
Bridging the Microwave Data Gap; Using Bayesian Deep Learning to “See” the Unseen

Pedro Ortiz

Department of Computer Science
Naval Postgraduate School
Monterey, CA 93943
pedro.ortiz@nps.edu

Eleanor Casas

Department of Meteorology
Naval Postgraduate School
Monterey, CA 93943
eleanor.casas@nps.edu

Marko Orescanin

Department of Computer Science
Naval Postgraduate School
Monterey, CA 93943
marko.orescanin@nps.edu

Scott W. Powell

Department of Meteorology
Naval Postgraduate School
Monterey, CA 93943
scott.powell@nps.edu

Abstract

Having microwave data with the spatial and temporal resolution of infrared data would provide a large positive impact on many climate and weather applications. We demonstrate that Bayesian deep learning is a promising technique for both creating and improving synthetic microwave data from infrared data. We report 0.7% mean absolute percentage error for 183 ± 3 GHz microwave brightness temperature. Based on quantified uncertainty metrics, we find that more training data is needed to achieve improved performance at 166 GHz, 37 GHz, and 23 GHz. Analysis of the spatial distribution of uncertainty reveals that additional cloud data will provide the greatest increase in skill, which will potentially allow for generation of many secondary products derived from microwave data in the future.

1 Introduction

Background. To fully understand and simulate how the Earth’s climate is changing, climate scientists need to understand end-to-end how energy is transferred into, within, and out of the Earth’s climate system at all wavelengths and time scales [23, 15, 22]. While much progress has been made towards closing this complete energy transfer “budget” since the development of weather satellites [6, 25, 23, 22], one of the larger remaining sources of uncertainty is the net amount of latent heat exchange [22], which occurs when water either absorbs or releases energy to the surrounding environment as it changes phase between solid, liquid, and gas in various weather phenomena.

However, uncertainty about the total amount of global atmospheric water at any given time remains because no single observing instrument can fully observe the distributions of each water phase over a large area at high temporal resolution. For example, infrared (IR) wavelengths measured by the Advanced Baseline Imager (ABI; [20]) on the geostationary GOES-16 satellite provide continuous observations with high spatial resolution over a large domain (e.g. Fig. 1a). However, ABI IR observations in cloudy or moist regions are mostly representative of emission from cloud tops or the middle troposphere due to absorption of IR radiation by water vapor and scattering by in-cloud hydrometeors. Since most latent heat exchange in precipitating clouds occurs within and below clouds, additional information provided by the longer, less scattered microwave (MW) wavelengths measured by the Global Precipitation Mission Microwave Imager (GMI; [7]), are needed to penetrate liquid water and ice clouds and retrieve precipitation structure [18]. However, spatial resolutions for

