2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Automated Testing of Spatially-Dependent Environmental Hypotheses through Active Transfer Learning

Nicholas Harrison, Nathan Wallace, and Salah Sukkarieh

***Abstract*— The efficient collection of samples is an important factor in outdoor information gathering applications on account of high sampling costs such as time, energy, and potential destruction to the environment. Utilization of available a-priori data can be a powerful tool for increasing efficiency. However, the relationships of this data with the quantity of interest are often not known ahead of time, limiting the ability to leverage this knowledge for improved planning efficiency. To this end, this work combines transfer learning and active learning through a Multi-Task Gaussian Process and an information-based objective function. Through this combination it can explore the space of hypothetical inter-quantity relationships and evaluate these hypotheses in real-time, allowing this new knowledge to be immediately exploited for future plans. The performance of the proposed method is evaluated against synthetic data and is shown to evaluate multiple hypotheses correctly. Its effectiveness is also demonstrated on real datasets. The technique is able to identify and leverage hypotheses which show a medium or strong correlation to reduce prediction error by a factor of 1.4–3.4 within the first 7 samples, and poor hypotheses are quickly identified and rejected eventually having no adverse effect.**

***Index Terms*— Autonomous Science, Robotic Sampling, Hypothesis Testing, Active Transfer Learning, Multi-Task Gaussian Process**

I. Introduction

Many motivations exist for gathering environmental data across various outdoor applications. These may be driven by 1) scientific objectives: a desire to learn and characterize an unknown environment or 2) operational concerns: leveraging this new knowledge to refine plans and increase efficiency. This process not only underpins effective planning and decision-making, but it also facilitates the minimization of resource expenditure, be it time, energy, or environmental impact.

A prime example of such strategies being applied in an agricultural context is our recent work on using an autonomous rover to collect soil samples and test them onboard for Nitrogen, Phosphorous, and Potassium content (Fig. 1). Understanding these values is important for farmers to more efficiently manage their fields and increase yields [1]. The process to extract the soil and then pass it through the stages of an onboard lab is complex and time-intensive. This results in only sparse data available and limited capacity to collect new data. In scenarios like these, it becomes a priority to gain the most information from the fewest samples as possible.

All authors are with the Australian Centre for Field Robotics, University of Sydney, NSW, AU. nicholas.harrison@sydney.edu.au, nathan.wallace@sydney.edu.au, salah.sukkarieh@sydney.edu.au

Fig. 1. The ACFR's Swagbot autonomously collecting and analyzing soil samples.

Incorporating prior information helps to address this problem. For any given quantity of interest (QOI), for example soil Nitrogen content, other sources of information may exist in the environment about its distribution. Environmental quantities are often spatially-dependent, meaning their measured values depend on where they are in space. A useful technique to exploit this is the Gaussian Process (GP), known as kriging in geostatistics. This nonparametric machine learning technique uses a covariance function to define how relations in space dictate relations in measurement values [2]. It can give an estimate for a quantity at any point in space conditioned on sparse samples. See example maps produced from a GP in Fig. 2.

Additionally, correlations often exist between the values of different quantities. One example is the relationship between Nitrogen, soil moisture, and temperature [3]. These relationships can be exploited through the technique of transfer learning. Though transfer learning is often applied to image and text classification [4], [5], the concept is also useful in domains such as this one. In the context of this work, transfer learning improves predictions of a QOI by using learned patterns in related quantities. This can also be realized through GPs through a "multi-task" covariance function that accommodates data from multiple sources.

To improve the efficiency of learning these relationships, the technique of active learning can be applied, which involves the process of choosing better samples to take based on their potential informativeness. GPs give values of uncertainty about the estimates they produce, which can be used to calculate which samples will yield the most information about the QOI.

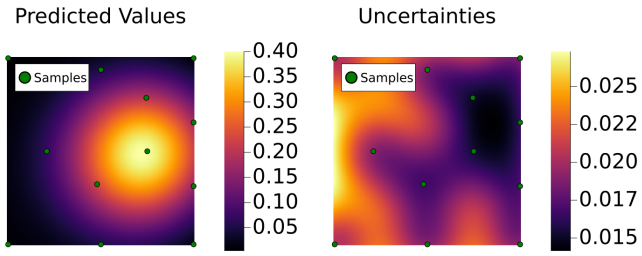

Fig. 2. Example GP maps. Yellow and black indicate high and low values, respectively, and green dots show sample locations. GPs produce smooth interpolations between measurements and higher uncertainty away from sample locations.

