# Light Curve Forecasting and Anomaly Detection Using Scalable, Anisotropic, and Heteroscedastic Gaussian Process Models

**Imène R. Goumiri**
*Physics Division, Lawrence Livermore National Laboratory*

**Amanda L. Muyskens**
*Computational Engineering Division, Lawrence Livermore National Laboratory*

**Benjamin W. Priest**
*Center for Applied Scientific Computing, Lawrence Livermore National Laboratory*

**Robert E. Armstrong**
*Physics Division, Lawrence Livermore National Laboratory*

LLNL-CONF-853760

## ABSTRACT

Light curves trace apparent visual light magnitudes of space objects over a given time horizon. One can use light curves to formulate latent variable inference problems that solve Space Domain Awareness (SDA) tasks such as satellite identification, pose estimation, or maneuver detection. Ground-based light curve observations from commercial off the shelf (COTS) cameras remain inexpensive compared to higher precision instruments; however, issues with COTS cameras such as limited sensor availability combined with noisier observations can produce sparse, corrupted time-series data. These external factors hinder the automation of light curve analysis, meaning that improvements to light curve anomaly detection and forecasting pipelines are crucial for SDA applications.

This paper investigates Gaussian Processes (GPs) as a tool for light curve forecasting and anomaly detection with SDA applications. GPs are priors over function spaces and universal function approximators that compute Gaussian distributed estimates of response variables, providing both a mean estimate and "error bars" around the mean. This native uncertainty quantification is one of their most important features, which distinguishes them from most other popular machine learning models. Uncertainty quantification in GP models arises directly from computable posterior variances, allowing for rigorous assessment of uncertainty in light curve forecasting settings. Moreover, posterior means and uncertainties computed from GP models can be used in anomaly detection: a real-time measurement with abnormally high deviation from the predicted mean based upon prior measurements could be the result of corruption, interference, sensor failure, or an unexpected rotation or orbit change. Automating the detection of these aberrancies is therefore of high interest to the SDA community.

This work employs MuyGPs, a scalable Gaussian process hyperparameter training and regression framework that can train large-scale light curve models in seconds on a standard workstation. We show that the native uncertainty quantification provided by MuyGPs is an invaluable tool for SDA practitioners through numerical experiments using real-world data.

Gaussian process models provide posterior means and variances via kernel functions, which are bivariate positive semi-definite functions that model the covariance between data points. The kernels and covariance structure of our model incorporate three primary features: (1) anisotropic kernels to account for different correlative length scales across different features in the training data (2) non-temporal features, e.g., solar phase angles, to yield more accurate apparent magnitude predictions by exploiting correlations between light curve magnitudes and other measurements (3) heteroscedastic noise parameters to account for the temporal variation of the response noise.

To demonstrate the efficacy of our approach, we devise a classification task where previously unseen measurements are matched against a database of past observations of several light curves to determine satellite identity. As the past

observations are sparse and irregularly timed, we use MuyGPs to interpolate our training data to a complete grid in time so that subsequence observations (a contiguous stream of observations over the course of one night) are directly comparable in a metric space. We also compare the performance of anisotropic and heteroscedastic MuyGPs models on light curve forecasting and near-real-time anomaly detection. We examine the uncertainty quantification provided by each model, with special emphasis on anomaly and maneuver detection, providing insight into its utility for machine learning practitioners in the SDA community.

## 1. INTRODUCTION

Light curves of Resident Space Objects (RSOs) are time series of observed brightness that provide a wealth of information. The brightness of an RSO depends on several components including: size, shape, reflective material, distance from observer, and solar phase angle. A light curve is also sensitive to dynamic quantities like spin-rate, tumbling motion, or maneuvers. A sudden change in the brightness of a satellite could indicate that it has been damaged or has had its performance degraded. Several studies [11, 5, 1, 10, 8, 4] have successfully demonstrated the use of light curves to infer RSO properties.

The amount of RSO data will grow dramatically in the near future due to the increase in satellite launches and the proliferation of low-cost cameras that can easily collect large samples of data. The identification and characterization of RSOs for Space Domain Awareness (SDA) will be increasingly important for space traffic management to monitor health and avoid future collisions. These tasks will require new methods that effectively handle large numbers of objects and give reliable predictions with low computational expense.

The classification of newly discovered RSO observations and the detection of maneuvers of existing objects are critical tasks that will grow more challenging as RSO populations increase. Accurate classification of new unseen data often requires interpolation of previous, potentially sparse, data into a common reference frame and an understanding of the uncertainties associated with predictions. While neural network-based models are currently popular, in general, they do not provide prediction uncertainties directly. Attempts to compute uncertainties resort to approximate Bayesian methods such as bootstrapping [2, 3] or dropout [6, 9].