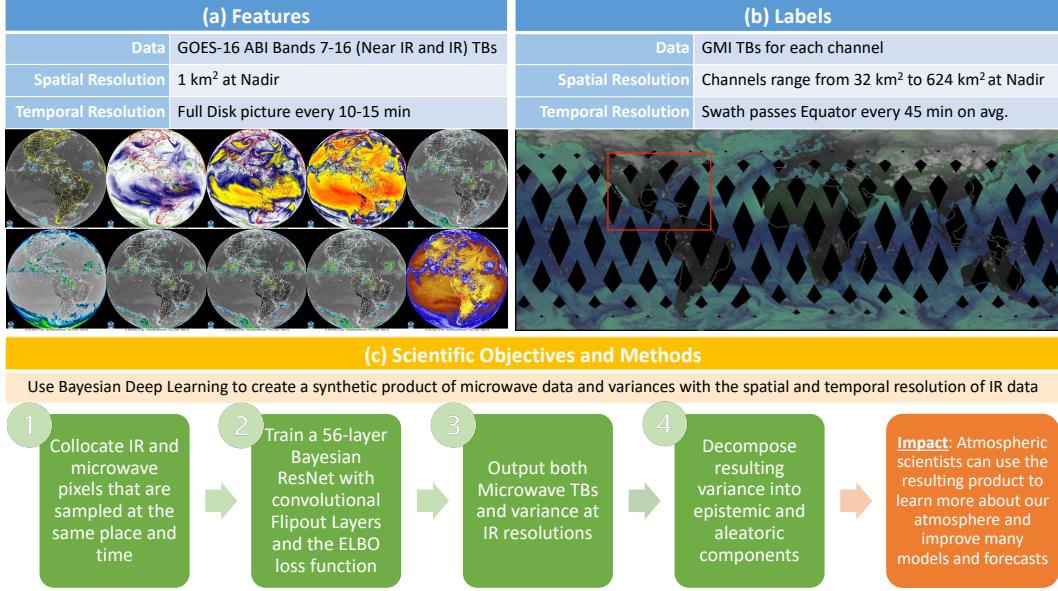


Figure 1: Executive summary of this study, where: (a) training dataset features are the brightness temperatures (TBs) measured by GOES-16 ABI Bands 7–16 (near IR and IR wavelengths) and are exemplified by full disk imagery from 17:40 UTC on 12 Sep. 2022 [2]; (b) training dataset labels are the microwave TBs from GMI, and are exemplified by the global, full-day GMI coverage on 01 Feb. 2020, where shading denotes data coverage [1, 16], black denotes data gaps, and the red box denotes the domain in Fig. 2; and (c) a flowchart outlines the methods and scientific impact.

MW data are much coarser, and there are very large gaps between the narrow data “swaths” due to the low-earth orbits of MW satellites (Fig. 1b). This means that a large portion of global atmospheric water is persistently unsampled, which negatively impacts a wide variety of climate and weather applications, including models, forecasts, and scientific understanding [18, 24, 22, 21, 19].

Scientific Goals and Impact. In this study, we seek to use Bayesian deep learning to create a new product of synthetic MW brightness temperatures (TBs) that combines both the high spatial and temporal resolutions of IR data and information of water distributions below cloud tops (Fig. 1c). Additionally, we seek to provide an estimate of the variance of predicted TB values at each pixel to help both scientists and downstream applications discern whether or not each predicted MW TB has skill. Furthermore, by decomposing variance into its aleatoric and epistemic components, we seek to help scientists discern whether predictive skill from our product can be improved with additional training data (e.g. high epistemic uncertainty) or if there is inherent atmospheric variability that warrants further scientific investigation into the utility of the model (e.g. high aleatoric uncertainty).

This new product of both synthetic MW TBs and estimated variance could then be utilized by atmospheric scientists to generate far-reaching, downstream impacts wherever existing MW data are currently being used in models, forecasts, and scientific analysis. With this new synthetic product, atmospheric scientists would be able to investigate individual weather processes using synthetic MW data that has a temporal resolution of 10–15 min over entire ocean basins for the first time. This would allow for near real-time analysis and characterization of high-impact, quickly evolving mesoscale phenomena. In comparison, current GMI MW data swaths typically only partially sample a tropical cyclone approximately 3–4 times per day. While optimizing the synthetic MW product’s skill will require additional training data, we demonstrate a pathway towards an ideal product.

2 Data and Methods

Data and Dataset Construction. To create our dataset, we first selected 10% of all GMI swaths (every tenth file from an S3 bucket) during the study period, which kept the training dataset manageable

GMI Frequency	Error Metrics			Uncertainty Metrics		
	RMSE	MAE	MAPE	MSD	MSRE	MSRA
183 \pm 3 GHz V	3.356 K	1.657 K	0.7%	2.307 K	0.352 K	2.276 K
166 GHz V	6.692 K	3.049 K	1.2%	3.857 K	0.625 K	3.795 K
37 GHz V	7.995 K	4.606 K	2.1%	4.758 K	1.013 K	4.638 K
23 GHz V	9.999 K	6.329 K	2.9%	6.940 K	1.323 K	6.802 K

Table 1: Results using models trained on 11,844,000 sample from Jan-Apr to generate synthetic GMI data using the May test dataset. RMSE = Root Mean Squared Error, MAE = Mean Absolute Error, MAPE = Mean Absolute Percentage Error, MSD = Mean Standard Deviation, MSRE = Mean Square Root Epistemic Uncertainty, MSRA = Mean Square Root Aleatoric Uncertainty, K = Kelvin.

given our computing resources. Next, we collocated ABI and GMI observations by labeling 39×39 pixel ABI patches from bands 7–16 (near-IR and IR) with the GMI T_b^{mw} corresponding to the center ABI pixel location. Any records containing an ABI patch with a pixel over land were discarded.

We trained our models using data collected during the first four months of 2020 (January through April). We used the 1st, 10th, and 20th days from January through April for the validation set; we used the remaining days as a training set. We held out the first 12 days of May as a test set.

Bayesian Deep Learning (BDL). In BDL, we assume distributions over weights, \mathbf{w} , of a neural network. If $p(\mathbf{w})$ is a prior distribution over \mathbf{w} and $p(\mathcal{D})$ is a distribution over the training data, then the goal of a Bayesian deep learning is to quantify a posterior distribution over the network weights conditioned on the distribution of the training data, $p(\mathbf{w} \mid \mathcal{D})$. For high dimensional problems, the calculation of $p(\mathbf{w} \mid \mathcal{D})$ has no closed formed solution making it intractable [4].