Beyond characterization efficiency, combining these two learning techniques has a valuable scientific extension: learning the true global relationships between quantities in the environment. Relationships can often be guessed (i.e. hypothesized) intuitively by a human, but they may or may not be valid for the current region. In this work, hypotheses are defined as a linear dependence between a QOI and another quantity. Transfer learning can represent such hypotheses, and the combination with active learning can test them and learn more accurate ones. Accurately-learned relationships can be confidently re-used in future applications.

One additional and significant aspect of dealing with sparse data from multiple quantities is that samples from the different quantities may not line up in space, which is known as non-collocated data. This aspect makes calculating correlations directly between quantities impossible. An intelligent use of GPs can handle this too by exploiting their property to interpolate values continuously through space. In effect, a spatially-dependent correlation is calculated.

A. Problem Definition and Contributions

In this work, we consider the task of autonomously characterizing a single QOI within a constrained area, such as a section of rangeland. The ground truth data for this quantity are not known by the robot, but it does have sparse data of other quantities in the environment that are hypothesized to be linearly correlated with the QOI. The robot can travel within the region, take samples of the QOI, and calculate predictions of the QOI's distribution as well as the quality of the hypotheses. The robot must sequentially select which sample locations will yield the most information about the QOI and proposed hypotheses.

The main contribution of this work is a method which

- calculates the quality of multiple environmental hypotheses simultaneously and in real-time from sparse, non-collocated data
- learns from good hypotheses, improving prediction accuracy
- disregards poor hypotheses, incurring no loss

After a review of related research, this article will explain the methods for the combined active transfer learning.

II. Related Work

Within the domains of information gathering in natural environments and automated hypothesis testing, a few works are especially relevant.

Furlong investigated a form of hypothesis testing, picking new samples based on greater entropy in the hypothesis space [6]. Their algorithm identified which out of a number of hypotheses was the most true for predicting a spatial map of binary variables. Since they were binary rather than continuous, each hypothesis was either completely true or completely false at each location. They assumed locations in space were independent and compared hypotheses point-by-point, adding a spatial reward to spread out samples.

Also in the area of hypothesis testing, Burger [7] developed a robotic chemist: a system to find an optimal chemical solution using a batched Bayesian search algorithm. The algorithm used GP mean and variance to pick potentially favorable combinations of compounds to test. It indirectly tested hypotheses related to the compounds it was given, but it didn't produce a measure of the validity of each hypothesis. This is in part because it didn't include a model for the relationship between the resulting compound and its constituent parts.

For information gathering, a number of works use information-theoretic measures to guide a robot to locations of greatest informativeness. A few examples of these include information gain [8], upper confidence bound [9], ergodic theory [10], and upper-bounded differential entropy [11]. GPs are common in these scenarios because they estimate both values and uncertainties — useful for quantifying current and potential information and defining objective functions [12], [13].

In addition, many works use various forms of multi-quantity GPs to share information and improve predictions of natural phenomena. For example, [10] and [11] incorporated remote spectroscopic data as part of the input space in the GP covariance function to enhance local predictions. In contrast, [14] and [15] use forms that predict multiple related output values from a common input space. In all of these, the structure chosen allowed correlations, but they didn't provide a way to globally evaluate the relationship between the quantities, whether strong or weak.

Overall, a method is needed to autonomously measure the spatial dependence of two quantities in a natural environment and adapt learning to that measure.

III. Methods

A. Model Learning

Predictions and uncertainties across the search region are modeled using a GP, which is defined by its mean and covariance functions. For the base of the covariance function, we use the squared exponential. Its simplicity leads to easier derivation and lower computational cost, and it is applicable to spatially-continuous phenomena found in nature. It takes the form

$$cov(\mathbf{x}_i, \mathbf{x}_j | \sigma_f^2, \sigma_l^2) = \sigma_f^2 \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma_l^2}\right), \quad (1)$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are sample locations, and σ_f and σ_l act as the signal standard deviation and characteristic length-scale, respectively. This is a common single-quantity covariance function. It causes locations closer in space to have greater covariance, which means their sample values are assumed to be closer.

