Towards model-based automation of plant-specific weed regulation

Marian Renz[1], Mark Niemeyer[1] and Joachim Hertzberg[1,2]

Abstract: Weeds are commonly known as a major factor for yield losses in agriculture, competing with crops for resources like nutrients, water, and light. However, keeping specific weeds could benefit agricultural sites for example by nitrogen fixation, erosion protection, or increasing biodiversity. This comes with technological challenges like plant detection and classification, damage estimation, and selective removal. This paper presents a model-based approach to the problem of damage estimation of perceived plants. The system uses contextual and background knowledge in the form of rules about the plant count per square meter and the distance to the nearest crop together with thresholds for each weed species. The functionality is demonstrated using an artificial dataset and exemplary thresholds, showing the potential of using knowledge about plant-crop interactions for more sophisticated weed control systems.

Keywords: selective weeding, biodiversity, rule-based reasoning

1 Introduction

Within the yield-reducing factors in agriculture, weeds have the biggest damage potential when not mitigated [Oe06]. However, the extensive use of herbicides brings further problems. Selective herbicides can increase the occurrence of resistant weeds [No12], while a decline in biodiversity is observable. Furthermore, herbicide accumulation in soil and water bodies negatively impacts the ecosystem and increases health risks in humans [Al21].

In Germany, however, only 20 of the 350 species of spontaneously growing plants on fields are considered harmful [Le13]. Keeping specific weeds may have beneficial effects like nitrogen fixation, reduction of erosion, or serving as a host and food source for birds and beneficial insects [We14]. Of course, the benefit and the damage from specific weeds are circumstantial, depending for example on species and inter-species relations, degree of coverage, or periods, where the crop is more vulnerable to weeds [Ma21].

Taking the benefits and damage potential into consideration for automated weed regulation is challenging. Besides automated weed detection, classification, and removal

[1] German Research Center for Artificial Intelligence, Plan-Based Robot Control Group, Berghoffstr. 11, 49090 Osnabrück, firstname.lastname@dfki.de
[2] Osnabrück University, Knowledge-Based Systems Group, Wachsbleiche 27, 49090 Osnabrück, joachim.hertzberg@uos.de

a model for estimating the potential damage is needed. The applied principle of weed-specific damage thresholds falls short when it comes to inter-species interactions and is only available for a limited number of crops and weeds [Ma21].

Progress has been made in selective weed control due to the use of artificial neural networks for weed detection, some systems already combine deep learning-based detection and selective regulation using smart sprayers [PCA19; LAN21; Hu20; Ab20]. In contrast, the approaches for plant-specific damage estimation are scarce. While a learning-based approach might be possible, collecting, and labeling datasets in the required size for damage estimation for individual weeds is impractical. A model-based approach using context and background knowledge has the advantage of not depending on high amounts of data. Already existing knowledge about the agricultural domain can be used to formulate rules from which plant-specific decisions can be made.

Furthermore, the importance of transparency and trust in AI is increasing in Europe [Eu20]. Rule-based systems have some advantages when it comes to explainability since all results are inferred deterministically from explicitly formulated rules [va21].

A model-based approach to the challenge of plant-specific weed removal is presented by using context- and background knowledge in the form of rules to make plant-specific decisions about the potential benefit of a weed. The resulting rule-based system generates an application map, which contains the position of harmful weeds on the field. The presented system can be seen as a general pipeline, which leaves the possibility for more sophisticated knowledge to be integrated. The functionality is demonstrated on a synthetic dataset containing RGB images of maize plants and 5 different weed species.

2 Related work

Integrated smart spraying systems with deep learning-based weed detection technology already exist. These systems usually aim to reduce the applied amount of spraying liquid. Patel et al. [PCA19] present a low-cost smart sprayer system utilizing deep convolutional neural networks for the distinction between artificial targets and non-targets. Similarly, Hussain et al. [Hu20] present a smart sprayer detecting weed-infected potato plants, reducing the sprayed amount of liquid by up to 42% in a laboratory setting. Comparable systems are developed for crop-specific applications like strawberry fields [LAN21] or citrus trees [PCA21].

While all these approaches successfully integrate deep learning-based weed- or target detection with variable spraying systems, none of them consider including domain knowledge for a more integrated approach to weed control. They limit their scope to the discrimination between targets and non-targets. The same situation can be found in non-AI-based smart spraying solutions [Ab20].