Here we present an approximate Gaussian Process based method for SDA object classification. This expands on our previous result [7] where we showed that Gaussian Processes can effectively be used to interpolate and forecast RSO light curves. In this paper, we explore the implementation of two new features: heteroscedastic noise and anisotropic kernels and test them in the scenario of classifying light curves of various durations. The purpose of the preliminary work presented here is to establish an analysis pipeline we believe will be important to future light curve data.

Traditional implementation of Gaussian Process Models employ stationary and isotropic kernel matrices. However, those assumptions are very rarely met in real world datasets. In [7], we modeled time along two dimensions: time-of-night and day-of-observation-interval instead of treating the data as a raw time series in order to capture periodicity inherent in the data. However, the isotropic kernel assumption enforces that the change in correlation across the time-of-night and day-of-observation-interval dimensions are the same. There is no justification for this assumption to be utilized other than convenience, and it results in predictions that greatly depend upon how the dimensions are normalized. In this manuscript we explore loosening this assumption using an anisotropic covariance function that allows for training different length scales scales in each of these dimensions.

It is common to assume a homoscedastic noise model, wherein each observation is assumed to be perturbed by i.i.d. Gaussian noise. However, [7] demonstrated that the homoscedastic assumption results in poorly calibrated uncertainties, because in fact higher magnitude observations (which tend to occur near dusk or dawn) have much larger noise variance than the low-magnitude observations taken during the darkest parts of the night. Enforcing homoscedasticity prevents the model from modeling non-stationarity in the noise prior, and contributes to posterior variances that are overconfident in the high-magnitude, noisy sections of the curve and underconfident in the low-magnidue, more stable sections of the curve. This manuscripts uses a heteroscedastic noise model, wherein each observation as a separate noise prior that is derived from on-sensor measurements taken at the time of observation.

This manuscript outlines a GP model workflow that is useful for future SDA problems based upon light curve observations. section 2 describes the data and GP methods, including descriptions of the heteroscedastic and anisotropic model features and the reference frame interpolation that we use for preparing data for classification. section 3 details how these methods work in practice, both on forecasting and anomaly detection tasks as well as on a classification

task that requires interpolated data. Finally, section 4 explains the implications of our results and future directions for this work.

## 2. METHODOLOGY

### 2.1 Light Curve Data

A ground-based commercial off the shelf (COTS) camera pointed at geostationary orbits can accumulate a large amount of RSO data for relatively little cost. However, there are limitations working from the ground such as changing lighting conditions, variable sensor availability, and poor or missing observations resulting from bad weather. Light curves from these objects will routinely include regions of noisy data and large gaps that must be addressed if the data is to be used for training machine learning models.

We use a dataset provided by Dave Monet [12] that consists of 43 known satellites measured from a single camera in Flagstaff, AZ between September 2014 and September 2018. This is the same dataset used in our previous paper [7]. Each measurement consists of a measured time, position on the sky, brightness and error in a filter close to Johnson V band. There are approximately 500,000 data points per object resulting in a dataset of $\sim$6.5 million data points. We limit our analysis to 13 satellites labeled as "Nominal" in the dataset (Fig. 1). These were selected because they have data over the full time period and appear to be active satellites.

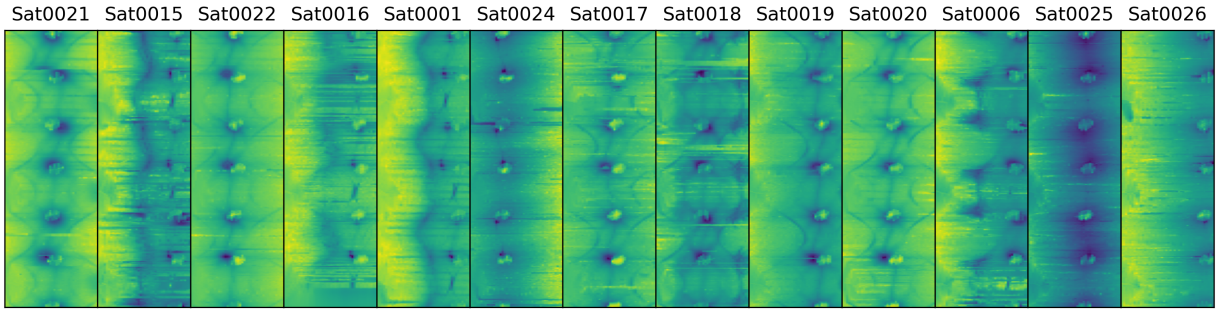


Fig. 1: All nominal light curves in our dataset, interpolated (see subsection 2.3) to denoise and complete missing data. The magnitude is plotted as a function of days (rows) and times of day (columns) for a total duration of 4 years.

