histogram analysis in Fig. 9. As previously discussed, these discrepancies are consistent with the explanation that the physics and chemistry of molecular emission in these outer regions might not be solely driven by FUV irradiation from the Trapezium cluster.

7. Conclusion

This paper introduces BEETROOTS, an open-source software that reconstructs maps of physical parameters from observation maps and an astrophysical model. This software leverages a realistic noise model and exploits spatial regularization to guide the estimations towards physically consistent reconstructions. It uses state-of-the-art reconstruction algorithms, both for optimization (for faster results) and for MCMC sampling (to produce uncertainty quantifications), which can escape from local modes thanks to global exploration steps. These algorithms are able to scale to high dimensions to analyze large observation maps in a reasonable time. Finally, the software quantifies and assesses the ability of the astrophysical model to explain the observations, providing feedback to improve ISM models.

After demonstrating the power of BEETROOTS on synthetic data, we used it to deduce physical parameter maps ($P_{\rm th}$, G_0, $A_V^{\rm tot}$) from the large observation map of the Orion molecular cloud 1 (OMC-1) of Goicoechea et al. (2019). To the best of our knowledge, this inference is the first performed thus far on such a large map (that is, with 1 000 pixels or more) for multiple physical parameters from molecular tracers. It is also the first time that the uncertainties have been quantified in an inference of maps of physical parameters from molecular tracers. We show that our PDR model successfully reproduces the observed line intensity maps, with parameter maps that are spatially coherent and in which all of the main features of the region can be recognized. We demonstrate, in particular, that our spatially regularized Bayesian approach is key in achieving this success, as simpler approaches yield unusable noise-dominated parameter maps. Our uncertainty quantification also shows that the five considered molecular lines are able to constrain G_0 and the thermal pressure over a wide range of conditions. Moreover, we demonstrate that our G_0 reconstructions are consistent with those obtained from FIR luminosities (i.e., dust thermal emission) over most of the central part of the map. This indicates that warm molecular emission over a large area of the map is controlled by UV irradiation from the Trapezium cluster. We find high $P_{\rm th}$ values in a significant portion of the map, including the full Orion Bar region as well as the eastern PDR, confirming that the high pressure values found by Joblin et al. 2018 at one position in the Orion Bar are widespread in the dense PDRs of OMC-1. In addition, we find the thermal pressures to be correlated to the incident G_0 with a near-linear relationship, thereby confirming and extending a finding of several previous studies (Joblin et al. 2018; Wu et al. 2018; Seo et al. 2019; Pabst et al. 2022). We ascribe this to the photoevaporation-driven compression of the PDRs surrounding a blister-type HII region.

This work demonstrates how much can be gained by using specially built state-of-the-art Bayesian inversion methods to derive physical parameter maps from spatially resolved line intensity maps of interstellar clouds. This paves the way to a systematic and rigorous analysis of observations produced by past, current, and future instruments. Several further improvements of the method proposed here can also be foreseen. As shown in Sect. 5, the current implementation of BEETROOTS can perform inference on maps of up to about 10^4 pixels on a personal laptop. However, it does not scale to larger Galactic observations containing hundreds of thousands or millions of pixels, such as the Orion B dataset (Pety et al. 2017) obtained at IRAM. Analyzing such datasets at once requires a distributed version of BEETROOTS. In addition, such observation maps often cover a variety of environments that cannot be accurately simulated with a single astrophysical model; for instance, some areas might contain astrophysical shocks, while other dense and dark regions might be governed by slow, time-dependent chemical evolution with little influence from UV irradiation. To analyze such clouds, BEETROOTS would need to combine a number of models describing multiple environments (e.g., PDR models, shock models, HII region models, time-dependent chemistry models, etc.). This would enable the determination of which environment best describes each pixel, while inferring the associated physical parameters. Both improvements are left for a future work.

Acknowledgements. This work received support from the French Agence Nationale de la Recherche through the DAOISM grant ANR-21-CE31-0010, and from the Programme National "Physique et Chimie du Milieu Interstellaire" (PCMI) of CNRS/INSU with INC/INP, co-funded by CEA and CNES. It also received support through the ANR grant "MIAI @ Grenoble Alpes" ANR-19-P3IA-0003. This work was partly supported by the CNRS through 80Prime project OrionStat, a MITI interdisciplinary program, by the ANR project "Chaire IA Sherlock" ANR-20-CHIA-0031-01 held by P. Chainais, and by the national support within the *programme d'investissements d'avenir* ANR-16-IDEX-0004 ULNE and Région HDF. JRG and MGSM thank the Spanish MCINN for funding support under grants PID2019-106110G-100 and PID2023-146667NB-I00. MSGM acknowledges support from the NSF under grant CAREER 2142300. Part of the research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). D.C.L. acknowledges financial support from the National Aeronautics and Space Administration (NASA) Astrophysics Data Analysis Program (ADAP).

References

Andrae, R., Schulze-Hartung, T., & Melchior, P. 2010, arXiv e-prints [arXiv:1012.3754]
Asensio Ramos, A., & Elitzur, M. 2018, A&A, 616, A131
Behrens, E., Mangum, J. G., Holdship, J., et al. 2022, ApJ, 939, 119
Bertoldi, F. 1989, ApJ, 346, 735
Blanc, G. A., Kewley, L., Vogt, F. P. A., & Dopita, M. A. 2015, ApJ, 798, 99
Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132
Brinch, C., & Hogerheijde, M. R. 2010, A&A, 523, A25
Bron, E., Agúndez, M., Goicoechea, J. R., & Cernicharo, J. 2018, arXiv e-prints [arXiv:1801.01547]
Bron, E., Roueff, E., Gerin, M., et al. 2021, A&A, 645, A28
Buchner, J. 2016, UltraNest: Pythonic Nested Sampling Development Framework and UltraNest, Astrophysics Source Code Library [record ascl:1611.001]
Buchner, J. 2021, J. Open Source Softw., 6, 3001
Chevallard, J., & Charlot, S. 2016, MNRAS, 462, 1415
Chevallard, J., Charlot, S., Wandelt, B., & Wild, V. 2013, MNRAS, 432, 2061
Chevance, M., Madden, S. C., Lebouteiller, V., et al. 2016, A&A, 590, A36
Christensen, N., Meyer, R., Knox, L., & Luey, B. 2001, Class. Quant. Grav., 18, 2677
Ciurlo, A., Paumard, T., Rouan, D., & Clénet, Y. 2016, A&A, 594, A113
Da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
de Mijolla, D., Viti, S., Holdship, J., Manolopoulou, I., & Yates, J. 2019, A&A, 630, A117
Del Moral, P., Doucet, A., & Jasra, A. 2006, J. Roy. Statist. Soc. Ser. B: Statist. Methodol., 68, 411
Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium
Dullemond, C. P., Juhasz, A., Pohl, A., et al. 2012, RADMC-3D: A multi-purpose radiative transfer tool, Astrophysics Source Code Library [record ascl:1202.015]
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, Rev. Mexicana Astron. Astrofis., 53, 385

