Risk-aware Path Planning via Probabilistic Fusion of Traversability Prediction for Planetary Rovers on Heterogeneous Terrains

Masafumi Endo[1], Tatsunori Taniai[2], Ryo Yonetani[2], and Genya Ishigami[1]

Abstract— **Machine learning (ML) plays a crucial role in assessing traversability for autonomous rover operations on deformable terrains but suffers from inevitable prediction errors. Especially for heterogeneous terrains where the geological features vary from place to place, erroneous traversability prediction can become more apparent, increasing the risk of unrecoverable rover's wheel slip and immobilization. In this work, we propose a new path planning algorithm that explicitly accounts for such erroneous prediction. The key idea is the probabilistic fusion of distinctive ML models for terrain type classification and slip prediction into a single distribution. This gives us a multimodal slip distribution accounting for heterogeneous terrains and further allows statistical risk assessment to be applied to derive risk-aware traversing costs for path planning. Extensive simulation experiments have demonstrated that the proposed method is able to generate more feasible paths on heterogeneous terrains compared to existing methods.**

I. INTRODUCTION

Reliable rover autonomy is crucial to exploring celestial surfaces, as manual rover operation from Earth involves significant communication latency. Past rovers, such as those for missions on Mars [1], were thus equipped with onboard systems that perceive surrounding environments from stereo vision, assess geometric obstacles, and navigate themselves on collision-free trajectories. Nevertheless, many human interventions were required to assess *traversability* in unexplored environments, resulting in slow rover travel. For instance, the Mars Science Laboratory mission reports an average drive distance of the Curiosity rover being limited to 28.9 meters in one Martian solar day (24 hours 39 min [2]), even though it can travel up to 15.12 m/h [3]. On extraterrestrial terrain, what turned out to be hazardous for rovers other than apparent obstacles were deformable surfaces. As a known case, the Curiosity rover experienced significant slips on rippled sand at the Hidden Valley, forced to change its route for more solid terrain. Such excessive wheel slips degrade driving speed, increase energy consumption, and eventually cause permanent entrapment in loose, granular materials. Hence, reliable traversability assessment on deformable terrains is essential for autonomous rover operation and, in turn, for faster, more extended rover exploration.

Traversability assessment for deformable terrains is challenging due to complex dependencies between physical properties, surface geometry, and rover mobility mechanisms [4]. This problem becomes more complicated when the terrain is *heterogeneous*, *i.e.*, such dependencies even differ from place to place. Thus, instead of theoretical traversability modeling, existing studies attempt to predict traversability via appearance and geometry information using machine learning (ML) algorithms. The appearance information provides cues of surface characteristics, such as material composition, to symbolically categorize heterogeneous terrains via ML classifiers [5]–[8]. The geometry information is also given for learning correlations between terrain geometry and slip behavior to predict traversability for unexplored regions [9]–[13].

We here argue that for the sake of safe rover navigation, learning-based traversability assessment requires increased attention to potential error inevitably existing in ML predictions. While recent ML models demonstrate significant improvement for robotic perception and control [14]–[16], they cannot guarantee perfect predictions. This common issue of ML can cause unrecoverable rover immobilization. Moreover, highly accurate traversability prediction is expectedly difficult, particularly for distant planets, due to the presence of unobservable terrain conditions and limited observations for unexplored regions.

To address future *immobilization risk* for safe rover navigation, we devise a path planning algorithm incorporating rational risk assessment against potential error in traversability prediction for heterogeneous terrains. The key idea is the probabilistic fusion of prediction models while taking into account their uncertainties. Suppose we have a terrain classifier predicting heterogeneous terrain classes given appearance information and latent slip (LS) models mapping geometric features to slip behaviors depending on terrain classes via Gaussian processes (GPs) [17] (Fig 1a). Our strategy then integrates multiple LS models of all possible terrain classes to form a single, multi-modal slip distribution as mixtures of Gaussian processes (MGP; Fig. 1b). This retains the prediction uncertainties of each model and enables us to effectively evaluate the immobilization risk via conditional value at risk (CVaR) [18] for conservative path planning (Fig. 1c-d).

We extensively evaluate our algorithm in synthetic celestial environments, which pose various difficulties in the traversability prediction. Comparison with different prediction models and risk evaluation metrics validates that our algorithm can plan feasible and safer paths on heterogeneous deformable terrains.

