Combining Open Data and Formal Reasoning for Autonomously Controlled Spreading near Water Bodies

Ahmad Kadi[1], Nikolas Müller[1], Ansgar Bernardi[1], Federico Ulliana[2], and Guillaume Pérution-Kihli[2]

Abstract: Applying fertilizers and pesticides near bodies of water poses significant environmental risks, primarily due to the potential for chemical runoff to contaminate aquatic ecosystems. To avoid this, regulations establish prohibition zones based on environmental and application-specific parameters, such as terrain slope, wind speed, precipitation, and the type and composition of substances used. This paper presents an autonomous robotic system that was developed to comply with these regulations while maximizing usable agricultural land. The robot scans its environment with sensors, including LiDAR, to measure features such as the distance to nearby bodies of water and the slope of the ground underneath. The InteGraal reasoning framework uses samples of regulations encoded in machine-readable RDF formats (using PAM vocabularies) and sensor observations modeled by the Semantic Sensor Network Ontology (SSNO) to make real-time decisions about where to stop or resume spraying. We extend existing vocabularies to include fertilizer-specific regulations, ensuring a comprehensive, semantically rich decision-support system for autonomous farms.

Keywords: Plant Protection Agent, Pesticides, Fertilizer, RDF, Vocabularies, Autonomous Robots, Simulation, Open Data, PAM

1 Introduction

Spreading plant protection agents, such as pesticides and fertilizers, is challenging. It becomes even more complicated when spraying next to water bodies, such as rivers, streams, or ponds, because the chemicals can easily be washed away by rain or irrigation, polluting the water. That's why states and agricultural institutions have set up many regulations to avoid such scenarios. These regulations take many factors into consideration to establish a distance threshold from which a prohibited zone is formed toward the water body. These factors include the slope of the ground, wind speed, rainfall, and the order of the water body in terms of its size and streamflow. The regulations also depend on what is being spread (fertilizer or a plant protection agent) and vary depending on the chemical composition. Our paper addresses the scenario of an autonomous robot dedicated to applying fertilizer to fields bordering water bodies. The robot must scan the environment to measure related parameters using different sensors. Then, it must reason about the measurements and the rules to decide when to stop and resume spraying.

[1] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH - DFKI, Trippstadterstr. 122, 67633 Kaiserslautern, ahmad.kadi@dfki.de; nikolas.mueller@dfki.de; ansgar.bernardi@dfki.de

[2] LIRMM, Inria, University of Montpellier, CNRS, Montpellier, France, Federico.Ulliana@lirmm.fr; guillaume.perution-kihli@lirmm.fr

2 Regulations on Spraying Next to Water Body

The regulations define the distance from a body of water where it is not allowed to spread fertilizers or plant protection agents. This helps utilize as much of the land as possible for growing crops while ensuring that the boundaries where it might harm or pollute the water are not exceeded. There is no general distance that applies everywhere; it depends on several parameters. The PAM [Es21] [Sc16] service, provided by the project of the same name, offers regulations related to applying plant protection agents near water bodies in machine-readable format using controlled RDF vocabularies, based on the publicly available database for approved plant protection products in Germany [BV]. This service and its vocabularies formed the basis for our application. We expanded those vocabularies to model the regulations related to spraying fertilizers. This was necessary due to the different structure of fertilizer regulations.

3 Measurement of the Environment

In order to comply with regulations on spraying fertilizers or plant protection agents near water bodies, environmental measurements are needed to determine whether the spraying machine can operate. This requires attaching many sensors and measuring devices, and feeding the data to a decision-making machine along with the regulations. To this end, we modeled the output of a robot and a drone in a simulated environment of a field bordering a stream of water. The output is distance from the robot to the water body, and the slope of the ground underneath, both calculated from the *LiDAR* sensors in the simulated environment.

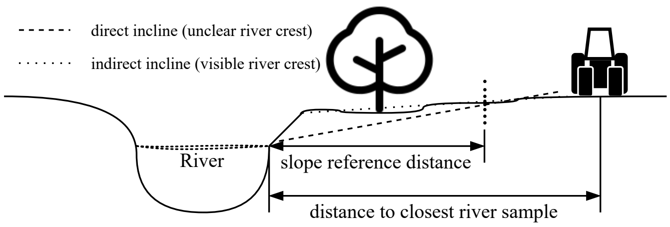

Fig. 1: Estimation of slopes next to a water body

3.1 Measured Features

While advancing on the field, the drone and robot scan the environment using *LiDAR* sensors with GPS localization and generate an implicit map using a *Neural Distance Field* (NDF) [Pa24]. An implicit map is a functional definition of the map surface, which can be queried for encoded features like the closest distance to the surface, color or semantics. For this use case it is beneficial, as it naturally extrapolates to unobserved data.

