BAYESIAN ESTIMATION OF UNCERTAINTIES FOR REDSHIFT INDEPENDENT DISTANCE MEASUREMENTS IN THE NED-D CATALOG

G. Chaparro Molano[1], O. A. Restrepo Gaitán[1,2], J. C. Cuervo Marulanda[3], and S. A. Torres Arzayus[4]

Obtaining individual estimates for uncertainties in redshift-independent galaxy distance measurements can be challenging, as for each galaxy there can be many distance estimates with non-gaussian distributions, some of which may not even have a reported uncertainty. We seek to model uncertainties using a bootstrap sampling of measurements per galaxy per distance estimation method. We then create a predictive bayesian model for estimating galaxy distance uncertainties that is better than simply using a weighted standard deviation. This can be a first step toward predicting distance uncertainties for future catalog-wide analysis.

We selected the Tully-Fisher method for estimating distances as it has the largest number of galaxies and samples in the NED-D redshift-independent galaxy distance catalog (Steer 2016). In order to obtain a distance measurement distribution we do a bootstrap sampling drawn from different distance modulus measurements per galaxy. From the distribution resulting from the bootstrap draws we obtain a mean distance error D_E and an uncertainty for this error Δ (i.e. an uncertainty for the distance error) for each galaxy.

As a second step, we created a Bayesian model that consistently predicts the distance error from D_E and Δ. Our predictive model will be built around the posterior distribution for the parameter estimators. We sample this distribution using the `emcee` Monte Carlo Markov Chain Python package (Foreman-Mackey 2013). For galaxies with more than 14 distance modulus measurements, we were able to create a 6-parameter (g, o, n, f, m, b) second-order predictive Bayesian model (Figure 1):

$$\sigma_{\text{Bayesian}} = gD_E^2 + o\Delta^2 + n\Delta D_E + f\Delta + mD_E + b. \quad (1)$$

The third step is model checking. We validated our model with the help of the Bayesian discrepancy measure test (Gelman 1996). We draw expected values from the posterior predictive distribution and compare those to the original dataset and a model-obtained synthetic dataset simultaneously. Using the Freeman-Tukey statistic for this measure, we obtained a Bayesian p-value of 0.18. Since this value is within the range (0.025, 0.975) we can say that the null hypothesis (the model is inconsistent with the data) cannot be rejected (Brooks 2000). However, for galaxies with a lower number of distance measurements or for lower-order models, the null hypothesis is rejected.

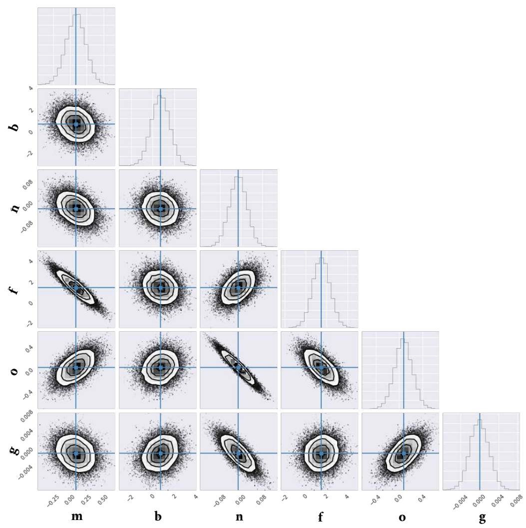

Fig. 1. Monte Carlo Markov Chain obtained posterior probability projections for our accepted predictive model estimators of the parameters in Equation (1).

CONCLUSION

Our model reproduces well the bootstrap-sampled distance-error data from the NED-D catalog. However, we need to extend this model to less-sampled galaxies and to other extragalactic distance estimation methods besides Tully-Fisher.

REFERENCES

Brooks, S. P., Catchpole, E. A., Morgan B. J. T. F. 2000, Statist. Sci., 15, 357

Foreman-Mackey, D., et al. 2013, PASP, 125, 306

Gelman, A., Meng, X.L., Stern, H. 1996, Stat. Sin., 6, 733

Steer. I. et al, 2017, AJ, 153, 1

[1]Grupo de Simulación, Análisis y Modelado (SiAMo), Universidad ECCI, Bogotá, Colombia (gchaparrom@ecci.edu.co).

[2]Radio Astronomy Instrumentation Group, Universidad de Chile, Santiago de Chile, Chile.

[3]Observatorio Astronómico, Universidad Sergio Arboleda, Bogotá, Colombia.

[4]Centro Internacional de Física, Bogotá, Colombia.