MAGMA Gaussian Processes to complete Venusian atmospheric profiles from SOIR

Simon Lejoly[1], Arianna Piccialli[2], Arnaud Mahieux[2,3], Ann Carine Vandaele[2], and Benoît Frénay[1]

[1]University of Namur, NaDI, Faculty of Computer Science, Belgium (simon.lejoly@unamur.be)

[2]Royal Belgian Institute for Space Aeronomy - Planetary Atmosphere Department, Belgium

[3]The University of Texas at Austin - Department of Aerospace Engineering and Engineering Mechanics, USA

1. Introduction

When working with atmospheric data, researchers often face partial datasets. This is the case of the SOIR instrument on board Venus Express [1], which measured the temperature of the Venus mesosphere, inferred via solar occultations. Due to inherent limitations in the measurement process, temperature measurements are missing at varying altitudes in the profiles. While altitudes of measurements across the whole dataset can span from 60 km to 160 km, most of the profiles only contain a range of 10 to 50 consecutive measurements. Figure 1 shows examples of such profiles. To fill this data gap, machine learning, and specifically probabilistic models are promising candidates.
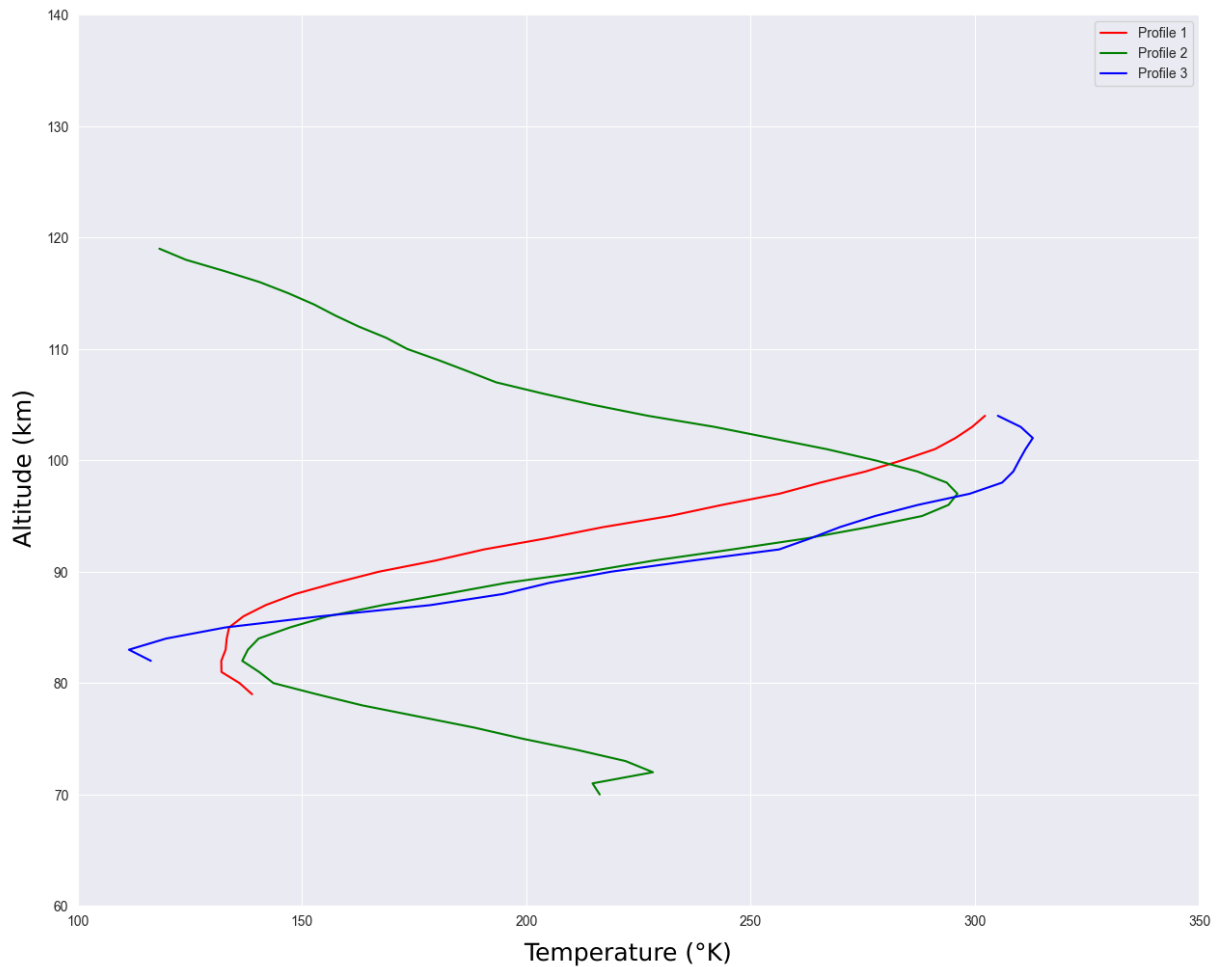
Fig. 1: Three temperature profiles from SOIR, with varying ranges of altitude.