After generating the NDF, the reference to the river is estimated by querying the points with the lowest z-value. Since most terrain is above the water level, it can be assumed that the water has the lowest reference points in the spatial map. Once the reference to the river is known, the relevant environmental measurements, namely the distance and slope to the river are estimated as visualized in Fig. 1. The distance is simply the horizontal distance to the closest reference point on the river. As the localization of the robot can experience inaccuracies, this distance can be modeled statistically with mean and standard deviation.

For the estimation of the slope, it first needs to be known, if the river crest is clearly perceivable, like on the right river bank in Fig. 1, or not perceivable, like on the left river bank. In case of the former, the distance has to be computed to the upper bound of the river bank, while in the latter case the distance is computed to the mean water level. This can be efficiently detected in the spatial map using a line-fitting algorithm like RANSAC or least-squares. The incline of the slope is then derived from the surface elevation around the reference point, which is n metres away from the river in direction of the robot. In the next step, the computed distance and slope are passed to the data model for the measurements.

The measured features of distance and slope are modeled using the Semantic Sensor Network Ontology (SSNO) vocabulary. This ontology describes sensors, the observations they make, the procedures involved, the features of interest under investigation, the samples used, the observed properties, and the actuators. The core ontology of SSNO is called SOSA (Sensor, Observation, Sample, and Actuator) for its elementary classes and properties [W3b]. Those elementary elements are depicted in Fig. 2. We also used QUDT vocabularies to model the values of the parameters, which include the proper units, as seen in Fig. 3. QUDT, which stands for (Quantities, Units, Dimensions, and Type) is an ontology and data model developed to represent and manage physical quantities and their associated units in a structured, machine-readable way—particularly for use on the Semantic Web and in Linked Data applications [W3a].

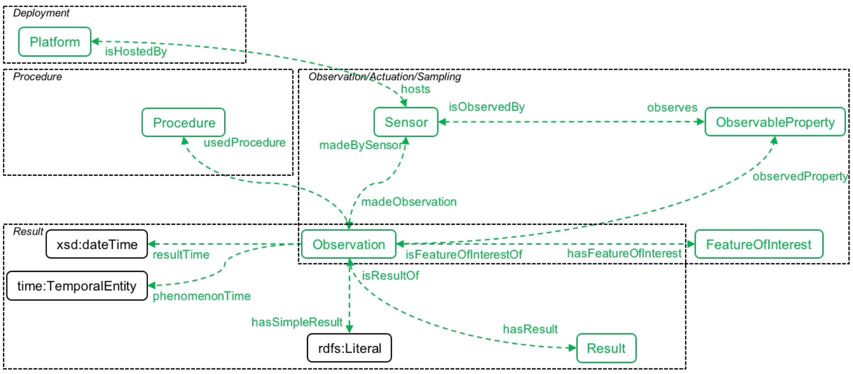

Fig. 2: Overview of the SOSA classes and properties from its official website

```
ex:waterbody a sosa:FeatureOfInterest ; rdfs:label "waterbody" .
ex:distance a sosa:ObservableProperty ; rdfs:label "distance" .
ex:robot a sosa:Platform ;
    rdfs:label "Robot" ;
    sosa:hosts ex:lidar .
ex:lidar a sosa:Sensor ;
    rdfs:label "lidar" ;
    sosa:observes ex:distance .
ex:1704895867941880882_waterbody_distance
        sosa:hasFeatureOfInterest ex:waterbody ;
    sosa:hasResult [ a qudt:Quantity ;
            qudt:unit unit:M ;
            qudt:value 2.570299e+01 ] ;
    sosa:madeBySensor ex:lidar ;
    sosa:observedProperty ex:distance ;
    sosa:resultTime "2024-01-10 14:11:07.941881"^^xsd:dateTimeStamp .
```

Fig. 3: Modeling distance to water body measurement in our application

4 Decision Making and Application Structure

The *InteGraal* reasoning framework [Ba23] will be responsible for making real-time decisions about whether or not a violation of regulations is occurring, and thus whether to continue or stop spraying. It can build a knowledge base from different datasources, and allows querying it while taking rule reasoning into account. To do so, it receives the modeled regulations a priori, and the environmental measurements in real time. The regulations are formalized as logic rules and stored in a deductive knowledge graph, while real-time sensor observations (such as slope and distance) are continuously published and ingested using a publisher-subscriber architecture. The model compares the current measurements with the related rules, made out from the modeled regulations, and makes a decision in real time, using forward-chaining rule evaluation over live data. This approach ensures explainability, as decisions can be traced back to specific rules and measurements within the knowledge graph. The scenario involves a robot moving on a field located next to a water body. The robot advances on the field while scanning the surroundings with LiDAR sensors. The robot performs a computation to calculate two important parameters: the distance of the robot from the border of the water body and the slope of the ground where the robot is currently standing. These values are measured many times per second, modeled using SOSA and QUDT vocabularies, timestamped, and published to the reasoner, which listens to the same topic using a subscriber. The reasoner applies the regulations' rules and determines whether spreading chemicals is permitted at this timestamp.