An extensive model for integrated weed management is FLORSYS, a biophysical simulation environment for cropping systems [Co21]. FLORSYS enables predictions for weed development based on inputs like crop succession, weed management techniques, soil characteristics, and weather data. Based on FLORSYS, a decision support system for integrated weed management has been developed [Co20]. The decision support system utilizes the FLORSYS simulation and a learned meta-model to communicate the effects of different cropping systems on weed development.

Extensive background knowledge is integrated into the biophysical models of FLORSYS. This can provide high-quality decision support for the design of integrated weed management but cannot be integrated into an automated weed regulation system working on sensor data due to missing interfaces. This paper presents an approach using a knowledge model based on rules as well as the necessary interfaces, which can be used in a smart spraying system.

In general, there are plenty of applied model- and rule-based systems for agricultural applications. Different systems have been developed for crop, pest, and disease management [Gh18; MRS11] but rule-based systems for weed management or even integrated weed management are scarce.

Typically, weed related approaches only consider the negative effects of weeds and are targeted at reducing effort and resources for weed management. This approach considers a system that not only detects and removes weeds but rather integrates background and domain knowledge. The system can be integrated into a smart weed regulation system with interfaces for deep learning-based plant detection and classification as well as selective actuators. In contrast to other model-based systems in agriculture, the knowledge is modeled for actuator control and not for decision support.

3 System architecture

The overall system for weed damage estimation consists of four major parts (Fig. 1): the semantic environment representation (SEEREP) [NPH22] for storing labeled sensor data, a grid map for spatial representation and attribute calculation, a rule engine including a knowledge base of rules for inference and the generation of an application map as the interface to the weeding system. In this section, each part is described in more detail. The plant detection and classification are considered as an external module and are summarized with other potential sources of information as “Input Data”.

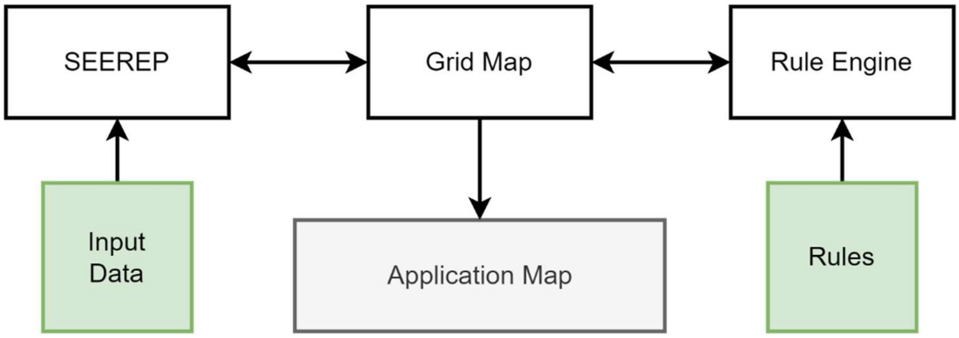

Fig. 1: Architecture diagram and data flow of the model-based weed regulation system

3.1 Spatio-temporal-SEmantic Environment REPresentation – SEEREP

The weed damage estimation cannot be performed on a sensor data stream directly. The sensor data must be stored and processed so that the model-based system can access the data in an efficient and unified way. To achieve this, the spatio-temporal-semantic environment representation (called SEEREP) is used [NPH22]. SEEREP stores the sensor data and creates indices to enable queries in the modalities of space, time, and semantics.

For the spatial modality, the geo-location, and the spatial extent of each sensor reading are indexed so that data queries regarding a specific position can be executed. Additionally, the point in time of the sensor reading is indexed in the temporal domain so that it is possible to store and access historical data. The temporal index is not used in the use-case of this paper but may be used in the future. The last modality, the semantics, is important for the model-based system because it enables symbolic reasoning. The semantic annotations in the data (for example bounding boxes in images with the semantic label of the plant shown) are also indexed, which enables queries for specific labels.

The usage of the three described modalities enables the analysis of the environment and the efficient data query in all three modalities. The model-based system cannot handle the annotated sensor data directly. It requires the position and the semantic label of specific plant instances. Thus, SEEREP can link the semantic annotations of the same object in multiple datasets to a common instance. This instance is not only linked to the sensor data but the position of the instance, which may be calculated based on the sensor data, can also be stored in SEEREP and be linked to the instance. This enables the query of an instance based on its position. Additionally, triple-based attributes can be stored alongside the instance enabling the storage of reasoning results.