### 2.2 Gaussian Processes

Gaussian processes (GPs) are a powerful tool for modeling and predicting noisy data that can be used to learn non-linear relationships between features and responses. These responses can be a continuous variable, such as the brightness of the light curve at a particular time, or discrete labels such as the identification of light curve observations with individual RSOs or classes of object or orbital regimes. We use GPs for both regression and classification tasks in this paper.

A GP parameterizes a collection of random variables such that any subset of these variables is jointly multivariate Gaussian with a known mean function $m_\theta(\cdot)$ and covariance function $k_\theta(\cdot,\cdot)$, where $\theta$ is a set of hyperparameters that control the behavior of the GP. $m_\theta(\cdot)$ is usually taken to be 0 without a loss of generality. As such, the joint probability distribution of any set of GP outputs can be completely specified by its mean vector and covariance matrix, and we can estimate the conditional (posterior) distribution of predictions by application of Bayes' rule.

GPs naturally incorporate uncertainty into their predictions, and model decisions including the form of $k_\theta(\cdot,\cdot)$ have a natural statistical interpretation. These features are advantagous compared to other popular machine learning techniques such as Deep Neural Networks (DNNs), although uncertainty quantification and intepretation of DNNs is an active area of research. The covariance matrix of a GP specifies the degree of correlation between different outputs. Being able to quantify the uncertainty of predictions without training specifically for it is important for many applications.

The authors initially explored the GP modeling of RSO light curves in [7]. That work used isotropic and homoscedastic assumptions in its model, which is to say that covariances were taken to be a function of euclidean distance in the input

space and the noise of each measurement was assumed to be i.i.d. Gaussian. However, the resulting posterior variances were misspecified, as the uncertainties around the relatively stable measurements taken in the middle of the night were the same width as the much noisier measurements taken near dusk or dawn. This manuscript attempts to correct these model discrepancies to more accurately measure the uncertainties associated with RSO light curve prediction.

The model in [7] assumed measurement noise for observations was drawn i.i.d. from $\mathcal{N}(0,\varepsilon)$. That is, given time dimensions $X = (\mathbf{x}_1, \ldots \mathbf{x}_n)^\top$ we model the response $Y(X)$ as

$$Y(X) = (Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_n))^\top \sim \mathcal{N}\left(\widetilde{0}, \sigma^2 \left(K_\theta(X,X) + \varepsilon I_n\right)\right), \tag{1}$$

where $\mathcal{N}$ is the multivariate Gaussian distribution, $\widetilde{0}$ is the $n$-dimensional zero vector, $\sigma^2$ is a variance scaling term, $I_n$ is the $n \times n$ identity matrix, and $K_\theta(X,X)$ is an $n \times n$ positive definite, symmetric covariance matrix between the elements of $X$ that is controlled non-linearly through kernel function $k_\theta(\cdot,\cdot)$ with hyperparameters $\theta$.

As noted in [7], using a homoscedastic noise model implies that the noise prior is constant over all magnitudes. However in practice the data at high magnitudes tend to have more variance than the data collected at lower magnitudes. The model considered in this manuscript uses a vector of noise prior variances $\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \boldsymbol{\varepsilon}_i)$ for each observation $i$. We take these variance priors from measurements collected from the sensor at the same time as the corresponding magnitude. This changes the prior in (1) to

$$Y(X) = (Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_n))^\top \sim \mathcal{N}\left(\widetilde{0}, \sigma^2 \left(K_\theta(X,X) + \mathrm{diag}(\boldsymbol{\varepsilon})\right)\right), \tag{2}$$

where $\mathrm{diag}(\boldsymbol{\varepsilon})$ is the diagonal matrix with $\boldsymbol{\varepsilon}$ along the diagonal.

We further generalize our prior model by utilizing an anisotropic distance model. Our prior model used the isotropy assumption, that $k_\theta(\mathbf{x},\mathbf{z}) = p_\theta\left(\frac{\|\mathbf{x}-\mathbf{z}\|_2^2}{\ell^2}\right)$ for some function $p_\theta(\cdot)$ and a single length scale parameter $\ell$. However, our embedding of time into multiple dimensions imposes an arbitrary relationship between displacement in time-of-day and displacement in day-of-observation-interval that is sensitive to the units of the observations and how the data is normalized. While our initial exploration disregarded this detail, in this manuscript we explicitly model this relationship by instead considering functions of the form $k_\theta(\mathbf{x},\mathbf{z}) = p_\theta\left(\sum_{j=1}^d \frac{(\mathbf{x}_j-\mathbf{z}_j)^2}{\ell^2})\right)$, for the same function $p_\theta(\cdot)$ and separate length scale parameters $\ell_j$. Explicitly modeling length scales along each feature dimension makes us less sensitive to data normalization choices and makes the covariance scales along each dimension clear.