To transfer information between quantities, something must additionally represent the amount of covariance between them. We follow the pattern of Bonilla in replacing the signal variance with a free-form covariance matrix [16],

$$cov(\mathbf{x}_{i,u}, \mathbf{x}_{j,v} | A, \sigma_l^2) = A_{uv} \exp\left(-\frac{||\mathbf{x}_{i,u} - \mathbf{x}_{j,v}||^2}{2\sigma_l^2}\right), \quad (2)$$

where u and v are quantity indices, and the matrix element A_{uv} is the covariance between two quantities. The covariance function's two factors each serve a role. The exponential factor serves as a spatial correlation between samples ranging from 0 to 1. The matrix A acts as a covariance between entire quantities, i.e. a "quantity-covariance".

For the mean function, we define a quantity-wise constant mean. That is, for each quantity, it will return the average of all the sample values of that quantity:

$$m(\mathbf{x}_{i,u}) = \mathbb{E}\left[y_{\cdot,u}\right] \quad (3)$$

This piece, along with the covariance, is essential to producing hypothesis evaluations as described in the next section. See Fig. 3 for an abstract idea of the Multi-Task Gaussian Process (MTGP) architecture.

To produce a valid positive semidefinite covariance matrix for A, transpose multiplication with a lower-triangular matrix is used (example with 4 quantities):

$$\begin{aligned} A &= LL^T \\ L &= \begin{bmatrix} a_1 & 0 & 0 & 0 \\ a_2 & a_3 & 0 & 0 \\ a_4 & a_5 & a_6 & 0 \\ a_7 & a_8 & a_9 & a_{10} \end{bmatrix}, \end{aligned} \quad (4)$$

These hyperparameters a_i can be chosen arbitrarily to cause or prevent correlations between the quantity predictions. All hyperparameter values are trained by maximizing the log marginal likelihood of the data given the hyperparameters as described in [2],

$$\begin{aligned} \log p(\mathbf{y}|X,\theta) &= -\frac{1}{2}\left[\tilde{\mathbf{y}}^T\Sigma^{-1}\tilde{\mathbf{y}} + \log|\Sigma| + n\log 2\pi\right] \\ \Sigma &= K(X,X|\theta) + \sigma_n^2 I \\ \tilde{\mathbf{y}} &= \mathbf{y} - m(X), \end{aligned} \quad (5)$$

where $K(X,X|\theta)$ is the covariance matrix between all training data produced using (2), $m(X)$ is the array of means produced using (3), and σ_n^2 is the hyperparameter for the noise variance and is also trained.

B. Hypothesis Evaluation

Each hypothesis tested in this work is defined as the claim that a prior quantity is linearly dependent with the QOI. To evaluate the validity of a hypothesis, we compute a coefficient of determination using the quantity-covariance matrix part of the covariance function (4). The correlation coefficient between two quantities can be calculated as

$$r_{uv} = \frac{\sigma_{uv}}{\sigma_u \sigma_v} = \frac{A_{uv}}{\sqrt{A_{uu}}\sqrt{A_{vv}}}, \quad (6)$$

where σ_{uv} is the covariance, and σ_u and σ_v are the standard deviations of the two quantities. The square (r_{uv}^2) represents what fraction of the variation in one quantity is explained by the other and gives a score for the validity of the hypothesis from 0 to 1. r_{uv}^2 values close to 1 indicate a good hypothesis, between 0.5 and 1 a possibility, and below 0.5 very unlikely.

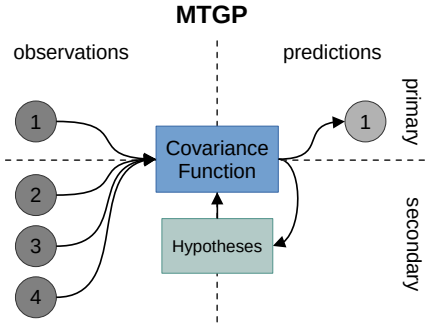

Fig. 3. Abstract depiction of the MTGP architecture. Gray circles represent values for each of the quantities, with 1 labeling the QOI. The covariance function allows transfering information between quantities and representing how much they correlate (hypotheses).