Based on these capabilities of SEEREP, the following data handling steps are needed. (1) The raw sensor data is stored in SEEREP and the indices for the spatial and temporal modality are created. (2) The raw sensor data is processed, for example by convolutional neural networks, so that semantic annotations can be added to the data. For those annotations, a common vocabulary, which the model-based system understands, defined by an ontology must be used. (3) In addition to the general semantic annotations, the annotations of the same real-world object (plant) must be linked to an instance. For this anchoring problem, the assumption can be used that the plants do not move. (4) When all the annotations of the same plant are linked to the same instance the number of the instances is known. For the rule-based system also the position of each instance must be known and linked to the instance. For this, the sensor data must be processed to get the position. This may be done via a projection of image data to the ground plane or via 3D information from point clouds or depth images. (5) The instances with their semantic annotations and their positions are obtained via the spatio-temporal-semantic query interface. This enables queries for plants of a specific type in a specific region. The obtained instances with semantic annotation and with their position can be handled by the rule-based system for decision generation. (6) Finally, the triple-based attribute store of the instances can be used to store the decision results.

3.2 Grid map value calculation

For the convenient representation of the agricultural field and value calculation, a grid map is used. Plant instances are inserted into the grid map based on their global position in the field. A plant instance has the following attributes: a unique identifier (UUID), a species name, a global 2D position, a category as result of the rule inference, and numeric values for the rule inference, being the count of the species per square meter and the distance to the nearest crop. The global position can be determined using high-precision GNSS during recording.

The grid map has three internal representations:

- a 2D grid map of plant instances,
- a 3D tensor with the same size for the first 2 dimensions as the 2D map and a third dimension with the number of known species as size, used for counting plant instances,
- a key-value map with the instance identifier as the key and the plant instance as the value.

When inserting a plant into the grid map, the plant instance and all attributes known beforehand are inserted into the 2D grid map based on the plant's global position. Accordingly, the respective counter in the 3D tensor is increased by one and the instance is added to the key-value map.

After retrieving all available plant instances from SEEREP, the numeric attributes are calculated for each plant. The count per square meter is calculated using a sliding window with the size of 1m x 1m. For each grid cell, the plants within the window are summed up per species, using the plant counts from the 3D tensor. The accumulated sum per species is then assigned to all plant instances in the grid cell under consideration. The distance to the nearest crop is calculated by searching for the closest cell with crops for each cell with weeds. The distance is then calculated by using the L2 norm on the plant's global position.

After calculating and assigning the numeric values for each plant instance they are sent into the rule-based system's working memory for inference about the plant's categories.

3.3 Rule-based system

For rule-based inference Drools is used as the backend, an open-source Business Rules Management System [Re22]. The modeled rules categorize whether a plant is considered a crop, a harmful weed, a harmless weed, or an endangered species.

A total of five rules are currently implemented as a working example:

- If the plant's name is listed as a crop, it is categorized as crop.
- If a plant's name is listed as endangered species, it is categorized as endangered species.
- If a plant's count per m^2 exceeds a species-specific threshold, it is categorized as harmful weed.
- If a plant's distance to the nearest crop falls below a species-specific threshold, it is categorized as harmful weed.
- If a plant is not categorized by any of the rules above, it is considered a harmless weed.

Even though these rules are rudimentary, they already allow for a more plant-specific weed regulation strategy, considering species-specific background knowledge. The specific thresholds as well as more sophisticated rules need to be estimated by more extensive research for weed management. An Excel file is used as an interface to input the crop, the species-specific thresholds, and the list of endangered species.

The declarative nature of rule-based systems allows for easy extensibility. Rules and data sources can be added and modified without modifying the program flow. More sophisticated sources of knowledge like weather forecasts, geometric attributes about the plants, or geographical information can be included.

3.4 Output

To utilize the system together with a selective sprayer, an application map is generated from the grid map. Weeds categorized as harmful are stored as points in shapefiles. To estimate the area for the sprayer to cover, the bounding box from the plant detection is converted into an area estimate in cm and added as an attribute to each point. The shapefile therefore contains a list of coordinates and areas to which chemicals are to be applied.