### 2.3 Interpolation with Gaussian Processes

In part because light curves are often noisy and intermittent, and in part because many applications require complete and smooth data, interpolation is key to being able to de-noise and further process light curve data. In [7], we demonstrated that Gaussian processes were very performant at interpolating light curves provided that the data was properly embedded in a low dimensional space that reflected the periodicity of the data. In this paper we use the 2D embedding where each data point is replaced by a 2D vector with the first dimension representing the time of day as a real number in the $[0,1]$ interval, and another dimension representing the day of the observation period (4 years) as an integer, starting at zero on the first day of observation, and further normalized to the interval $[0,1]$. We also subtract the constant mean from the responses and add it back to the predictions, as is customary.

We use MuyGPs [13] with a Matérn kernel and a fixed smoothness hyperparameter of $\nu = 0.1$ to implement our new models. This smoothness hyperparameter was selected to ensure that predicted curves have the desired degree of smoothness and to fixed to limit the cost of optimizing other hyperparameters. The MuyGPs algorithm depends upon identifying nearest neighbors of feature points, both for optimization where the queried points are sampled batch elements and for prediction where the queried points are unobserved test elements. Hence, we need to ensure that our two dimensions are properly scaled such that the nearest neighbors correspond with the points that are most correlated with the query point. We normalized our data so that the typical nearest point in one dimension is equidistant to the typical nearest point in the other dimension. So in addition to the initial normalization of each feature between 0 and 1, we compute the medians of the inter-feature distances in each dimension and renormalize the first dimension so that they match. This ensures that one dimension does not overwhelm the neighborhoods while indexing nearest neighbors before training. The result is illustrated in Fig. 2. Another possible approach, which we will explore in future work, is to recompute the nearest neighborhoods during training based upon the distance distortions introduced by the length

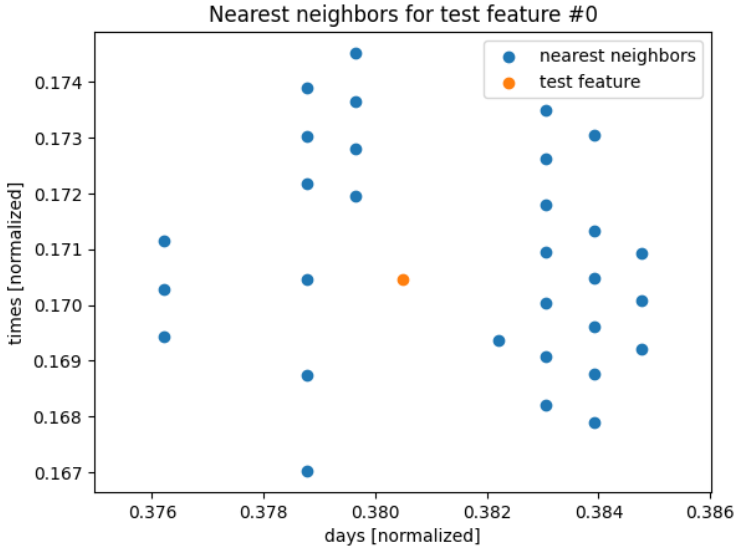


Fig. 2: Nearest neighbors for a sample test feature when interpolating an entire day of missing data, which is why there are no nearest neighbors on the same day as the test feature. The normalization is calibrated such as nearest neighbors are approximately equidistant whether they are from a surrounding day or time of day. (Equal axis plot.)

scale parameters $\ell_j$. This will result in a more accurate and evolving encapsulation of the effect of the length scales on correlation at the expense of greater computational overhead per evaluation of the objective function.

We use four different models for interpolating light curves and compare the results. For the isotropic models, after optimizing the length scale hyperparameter $\ell$ in the interval $[10^{-3}, 10^2]$, we use the optimal value $\ell = 0.0025$. For the anisotropic models, the two length scale hyperparameters $\ell_1$ and $\ell_2$ are optimized in the interval $[10^{-3}, 10^2]$ and we use the optimal values $\ell_1 = 50$ and $\ell_2 = 100$. We use the median of the measured magnitude error as the homoscedastic noise $\varepsilon$ and a polynomial fit of order 2 of that same measured magnitude error to smooth the heteroscedastic noise $\boldsymbol{\varepsilon}$.