C. Sample Selection

Given the GP model from the previous section, the algorithm searches within the region's bounds for the most informative location to sample. Sample locations are evaluated by an objective function based on Expected Improvement for Global Fit [17],

$$\max_{\mathbf{x}} \quad (\mu(\mathbf{x}) - y(\mathbf{x}^*))^2 + \alpha\,\sigma^2(\mathbf{x}), \quad (7)$$

where $\mathbf{x}$ is the chosen location, $\mu(\mathbf{x})$ and $\sigma^2(\mathbf{x})$ are respectively the GP expected value and variance at the location, and $y(\mathbf{x}^*)$ is the value of the nearest sample location. This function favors new samples that reduce a balance of local and global uncertainties. α is a weight to choose this balance. A particle swarm optimization routine is used for the search.

IV. Results and Discussion

Synthetic data were used to focus on the methods being tested and to directly control relationships between quantities. Data for the QOI were created by generating three to five Gaussian peaks located randomly within the search region. Secondary quantities were then generated from the QOI using three methods:

- **High dependence (H)**: the entire map multiplied by a single value = $QOI_i * \eta \;:\; \eta \sim \mathcal{N}(0,1),\ \eta \neq 0$
- **Medium dependence (M)**: a different random value added at each point = $QOI_i + 0.2\eta_i \;:\; \eta_i \sim \mathcal{N}(0,1)$
- **Low dependence (L)**: a completely new data map generated using the same randomized technique as the QOI

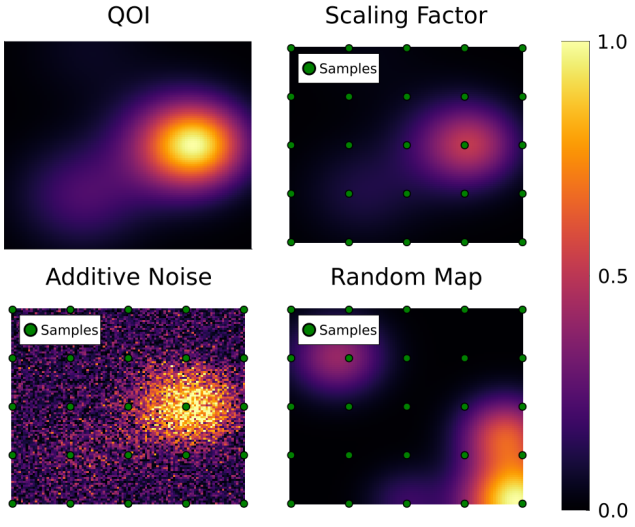

Fig. 4. Maps of the QOI and prior quantities with varying levels of dependence. The top-right has high dependence with the QOI (H), the bottom-left has medium dependence (M), and the bottom-right has low dependence (L).

To simulate having only sparse prior data, a 5x5 grid of samples evenly spread out over the region was extracted. See example maps in Fig. 4.

The simulation was run as follows:

1) Initialize QOI map, prior samples, search bounds, and starting location
2) Take sample at current location and train the model
3) Choose next sample location from the model
4) Continue taking samples and training until a budget of 30 samples is reached

Simulations were run on all possible combinations of prior data from none to all three classes, making eight tests in all (e.g. see Fig. 6). This was repeated over 18 runs of randomized map data from which average performance was calculated. This was enough runs to get statistically significant differences between hypothesis groups ($p < 0.001$). A value for α of 100 was found to be a good balance with slightly more global exploration, as recommended in [12].

On a laptop AMD Ryzen 9 6900HS CPU, the simulation took 0.26 seconds on average to calculate both the belief model and next location after each sample. When run with all 3 priors with 25 samples each, it took 2.1 seconds on average. This can be reduced to below a second if less model refinement is chosen. These are acceptable times for in-situ prediction and decision-making and make real-time use possible.

A. Hypothesis Evaluations

In Fig. 5 we show the calculated hypothesis scores averaged over all simulation runs. For comparison, calculating corresponding scores for all true map values gave $[1.0 \pm 0.0, 0.61 \pm 0.03, 0.06 \pm 0.06]$ for high, medium, and low across all the runs.