[1]Masafumi Endo and Genya Ishigami are with the Space Robotics Group, Department of Mechanical Engineering, Keio University, Kanagawa 223-8522, Japan `masafumi.endo@keio.jp`, `ishigami@mech.keio.ac.jp`

[2]Tatsunori Taniai and Ryo Yonetani are with OMRON SINIC X Corporation, Tokyo 113-0033, Japan `{tatsunori.taniai, ryo.yonetani}@sinicx.com`
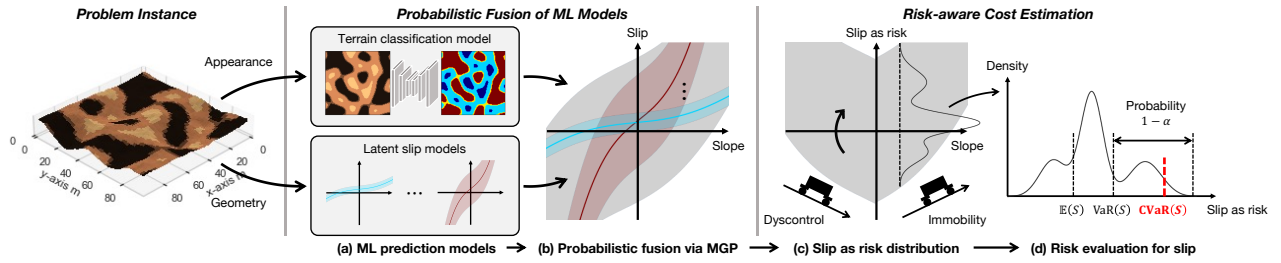
Fig. 1. Risk-aware path planning via probabilistic fusion of traversability prediction models and conservative risk evaluation.

II. RELATED WORK

Numerous studies have tackled terrain traversability prediction by employing ML algorithms with visual and geometric information. Classification-based approaches categorically distinguish terrain types, typically traversable or non-traversable regions, from appearance [5] and infrared [6] imagery. Combining vision and proprioceptive (*e.g.*, vibration) sensing has also been studied for increasing robustness against appearance changes [7], [8]. Regression-based approaches assess traversability to a more detailed degree. These approaches model (non-)traversability as the degree of wheel slip and estimate a mapping from terrain inclination to slip via regression algorithms such as GPs [9]–[13]. To account for multi-modal slip behaviors with regression-based approaches, classify-then-regress-based methods first classify terrains from appearance imagery and then exploit class-dependent regressors for geometry-based slip prediction [9]–[11]. Other work exploits a mixture-of-experts (MoE) framework to incorporate terrain class likelihoods for multi-modal slip prediction [12], [13].

These traversability prediction techniques are further utilized in rover navigation studies for planetary exploration. For example, Ono *et al.* [19] use a terrain classifier to detect non-traversable regions for collision-free path planning using the A* algorithm. Similarly, a path planner by Helmick *et al.* [20] exploits a traversability map predicted by an MoE-based regressor. Such direct use of traversability prediction results for rover navigation, however, exposes rovers to the immobilization risk owing to ML-inherent prediction error.

Such prediction error in ML has also been studied under the concept of *uncertainty*. When quantized via probabilistic models, uncertainty provides useful risk assessment tools in ML applications. Particularly in the context of risk-aware path planning for planetary environments, planners based on chance constraint formulations have been proposed by exploiting the variance of GPs as traversability prediction uncertainties [21]–[23]. However, these methods rely on single GP models and thus assume homogeneous materials covering given environments or accurate terrain classification results in advance. Moreover, although chance constraints are suitable for capturing risk corresponding to boolean events (*e.g.*, collisions) [18], the immobilization risk incurs a range of costs corresponding to the degree of slip.

Conditional value at risk (CVaR), defined to provide a pessimistic estimate of a random variable by accounting for its uncertainty, is another statistical technique for risk assessment. CVaR satisfies certain axioms required for rational risk assessment for robotics applications under uncertainty [18]. For risk-aware planning, CVaR has been applied to avoid various risks such as moving obstacles in robot navigation [24], [25], uncertain regions in semantic segmentation of road scenes [26], or an abstract risk in a hazardous process [27]. While these studies focus on either collision avoidance [24], [25], classification-based traversability modeling [26], or a single-modal risk behavior [27], our focus is on handling the immobilization risk that requires continuous and multi-modal modeling of rover slip on various kinds of deformable terrains. Risk-aware off-road navigation by Cai *et al.* [28] uses CVaR to estimate pessimistic robot speed in environments with dirt and vegetation. Although their use of CVaR for velocity prediction is similar to ours, their semantic-based scene understanding without considering terrain geometry is unsuitable for capturing the immobilization risk owing to rugged celestial surfaces.

In summary, our work focuses on path planning of rovers in the face of the immobilization risk due to heterogeneous and deformable terrains. Compared to existing planners for planetary rover navigation [19]–[23], our method simultaneously considers multi-modal slip behaviors due to heterogeneous terrains and the immobilization risk. The proposed solution (*i.e.*, MGP-based traversability model integrating visual and geometric information as well as CVaR-based continuous (non-boolean) risk assessment) addresses challenges that are only partly handled by existing CVaR-based risk-aware planners in other areas [24]–[28].

III. PRELIMINARIES

This section describes a general path planning formulation, GPs, and CVaR as technical foundations of our algorithm.