### 2.4 Classification with Gaussian Processes

To demonstrate an example of a task that requires properly interpolated data rather than raw time series of magnitudes, we use GPs as a tool for identifying unseen light curves based on a database of know ones. This classification task maps vectors of magnitudes to light curves labels hence requires that the magnitudes for all different light curves be interpolated to a common set of time points. We choose to interpolate the light curves using the methods described in subsection 2.3 on a regularly-spaced grid of time points (10 minutes apart) spanning the longest common interval of time to all light curves in the dataset (about 4 years). That common interval guarantees that we are strictly interpolating and not extrapolating, since extrapolation is susceptible to larger errors, and the use of a regularly-spaced grid for times makes plotting easier (see Fig. 1). Next we split the interpolated data into 1-day chunks and each chunk of interpolated magnitudes forms a feature vector. Each feature vector is then associated with a one-hot encoded label with zero mean, meaning light curve $i$ has a vector label that is all $-1/N$ except for component $i$ which is $\frac{N-1}{N}$ where $N = 13$ is the number of light curves in the dataset.

To make the task realistic, we reserve the first three years for training the classifier, an isotropic GP with homoscedastic noise similar to the one described in subsection 2.3, and use the fourth year as the test data. We make sure to interpolate the training and test data independently. Predictions are then compared to actual measurements to determine the accuracy. This is a realistic SDA scenario since it forecasts in the future, and can be readily applied given historical data.

## 3. RESULTS

As a way to compare the different models described in subsection 2.3, we replicate the analysis done in [7] where we interpolate a day of artificially missing data in a given light curve and compare the results to the actual measurements.

Fig. 3 illustrates the results for day 445 of light curve "Sat0019". For all variants of the model, we are able to predict the entire day of missing data with great accuracy while removing most of the noise. The de-noising effect is most apparent near dusk and dawn where both the magnitude and the measurement noise are highest. The root mean square error (RMSE) of the predictions compared to the measurements ranges from 0.161 (units of magnitude) for the isotropic homoscedastic model to 0.158 for the anisotropic heteroscedastic model.