4 Example with synthetic data

The system is tested using a synthetic dataset consisting of RGB images of maize plants and 5 different weed species (Tab. 1) with ground truth labeling. The images are created using Blender [B.o.J.]. The simulated scene consists of a field of maize plant models placed in rows and the randomly distributed weed models.

The images are rendered from a virtual camera (1280*960 px), which is placed approximately 2 m above the ground facing down orthogonally. For each image, the camera's position is changed by 0.5 m along the maize row, rendering 40 images over 20 m. 2 rows of images are rendered next to each other, resulting in 80 images covering 3 adjacent maize rows and 485 plants in total.

Species name	Plant count per m²	Crop distance	Total count	% categorized as weed
Maize	-	-	200	0 %
Autumn Hawkbit	-	-	7	0 %
Common Daisy	6	0.04 m	157	48 %
Curly Duck	1	0.2 m	7	100 %
Ground Elder	2	0.2 m	34	79 %
Plantain Ribwort	5	0.15 m	80	65 %

Tab. 1: Simulated plant species with thresholds for plant count per m², distance to the nearest crop, total count within the simulation, and percentage of plants categorized as weed

Using synthetic data enables pixel-correct classification results for the input data, delivering ground truth quality bounding boxes for each plant. In a real-world use case, this would be substituted by a plant perception pipeline. However, for demonstrating the functionality of the model-based system, the simulated data is sufficient.

The images and bounding boxes are stored in SEEREP and retrieved using a query, obtaining all 485 detected plants with species names and global positions. Then, the plant instances are inserted into the grid map for the attribute calculation. The used grid cell size is 10 * 10 cm. The plant count per m² and the distance to the nearest crop are calculated for each plant, which are then submitted into the rule-based system's working memory.