A. Path Planning

The path planning objective is to find a sequence of feasible robot state transitions that navigates a robot to a destination safely and efficiently. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a grid map representing an environment in a 2D space, where $\mathcal{V}$ is a finite set of vertices v enumerating possible robot states and $\mathcal{E}$ is a set of edges representing state transitions from each vertex to its eight neighbors. Each edge is associated with a strictly positive cost calculated by a movement cost function $f_{\text{cost}}(v, v')$. The optimal path planning problem is formulated to find a path $\mathcal{P} = \{v_1, v_2, ..., v_{|\mathcal{P}|}\}$, from the

start $v_1 = v_{\text{start}} \in \mathcal{V}$ to the goal $v_{|\mathcal{P}|} = v_{\text{goal}} \in \mathcal{V}$, with the minimum possible total cost defined as follows:

$$\min_{\mathcal{P}} \sum_{i=1}^{|\mathcal{P}|-1} f_{\text{cost}}(v_i, v_{i+1}). \tag{1}$$

B. Gaussian Process for Traversability Modeling

Wheeled rovers experience slip when traveling on deformable terrains. We consider the longitudinal slip ratio s to quantitatively describe the rover traversability as follows:

$$s = \begin{cases} (u_{\text{ref}} - u)/u_{\text{ref}}, & u \leq u_{\text{ref}} : \text{driving}, \\ (u_{\text{ref}} - u)/u, & u > u_{\text{ref}} : \text{braking}, \end{cases} \tag{2}$$

where u and u_{ref} are the actual and reference velocities in the longitudinal direction, respectively [4]. A positive s represents a traverse slower than commanded, with $s = 1$ representing complete rover entrapment in deformable terrains. Conversely, a negative s expresses a traverse faster than commanded.

To model the unknown relationship between terrain geometry and wheel slip s, we exploit GP, a non-parametric regression approach employing statistical inference to learn dependencies between points in a dataset [17]. The use of GPs is justifiable as 1) they can handle the effects of unobservable terrain conditions as uncertainty, and 2) they can express nonlinearity in rover-terrain interactions. A training dataset for GPs may be collected as rovers measure terrain pitch angles ϕ and corresponding longitudinal slips s in the past traverse experience. To account for various slip trends on heterogeneous terrains, suppose terrain classes dominate these slip trends. We then introduce a class-dependent GP that probabilistically models the traversability (*i.e.*, slip s) at an edge $e \in \mathcal{E}$ for a terrain class c as

$$\mathbb{P}_c(s|\phi_e) := \mathcal{N}\left(\mu_c(\phi_e; \boldsymbol{\phi}, \boldsymbol{s}), \sigma_c^2(\phi_e; \boldsymbol{\phi}, \boldsymbol{s})\right), \tag{3}$$

where the predictive mean μ_c and and variance σ_c^2 are parameterized with training samples of input pitch angles $\boldsymbol{\phi}$ and output slip measurements $\boldsymbol{s}$ (see [17] for details).

C. Conditional Value at Risk

Rovers must measure uncertainty in traversability to assess risk during planning. Unlike collision risk considered in typical planning problems [29], our immobilization risk by deformable terrains gradually increases with slip magnitude. We exploit CVaR to quantify such risk through probabilistic traversability modeling. CVaR measures a conservative expected value accounting for tail events in a given distribution, enabling rational risk assessment without being too optimistic or pessimistic [18]. Let S be a random variable of the longitudinal slip ratio at an arbitrary rover pitch. CVaR at level $\alpha \in [0, 1]$ denotes the expected value of S in the upper $(1-\alpha)$-tail, subject to the conditional distribution of S as follows:

$$\text{CVaR}_\alpha(S) := \mathbb{E}\left[S|S > \text{VaR}_\alpha(S)\right], \tag{4}$$

where VaR_α is the value at risk (VaR) expressing $(1-\alpha)$-quantile of S as $\text{VaR}_\alpha(S) := \min\{s|\mathbb{P}(S > s) \leq \alpha\}$.

Intuitively, CVaR provides pessimistic risk estimates from the given traversability prediction, as illustrated in Fig. 1d, factoring in its uncertainty for risk-aware path planning. Given α controls risk assessment behavior from aggressive to conservative throughout the above formulations. If $\alpha = 0$, CVaR returns the expectation. If $\alpha = 1$, the CVaR outcome matches the worst case.

IV. RISK-AWARE PATH PLANNING ALGORITHM

A. Overview

This section presents the unified algorithm translating ML-based traversability predictions and their uncertainties as risk-aware traveling costs evaluating immobilization risk in planning. Unlike typical path planning problems, such as shortest pathfinding in 2D maps, our formulation for rover traveling requires variable traversability costs to account for slip behaviors on heterogeneous deformable terrains. Although the actual slip trend is unknown, it chiefly depends on the material composition and surface geometry in given environments. We thus associate the environment graph with appearance and geometry information, particularly the colors and 3D positions of terrain surfaces, to be utilized for traversability cost prediction via ML models. Projecting such information into the graph vertices is a durable assumption for actual exploration missions, as the HiRISE camera used for Mars exploration can capture overhead imagery with at most 0.25 m/pixel resolution [30].

The overall method pipeline is illustrated in Fig. 1 and summarized as follows. We employ two kinds of pre-trained models: 1) a terrain classifier to pixel-wisely predict terrain classes from appearance imagery and 2) GPs to model class-dependent LS functions with geometry information. These models are fused into a single, multi-modal probabilistic slip distribution via mixtures of GPs (MGP) [31] (Section IV-B). We then extract CVaR estimates from this MGP to derive an edge-wise cost for planning (Section IV-C). We here essentially translate a probabilistic slip prediction with uncertainty into a deterministic, effectively conservative cost by considering immobilization risk due to the potentially erroneous classifier and GPs. Finally, we perform path planning procedures, such as A*, to find a trajectory that will not induce rover entrapment.