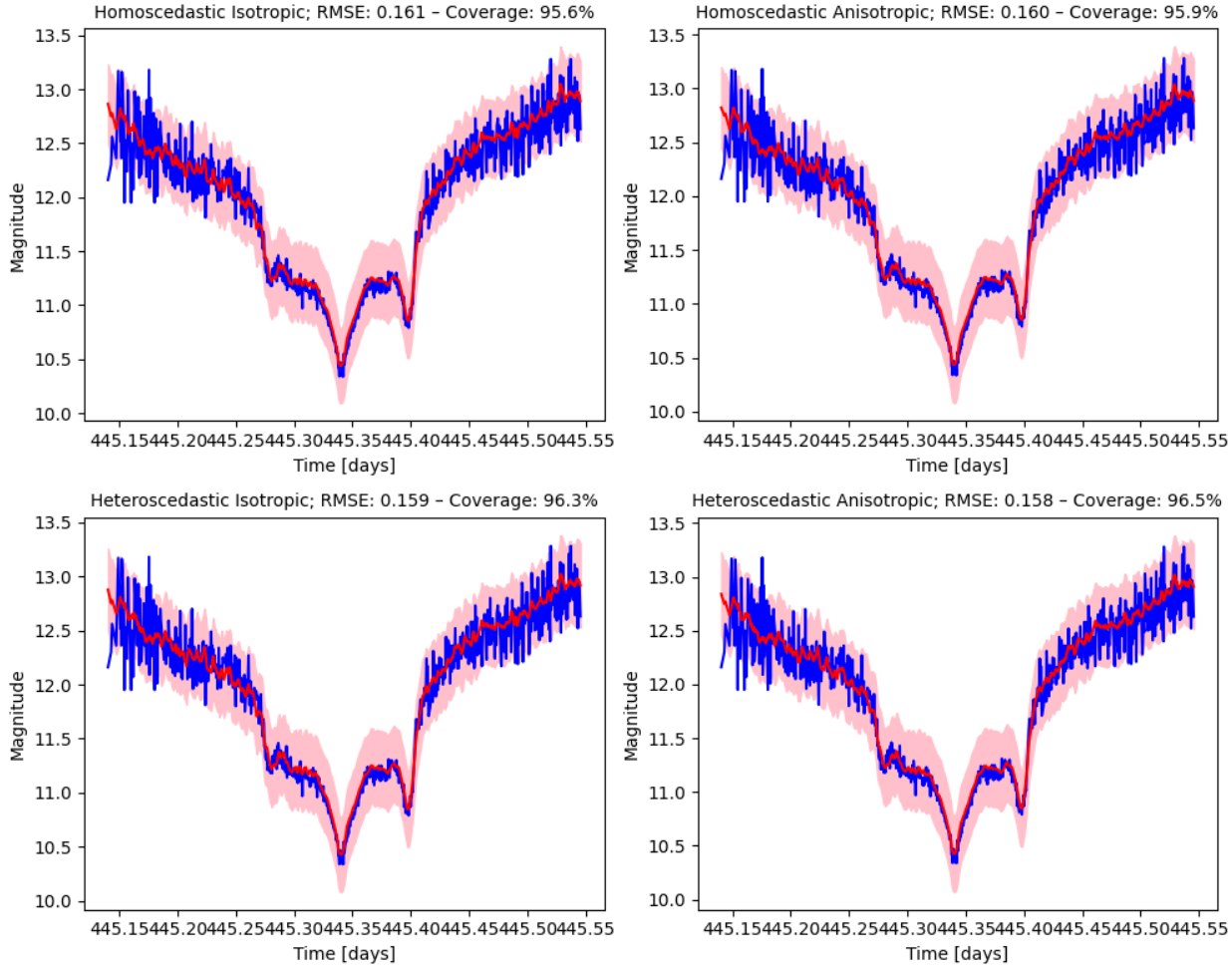


Fig. 3: GP prediction (red) compared to measurement (blue) for one day of missing data (day 445 of Sat0019). The pink shaded region is the 95% confidence interval of the prediction. In terms of RMSE, for the selected hyperparameters ($\nu = 0.1$ for all, $\ell = 100$ for isotropic, $\ell_1 = 50$, $\ell_2 = 100$ for anisotropic), anisotropic models (right) perform slightly better than isotropic models (left), and heteroscedastic models (bottom) perform slightly better than homoscedastic models (top), but the differences are marginal. The coverage is close to the target of 95%.

GPs also output the variance of their predicted measurements and we use it to plot the 95% confidence interval, defined as $\mu \pm 1.96 \cdot \sigma$ where $\mu$ is the predicted mean and $\sigma^2$ is the variance. The coverage, which is the proportion of points falling inside the 95% confidence interval, is very close to 95%. For this set of hyperparameters ($\nu = 0.1$, $\ell = 100$, $\ell_1 = 50$, $\ell_2 = 100$) the coverage is slightly higher, meaning the model is less confident about its prediction, for the anisotropic and heteroscedastic models than for isotropic and homoscedastic ones. This is better visible on Fig. 4.

Fig. 4 also shows the effect of taking into account the heteroscedasticity of the noise, a positive correlation between magnitude and variance, that is apparent in the raw data. Both homoscedastic models have a fairly flat variance, except for the hours closest to dawn and dusk where the variance spike up, meaning the models are equally confident for most predictions. By contrast, both heteroscedastic models have a variance that gradually decreases until the middle of the night where they start gradually increasing again, in an almost parabolic way. This shape matches the both the

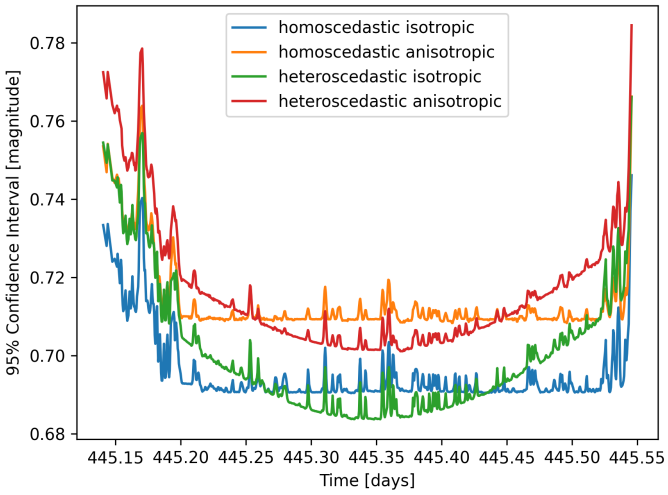


Fig. 4: Comparison of the half-width of the 95% confidence interval (i.e. $1.96 \cdot \sigma$) of the prediction for the four model variants. The heteroscedastic models have a reduced uncertainty near the middle of the night where magnitudes are lower. For those hyperparameters ($\nu = 0.1$ for all, $\ell = 100$ for isotropic, $\ell_1 = 50$, $\ell_2 = 100$ for anisotropic), selected for consistency across the models, the coverage happens to be higher for anisotropic models.

apparent variation of noise in the measurements and the error in measurements provided in the raw data.

To demonstrate the necessity of anisotropic modeling, we computed the RMSE of predictions for a range of length scales. The results, presented in Fig. 5, show that in order to minimize the RMSE the length scales have to differ. In particular, the length scale corresponding to the first dimension ($\ell_1$, days) must be about 10 times smaller than the length scale corresponding to the second dimension ($\ell_2$, times of day).