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Fitzpatrick, E. L., & Massa, D. 2007, ApJ, 663, 320
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Galliano, F. 2018, MNRAS, 476, 1445
Galliano, F., Madden, S. C., Jones, A. P., et al. 2003, A&A, 407, 159
Galliano, F., Nersesian, A., Bianchi, S., et al. 2021, A&A, 649, A18
Gelman, A., Meng, X.-L., & Stern, H. 1996, Statist. Sinica, 6, 733
Godard, B., Pineau des Forêts, G., Lesaffre, P., et al. 2019, A&A, 622, A100
Goicoechea, J. R., & Le Bourlot, J. 2007, A&A, 467, 1
Goicoechea, J. R., Teyssier, D., Etxaluze, M., et al. 2015, ApJ, 812, 75
Goicoechea, J. R., Santa-Maria, M. G., Bron, E., et al. 2019, A&A, 622, A91
Goodman, J., & Weare, J. 2010, Commun. Appl. Math. Computat. Sci., 5, 65
Gordon, K. D., Roman-Duval, J., Bot, C., et al. 2014, ApJ, 797, 85
Grassi, T., Merlin, E., Piovan, L., Buonomo, U., & Chiosi, C. 2011, arXiv e-prints [arXiv:1103.0509]
Grassi, T., Nauman, F., Ramsey, J. P., et al. 2022, A&A, 668, A139
Gratier, P., Majumdar, L., Ohishi, M., et al. 2016, ApJS, 225, 25
Habart, E., Peeters, E., Berné, O., et al. 2024, A&A, 685, A73
Heays, A. N., Bosman, A. D., & van Dishoeck, E. F. 2017, A&A, 602, A105
Holdship, J., Viti, S., Jiménez-Serra, I., Makrymallis, A., & Priestley, F. 2017, AJ, 154, 38
Holdship, J., Jeffrey, N., Makrymallis, A., Viti, S., & Yates, J. 2018, ApJ, 866, 116
Holdship, J., Viti, S., Haworth, T. J., & Ilee, J. D. 2021, A&A, 653, A76
Indriolo, N., Geballe, T. R., Oka, T., & McCall, B. J. 2007, ApJ, 671, 1736
Joblin, C., Bron, E., Pinto, C., et al. 2018, A&A, 615, A129
Jóhannesson, G., Ruiz de Austri, R., Vincent, A. C., et al. 2016, ApJ, 824, 16
Juvela, M. 2020, A&A, 644, A151
Juvela, M., Montillaud, J., Ysard, N., & Lunttila, T. 2013, A&A, 556, A63
Kamenetzky, J., Rangwala, N., Glenn, J., Maloney, P. R., & Conley, A. 2014, ApJ, 795, 174
Keil, M., Viti, S., & Holdship, J. 2022, ApJ, 927, 203
Kelly, B. C., Shetty, R., Stutz, A. M., et al. 2012, ApJ, 752, 55
Le Petit, F., Roueff, E., & Herbst, E. 2004, A&A, 417, 993
Le Petit, F., Nehme, C., Le Bourlot, J., & Roueff, E. 2006, ApJS, 164, 506
Lebouteiller, V., & Ramambason, L. 2022, A&A, 667, A34
Lee, M.-Y., Madden, S. C., Le Petit, F., et al. 2019, A&A, 628, A113
Lewis, A., & Bridle, S. 2002, Phys. Rev. D, 66, 103511
Makrymallis, A., & Viti, S. 2014, ApJ, 794, 45
Marchal, A., Miville-Deschenes, M.-A., Orieux, F., et al. 2019, A&A, 626, A101
Marconi, A., Testi, L., Natta, A., & Walmsley, C. M. 1998, A&A, 330, 696
Mathis, J. S., Mezger, P. G., & Panagia, N. 1983, A&A, 128, 212
McElroy, D., Walsh, C., Markwick, A. J., et al. 2013, A&A, 550, A36
Menten, K. M., Reid, M. J., Forbrich, J., & Brunthaler, A. 2007, A&A, 474, 515
Möller, T., Bernst, I., Panoglou, D., et al. 2013, A&A, 549, A21
Pabst, C. H. M., Goicoechea, J. R., Hacar, A., et al. 2022, A&A, 658, A98
Pacifici, C., Charlot, S., Blaizot, J., & Brinchmann, J. 2012, MNRAS, 421, 2002
Palud, P. 2023, PhD thesis, Lille, France, https://theses.hal.science/tel-04424965
Palud, P., Chainais, P., Le Petit, F., Thouvenin, P.-A., & Bron, E. 2023a, in 29° Colloque sur le traitement du signal et des images (Grenoble: GRETSI – Groupe de Recherche en Traitement du Signal et des Images), 705
Palud, P., Einig, L., Le Petit, F., et al. 2023b, A&A, 678, A198
Palud, P., Thouvenin, P.-A., Chainais, P., Bron, E., & Le Petit, F. 2023, IEEE Trans. Signal Process., 71, 2491
Panter, B., Heavens, A. F., & Jimenez, R. 2003, MNRAS, 343, 1145
Paradis, D., Veneziani, M., Noriega-Crespo, A., et al. 2010, A&A, 520, L8
Paumard, T., Ciurlo, A., Morris, M. R., Do, T., & Ghez, A. M. 2022, A&A, 664, A97
Paumard, T., Morris, M. R., Do, T., & Ghez, A. 2014, in IAU Symposium, 303, The Galactic Center: Feeding and Feedback in a Normal Galactic Nucleus, eds. L. O. Sjouwerman, C. C. Lang, & J. Ott, 109
Pérez-Montero, E. 2014, MNRAS, 441, 2663
Pety, J., Guzmán, V. V., Orkisz, J. H., et al. 2017, A&A, 599, A98
Pinte, C., Ménard, F., Duchêne, G., et al. 2022, MCFOST: Radiative transfer code, Astrophysics Source Code Library [record ascl:2207.023]
Ramambason, L., Lebouteiller, V., Bik, A., et al. 2022, A&A, 667, A35
Robert, C. P., & Casella, G. 2004, Monte Carlo Statistical Methods, Springer Texts in Statistics (New York, NY: Springer New York)
Roelfsema, P. R., Helmich, F. P., Teyssier, D., et al. 2012, A&A, 537, A17
Röllig, M., & Ossenkopf-Okada, V. 2022, A&A, 664, A67
Röllig, M., Abel, N. P., Bell, T., et al. 2007, A&A, 467, 187
Roueff, A., Gerin, M., Gratier, P., et al. 2021, A&A, 645, A26
Schilke, P., Comito, C., Müller, H. S. P., et al. 2010, A&A, 521, L11
Seo, Y. M., Goldsmith, P. F., Walker, C. K., et al. 2019, ApJ, 878, 120
Sheffer, Y., & Wolfire, M. G. 2013, ApJ, 774, L14
Sheffer, Y., Wolfire, M. G., Hollenbach, D. J., Kaufman, M. J., & Cordier, M. 2011, ApJ, 741, 45
Skilling, J. 2004, in Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, eds. R. Fischer, R. Preuss, & U. V. Toussaint (AIP), American Institute of Physics Conference Series, 735, 395
Smirnov-Pinchukov, G. V., Molyarova, T., Semenov, D. A., et al. 2022, A&A, 666, L8
Speagle, J. S. 2020, MNRAS, 493, 3132
Sutherland, R., Dopita, M., Binette, L., & Groves, B. 2018, MAPPINGS V: Astrophysical plasma modeling code, Astrophysics Source Code Library [record ascl:1807.005]
Thomas, A. D., Dopita, M. A., Kewley, L. J., et al. 2018, ApJ, 856, 89
Tielens, A. G. G. M., & Hollenbach, D. 1985, ApJ, 291, 722
Thrane, E., & Talbot, C. 2019, PASA, 36, e010
Vale Asari, N., Stasińska, G., Morisset, C., & Cid Fernandes, R. 2016, MNRAS, 460, 1739
van der Tak, F., Black, J., Schoeier, F., Jansen, D., & van Dishoeck, E. 2007, A&A, 468, 627
Veneziani, M., Piacentini, F., Noriega-Crespo, A., et al. 2013, ApJ, 772, 56
Vidal, A. F., De Bortoli, V., Pereyra, M., & Durmus, A. 2020, SIAM J. Imaging Sci., 13, 1945
Villa-Vélez, J. A., Buat, V., Theulé, P., Boquien, M., & Burgarella, D. 2021, A&A, 654, A153
Wakelam, V., Herbst, E., Loison, J. C., et al. 2012, ApJS, 199, 21
Wolfire, M. G., Vallini, L., & Chevance, M. 2022, ARA&A, 60, 247
Wu, R., Madden, S. C., Galliano, F., et al. 2015, A&A, 575, A88
Wu, R., Bron, E., Onaka, T., et al. 2018, A&A, 618, A53
Yang, C., Omont, A., Beelen, A., et al. 2017, A&A, 608, A144
Yu, Y., & Meng, X.-L. 2011, J. Computat. Graph. Statist., 20, 531
Zucker, C., Schlafly, E. F., Speagle, J. S., et al. 2018, ApJ, 869, 83
Zucker, C., Speagle, J. S., Schlafly, E. F., et al. 2019, ApJ, 879, 125
Zucker, C., Goodman, A., Alves, J., et al. 2021, ApJ, 919, 35