We use variational inference to approximate $p(\mathbf{w} \mid \mathcal{D})$ via the evidence lower bound (ELBO) [4] by implementing a 56-layer Bayesian Residual Network (ResNet) version 2 [11] model architecture, using convolutional Flipout layers [26]. This model architecture performs heteroscedastic regression, where the model outputs the parameters of a Gaussian distribution, μ and σ^2 , using a pre-built distribution layer from the Tensorflow Probability library version 0.12.1 [5].

Since our models have distributions over weights, we use Monte Carlo integration to conduct model inference [8, 9], sampling from the weight distribution T times and averaging over the output values. Following the approach described in [9], we chose $T = 100$ to calculate \hat{y}_n . As the output of our regression models represents the parameters of a Normal distribution, we calculate \hat{y}_n as a mixture of Normal distributions [14] with uniform mixture weights:

$$\hat{y}_n = \frac{1}{T} \sum_{t=1}^T \mathcal{N}(\mu_t(x_n, \mathbf{w}_t), \sigma_t^2(x_n, \mathbf{w}_t)) = (\hat{\mu}_n, \hat{\sigma}_n^2) \quad (1)$$

where $\frac{1}{T} \sum_{t=1}^T \mu_t(x_n, \mathbf{w}_t) = \hat{\mu}_n$ is the model’s predicted T_b^{mw} based on the T_b^{ir} input features. Following the law of total variance, we calculate the variance of the \hat{y}_n mixture distribution as:

$$\hat{\sigma}_n^2 = \frac{1}{T} \sum_{t=1}^T \underbrace{\hat{\sigma}_t^2(x_n, \mathbf{w}_t)}_{\text{Aleatoric Uncertainty}} + \underbrace{\hat{\mu}_t^2(x_n, \mathbf{w}_t) - \hat{\mu}_n^2}_{\text{Epistemic Uncertainty}} \quad (2)$$

where $\hat{\sigma}_t^2(x_n, \mathbf{w}_t)$ and $\hat{\mu}_t^2(x_n, \mathbf{w}_t)$ are the outputs of the last layer of the model (heteroscedastic regression).

Our BDL models quantify both aleatoric and epistemic uncertainty (see Eq. 2). Aleatoric uncertainty is irreducible [13], and epistemic uncertainty can be reduced given enough data [12]. Quantifying heteroscedastic aleatoric uncertainty using deterministic models was previously demonstrated to lead to better performance on geoscience regression tasks [3, 10]. The addition of epistemic uncertainty through BDL can help identify out-of-distribution samples (concept drift) [17] and data distributions that would be beneficial for additional training to improve a model’s skill on a given task.

3 Results and Discussion

Table 1 shows the mean error and uncertainty statistics for the vertically polarized GMI channels of 183 \pm 3 GHz, 166 GHz, 37 GHz, and 23 GHz. The less than 1% MAPE produced at 183 \pm 3 GHz is an

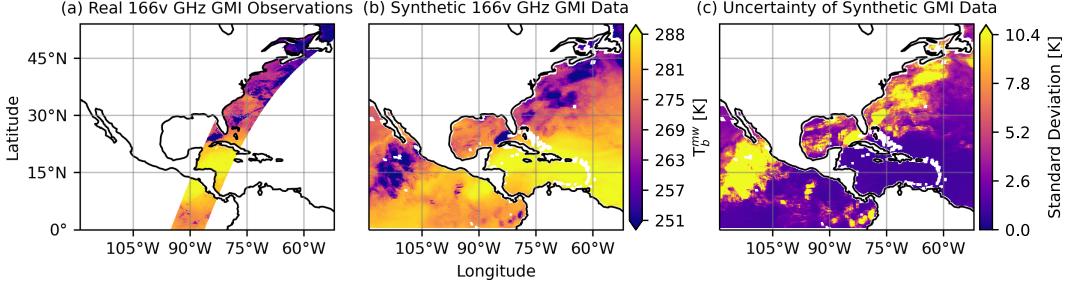


Figure 2: (a) ~ 15 minutes of GMI Observations (b) Synthetic T_b^{mw} generated from 15-minutes of ABI data (GPM orbit number 33679 at 1440 UTC 1 Feb 2020) corresponding to the box in Fig. 1. (c) Standard deviation of each T_b^{mw} in (b).

example of how well neural networks can learn any non-linear function with a large enough training dataset (~ 12 million examples), and is expected to have the highest skill since, like the ABI Band 8 where water vapor is present, this GMI channel also primarily senses the upper troposphere. The 1.2% MAPE at 166 GHz is also encouraging, since it primarily senses the mid- to upper-troposphere. The MAPE for 37 GHz and 23 GHz indicates that the model did not learn as well at these frequencies and may require more training data. This skill degradation is also consistent with remote sensing characteristics, since these channels “see” lower in the troposphere and are less similar to the IR features in cloudy regions. However, just having the T_b^{mw} output does not indicate whether influences from the surface or within clouds are the primary source of error. We can make this determination by reviewing the uncertainty metrics.