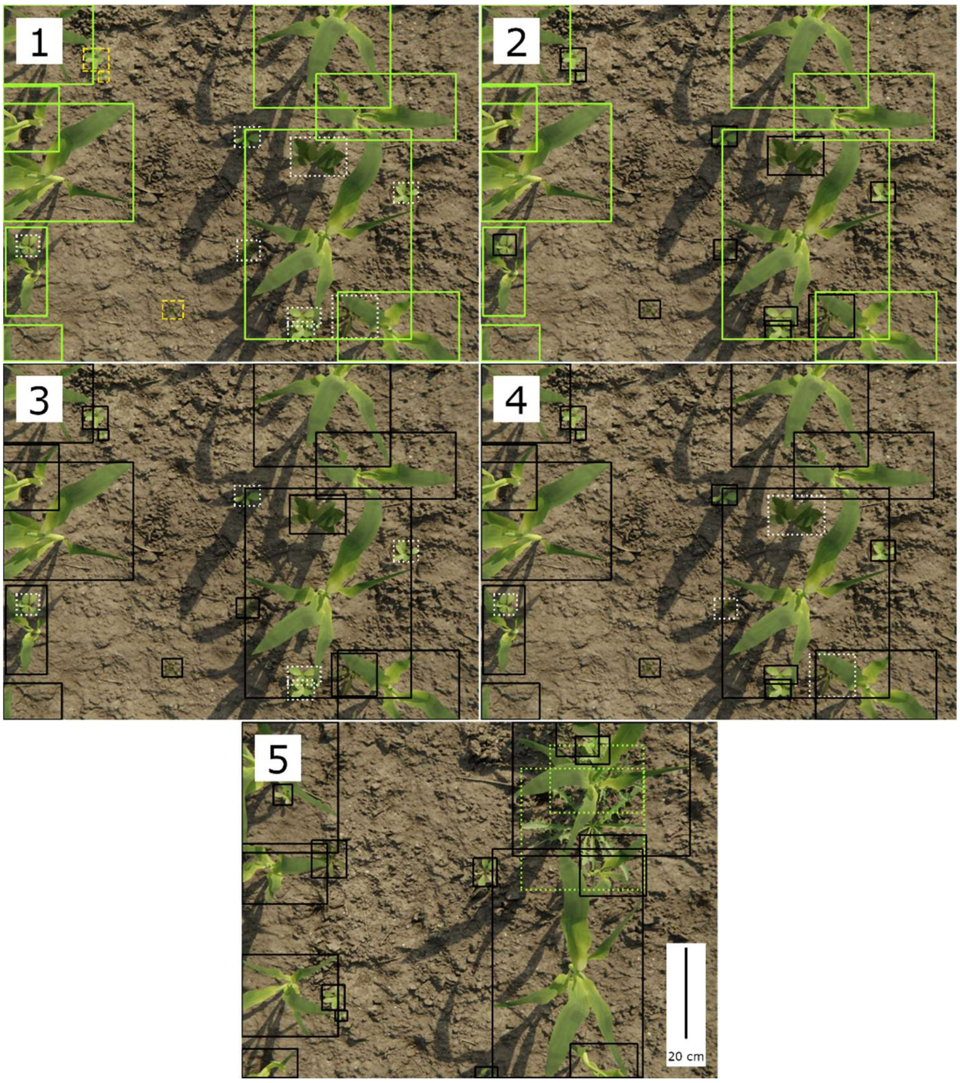

Fig. 2: Result of the rule-based system marking plants as crop (green, solid line), harmless weed (yellow, dashed line), harmful weed (white, dotted line), endangered species (green, dotted line) or as not categorized (black, solid line). Image 1 shows the total result. Image 2 displays the result of the effect of the crop rule, all maize plants are marked with a green bounding box. In Image 3, weeds categorized as harmful by the plants per m² rule are marked white. Image 4 shows weeds marked as harmful by the distance rule. Image 5 displays a different section of the field, showing the Autumn Hawkbit categorized as endangered species.

Based on the plant's species and the respective thresholds (Tab. 1), the rule-based system infers whether a plant is to be categorized as crop, harmful or harmless weed, or endangered species. Plants exceeding the plant count threshold as well as plants falling below the distance threshold are considered harmful weeds, otherwise, they are considered harmless weeds. Maize and Autumn Hawkbit have no threshold values since Maize is listed as a crop and Autumn Hawkbit as an endangered species. The threshold values and the listing as endangered species are exemplary due to the limited availability of other plant models.

After inference, the result is written back into the grid map. Based on this, the application map can be generated, which is not necessary for the example with synthetic data.

For transparency purposes, the resulting category of each rule can be displayed in the original images (Fig. 2).

5 Discussion

The percentages of each plant species categorized as a weed (Tab. 1) give an impression of the functionality of the model-based system.

Maize and Autumn Hawkbit are never categorized as weed at all, as they should not. The other plants, however, are categorized as weeds to different proportions. It is worth mentioning, that stricter thresholds lead to a higher number of weeds within a species. The Common Daisy for example has the least strict thresholds and therefore only 48 % of all Daisies are categorized as harmful weeds. The Curly Duck on the other hand shows the strictest rule with 1 plant per m² as threshold, thus all Curly Ducks are categorized as harmful weeds.

This was expected, but it confirms the effect the rules have on the categorization and therefore the overall functionality of the system to mark some weeds as worthy of maintaining.

The rules and thresholds must be evaluated carefully though. While keeping individual weeds is beneficial for biodiversity, this must not result in significant crop losses.

The quality and design of the plant detection and classification model also limit the effectiveness of the system. The design of the rules is dependent on the types of plants the classification model can identify. Nevertheless, the presented system works with any kind of plant perception if the detections are stored in SEEREP in the presented manner. A further constraint for this would be precise plant localization, so plants appearing in multiple images can be fused to a single instance.

The current implementation can be used by any precision weeding system that works on in advance generated application maps but not in an online fashion, where plant detection, categorization, and removal are done simultaneously.

The visualization of each rule effect (Fig. 2) provides an intuitive way to make the internal processes of the system more transparent. It must be mentioned though, that the presented images were ground truth data and that a deep learning-based plant perception system adds a non-transparent black box to the system.

6 Conclusion

A model-based system for plant-specific automated weed regulation is presented which formulates background knowledge about the damage potential of weeds as rules. The system infers plant-specific information about weeds and crops, categorizing plants as crops, harmful or harmless weeds, and endangered species.

The presented rule set does not integrate much background or domain knowledge, but the rule base and the system architecture can easily be extended. Thus, additional rules and information sources such as weather, data about the agricultural site, soil analysis results, or geographical data can be added, depending on the recommendations in integrated weed management. This work is providing a pipeline, which enables the integration of sensor data as well as symbolic knowledge for integrated weed management.