[1] LUX, UMR 8262, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, 92190 Meudon, France
[2] Univ. Lille, CNRS, Centrale Lille, UMR 9189 – CRIStAL, 59651 Villeneuve d'Ascq, France
[3] Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
[4] Instituto de Física Fundamental (CSIC), Calle Serrano 121, 28006, Madrid, Spain
[5] Department of Astronomy, University of Florida, PO Box 112055, Gainesville, FL 32611, USA
[6] LUX, UMR 8262, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, 75014 Paris, France
[7] IRAM, 300 rue de la Piscine, 38406 Saint-Martin-d'Hères, France
[8] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, GIPSA-Lab, Grenoble 38000, France
[9] Univ. Toulon, Aix Marseille Univ., CNRS, IM2NP, Toulon, France
[10] Department of Earth, Environment, and Physics, Worcester State University, Worcester, MA 01602, USA
[11] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
[12] Institut de Recherche en Astrophysique et Planétologie (IRAP), Université Paul Sabatier, Toulouse cedex 4, France
[13] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
[14] Laboratoire d'Astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, Allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[15] Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile
[16] Laboratoire de Physique de l'École normale supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, Sorbonne Paris Cité, Paris, France
[17] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[18] School of Physics and Astronomy, Cardiff University, Queen's buildings, Cardiff CF24 3AA, UK

Appendix A: How to use BEETROOTS in practice

This section describes the steps to take in practice to run BEETROOTS on an observation, $\mathbf{Y}$. Figure A.1 summarizes the full process in the case of an observation map affected by both a multiplicative noise and an additive noise, regularized with a spatial prior. The STD of the multiplicative noise, σ_m, and the STD map of additive noise, $(\sigma_{a,n\ell})$, that affect the observation are assumed to be provided by observers, which is generally the case in ISM.