2. Gaussian Processes and MAGMA

Gaussian processes (GPs) are probabilistic models mainly used for predictions based on empirical observations. The capacity of GPs to estimate the uncertainty in their predictions makes them particularly appropriate for extrapolating atmospheric data. Since traditional GPs only try to fit a single function, adaptations should be implemented for the case of the SOIR dataset, which contains numerous profiles. Using one GP for each profile would not allow models to leverage information across the whole dataset to make better predictions. This motivates the use of so-called multi-task GPs, where "task" refers to an atmospheric profile in our context. The literature in the field of multi-task GPs is extensive [2, 3, 4], but most models either struggle with large datasets or have difficulties learning the covariance between the tasks. The MAGMA algorithm [5] is a recent advancement in the field of multi-task GPs that solves the above-mentioned issues. To complete a gap at a specific altitude in a profile, MAGMA uses both the profile values that are close to the gap and the mean value measured at that altitude in all other profiles. This results in better confidence intervals, even far from known observations. Previous to this work, MAGMA had never been applied to atmospheric datasets.

3. Profile Extrapolation

To assess the performances of MAGMA, we compare this algorithm with a traditional GP working on each profile individually. First, each profile is preprocessed to standardise each temperature value and use a logarithmic pressure scale as a height indicator rather than altitude. As MAGMA needs profiles to share common inputs, each pressure measurement is mapped to its closest bin on a 250 discrete bin scale. We then divide the dataset into train and test sets. Each test profile is further divided into test observations and validation observations. During experiments, test observations are given to the models as a starting point. The models can then make predictions, providing an estimated mean value and a confidence interval for each missing pressure bin. The validation observations are compared with their corresponding predictions to evaluate the performances of each model. Experimental results show that MAGMA has significantly better performances than a traditional GP, as seen in Figure 2. It provides estimations that are closer to the actual measurement and confidence intervals that are more precise, as shown in Figure 3.
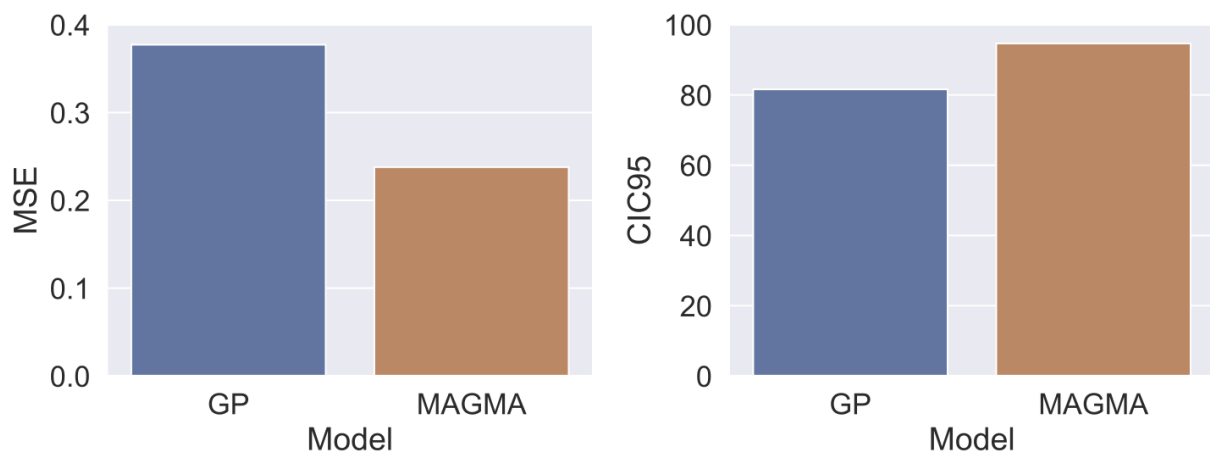

