Machine Learning Algorithm for Predicting Astronomical Observations

Principal Investigator: Karen Willacy (326); Co-Investigators: Youngmin Seo (326), Umaa Rebbapragada (398)

Program: FY21 R&TD **Strategic Focus Area: Innovative Spontaneous Concepts**

Objectives Develop an algorithm that makes observational predictions without undergoing the detailed and time-consuming modeling step.

Background Making accurate predictions of scientific measurements is crucial in planning missions and writing successful research proposals. In astronomy, a prediction is typically a synthesized image of the target object at the desired wavelength. Such an image is a requirement for justifying feasibility in observational proposals using NASA facilities such as SOFIA and JWST. However, making such an image has always been challenging since it frequently involves heavy theoretical modeling of target objects, which consumes significant resources (time, money, and labor). The theoretical modeling can be avoided when one can use an existing image at a different wavelength to create a predicted image at the desired wavelength if there is a well-established observational and/or theoretical model defining a relationship between observations at the two wavelengths. Up to date, we only have limited knowledge on such relationships but probabilistic approach along with neural networks provide a efficient solution to such problem.

Approach and Results The first task is to estimate the multivariate correlations between the input dust continuum images at 8, 70, 160, 250, and 350 μm to the output image of [CII] emission. In our problem, the correlation between the a single input image and the output image is not sufficiently tight (Figure 1). On the other hand, using five different dust continuum images simultaneously delivers a promising way to constrain the [CII] intensity (Figure 2). The 1-σ width of probability density is ±50 K km/s in Figure 2 from ±250 K km/s in Figure 1. The second task is to build a neural network that can estimate the most probable [CII] intensity when intensities at the five wavelengths are given as an input data set. In principle, we may estimate the most probable [CII] intensity by looking up the table made in Task 1. However, looking up the table having ~0.1 billion rows for every data set consumes significant computation resources as well as time (100 days per image of 1M pixels). Figure 3 shows the comparison of the [CII] images of the RCW120 star-forming region that are predicted using our algorithm and the actual SOFIA observation. The median error of the predicted image to the observed image is 50% for the area with [CII] intensity above 25 K km/s and 24% for the area with [CII] intensity above 50 K km/s. These error values are comparable to the calibration uncertainty in [CII] observations (30 – 70%) and are sufficiently small for the predicted images to be used to check observation feasibility.

Significance/Benefits to JPL and NASA This work demonstrates the proof-of-concept for making observational predictions and gauging mission feasibility through multivariate correlation analysis using machine learning algorithms with sufficient accuracy but without carrying out extensive theoretical modeling for each pixel in the images. [CII] observations are one of the key goals for multiple NASA missions, including SOFIA and GUSTO. Up to date, SOFIA [CII] observations cover less than 1% of area that *Herschel* and *Spitzer* have observed, and there are high demands in observing [CII]. The trained neural network for [CII] prediction is fully ready to be used to make prediction of [CII] observations for these missions. Beyond making prediction of observations, the algorithm can be further expanded to a scientific analysis. A large database of theoretical models may provide multivariate correlations among observables and physical quantities. Our algorithm provides an efficient way to deduce the physical quantities from observables without extensive modeling, which may increase the efficiency of science returns significantly.