Further steps include integration and testing into a working perception and weed regulation pipeline, expanding the rule set and the utilized background knowledge, and extending the explanatory potential of the rule-based system.

Acknowledgements: The DFKI Niedersachsen (DFKI NI) is sponsored by the Ministry of Science and Culture of Lower Saxony and the VolkswagenStiftung. This work is supported by the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) within the CognitiveWeeding project (grant number: 67KI21001B).

Bibliography

[Ab20] Abbas, I. et al.: Different sensor based intelligent spraying systems in Agriculture. Sensors and Actuators A: Physical 316, p. 112265, 2020.

[Al21] Alengebawy, A. et al.: Heavy Metals and Pesticides Toxicity in Agricultural Soil and Plants: Ecological Risks and Human Health Implications. Toxics 3/9, 2021.

[B.o.J.] B. Foundation: blender.org. Home of the Blender project - Free and Open 3D Creation Software. https://blender.org/, accessed 20 Oct 2022.

[Co20] Colas, F. et al.: Co-development of a decision support system for integrated weed management: Contribution from future users. European Journal of Agronomy 114, p. 126010, 2020.

[Co21] Colbach, N. et al.: The FLORSYS crop-weed canopy model, a tool to investigate and promote agroecological weed management. Field Crops Research 261, p. 108006, 2021.

[Eu20] White Paper on Artificial Intelligence. A European approach to excellence and trust, Brussels, 2020.

[Gh18] Ghosh, I. et al.: Artificial Intelligence in Agriculture: A Literature Survey. International Journal of Scientific Research in Computer Science Applications and Management Studies IJSRCSAMS 3/7, pp. 1-6, 2018.

[Hu20] Hussain, N. et al.: Design and Development of a Smart Variable Rate Sprayer Using Deep Learning. Remote Sensing 24/12, p. 4091, 2020.

[LAN21] Liu, J.; Abbas, I.; Noor, R. S.: Development of Deep Learning-Based Variable Rate Agrochemical Spraying System for Targeted Weeds Control in Strawberry Crop. Agronomy 8/11, p. 1480, 2021.

[Le13] Leuschner, C.: Vorwort. In (Meyer, S., Hilbig, W, Steffen, K., Schuch, S. Ed.): Ackerwildkrautschutz – eine Bibliographie. Ergebnisse aus dem F+E-Vorhaben (FKZ 3512 86 0300). BfN Bundesamt für Naturschutz, Bonn, 2013.

[Ma21] Masson, S. et al.: Neue Entscheidungshilfen für eine nachhaltige Unkrautbekämpfung. Agrarforschung Schweiz 12, pp. 78-89, 2021.

[MRS11] Mercy Nesa Rani, P.; Rajesh, T.; Saravanan, R.: Expert Systems in Agriculture: A Review. Journal of Computer Science and Applications 1/3, pp. 59-71, 2011.

[No12] Norsworthy, J. K. et al.: Reducing the Risks of Herbicide Resistance: Best Management Practices and Recommendations. Weed Science SP1/60, pp. 31-62, 2012.

[NPH22] Niemeyer, M.; Putz, S.; Hertzberg, J.: A Spatio-Temporal-Semantic Environment Representation for Autonomous Mobile Robots equipped with various Sensor Systems: 2022 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI). IEEE, pp. 1-6, 2022.

[Oe06] Oerke, E.-C.: Crop losses to pests. The Journal of Agricultural Science 1/144, pp. 31-43, 2006.

[PCA19] Partel, V.; Charan Kakarla, S.; Ampatzidis, Y.: Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence. Computers and Electronics in Agriculture 157, pp. 339-350, 2019.

[Re22] Red Hat: Drools. https://www.drools.org, accessed 20 Oct 2022.

[va21] van der Waa, J. et al.: Evaluating XAI: A comparison of rule-based and example-based explanations. Artificial Intelligence 291, p. 103404, 2021.

[We14] Weiss, M., Weiss, K., Krebs, S.: Ackerwildkräuter am Württembergischen Riesrand. In (Naturschutz und Landschaftspflege Baden-Württemberg Band 76. Systemedia GmbH Ed.): Landesanstalt für Umwelt; Messungen und Naturschutz Baden-Württemberg (LUBW), pp. 112-151, 2014.