The epistemic (model parameter) uncertainty for the 183 ± 3 GHz model is approaching zero (Table 1), indicating that the training data is representative of the underlying data distribution. Conversely, the 37 GHz and 23 GHz models have epistemic uncertainty that is more than three times that of the 183 ± 3 GHz model. Therefore, the error metrics for these models are likely to decrease if they are trained on more data that is representative of the underlying distribution. Furthermore, examination of the spatial distribution of the predicted standard deviations in Fig. 2 shows that a greater source of uncertainty results from clouds, not the surface. Thus, our synthetic MW product directly shows that providing additional cloud training data would most benefit (of the ones shown here) the 23 GHz channel and least benefit the 183 ± 3 GHz channel.

4 Conclusion

Without the context of the decomposed uncertainty metrics, the error metrics for 37 GHz and 23 GHz could be interpreted to mean that our network architecture is not suitable for this regression task. However, the MSRE (model parameter uncertainty) indicates that the error metrics for 37 GHz and 23 GHz could be reduced by training on additional data, and the spatial distribution of uncertainty shows that specifically cloud data would provide the most benefit. Thus, we find BDL a promising technique for creating more synthetic MW data from 15 minutes of ABI IR data (e.g. Fig. 2b) than is available from 24 hours of GMI MW data collection (e.g. Fig. 1b).

As we continue to utilize our uncertainty decomposition methodology to improve the predictive skill of MW TBs from IR TBs, we hope to provide a highly useful synthetic product for atmospheric scientists. The extreme boost in both spatial and temporal resolution provided by our synthetic product allows atmospheric scientists to perform previously impossible analyses in the MW spectrum, such as how individual oceanic storms contribute towards the energy budget. Furthermore, the large, continuous spatial coverage would allow scientists to investigate interactions across many time and space scales, which could lead to increased scientific understanding of weather and climate interactions and more accurate model physics. Additionally, MW data and uncertainties can be assimilated into models to improve simulated structures, which could lead to more accurate forecasts. Furthermore, since MW imagery is also used in real-time forecasts of hurricane intensity and other extreme weather phenomena, forecast lead times for high-impact events could also be increased, which helps mitigate loss of life and property. Finally, the uncertainty metrics provided by our product allows atmospheric scientists to assess which predictions are most trustworthy, which can further improve interpretation of the product and facilitate widespread adoption.

Acknowledgments and Disclosure of Funding

This work has been supported by: Office of Naval Research grants N0001421WX00575 and N0001422WX01251.