Before performing the inference, the user needs to choose an astrophysical model that can predict the observables contained in $\mathbf{Y}$ from the physical parameters of interest, $\mathbf{\Theta}$. To fully exploit the potential of BEETROOTS, this model should be fast to run and give access to its gradient, $\nabla\mathbf{M}$. With ANNs, this gradient is simple to obtain using auto-differentiation techniques. To fully exploit the ANN characteristic (speed, accuracy, and gradient), BEETROOTS has an interface with NNBMA[8] (Palud et al. 2023b). Other astrophysical models can also be interfaced with BEETROOTS, even those that do not allow for a gradient evaluation. Then, the user should select the noise model to be used – see discussion in Sect. 2.3. In a first run, a purely Gaussian approximation can be used as a default choice. Finally, the user can manually set the value of the spatial regularization prior, τ_d, or automatically adjust it during sampling. If manually set, low values should initially be preferred to avoid biasing the estimation towards over-smoothed maps.

The sampling or optimization parameters are introduced in Sect. 3.3. BEETROOTS produces a list of posterior samples (in sampling mode) or iterates (in optimization mode), denoted $\mathbf{\Theta}^{(t)}$, which are used to derive estimators. To avoid misinterpretations, we recommend performing some sanity checks before analyzing the reconstructions. First, it is necessary to check whether the algorithm converged using the figure of the loss function evolution with iterations (produced automatically). For optimization, the loss function should be constant during the last iterations. For a sampling procedure, it should look noisy around a constant value. Then, for the sampling procedure, the exploration of the posterior can be checked with the average acceptance frequencies of the kernels. Ideally, this acceptance frequency should be around 60% for PMALA and non-zero for MTM (Palud et al. 2023) (the higher the better). The effective sample size (ESS) quantifies how efficiently the sampler explored the posterior distribution by indicating the effective number of independent samples within the returned Markov chain (ESS maps are produced automatically). For a sampling procedure with $T_{\mathrm{MC}} = 10^4$ iterations, the ESS should ideally be larger than one hundred on most pixels of the map. See Robert & Casella (2004, chapter 12) for more information on MCMC convergence diagnosis If the posterior samples pass the sanity checks, then the ability of the model to reproduce the observations should be tested, using the model checking approach presented in Sect. 6.3.1 (maps of p-value and of results of the pixelwise hypothesis testing are produced automatically). If the observations are well reproduced, then the reconstructions and associated uncertainty quantifications can be analyzed from automatically produced maps of MMSE and UFs.

BEETROOTS was built to automatically perform all these steps for each new run. For this reason, we recommend it as an all-in-one tool for future analyses of ISM observations.

[8] https://github.com/einigl/ism-model-nn-approximation

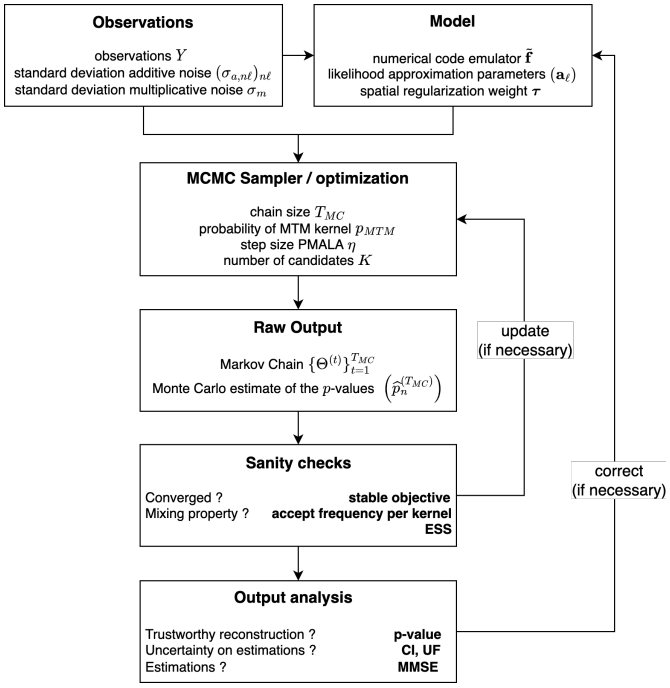


Fig. A.1: Full inference workflow with BEETROOTS.

Appendix B: More on statistical modeling and estimation

B.1. Likelihood for upper limits

Upper bounds on observations are quite common in ISM studies. The corresponding observables are sometimes discarded in the inversion process. Omitting observables leads to a loss of information that could damage the inference results. Including these upper bounds on observations in the likelihood function enables us to account for all constraints provided by the observations. In statistics, censorship enables us to include such upper limits $\omega \geq 0$. In case of Gaussian noise, the observation model is

$$y_{n\ell} = \max\left\{\omega,\ M_\ell(\boldsymbol{\theta}_n) + \varepsilon_{n\ell}^{(a)}\right\} = \begin{cases} \omega & \text{if } M_\ell(\boldsymbol{\theta}_n) + \varepsilon_{n\ell}^{(a)} \leq \omega \\ M_\ell(\boldsymbol{\theta}_n) + \varepsilon_{n\ell}^{(a)} & \text{otherwise} \end{cases} \tag{B.1}$$

In (Ramambason et al. 2022), censorship is modeled with a half-normal distribution $\mathcal{N}_-(\omega, \sigma^2)$. This approach prohibits all values above the censoring threshold ω, while values slightly above ω should also be allowed. Besides, it biases $M_\ell(\boldsymbol{\theta}_n)$ towards values close to the threshold ω, while $M_\ell(\boldsymbol{\theta}_n) \ll \omega$ also satisfies the upper bound constraint. Statistically, a better encoding of censorship in the likelihood is (Robert & Casella 2004, chapter 1)

$$-\log \pi(\omega|\boldsymbol{\theta}_n) = -\log \int_{-\infty}^{\omega} \pi(y_{n\ell}|\boldsymbol{\theta}_n)\ \mathrm{d}y_{n\ell}. \tag{B.2}$$

The integral covers all the integrated intensities $y_{n\ell} \leq \omega$, and sums the associated uncensored likelihoods $\pi(y_{n\ell}|\boldsymbol{\theta}_n)$.