Fig. 2: Average performances of MAGMA and the baseline model on two distinct metrics. Mean Squared Error (MSE) measures the distance between predictions and true values (to be minimised). CIC95 is the ratio of validation observations actually sitting in the predicted 95% confidence interval (should be close to 95%).
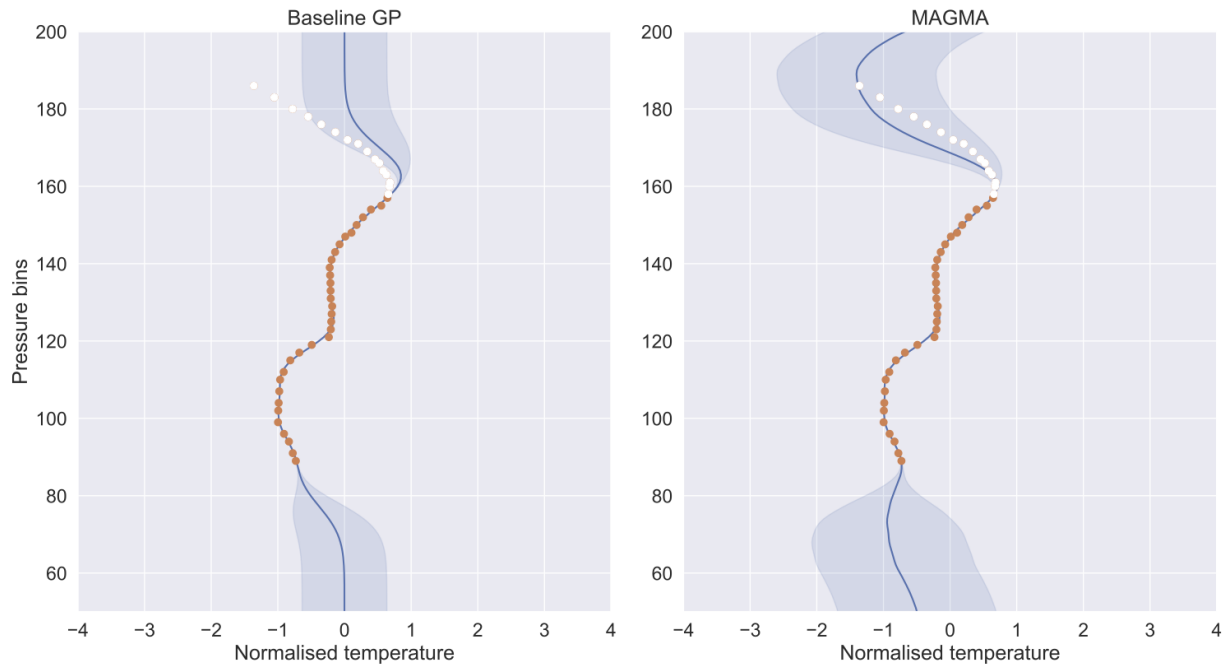
Fig. 3: Predictions from the baseline and MAGMA, for different training settings. The test observations and validation observations are represented in orange and white, respectively. The grey-shaded area corresponds to the 95% confidence interval.

4. Conclusion

We apply MAGMA, a novel probabilistic learning algorithm, to the SOIR dataset, enhancing predictions for missing observations. This contribution is a first step toward a practical application of GPs to planetary aeronomy datasets. Future research will explore possible enhancements to the data preprocessing and model architectures to complete the SOIR dataset with more precise and credible estimations.

References

- [1] L. Trompet, Y. Geunes, T. Ooms, et al. Description, accessibility and usage of SOIR/Venus Express atmospheric profiles of Venus distributed in VESPA (Virtual European Solar and Planetary Access). Planetary and Space Science, 150:60–64, 2018.

- [2] Edwin V Bonilla, Kian Chai, and Christopher Williams. Multi-task Gaussian Process Prediction. In Advances in Neural Information Processing Systems, volume 20, 2007.

- [3] Edwin V. Bonilla, Felix V. Agakov, and Christopher K. I. Williams. Kernel Multi-task Learning

using Task-specific Features. In Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics, pages 43–50. PMLR, 2007.

- [4] Carlos Ruiz, Carlos M. Alaiz, and José R. Dorronsoro. A survey on kernel-based multi-task learning. Neurocomputing, 577:127255, 2024.

- [5] Arthur Leroy, Pierre Latouche, Benjamin Guedj, and Servane Gey. MAGMA: inference and prediction using multi-task Gaussian processes with common mean. Machine Learning, 111(5):1821–1849, 2022.