B. Probabilistic Fusion of ML Models

Here, we elaborate on MGP representing the probabilistic fusion of the appearance-based terrain classifier and geometry-based slip regressors (GPs). The MGP estimates $\mathbb{P}_{e:v \to v'}(s)$, a probability distribution of the slip s in the transition e from a vertex v to its neighbor v'. This distribution is formulated as the sum of multiple class-dependent GPs $\mathbb{P}_e(s|c)$, weighted by classification likelihoods $\mathbb{P}_v(c)$ as

$$\mathbb{P}_{e:v \to v'}(s) = \sum_{c \in C} \mathbb{P}_v(c)\mathbb{P}_e(s|c), \tag{5}$$

where $\mathbb{P}_e(s|c) = \mathbb{P}_c(s|\phi_e)$ in eq. (3). This single MGP distribution in eq. (5) can explain various slip trends on heterogeneous terrains and can simultaneously express overall

uncertainties from both appearance-based terrain classification and geometry-based slip regression.

For terrain classification, any model taking appearance imagery and predicting a categorical distribution at each pixel can be used. In this work, we leverage the U-Net [32], one of the popular computer vision models for pixel-wise semantic segmentation tasks. Applying this model results in per-pixel class probabilities expressed via the softmax function as $\mathbb{P}_v(c) = \exp(a_c(v))/\sum_{c' \in C} \exp(a_{c'}(v))$, where $a_c(v)$ is the scalar output of U-Net for a vertex v and a terrain class c. Note that selecting a single GP corresponding to the most likely class $c^\star = \arg\max_c \mathbb{P}_v(c)$ in eq. (5), instead of summing multiple GPs, reduces the presented traversability prediction to existing one [9]. However, such a formulation ignores classification-owing uncertainties in the risk assessment described next.

C. Risk-aware Cost Estimation

To perform path planning using eq. (1), we derive its traversability costs f_{cost} from the slip distribution given by MGP. We design the cost function as *risk-aware travel time*, which considers both traversability in terms of risk endangering rovers and traverse efficiency. Travel time is a desirable criterion for the cost function since it encompasses efficiency and safety in terms of maintaining a higher traverse rate with lower slip. The time to traverse an edge $e : v \to v'$ is given as $t = ||\boldsymbol{x}(v) - \boldsymbol{x}(v')||/u(e)$, where $\boldsymbol{x}(v)$ denotes the 3D position at a vertice v and $u(e)$ is the rover velocity when traversing an edge. Here, $u(e)$ is derived with slip s by converting eq. (2) as follows:

$$u(e) = \begin{cases} (1-s)\,u_{\text{ref}}, & \phi_e \geq 0 : \text{ascending}, \\ u_{\text{ref}}/(1+s), & \phi_e < 0 : \text{descending}, \end{cases} \quad (6)$$

where ϕ_e is the pitch angle at the edge e. In the ascending case in eq. (6), a larger slip $s \in [0, 1]$ can be interpreted as risk since it slows down rover velocity, increases traverse time, and eventually triggers rover immobilization. In the descent case, a negatively larger slip $s \in (-1, 0]$ may lead to faster velocity and shorter travel time. However, this seemingly 'preferable' state should be actually regarded as another risk because it induces rover dyscontrol, such as deviation from the planned trajectory and tip-over situation.

To comprehensively evaluate these two types of risks from slip phenomena, we introduce *slip as risk*. Given a random variable of slip $S(\phi)$ parameterized by a pitch angle ϕ, the slip as risk is defined as another random variable as

$$S_{\text{risk}}(\phi) = \begin{cases} S(\phi), & \phi \geq 0 : \text{ascending}, \\ 2S(0) - S(\phi), & \phi < 0 : \text{descending}. \end{cases} \quad (7)$$

Here, the distribution of $S(\phi)$ follows MGP in eq. (5). Essentially, the slip as risk in eq. (7) evaluates the deviation from the state of traversing on the flat ground (i.e., $\phi = 0$), by vertically flipping $S(\phi)$ in $\phi < 0$ (Fig. 1c), assuming that the flat-ground traversing is most stable and least risky. We then extract a pessimistic estimate of $S_{\text{risk}}(\phi_e)$ via CVaR as

$$s_{\text{risk}}^+(e) = \text{CVaR}_\alpha \left(S_{\text{risk}}(\phi_e) \right). \quad (8)$$

By substituting $s = s_{\text{risk}}^+$ into eq. (6), we obtain a rover velocity accounting for the comprehensive risks, denoted as $u_{\text{risk}}(e)$. The cost function is derived as the risk-aware travel time that evaluates time efficiency and rover traverse stability in both ascent and descent directions, as follows:

$$f_{\text{cost}}(e : v \to v') = \frac{||\boldsymbol{x}(v) - \boldsymbol{x}(v')||}{u_{\text{risk}}(e)}. \quad (9)$$

Note that MGP in eq. (5) can be alternatively expressed with $\mathbb{P}_{v'}(c)$ instead of $\mathbb{P}_v(c)$. We thus compute costs in two ways using both definitions and use their average for planning.