Figure B.1 illustrates this censorship modeling for a Gaussian additive random deterioration $\varepsilon_{n\ell}^{(a)} \sim \mathcal{N}(0, \sigma^2)$. In this case, the integral in Eq. B.2 can be evaluated efficiently for Gaussian noise using the Gaussian cumulative density function (CDF) $\Phi(\omega|M_\ell(\boldsymbol{\theta}), \sigma^2)$. Unlike the half Gaussian mentioned above, it does not forbid all physical parameters $\boldsymbol{\theta}$ such that $M_\ell(\boldsymbol{\theta}) > \omega$.

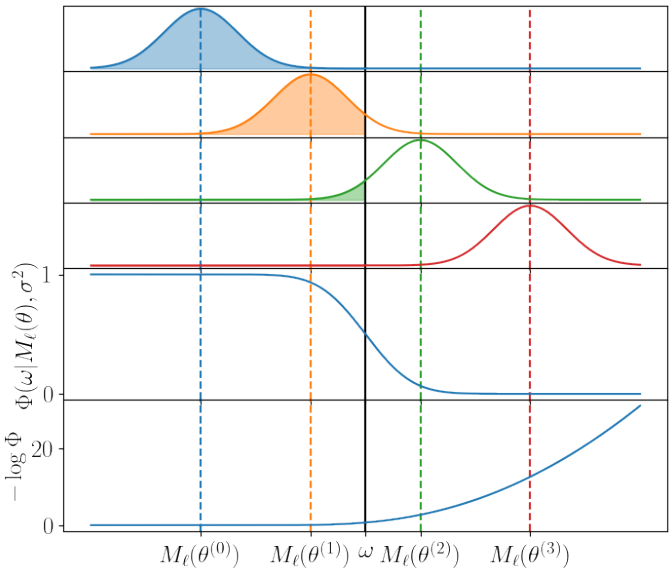


Fig. B.1: Illustration of censorship in likelihood for Gaussian additive random deteriorations. The first four panels show four values of $M_\ell(\boldsymbol{\theta}^{(i)})$, $i = 0,\ldots,3$ and the associated Gaussian additive PDF. The censored portion, below the censorship threshold ω, is highlighted. The fifth panel shows the likelihood function taking censorship into account, that is, the Gaussian CDF Φ. The sixth panel shows the corresponding negative log-likelihood.

Instead, it smoothly and increasingly penalizes them as $M_\ell(\boldsymbol{\theta})$ gets farther to the threshold limit, that is, as the probability of $y_{n\ell} \leq \omega$ becomes smaller. Some ISM inversion codes resort to this censorship modeling (see, e.g., Blanc et al. 2015; Holdship et al. 2018; Thomas et al. 2018).

B.2. Other sampling methods used in ISM studies

Among ISM studies, inference is often performed with one of four classes of methods in addition to RWMH.

The affine-invariant MCMC sampler (Goodman & Weare 2010) and the associated EMCEE package (Foreman-Mackey et al. 2013) is a very popular MCMC algorithm in astronomy (Thrane & Talbot 2019). In ISM studies, it was applied for instance in astrochemistry (Gratier et al. 2016; Holdship et al. 2018; Keil et al. 2022) and extragalactic molecular gas (Yang et al. 2017). However, this sampler exclusively addresses low dimensional problems. Besides, it fails to explore multimodal distributions and requires to be initialized close to a mode.

Nested sampling (Skilling 2004) is also a very popular Bayesian framework in astrophysics and ISM studies. Popular algorithms include (Feroz et al. 2009), DYNESTY (Speagle 2020) and ULTRANEST (Buchner 2016, 2021), that were developed by and for astrophysicists. MULTINEST was applied to study extragalactic molecular clouds (Kamenetzky et al. 2014; Chevallard & Charlot 2016) and cosmic ray propagation (Jóhannesson et al. 2016). DYNESTY was applied to evaluate distances between the Earth and nearby molecular clouds (Zucker et al. 2018, 2019) or to reconstruct their 3D structure (Zucker et al. 2021). ULTRANEST was applied on extragalactic spectral energy densities (SEDs) (Behrens et al. 2022). These codes handle multimodal distribution but are limited to low dimensional distributions – up to 30 in Feroz et al. (2009) and Jóhannesson et al. (2016).

Sequential MC (Del Moral et al. 2006) is quite rare in ISM. It was applied in Ramambason et al. (2022) in an inversion of integrated intensities of ionic, atomic and molecular emission lines. In that work, it is chosen for its ability to sample from multimodal distributions in low dimension settings – up to 28 in the paper – and to compute the Bayesian evidence $\pi(\mathbf{Y})$ (Eq. (2)).

One sampling algorithm based on a combination of Metropolis-Hastings, Gibbs sampling and the ancillarity-sufficiency interweaving strategy (ASIS) (Yu & Meng 2011) is widespread in dust studies that involve a hierarchical model. ASIS is a sampling strategy that relies on complementary data augmentation schemes to better explore the posterior, including along strongly correlated directions. This sampler was first used in a dust study in Kelly et al. (2012). It was then coupled with a Gibbs sampling approach to exploit the structure in the physical parameter $\boldsymbol{\Theta}$ in very high dimensional applications. For instance, Galliano (2018) involve a few thousand dimensions, and (Galliano et al. 2021) about 10^4.

Appendix C: More results for synthetic case

This section provides more details on the synthetic case application from Sect. 5. First, we display the results on the $N_{\text{side}} = 10$ and 30 use cases. Then, the effect of the regularization weight is studied in more detail. Finally, the model checking results on the three cases are shown and discussed.

C.1. Estimation results for more spatial resolutions

Figures C.1 and C.2 display the reconstructions for all considered estimators in the $N_{\text{side}} = 10$ and 30 use cases, respectively. Results are commented in Sect. 5.3. Table 1 displays the quantitative results associated with the reconstructions.