The algorithm had no problem in correctly evaluating the multiple hypotheses simultaneously. The three classes

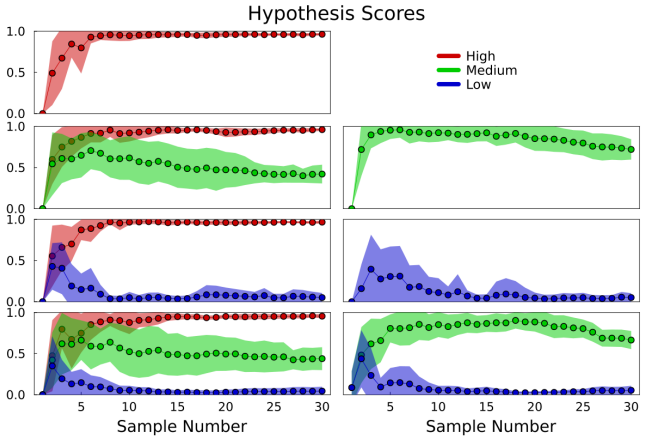

Fig. 5. Mean hypothesis scores between each quantity and the QOI throughout simulation runs of 30 samples. Each subplot is a different combination of High, Medium, and Low hypotheses with values averaged over all runs.

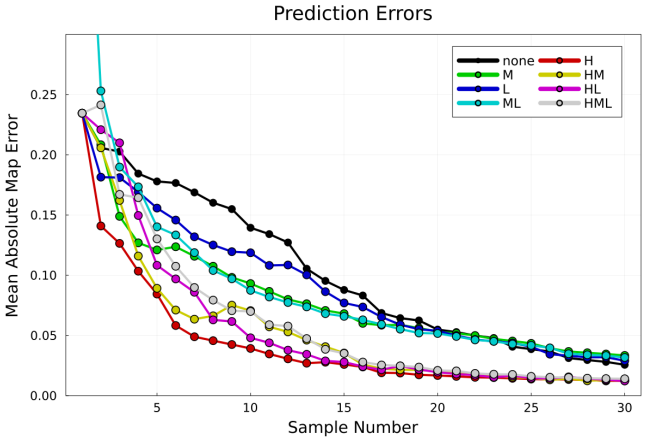

Fig. 6. Mean prediction errors averaged over all runs. Each series is a different combination of High, Medium, and Low hypotheses. A division into three performance groups based on best available hypothesis is visible after 7 samples. The value not shown from the ML series is 0.54.

of prior data (H,M,L) all resulted in consistently distinct hypothesis scores from each other, no matter what the combination. The specific combinations did have slight effects on the scores, especially during the first 5 or so samples. For any method to robustly and accurately accept or reject hypotheses, sufficient data must be collected.

H-priors were assigned near-perfect scores after the first 6 samples and maintained this with high confidence throughout the rest of the run. This trend was consistent through all the possible combinations.

When not combined with a H-prior, M-priors started with higher scores that eventually came down to around 0.83. When H-priors were present, they more quickly settled to around 0.45. The existence of a good hypothesis helped to learn the distinction between it and the medium one.

L-priors universally scored below 0.5 after a few uncertain predictions. This class naturally had the greatest variance since the random map generation had more possibility to result in partial correlations.

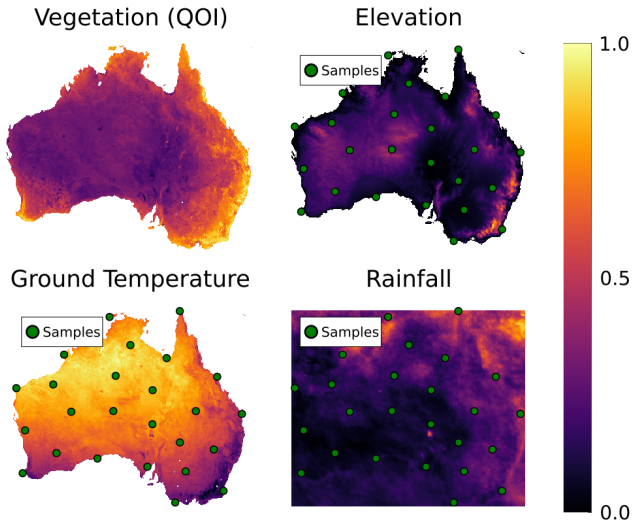

Fig. 7. Real data maps of vegetation, elevation, ground temperature, and rainfall throughout Australia and the sparse samples used. Values are averages over a year and have been normalized within each map.