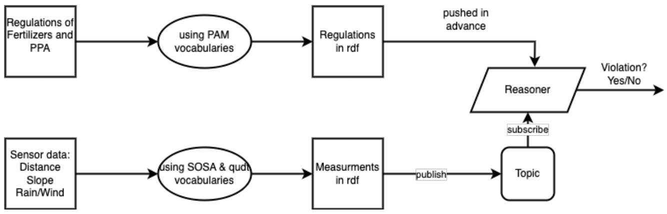

Fig. 4: Structure of the application. Both regulations and facts are being modeled in rdf and fed to the reasoner, but the top is in advance, the lower in a streaming way using publisher-subscriber.

5 Current Status and Next Steps

In the *ROS2* framework, an open-source platform for developing robot applications, we created a simulation file, *rosbag*, that records a drone flying alongside a river and a robot advancing toward the river. The file plays within a Docker container containing our application and publishes the measurements in real time to the reasoner, which listens to the file and performs its task by making decisions while considering the regulations and measurements (distance and slope). Currently, we are working on incorporating measurement uncertainties into our simulation, modeling its output in a standardized way, and defining rules in the reasoner that address the uncertainties. The recorded simulation publishes the same output as a real robot and drone operation, meaning that testing the application with real data is possible. For the same reason, the reasoner and decision-making calculations can also occur on a remote device or in the cloud. This reduces the amount of resources needed by the robot to make a decision but makes the system less resilient to connection drops. We also believe that system would work with the available tractors, who be made capable of sending the measurements of slope and distance in the standardized model way.

6 Evaluation

To assess the accuracy and reliability of our system, we propose a two-stage evaluation process. First, we evaluate the perception layer by comparing the estimated environmental features—specifically the slope and the distance to the water body—with ground truth values available from the simulation environment. Since the terrain and water body were artificially created, we have precise knowledge of their geometry, allowing us to quantify the error in LiDAR-based estimation using standard metrics such as mean absolute error (MAE) or root mean square error (RMSE). Second, we evaluate the decision-making component by comparing the system's spray/no-spray decisions with an expected decision derived from applying the regulation rules directly to the known ground truth measurements. This allows us to isolate errors introduced by either measurement inaccuracies or reasoning mismatches and determine the overall decision accuracy of the system under controlled, repeatable conditions.

7 Conclusion

In our application, we demonstrated how to autonomously spray plant protection agents or fertilizers next to a water body in a simulated environment. We demonstrated how to integrate environmental measurements and regulatory rules into the RDF world using well-known ontologies. We use publicly available open data that contains the rules and restrictions for applying such substances to fields next to water bodies. Since these regulations vary by state, we attempted to expand the available vocabulary to model French regulations, as an example. The application profits form the interoperability of the standard representation of the measurements, which ease its use with any current system who sends the data is this format. Spraying near water bodies is an example of applying open data, controlled vocabularies, and reasoning to agriculture. We believe our approach could be applied to similar use cases where regulations should also be considered when performing actions in the field. For example, protection of groundwater from contamination.

References

[Ba23] Baget, J.-F. et al.: InteGraal: a Tool for Data-Integration and Reasoning on Heterogeneous and Federated Sources. In: BDA 2023-39e Conférence sur la Gestion de Données–Principes, Technologies et Applications. Montpellier, France, 2023.

[BV] BVL: Online-Datenbank Pflanzenschutzmittel, Bundesamt für Verbraucherschutz und Lebensmittelsicherheit, https : / / www . bvl . bund . de / DE / Arbeitsbereiche / 04 _ Pflanzenschutzmittel / 01 _ Aufgaben / 02 _ ZulassungPSM / 01 _ ZugelPSM / 01 _ OnlineDatenbank/psm_onlineDB_node.html, accessed: 04/15/2025.

[Es21] Estel, S. et al.: Der Produktionsmittel-Anwendungs-Managers (PAM) – Ein exemplarisches Beispiel für verteilte internetbasierter Dienste in der digitalen Landwirtschaft. In. Julius-Kühn-Institut, Bundesforschungsinstitut für Kulturpflanzen, Quedlinburg, Germany, pp. 552–553, 2021.

[Pa24] Pan, Y. et al.: PIN-SLAM: LiDAR SLAM using a point-based implicit neural representation for achieving global map consistency. IEEE Transactions on Robotics, 2024.

[Sc16] Scheiber, M. et al.: Pflanzenschutz-Anwendungs-Manager (PAM): Automatisierte Berücksichtigung von Abstandsauflagen. Praktische Vorführung und Feldtestergebnisse. In: Informatik in der Land-, Forst-und Ernährungswirtschaft 2016. Gesellschaft für Informatik eV, pp. 177–180, 2016.

[W3a] W3C: QUDT Ontology, https://qudt.org/, accessed: 04/15/2025.

[W3b] W3C: Semantic Sensor Network Ontology, https://www.w3.org/TR/vocab-ssn/, accessed: 04/15/2025.