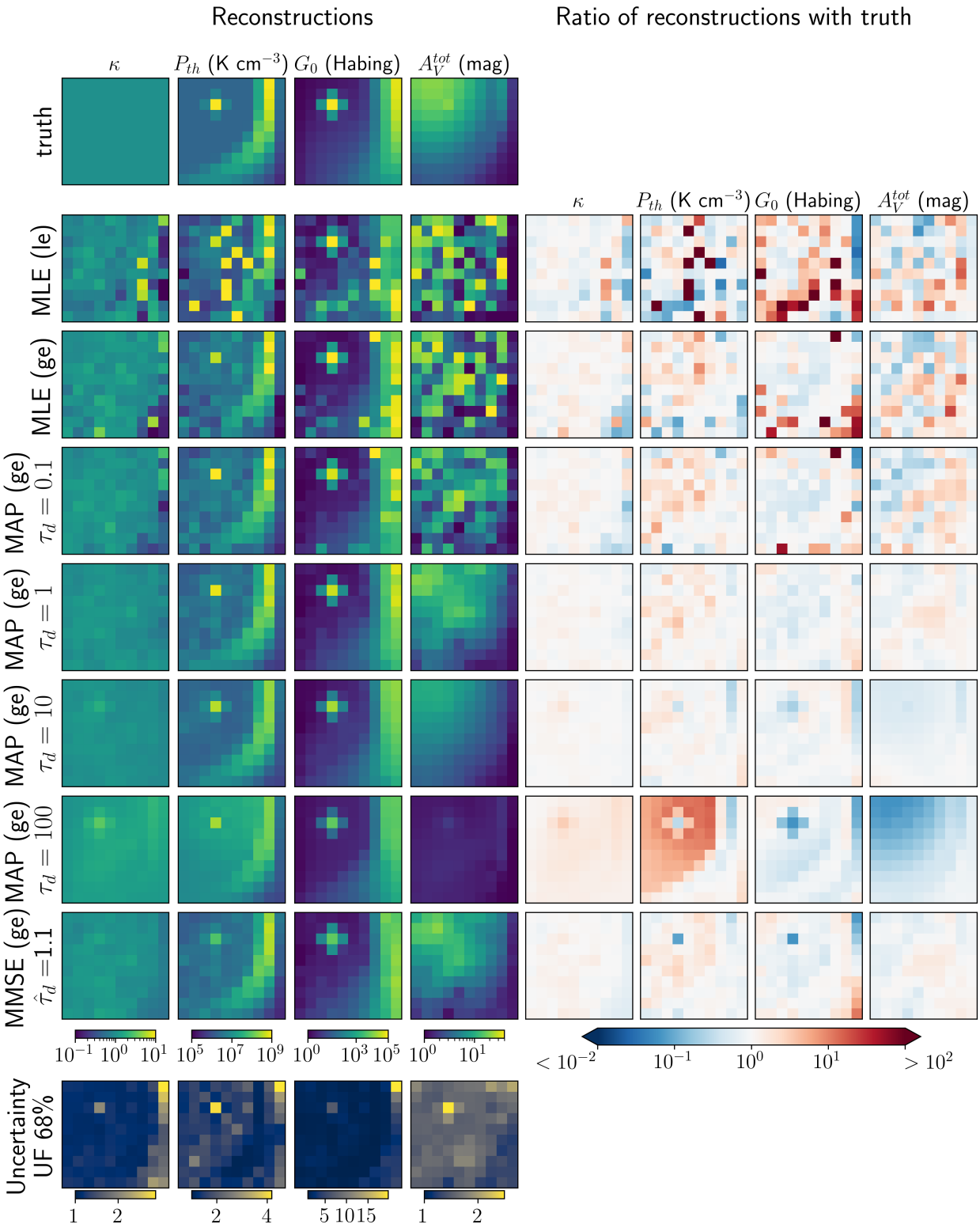


Fig. C.1: Reconstruction results for the $N_{\text{side}} = 10$ spatial resolution. The first row shows the $D = 4$ true maps that each estimator tries to reconstruct. The obtained reconstructions are shown below the true maps. The first estimator is "MLE (le)": the MLE evaluated using the PMALA kernel only, that is, only local exploration. The second estimator is "MLE (ge)": the MLE evaluated using both the PMALA and MTM kernels, that is, a combination of local and global exploration. The third to fifth estimators are MAP estimators (i.e., including spatial regularization) with different values for the regularization weight, τ_d. The sixth estimator is the MMSE, obtained with sampling. The last row, "UF 68%", quantifies the uncertainties associated with the MMSE by indicating the size of the 68% CI. The maps on the right display the ratios between the estimated and true maps, to better assess the quality of each estimation.