References

- [1] Eosdis worldview. https://worldview.earthdata.nasa.gov/?v=-245.09403723915682,-135.87380640621802,238.6758339845142,113.3700624319035&1=GMI_Brightness_Temp_Dsc,GMI_Brightness_Temp_Asc,Coastlines_15m,&lg=true&t=2020-02-01-T21%3A22%3A50Z. Accessed: 2022-09-12.
- [2] Goes image viewer. <https://www.star.nesdis.noaa.gov/GOES/fulldisk.php?sat=G17>. Accessed: 2022-09-12.
- [3] Elizabeth A. Barnes, Randal J. Barnes, and Nicolas Gordillo. Adding Uncertainty to Neural Network Regression Tasks in the Geosciences. *arXiv:2109.07250*, 2021.
- [4] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518):859–877, April 2017.
- [5] Joshua V. Dillon, Ian Langmore, Dustin Tran, Eugene Brevdo, Srinivas Vasudevan, Dave Moore, Brian Patton, Alex Alemi, Matt Hoffman, and Rif A. Saurous. TensorFlow Distributions. *arXiv:1711.10604*, November 2017.
- [6] W. H. Dines. The heat balance of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 43(182):151–158, 2007.
- [7] David W. Draper, David A. Newell, Frank J. Wentz, Sergey Krimchansky, and Gail M. Skofronick-Jackson. The Global Precipitation Measurement (GPM) Microwave Imager (GMI): Instrument Overview and Early On-Orbit Performance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(7):3452–3462, July 2015.
- [8] Runhai Feng, Niels Balling, Dario Grana, Jesper Sören Dramsch, and Thomas Mejer Hansen. Bayesian Convolutional Neural Networks for Seismic Facies Classification. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–8, 2021.
- [9] Angelos Filos, Sebastian Farquhar, Aidan N. Gomez, Tim G. J. Rudner, Zachary Kenton, Lewis Smith, Milad Alizadeh, Arnoud de Kroon, and Yarin Gal. A Systematic Comparison of Bayesian Deep Learning Robustness in Diabetic Retinopathy Tasks. *arXiv:1912.10481*, December 2019.
- [10] Emily M Gordon and Elizabeth A. Barnes. Incorporating Uncertainty into a Regression Neural Network Enables Identification of Decadal State-Dependent Predictability. Preprint, Climatology (Global Change), March 2022.
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity Mappings in Deep Residual Networks. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science, pages 630–645, Cham, 2016. Springer International Publishing.
- [12] Alex Kendall and Yarin Gal. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Neural Information Processing Systems*, volume 30. Curran Associates, Inc., October 2017.
- [13] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or Epistemic? Does it Matter? *Structural Safety*, 31(2):105–112, March 2009.
- [14] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[15] Norman G Loeb, Bruce A Wielicki, David R Doelling, G Louis Smith, Dennis F Keyes, Seiji Kato, Natividad Manalo-Smith, and Takmeng Wong. Toward optimal closure of the earth's top-of-atmosphere radiation budget. *Journal of Climate*, 22(3):748–766, 2009.

[16] Andrew J. Negri, Robert F. Adler, and Christian D. Kummerow. False-color display of special sensor microwave/imager (ssm/i) data. *Bulletin of the American Meteorological Society*, 70(2):146–151, 1989.

[17] Pedro Ortiz, Marko Orescanin, Veljko Petkovic, Scott W. Powell, and Benjamin Marsh. Decomposing Satellite-Based Classification Uncertainties in Large Earth Science Datasets. *IEEE Transactions on Geoscience and Remote Sensing*, 60(1):1–11, 2022.

[18] Veljko Petković and Christian D Kummerow. Understanding the sources of satellite passive microwave rainfall retrieval systematic errors over land. *Journal of Applied Meteorology and Climatology*, 56(3):597–614, 2017.

[19] Zhaoxia Pu, Chaulam Yu, Vijay Tallapragada, Jianjun Jin, and Will McCarty. The Impact of Assimilation of GPM Microwave Imager Clear-Sky Radiance on Numerical Simulations of Hurricanes Joaquin (2015) and Matthew (2016) with the HWRF Model. *Monthly Weather Review*, 147(1):175–198, January 2019.

[20] Timothy J. Schmit, Paul Griffith, Mathew M. Gunshor, Jaime M. Daniels, Steven J. Goodman, and William J. Lebar. A Closer Look at the ABI on the GOES-R Series. *Bulletin of the American Meteorological Society*, 98(4):681–698, 2017.

[21] Graeme L Stephens and Yongxiang Hu. Are climate-related changes to the character of global-mean precipitation predictable? *Environmental Research Letters*, 5(2):025209, 2010.

[22] Graeme L. Stephens, Juilin Li, Martin Wild, Carol Anne Clayson, Norman Loeb, Seiji Kato, Tristan L'Ecuyer, Paul W. Stackhouse, Matthew Lebsack, and Timothy Andrews. An update on Earth's energy balance in light of the latest global observations. *Nature Geoscience*, 5(10):691–696, 2012.

[23] Kevin E Trenberth, John T Fasullo, and Jeffrey Kiehl. Earth's global energy budget. *Bulletin of the American Meteorological Society*, 90(3):311–324, 2009.

[24] Kevin E. Trenberth, Lesley Smith, Taotao Qian, Aiguo Dai, and John Fasullo. Estimates of the Global Water Budget and Its Annual Cycle Using Observational and Model Data. *Journal of Hydrometeorology*, 8(4):758–769, 2007.

[25] Thomas H. Vonder Haar and Verner E. Suomi. Measurements of the Earth's Radiation Budget from Satellites During a Five-Year Period. Part I: Extended Time and Space Means. *Journal of the Atmospheric Sciences*, 28(3):305–314, April 1971.

[26] Yeming Wen, Paul Vicol, Jimmy Ba, Dustin Tran, and Roger Grosse. Flipout: Efficient Pseudo-Independent Weight Perturbations on Mini-Batches. In *International Conference on Learning Representations*, April 2018.