Finally, we used interpolated light curves to demonstrate the applicability of our methods to the realistic SDA task of identifying RSOs from measurements. The goal is to understand if we can, using a single day of new observations, identify an RSO based on three years of past observations of several candidates. Fig. 6 shows that we can with about 70% accuracy on average. The classification accuracy varies from light curve to light curve, and can be close to 100% for some of them (e.g. the one labelled 2: *Sat0022*). Using the anisotropic model with heteroscedastic noise during interpolation yields marginally better results, as seen above, which in turn improves the classification accuracy.

## 4. CONCLUSION

We have shown that interpolation of noisy, incomplete light curve data with GPs enables further applications that require smooth, de-noised, and complete data. We were able to classify "new", single-night observations to an existing catalog of light curves with decent accuracy. This task is difficult because the test data cannot be interpolated as accurately as the training data since there are fewer test points than training points. Furthermore, it is a multi-step process comprising of two independent interpolation phases (training and test data) followed by a classification phase and each step affects the overall accuracy. Fortunately, all three steps are similar in that they all use similar GPs with similar normalization and mostly automated hyperparameter optimization. The very small number of hyperparameters is one advantage of GP-based methods compared to other machine learning techniques like Deep Neural Networks (DNNs) and can be a selling point for SDA practitioners ramping up on these methods.

We implemented and demonstrated the use non-stationary kernels by introducing anisotropy and heteroscedastic noise. We showed how the predicted variance follows the right trend of being inversely correlated with the magnitude when using heteroscedastic noise, and how the anisotropic model unlocked improvements in prediction accuracy. We believe these non-stationary kernels are generically useful but realize that they only provided marginal improvements over the basic isotropic and homoscedastic model. We believe that hereroscedasticity in the noise model is insufficient to

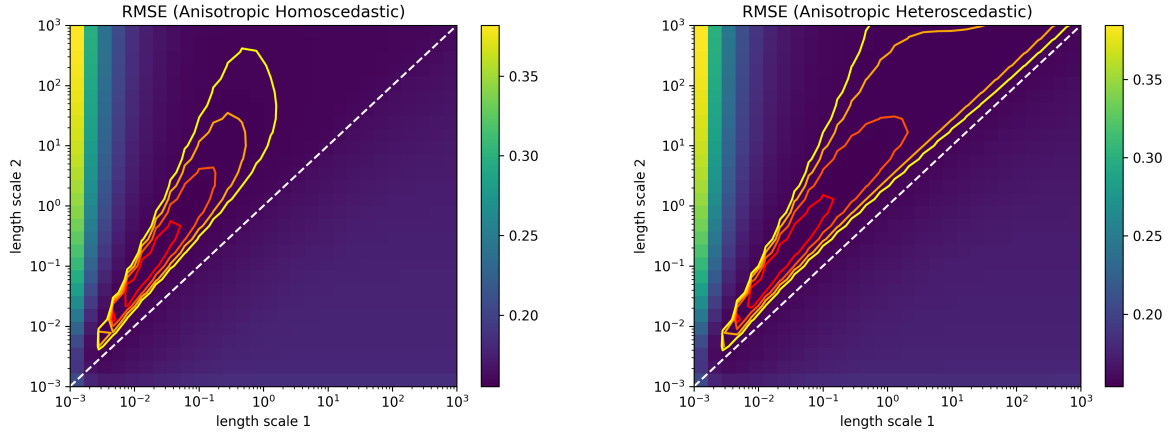


Fig. 5: RMSE after interpolating a single day of missing data in a light curve using an anisotropic model as a function of the two length scale hyperparameters. Length scale 1 corresponds to the first dimension (days) and length scale 2 corresponds to the second dimension (times of day). The minimal region is highlighted using contours and is clearly located above the diagonal illustrating how anisotropic modeling improves accuracy. The extent of the contours also show how using heteroscedastic over homoscedastic noise further improves the accuracy.