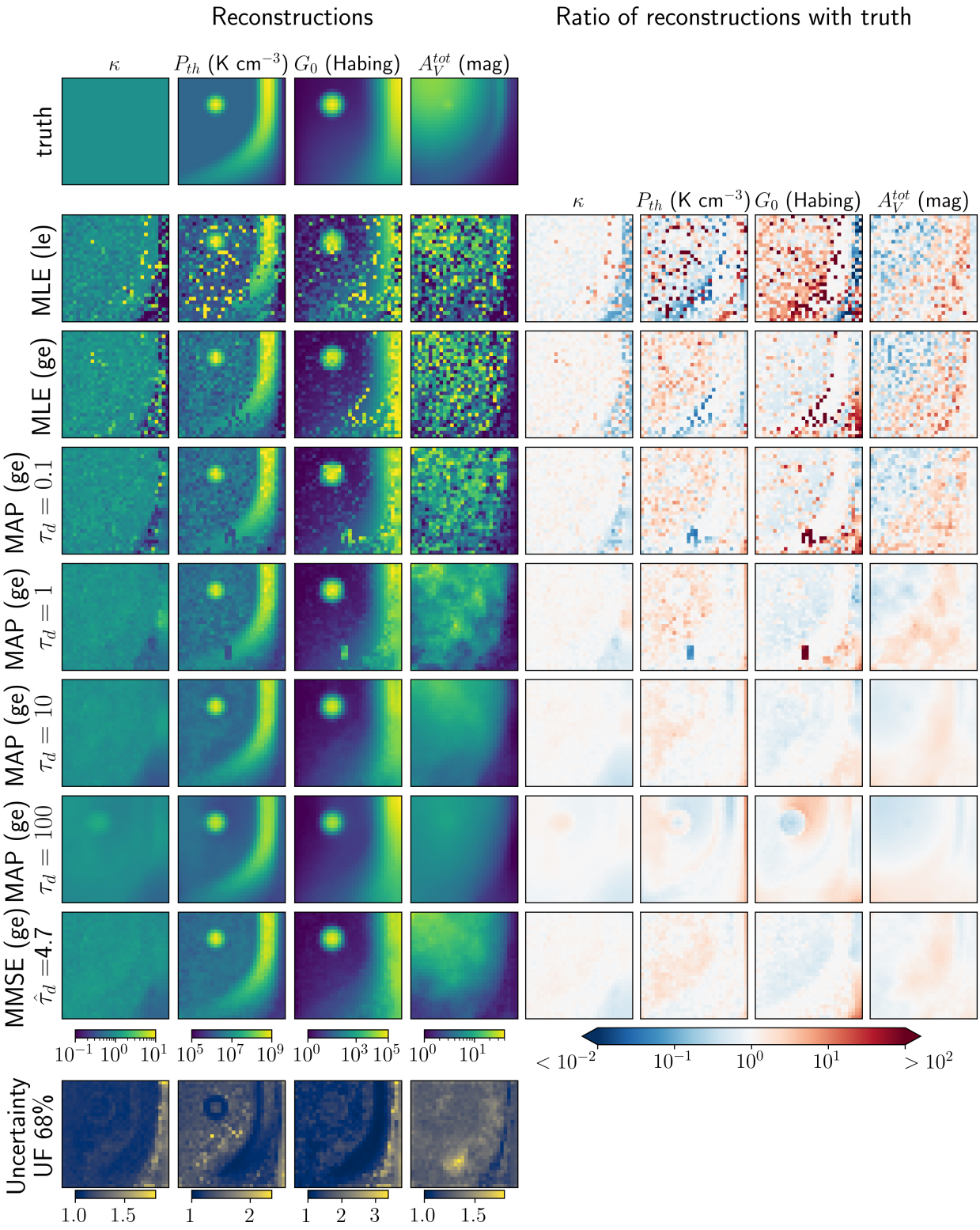


Fig. C.2: Reconstruction results for the $N_{\text{side}} = 30$ spatial resolution. The first row shows the $D = 4$ true maps that each estimator tries to reconstruct. The obtained reconstructions are shown below the true maps. The first estimator is "MLE (le)": the MLE evaluated using the PMALA kernel only, that is, only local exploration. The second estimator is "MLE (ge)": the MLE evaluated using both the PMALA and MTM kernels, that is, a combination of local and global exploration. The third to fifth estimators are MAP estimators (i.e., including spatial regularization) with different values for the regularization weight, τ_d. The sixth estimator is the MMSE, obtained with sampling. The last row, "UF 68%", quantifies the uncertainties associated with the MMSE by indicating the size of the 68% CI. The maps on the right display the ratios between the estimated and true maps, to better assess the quality of each estimation.

C.2. Optimization: Effect of the regularization weight

Figure C.3 shows how the regularization parameter τ_d influences the mean EF (Eq. (11)) for each physical parameter and their combination. In all three cases, very low regularization weights do not regularize enough and lead to large errors. They also complicate the escape from local minima, leading to an artifact in the $N_{\text{side}} = 90$ case, as shown in Fig. 4. Conversely, very high regularization weights bias the reconstructions towards over-smoothed maps. In all three cases, the automatic tuning reaches τ_d values leading to a good likelihood-prior trade-off.

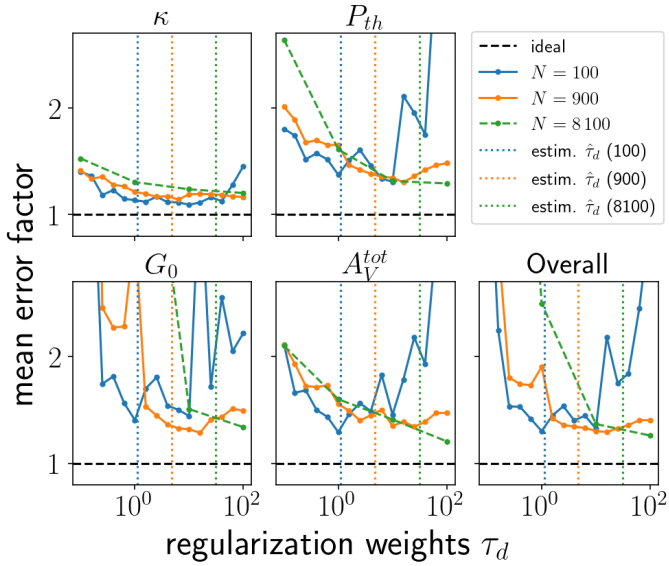


Fig. C.3: Regularization effects: mean EF achieved by the MAP (ge) estimator for different regularization weights τ_d.

C.3. Sampling approach: Model checking results

Figure C.4 shows the results of the model checking performed by BEETROOTS on the three synthetic cases. It shows that the null hypothesis, namely "the astrophysical model **M**, noise characteristics $\mathcal{D}$ and prior (i.e., spatial regularization) can reproduce the observations" is not rejected with the Bayesian hypothesis test for a wide majority of pixels. For each case, some pixels are rejected with confidence level 5%: 2 pixels for the $N_{\text{side}} = 10$ case, 19 pixels for the $N_{\text{side}} = 30$ case, and 117 for the $N_{\text{side}} = 90$ case. For each case, it means that 2%, 2.1% and 1.4% of the total number of pixels are rejected with confidence level 5%, respectively. These pixels are randomly distributed in the map, that is, they do not come from a specific structure in the image. Besides, if the null hypothesis is true, then the p-value has a 5% chance of being below the confidence level for each pixel. Therefore, this model checking analysis confirms that the reconstructions obtained with sampling explain the observations. This means that the reconstructions are meaningful and can be trusted.

In these synthetic cases where the true values are known, the validity of the reconstructions can be assessed directly. This model checking approach is thus secondary here. However, true values are unknown for real observations, as well as the best astrophysical model. Performing model checking is thus critical to assess the relevance of the reconstructions.