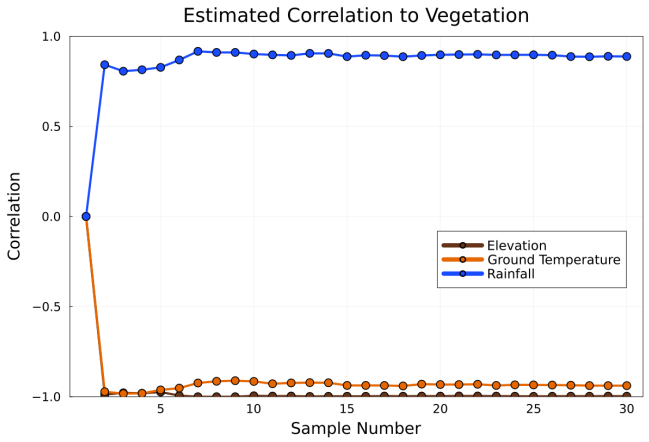

Fig. 8. Estimated correlations to Vegetation over a run of 30 samples for Australia. It is estimated that rainfall is positively correlated while ground temperature and elevation are negatively.

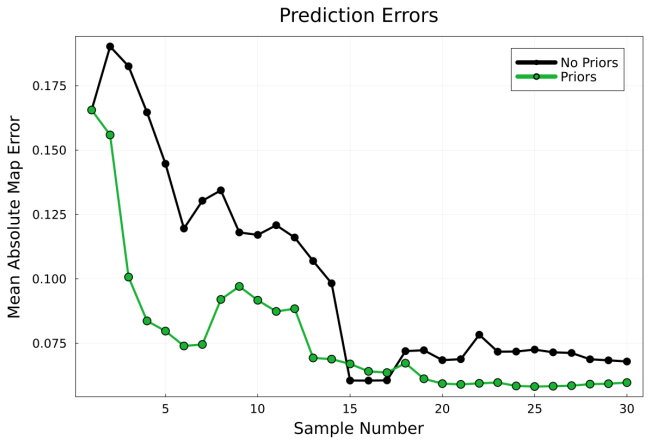

Fig. 9. Mean absolute prediction errors over a run of 30 samples for Australia. Using priors reduces errors over much of the run as compared to no priors.

B. Accuracy to ground truth

Fig. 6 shows average prediction errors for all 8 prior data combinations. The algorithm adapted its use of the prior data, and by 7 samples the errors settled into three major groups. These groups are defined by the highest-dependence prior that they include.

Runs with a H-prior had the lowest errors and maintained that gap for the rest of the 30 samples. Inclusion of lower-dependence priors slightly degraded performance at the start, but by sample 15 they were all effectively equal.

The ML-priors run performed similarly to the M-prior run, and the L-prior run performed at least as well as when including no priors. Overall, each run performed no worse than its best hypothesis and incorrect hypotheses were quickly disregarded. Adding moderate to good hypotheses reduced prediction error and thus the number of samples needed to characterize the QOI.

The principal benefits occur in the short to mid-term. Eventually, enough data of the QOI will be gathered so that prior hypotheses no longer make a difference. This can be seen for runs with only a M-prior or worse: after around 17 samples, they all converged. It is expected that the same thing will occur for all the runs after many more samples. In other words, the QOI will eventually be completely known, whether or not a high-quality hypothesis was used to start.

C. Real-World Examples

We further demonstrate the usefulness of these methods on two real-world datasets with real environmental relationships. This is to show the algorithmic methods of this paper working — the scale and specific quantities can change per outdoor robotic application. Also note that the only thing simulated in these tests is the action of sampling — the methods perform the same on the data regardless of how or when it was obtained. As a plus, having high-resolution offline data allows for ground truth comparisons not possible in the field.

The data were taken from satellite imagery provided by NASA Earth Observations [18]. The two regions are the whole of Australia (Fig. 7) and a 1000x1000 km patch of land between New South Wales and Queensland, Australia (Fig. 10). The QOI was chosen to be vegetation and 25 dispersed samples were extracted from each of the other quantities for prior data. The algorithm was set to characterize vegetation across the map during a run of 30 samples.

After only 2 samples on the full Australia dataset, positive and negative correlations were identified and maintained for the rest of the run. Elevation and ground temperature were found to be highly negatively correlated and rainfall highly positively correlated (Fig. 8). Using these good hypotheses improved the prediction accuracy throughout almost the entire run, including up to 75% better in the beginning (Fig. 9).