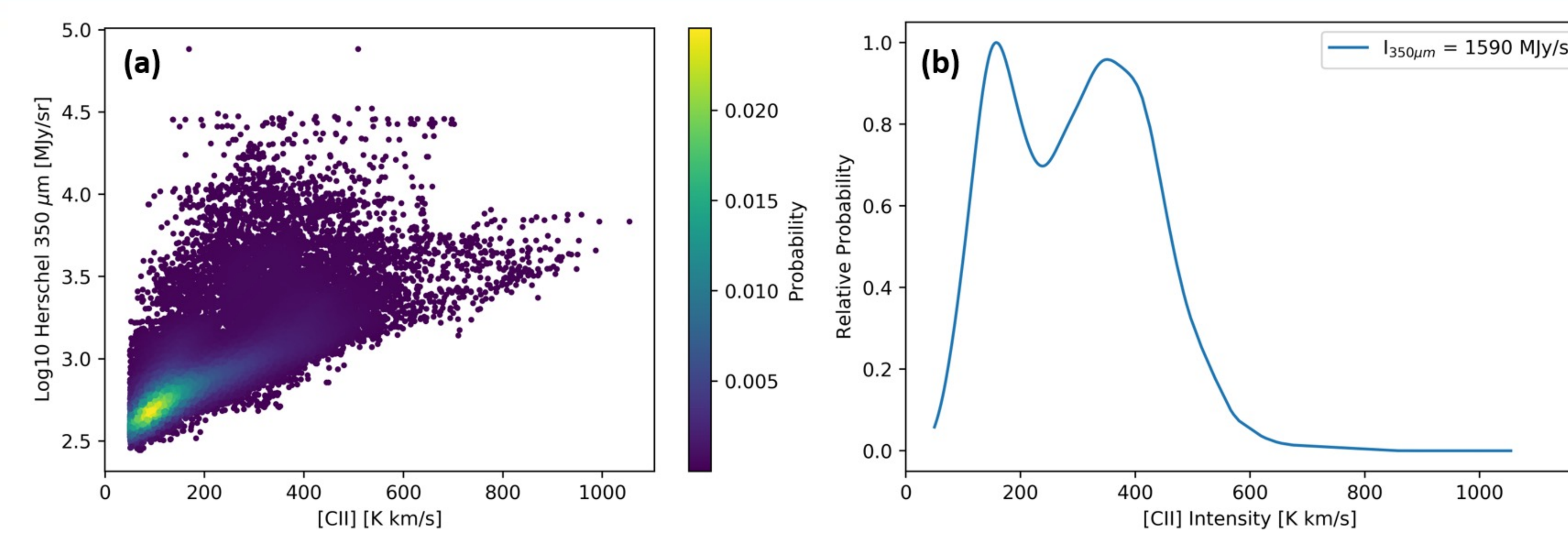

Figure 1. (a) Correlation between intensities of [CII] and dust continuum at 350 μm. Each point corresponds to each pixel in SOFIA [CII] images. (b) The likelihood of [CII] intensity when the dust continuum at 350 μm is 1590 MJy/sr. The relative probability is the probability density normalized by the peak probability.

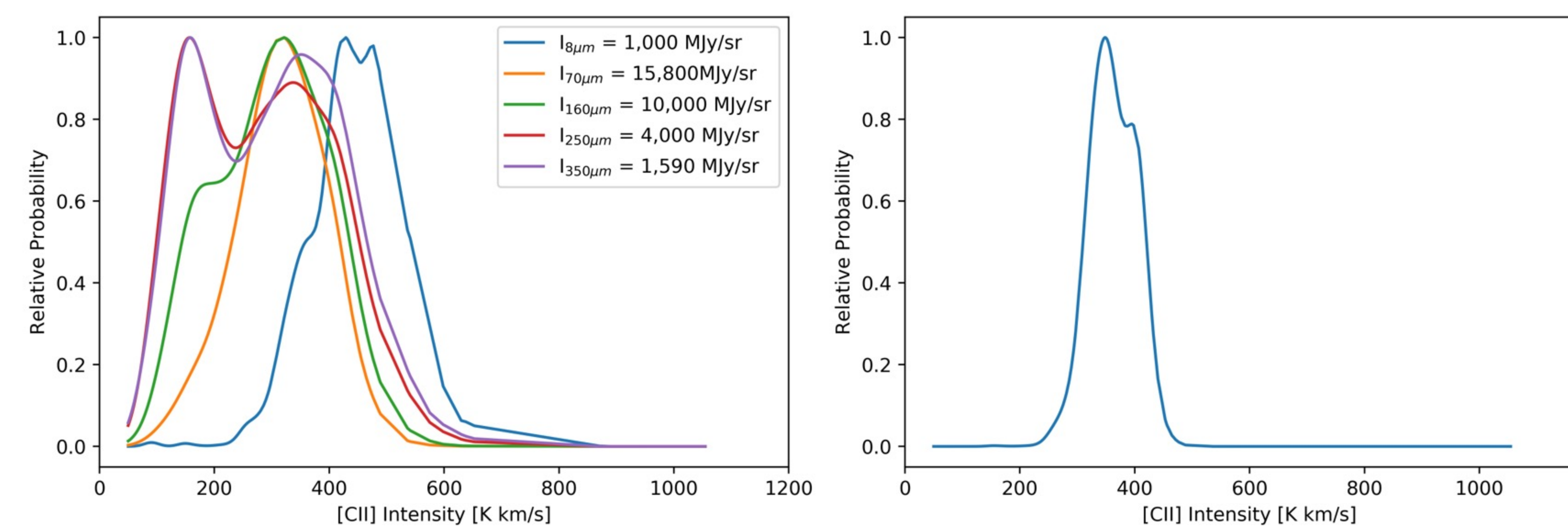

Figure 2. (a) The individual probability density of [CII] intensity for given dust continuum intensities at five different wavelengths. (b) The convolved probability density of [CII] intensity when the five intensities are constraining the [CII] intensity together.