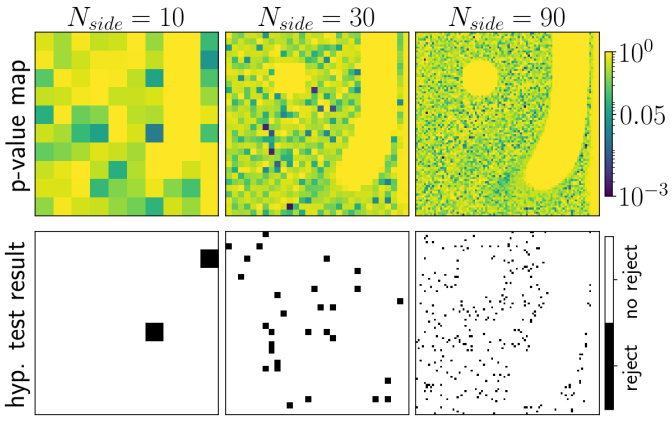


Fig. C.4: Model checking on the three synthetic cases.

Appendix D: More results on OMC-1

This section provides more details on the OMC-1 application from Sect. 6. First, we display the maps of Gaussian noise STD. Then, to detail the model checking results, the posterior predictive distribution is shown for the four highlighted positions.

D.1. Standard deviation maps for the additive Gaussian noise

Figure D.1 displays the maps of the STD $\sigma_{a,n\ell}$ of the additive Gaussian noise for the four considered lines in the OMC-1 inversion. In these maps, the white region in the middle corresponds to the BN/KL region. This region is dominated by shocks, while the Meudon PDR code does not model shocks. Therefore, this region is removed from the inversion. Some other regions are displayed in white on the map borders. In these regions, the STD is not defined. In such cases, an arbitrarily high value is used so that this line does not affect the reconstruction. The associated integrated intensities are shown in Fig. 6.

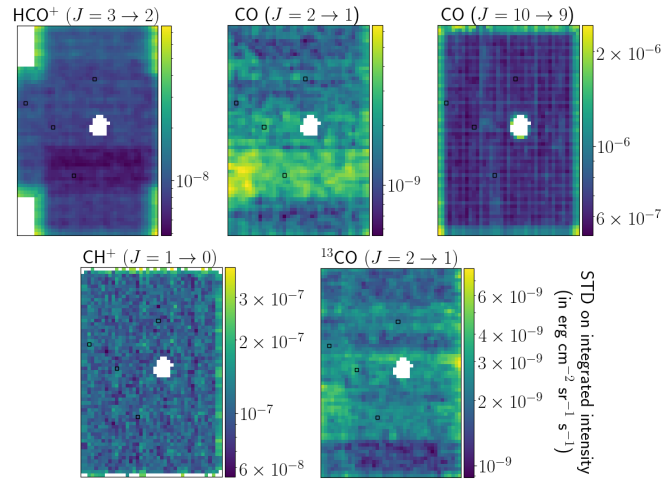


Fig. D.1: Maps of the STD $\sigma_{a,n\ell}$ of the additive Gaussian noise in the observations of OMC-1 used for inversion. The white region in the middle of the maps is the BN/KL region, dominated by shocks, and not considered in the inversion.

D.2. Model checking: Posterior predictive distribution

Figure D.2 compares the posterior predictive distribution (in blue) with the observations $y_{n\ell}$ (orange stars). The predicted integrated intensities show very good agreement with the actual observations. The two PDR pixels contain brighter lines than the two other pixels, which is also visible in Fig. 6.

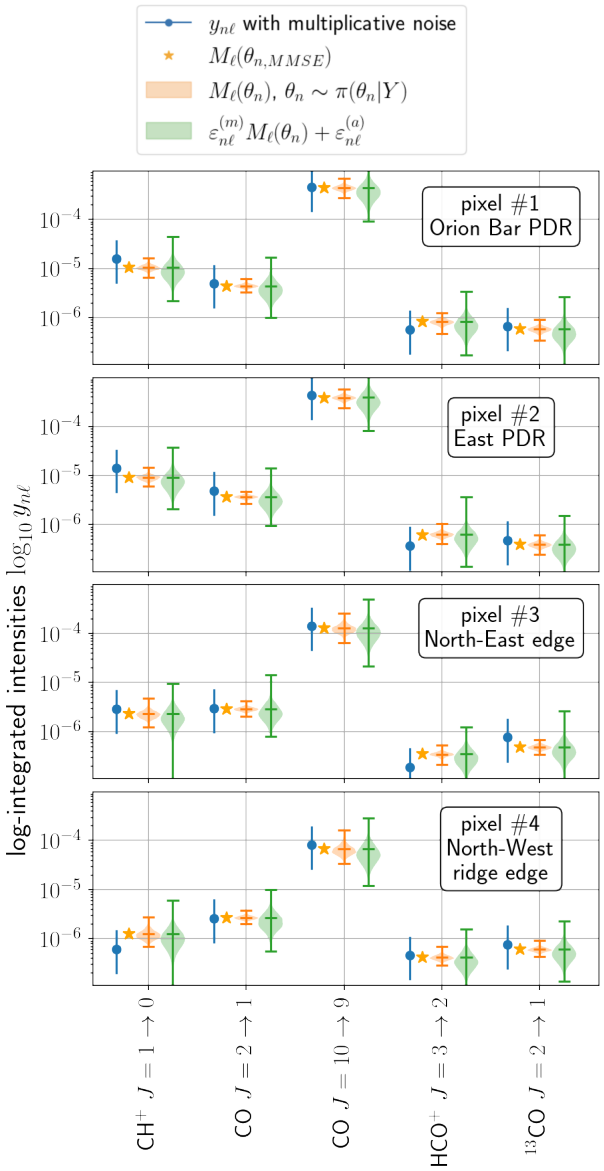


Fig. D.2: Visual diagnostic of the quality of the fit on the lines used for the inversion of OMC-1 for four pixels. The blue points are the actual observation with a 1σ error for the multiplicative noise. The orange stars are the prediction of the Meudon PDR code based on the MMSE. The orange violin plots show the predictions of the Meudon PDR code on samples of the posterior distribution. The green violin plots show the posterior predictive distribution by including addition and multiplicative noises.