Since the algorithm calculates a spatially-dependent correlation, a note on length-scale is important here. The amount

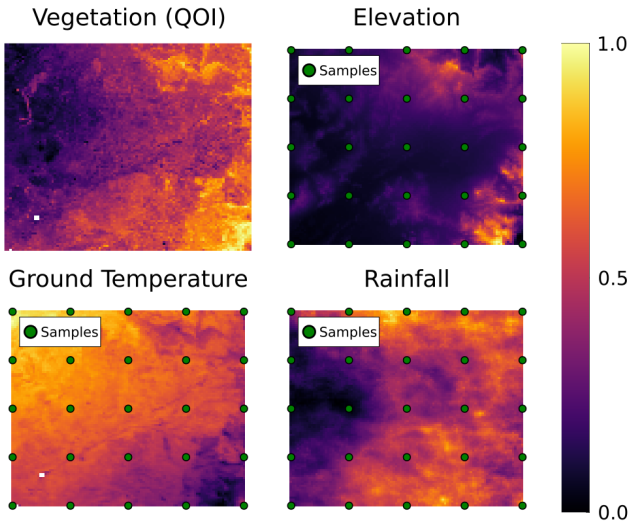

Fig. 10. Real data maps of vegetation, elevation, ground temperature, and rainfall in a section of east Australia and the sparse samples used. Values are averages over a year and have been normalized within each map.

of correlation in the data between two quantities changes depending on the length-scale of the trends examined. Our algorithm models this by the length-scale hyperparameter in the covariance function (1), and its value is mainly determined by the distance between samples and the variation in their values. Our algorithm focuses on best global fit, so it favors more space between samples and hence a larger, more global length-scale. With that in mind, there may be local areas that differ from the global trend, especially with real data. For example, the negative correlation found between vegetation and elevation will not hold in the mountainous regions near the southeast coast. Development of strategies to capture both local and global trends, i.e. correlations on multiple length scales, is of interest for future work.

For the local dataset, the method established correlation estimates by around 7 samples, and they remained mostly stable over the course of the run. The results were partial positive correlations with elevation and rainfall and a negative correlation with ground temperature (Fig. 11). Mean prediction errors showed slightly improved performance when using priors compared with using no priors (Fig. 12).

This used a subset of the overall map from the first example, and it demonstrates what was mentioned in the discussion: local correlations may be different than global ones. At this smaller-scale view, the effect of the mountains on the surrounding climate is more apparent in the data and higher elevations are found to be correlated with more vegetation. It's also important to note again that this is a spatially-dependent correlation rather than a point-by-point one. This means similarities are allowed to be found in regions around each point based on the learned length-scale. As with any model, the underlying assumptions must be understood and assessed as to their applicability for the given scenario. Deeper explanations still fall to the scientist, but the key here is that the algorithm aids them in their analysis.

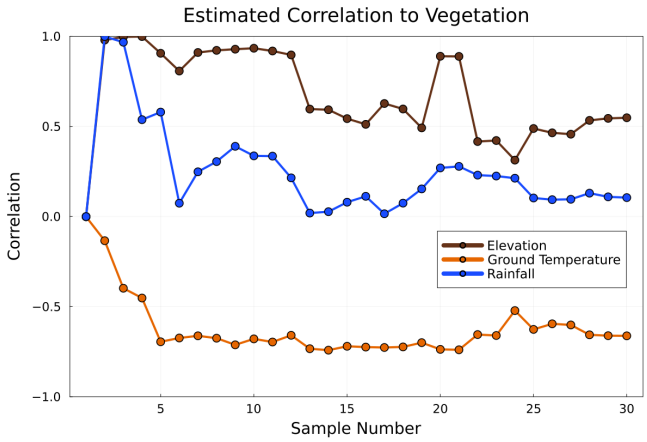

Fig. 11. Estimated correlations to Vegetation over a run of 30 samples. It is predicted that all quantities are partly correlated with vegetation.