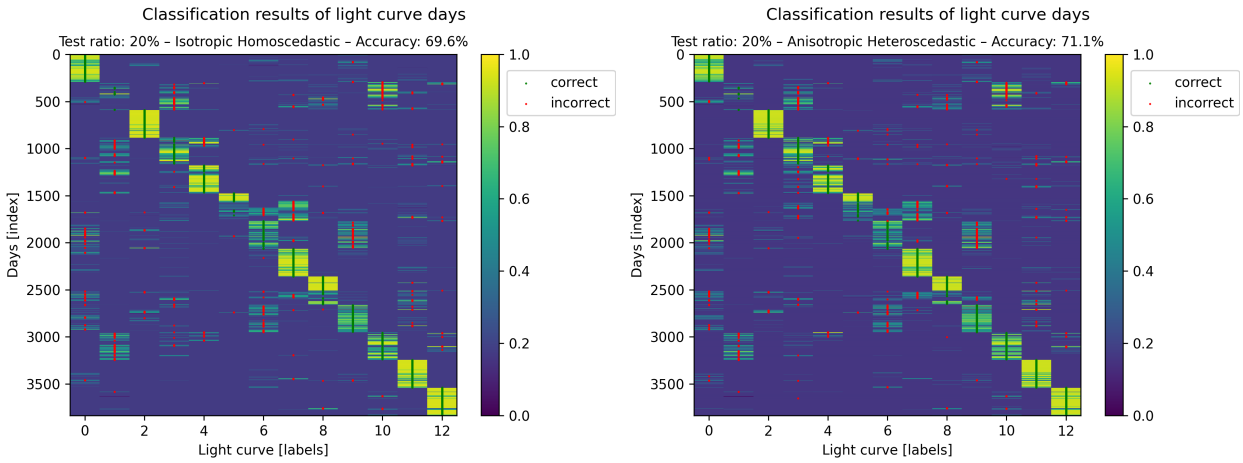


Fig. 6: Classification results. Each row represents a single day and rows corresponding to each light curve are stacked on top of each other. Columns represent the predicted weights associated with each light curve. For each row, the cell with the highest weight determines the predicted label. Green and red dots are superimposed to highlight correct and incorrect predictions respectively. Using the anisotropic model with heteroscedastic noise to interpolate the light curves yields a slightly better classification accuracy.

capture the phenomenon of necessitating a higher variance in high magnitude regions (near dusk and dawn) and that non-stationary variance parameters, like sigma squared or the length scale, would need to be allowed and implemented for the posterior variance to behave as expected. In addition, we think the anisotropic model is impaired by the isotropic nearest neighbors assumption. In further studies, we recommend refactoring the neighborhoods based upon the scaling factors.

The changes proposed above should allow for better uncertainty quantification, which in turns could improve decision making. Combined with the performance of MuyGPs, which allows all the analysis of this paper to run in minutes on a regular laptop, and is also suitable for running on supercomputers and GPUs for tackling even larger problems for which only other machine learning techniques like DNNs were considered, we believe our methods are a great fit for SDA applications. Beyond object identification, one could consider trying to detect and predict finer details of the whereabouts of RSOs like estimating their pose, the accumulation of dust on their reflective surfaces, and their various maneuvers.

## ACKNOWLEDGMENTS

The authors would like to thank Alec M. Dunton for code contributions, discussions, and prototypes that contributed to the analysis in this manuscript.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 with IM release number LLNL-CONF-853760.

Funding for this work was provided by LLNL Laboratory Directed Research and Development grant 22-ERD-028.



## REFERENCES

[1] Andrew D Dianetti and John L Crassidis. Space object material determination from polarized light curves. In *AIAA Scitech 2019 Forum*, page 0377, 2019.

[2] Bradley Efron. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, pages 569–593. Springer, 1992.

[3] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[4] Roberto Furfaro, Richard Linares, and Vishnu Reddy. Space objects classification via light-curve measurements: deep convolutional neural networks and model-based transfer learning. In *AMOS Technologies Conference, Maui Economic Development Board*, 2018.

[5] Roberto Furfaro, Richard Linares, and Vishnu Reddy. Shape identification of space objects via light curve inversion using deep learning models. In *AMOS Technologies Conference, Maui Economic Development Board, Kihei, Maui, HI*, 2019.

[6] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[7] Imene R Goumiri, Alec M Dunton, Amanda L Muyskens, Benjamin W Priest, and Robert E Armstrong. Light curve completion and forecasting using fast and scalable gaussian processes (muygps). *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS)*, 2022.

[8] Bin Jia, Khanh D Pham, Erik Blasch, Zhonghai Wang, Dan Shen, and Genshe Chen. Space object classification using deep neural networks. In *2018 IEEE Aerospace Conference*, pages 1–8. IEEE, 2018.

[9] Laurence Perreault Levasseur, Yashar D Hezaveh, and Risa H Wechsler. Uncertainties in parameters estimated with neural networks: Application to strong gravitational lensing. *The Astrophysical Journal Letters*, 850(1):L7, 2017.

[10] Richard Linares and Roberto Furfaro. Space object classification using deep convolutional neural networks. In *2016 19th International Conference on Information Fusion (FUSION)*, pages 1140–1146. IEEE, 2016.

[11] Richard Linares, Moriba K Jah, John L Crassidis, and Christopher K Nebelecky. Space object shape characterization and tracking using light curve and angles data. *Journal of Guidance, Control, and Dynamics*, 37(1): 13–25, 2014.

[12] Dave Monet. Dataset from flagstaff az, 2014–2018. Personal communication, 2022.

[13] Amanda Muyskens, Benjamin Priest, Imène Goumiri, and Michael Schneider. MuyGPs: Scalable gaussian process hyperparameter estimation using local cross-validation. *arXiv preprint arXiv:2104.14581*, 2021.