V. SIMULATION EXPERIMENTS

This section demonstrates the proposed path planning algorithm through extensive simulations and validates its robustness against unreliable traversability prediction.

A. Datasets

We prepare three synthetic datasets for path planning problems with varying difficulties in traversability prediction. Each dataset has the training, validation, and testing problem instances as 96×96 grid maps with a resolution of one grid per meter, providing RGB and height maps as inputs for planners. As hidden information, each instance has pixel-wise assignment of terrain classes $c \in C$ and a set of class-dependent slip functions $f_c(\phi)$. Throughout the dataset creation, fractal terrain modeling [33] is applied to generate randomly distributed rough terrain geometry.

To synthesize diverse but reasonably controlled datasets, we here define *environment groups*, a random subset of terrain classes with occupancy ratios for controlling the terrain class occurrence in each map (*e.g.*, 50% for class #1, 30% for class #4, and 20% for class #6). Training, validation, and testing subsets individually have ten distinct environment groups (*i.e.*, ten different random choices of terrain classes), from each of which we generate 100, 50, and 10 problem instances, respectively. This results in 1000 training, 500 validation, and 100 testing instances. For each problem instance, we first synthesize a random Perlin noise image, cluster the values based on the occupancy ratios specified by the environmental group, and assign a corresponding terrain class (*i.e.*, slip model) to each cluster. For each slip model, a slip measurement process is simulated via additive zero-mean Gaussian noise of a pre-defined variance as $s = f_c(\phi) + \epsilon_c$. Synthesized slip measurements are used to train each class-dependent GP. In-situ measurements are also taken during path executions by substituting calculated ϕ_e into s.

With the above approach in mind, here is the detail of the three datasets we have created.

- **Standard (Std) Dataset:** This dataset adopts a simple setting where pixel color information identifies terrain classes and slip measurements contain slight noise ϵ_c. Each environment group consists of four out of ten terrain classes. Fig. 2 (a) and (b) show an example terrain appearance image and a corresponding terrain class map coded by distinctive colors, respectively. Fig. 2 (c) shows learned slip models via GPs.
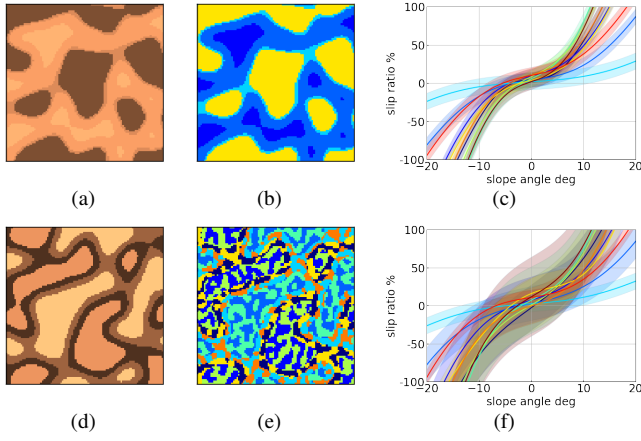
Fig. 2. Visualized examples of the (a)-(c) Std dataset and (d)-(f) AA dataset. Each dataset consists of color maps, corresponding terrain class maps, and GPs expressing learned latent slip functions (solid line: mean, shaded area: 95% confidence region). Color in terrain class maps and GPs denote distinct terrain classes.

- **Erroneous Slip (ES) Dataset:** A more complicated setting is considered in this dataset by imposing greater noise ϵ_c in slip measurements as the gradient magnitude of $f_c(\phi)$ increases. Hence, learned GPs result in erroneous slip models shown in Fig. 2 (f). As described in [9], such erroneous slip modeling can happen due to multiple factors, including insufficient slip measurements and unobservable terrain conditions.
- **Ambiguous Appearance (AA) Dataset:** We replicate another complicated setting where multiple terrain classes appear with the same appearance, assuming situations, *e.g.*, when distinctive terrain characteristics do not provide visually identifiable cues or one terrain material is covered by another. This dataset comprises four pairs of terrain classes, each with high-variance GPs (as in the ES dataset) but sharing the same color cues, resulting in eight classes with less appearance diversity. Each environment group is then defined by the random ordering of those classes such that four different colors always appear, as shown in Fig. 2 (d) and (e).

B. Implementation Details

As the path search algorithm, we use the A* search, which returns a resolution-optimal solution as long as it exists in a given discrete domain. As the terrain classifier, we adopt the U-Net model [32] with the ResNet-18 encoder [34] to provide classification likelihoods. This model was trained for each dataset separately using its training subset for 50 epochs using the Adam optimizer [35] with a learning rate of 0.001 and a batch size of 8. The model weight producing the lowest loss on the validation subset was used for evaluation on the test subset. Computing VaR and CVaR for random variables associated with the MGP distribution in eq. (5) was carried out approximately through a Monte Carlo simulation.