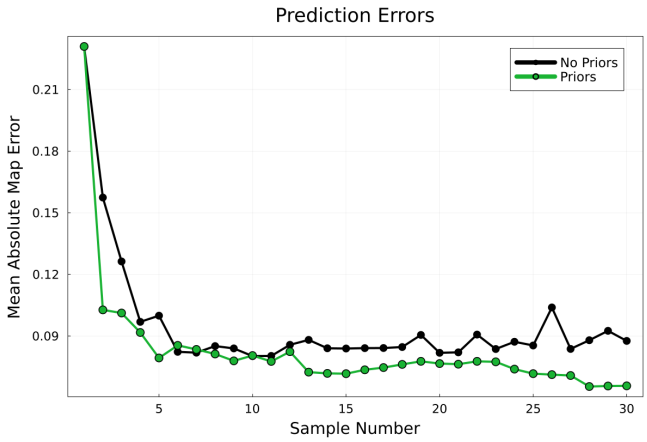

Fig. 12. Mean absolute prediction errors over a run of 30 samples. Using priors slightly improves over using no priors, particularly at the end.

V. Conclusion

This paper presented a method to address two related tasks in outdoor applications: evaluating spatially-dependent environmental hypotheses and reducing the number of samples to characterize a QOI. The method simultaneously evaluates the quality of multiple hypotheses about linear dependence between quantities after every sample. Furthermore, it adapts its learning to the quality of each hypothesis: good hypotheses increase the amount of information gained from each sample and poor hypotheses lose effect after a few samples. Hypotheses identified as high-quality can more confidently be used in further outdoor applications.

This system is suitable for in-situ information gathering on robotic platforms, though the use of this method with real data requires extra considerations. Effective operation is currently limited to linear correlations on a global scale. Future work will investigate nonlinear correlations and correlations on multiple length-scales.

References

[1] S.-J. Kwon, *et al.*, "Effects of nitrogen, phosphorus and potassium fertilizers on growth characteristics of two species of bellflower (platy-

codon grandiflorum)," *Journal of Crop Science and Biotechnology*, vol. 22, pp. 481–487, 2019.
[2] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Massachusetts Institute of Technology, 2006.
[3] M. Guntiñas, *et al.*, "Effects of moisture and temperature on net soil nitrogen mineralization: a laboratory study," *European Journal of Soil Biology*, vol. 48, pp. 73–80, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1164556311000732
[4] F. Zhuang, *et al.*, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.
[5] K. Weiss, *et al.*, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, May 2016. [Online]. Available: http://dx.doi.org/10.1186/s40537-016-0043-6
[6] P. M. Furlong, "Foraging, prospecting, and falsification-improving three aspects of autonomous science," 2018.
[7] B. Burger, *et al.*, "A mobile robotic chemist," *Nature*, vol. 583, no. 7815, pp. 237–241, Jul 2020.
[8] A. Arora, *et al.*, "Online multi-modal learning and adaptive informative trajectory planning for autonomous exploration," in *Field and Service Robotics*, M. Hutter and R. Siegwart, Eds. Cham: Springer International Publishing, 2018, pp. 239–254.
[9] R. Marchant and F. Ramos, "Bayesian optimisation for intelligent environmental monitoring," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 10 2012, p. nil.
[10] K. Edelson, "Ergodic trajectory optimization for information gathering," Master's thesis, Carnegie Mellon University, Pittsburgh, PA, October 2020.
[11] A. Candela, *et al.*, "Planetary rover exploration combining remote and in situ measurements for active spectroscopic mapping," in *2020 IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2020, pp. 5986–5993.
[12] J. N. Fuhg, *et al.*, "State-of-the-art and comparative review of adaptive sampling methods for kriging," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2689–2747, Aug 2020.
[13] B. Shahriari, *et al.*, "Taking the human out of the loop: a review of bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2016. [Online]. Available: http://dx.doi.org/10.1109/JPROC.2015.2494218
[14] H. Leenaers, *et al.*, "Employing elevation data for efficient mapping of soil pollution on floodplains," *Soil Use and Management*, vol. 6, no. 3, pp. 105–114, 1990.
[15] M. Mehta and C. Shao, "Adaptive sampling design for multi-task learning of gaussian processes in manufacturing," *Journal of Manufacturing Systems*, vol. 61, no. nil, pp. 326–337, 2021.
[16] E. V. Bonilla, *et al.*, "Multi-task gaussian process prediction," in *Advances in Neural Information Processing Systems*, J. Platt, *et al.*, Eds., vol. 20. Curran Associates, Inc., 2008.
[17] C. Q. Lam, "Sequential adaptive designs in computer experiments for response surface model fit," Ph.D. dissertation, The Ohio State University, 2008.
[18] NEO, "Dataset index." [Online]. Available: https://neo.gsfc.nasa.gov/dataset_index.php