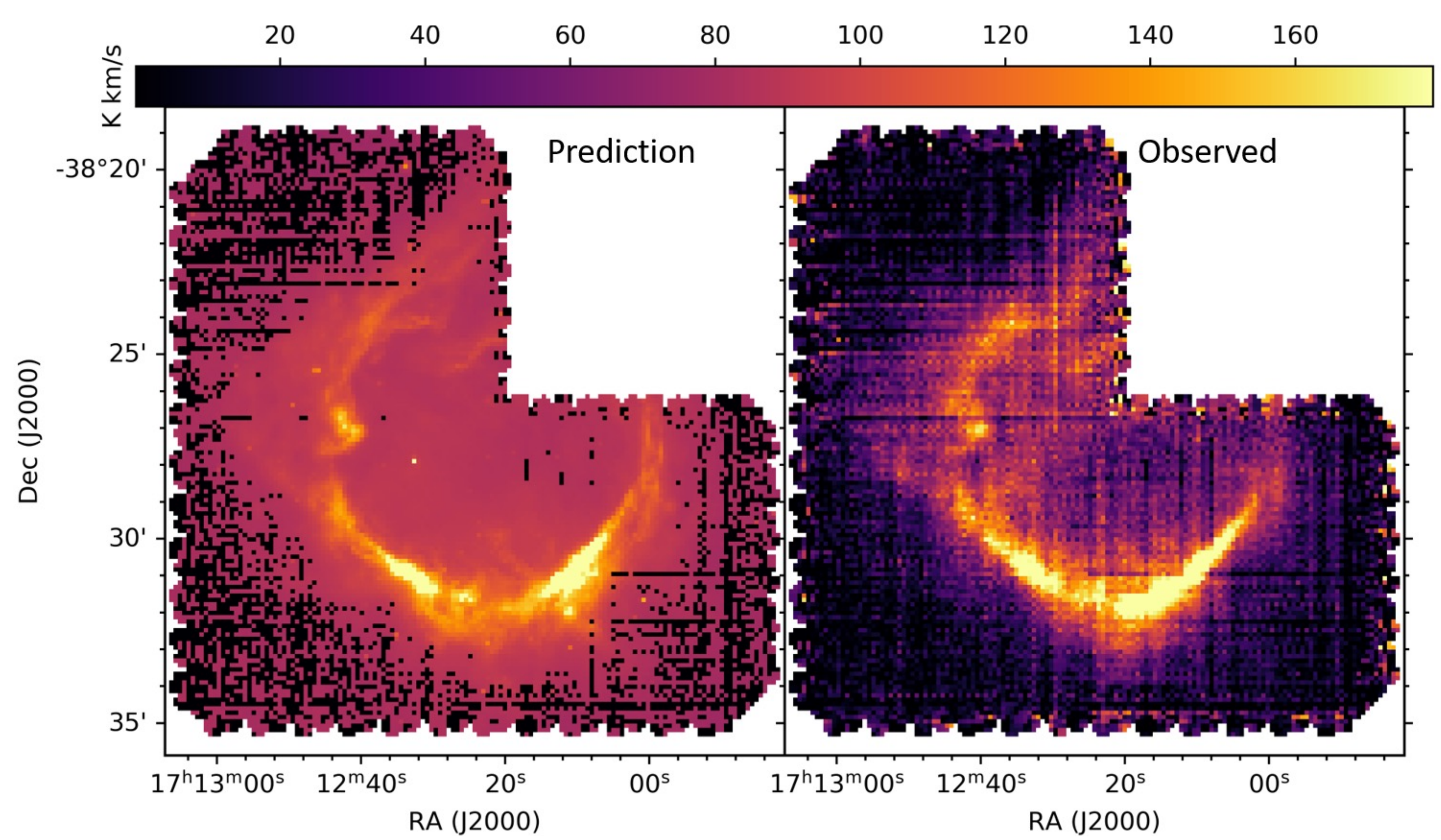

Figure 3. The example of [CII] intensity prediction using our algorithm toward the RCW 120 star-forming region. The left panel shows the prediction image while the right panel is the actual observation using the SOFIA GREAT instrument. The discrepancy between the two images are sufficiently small (<50% to the observed intensities in most of area), demonstrating our algorithm can predict the [CII] intensity accurately.

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

www.nasa.gov

PI/Task Mgr Contact
Email: Karen.Willacy@jpl.nasa.gov

Poster No.
Clearance No.
JPLTask#R21224