C. Experimental Setups

In every problem instance, planners search the path from the start position (8 m, 8 m) to the goal (88 m, 88 m).

TABLE I
QUANTITATIVE PATH PLANNING RESULTS OVER DIFFERENT DATASETS

		Sol.	Succ.	T_{total}	s_{max}
Standard (Std) Dataset					
SGP [9]	+EV	100	71	22.6 ± 3.7	58.3 ± 47.0
	+VaR	100	74	22.6 ± 3.9	55.6 ± 45.4
	+CVaR	100	75	22.6 ± 3.9	55.3 ± 45.8
MGP	+EV	100	75	$\mathbf{22.5} \pm 3.5$	56.1 ± 44.4
	+VaR	100	93	22.9 ± 4.4	47.5 ± 30.7
Ours (MGP+CVaR)		100	**96**	23.2 ± 5.2	$\mathbf{45.9} \pm 24.2$
Erroneous Slip (ES) Dataset					
SGP [9]	+EV	100	60	$\mathbf{23.0} \pm 3.5$	70.3 ± 40.7
	+VaR	98	65	24.3 ± 5.1	68.8 ± 44.7
	+CVaR	93	62	25.0 ± 5.7	67.9 ± 44.8
MGP	+EV	100	71	23.1 ± 4.4	68.2 ± 38.2
	+VaR	98	**77**	25.2 ± 12.7	63.1 ± 38.8
Ours (MGP+CVaR)		90	**77**	27.2 ± 37.4	$\mathbf{60.5} \pm 37.1$
Ambiguous Appearance (AA) Dataset					
SGP [9]	+EV	100	11	$\mathbf{22.1} \pm 1.7$	92.9 ± 27.4
	+VaR	100	7	24.8 ± 4.0	95.8 ± 19.3
	+CVaR	100	2	26.7 ± 1.8	96.3 ± 19.3
MGP	+EV	100	38	24.0 ± 5.6	81.4 ± 32.9
	+VaR	98	**95**	25.7 ± 5.3	64.5 ± 22.3
Ours (MGP+CVaR)		96	**95**	26.4 ± 6.3	$\mathbf{63.7} \pm 21.6$

After finding a solution, a rover follows the given path at $u_{\text{ref}} = 0.1$ m/sec while receiving an edge-wise noisy slip s_e. We regard the solution as a success path if we never observe $s_e \geq 1$ (complete rover immobilization) and $s_e \leq -1$ (uncontrollable situations such as tip-over) on the way. The following four metrics quantify algorithm performances.

- **Solved rate (Sol.) [%]** expresses the ratio of problem instances for which a planner finds solutions.
- **Success rate (Succ.) [%]** denotes the ratio of problem instances whose found solutions are successful.
- **Total time (T_{total}) [min]** provides path efficiency by calculating the total traversing time along the path when traversing is successful.
- **Maximum slip (s_{max}) [%]** evaluates path safety against rover entrapment by calculating $100 \times \max(\boldsymbol{s}_e)$ after path execution, where $\boldsymbol{s}_e$ denotes the experienced slips. A lower s_{max} indicates a safer traverse on terrains.

To see the effectiveness of our proposed approach, we also evaluated different traversability prediction models and risk evaluation metrics. As for the prediction, a single Gaussian process (SGP) approach is given by selecting a GP based on classified terrain types [9]. In addition to VaR and CVaR, the expected value (EV) quantifies the given slip distribution for cost calculation. With combinations of the above metrics, planning procedures run on the datasets for comparison.

D. Results

1) Quantitative Comparison: Table I summarizes quantitative path planning performances on the three datasets, with $\alpha = 0.99$ in VaR and CVaR. Throughout the datasets, our proposed approach owns the highest performance in terms of safety to avert immobilized situations, as shown by its success rates and maximum slips during path execution.
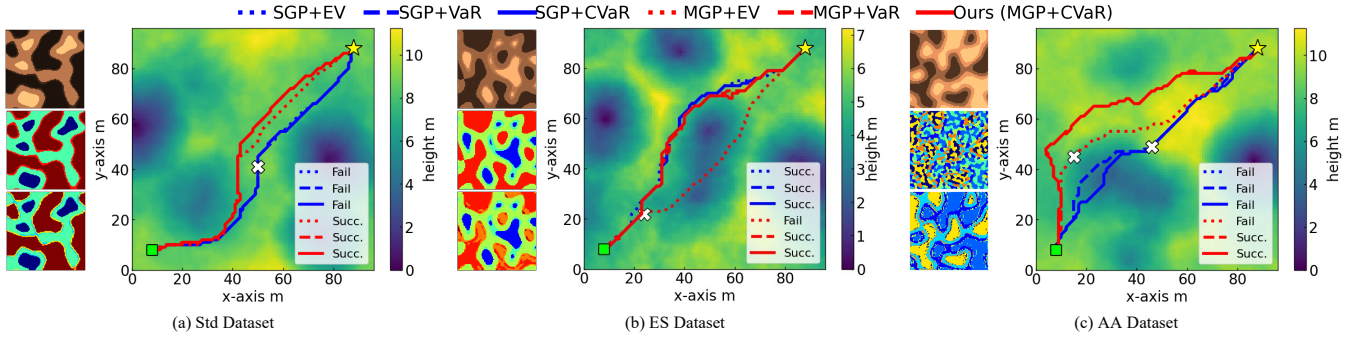
Fig. 3. Path planning results over different datasets. In the environments represented as heatmaps, green squares, and yellow stars are start and goal positions. White crosses denote positions where rovers fail to transit to the next states due to experiencing $s = 1$. Input colors, ground truth terrain classes, and terrain classification results are shown from top to bottom on the left side of each heatmap.
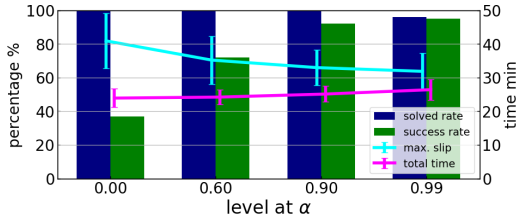

Fig. 4. Parameter study results over AA Dataset. Line graphs use mean and standard deviation format.

Having lower slip also contributes to retaining efficiency in time by maintaining rover velocity, as shown by the total time outcomes. These results are preferable for planetary exploration, where safety is of utmost importance, as rover entrapment leads to mission failure while overly conservative rover navigation delays mission schedules. Erroneous slip modeling, despite higher appearance diversity experienced in the ES Dataset, poses a relative performance decrease in planning outcomes compared to the other datasets. We found that the planner with VaR- and CVaR-based risk evaluations sometimes does not provide solutions. This can happen when the goal position is in hazardous areas. Dismissing such risk to prioritize finding solutions worsens the success rate, as indicated by the EV-based planners. The MGP formulation for traversability prediction greatly contributes to the success of rover navigation. The most effective cases are those provided by the AA Dataset, where terrain appearance cannot assure its terrain class. Without MGP, traversability prediction only refers to representative slip models according to identified terrain properties, resulting in failed traverses due to huge prediction errors from wrong slip models. Our proposed algorithm solves this issue by integrating multiple GPs with terrain likelihoods, thus enabling careful risk evaluation accounting for less-likely tail events that can lead to immobilized situations. Total computation time for planning, including cost estimation with MGP, is typically 10 to 15 minutes on a standard computer in Python implementation.

2) Qualitative Comparison: Fig. 3 illustrates how the planners with different risk metrics plan paths for a problem instance from each Dataset. MGP-based approaches tend to find trajectories along geometrically benign regions com-

pared to SGP-based ones, but the MGP+EV-based planner often seeks a too optimistic path leading to rover immobilization, as shown in Fig. 3 (b). This is because the MGP generates a slip distribution covering a broader range of slips compared to a single GP. VaR and CVaR are thereby necessary for rational risk evaluation to see less likely occurring slips in the MGP distribution, enabling successful rover traverses in challenging environments.

3) Parameter Study: Finally, a parameter study testing $\alpha \in \{0.00, 0.60, 0.90, 0.99\}$ for the AA Dataset is conducted to see its effects against the performance of the proposed algorithm. Fig. 4 indicates that α governs the trade-off between safety versus efficiency: a higher α leads to an increased success rate by navigating rovers on a less-slippery path, while a lower α allows a more aggressive maneuver to shorten the time to traverse. The controllability of planning behavior via simple parameter tuning is a beneficial aspect of our method, considering that priorities of safety and efficiency will change depending on the mission timeline; we may want to prioritize safety during the early exploration phase when most terrain properties are unknown.

4) Open Issues and Possible Extensions: Although our proposed algorithm can significantly improve safely traversing celestial environments, no theoretical guarantee is proven for avoiding rover entrapment. Possible extensions will complementarily exploit probabilistic constraints, such as chance constraints, by formulating reasonable interpretations for non-boolean risk of immobility.

VI. CONCLUSION

We presented a new path planning algorithm that can explicitly consider the uncertainty in ML-based traversability prediction for reliable rover operations on heterogeneous deformable terrains. Through the extensive simulation experiments, we confirmed that the proposed algorithm consistently outperformed existing methods in terms of planning success rates and the maximum degrees of wheel slip experienced during path execution. Future work will investigate the feasibility of the proposed method using actual data such as those for Mars [36]. Another possible future direction is to extend the algorithm to refine the prediction model through online learning [23].

REFERENCES

[1] S. Goldberg *et al.*, "Stereo vision and rover navigation software for planetary exploration," in *Proceedings of IEEE Aerospace Conference*, vol. 5, 2002, pp. 5–5.

[2] M. H. Hecht *et al.*, "Detection of perchlorate and the soluble chemistry of martian soil at the phoenix lander site," *Science*, vol. 325, no. 5936, pp. 64–67, 2009.

[3] A. Rankin *et al.*, "Driving curiosity: Mars rover mobility trends during the first seven years," in *Proceedings of IEEE Aerospace Conference*, 2020, pp. 1–19.

[4] J. Y. Wong, *Theory of ground vehicles*. John Wiley & Sons, 2008.

[5] B. Rothrock *et al.*, "Spoc: Deep learning-based terrain classification for mars rover missions," in *Proceedings of AIAA SPACE*, 2016, p. 5539–5551.

[6] Y. Iwashita *et al.*, "Virtual ir sensing for planetary rovers: Improved terrain classification and thermal inertia estimation," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6302–6309, 2020.

[7] C. A. Brooks and K. Iagnemma, "Self-supervised terrain classification for planetary surface exploration rovers," *Journal of Field Robotics*, vol. 29, no. 3, pp. 445–468, 2012.

[8] K. Otsu *et al.*, "Autonomous terrain classification with co- and self-training approach," *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 814–819, 2016.

[9] C. Cunningham *et al.*, "Locally-adaptive slip prediction for planetary rovers using gaussian processes," in *Proceedings of International Conference on Robotics and Automation*, 2017, pp. 5487–5494.

[10] M. Endo *et al.*, "Terrain-dependent slip risk prediction for planetary exploration rovers," *Robotica*, vol. 39, no. 10, p. 1883–1896, 2021.

[11] H. Inotsume and T. Kubota, "Adaptive terrain traversability prediction based on multi-source transfer gaussian processes," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021, pp. 2297–2304.

[12] A. Angelova *et al.*, "Learning and prediction of slip from visual information," *Journal of Field Robotics*, vol. 24, no. 3, pp. 205–231, 2007.

[13] C. Cunningham *et al.*, "Improving slip prediction on mars using thermal inertia measurements," *Autonomous Robots*, vol. 43, no. 2, pp. 503–521, 2019.

[14] J. Kober *et al.*, "Reinforcement learning in robotics: A survey," *International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[15] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019.

[16] S. Grigorescu *et al.*, "A survey of deep learning techniques for autonomous driving," *Journal of Field Robotics*, vol. 37, no. 3, pp. 362–386, 2020.

[17] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. Cambridge, MA, USA: MIT press, 2006.

[18] A. Majumdar and M. Pavone, "How should a robot assess risk? towards an axiomatic theory of risk in robotics," in *Robotics Research*, 2020, pp. 75–84.

[19] M. Ono *et al.*, "Risk-aware planetary rover operation: Autonomous terrain classification and path planning," in *Proceedings of IEEE Aerospace Conference*, 2015, pp. 1–10.

[20] D. Helmick *et al.*, "Terrain adaptive navigation for planetary rovers," *Journal of Field Robotics*, vol. 26, no. 4, pp. 391–410, 2009.

[21] H. Inotsume *et al.*, "Robust path planning for slope traversing under uncertainty in slip prediction," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3390–3397, 2020.

[22] A. Candela and D. Wettergreen, "An approach to science and risk-aware planetary rover exploration," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9691–9698, 2022.

[23] M. Endo and G. Ishigami, "Active traversability learning via risk-aware information gathering for planetary exploration rovers," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11 855–11 862, 2022.

[24] A. Hakobyan, G. C. Kim, and I. Yang, "Risk-aware motion planning and control using cvar-constrained optimization," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3924–3931, 2019.

[25] K. Ren *et al.*, "Chance-constrained trajectory planning with multimodal environmental uncertainty," *IEEE Control Systems Letters*, vol. 7, pp. 13–18, 2023.

[26] V. D. Sharma *et al.*, "Risk-aware planning and assignment for ground vehicles using uncertain perception from aerial vehicles," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 11 763–11 769.

[27] F. S. Barbosa *et al.*, "Risk-aware motion planning in partially known environments," in *Proceedings of IEEE Conference on Decision and Control*, 2021, pp. 5220–5226.

[28] X. Cai *et al.*, "Risk-aware off-road navigation via a learned speed distribution map," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2022, pp. 2931–2937.

[29] B. D. Luders *et al.*, "Robust sampling-based motion planning with asymptotic optimality guarantees," in *Proceedings of AIAA Guidance, Navigation, and Control Conference*, 2013, p. 5097.

[30] A. S. McEwen *et al.*, "Mars reconnaissance orbiter's high resolution imaging science experiment (hirise)," *Journal of Geophysical Research: Planets*, vol. 112, no. E5, 2007.

[31] V. Tresp, "Mixtures of gaussian processes," in *Advances in Neural Information Processing Systems*, vol. 13, 2000.

[32] O. Ronneberger *et al.*, "U-net: Convolutional networks for biomedical image segmentation," in *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[33] Y. Yokokohji *et al.*, "Evaluation of traversability of wheeled mobile robots on uneven terrains by fractal terrain model," in *Proceedings of International Conference on Robotics and Automation*, vol. 3, 2004, pp. 2183–2188.

[34] K. He *et al.*, "Deep residual learning for image recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of International Conference on Learning Representations*, 2015.

[36] M. Ono *et al.*, "Data-driven surface traversability analysis for mars 2020 landing site selection," in *Proceedings of IEEE Aerospace Conference*, 2016, pp. 1–12.