Popular written summary for **Estimating maximum reservoir volume with satellite data and machine learning***

Determining how much water a reservoir can store is challenging because direct measurements are difficult and expensive to obtain. Reservoirs are surface water bodies that are often associated with dams but can also include lakes. They play a crucial role in society and the environment by providing recreational opportunities for people and supporting biodiversity. Moreover, knowing the maximum reservoir volume is important for hydrological modeling, adaptive water management, e.g., for agricultural irrigation and power generation, and for risk assessment, e.g., when a dam is in danger of breaking.

Direct in-field measurements can measure water depth in a regular grid, but are related to high time and financial costs. These data are often lacking or not publicly available. Instead, over the years, different approaches were developed to estimate the reservoir volume. Traditional approaches are based on geometrical approximation, for instance, fitting the reservoir to a bowl shape (ellipsoid method) and calculating the reservoir volume by applying a geometrical formula. More recently, machine learning models were developed. Machine learning methods can learn from examples. By analyzing thousands of reservoirs, they can identify patterns that help estimate the maximum volume of other reservoirs. Unlike traditional methods, they do not require assumptions about the reservoir's shape and can learn various kinds of relationships.

In this study, the aim was to estimate the maximum reservoir volume as a single value by creating a machine learning model. The model used images containing information about the terrain surrounding reservoirs and the maximum extent of the water body. The model was created using around 5,200 reservoirs worldwide.

Throughout the process, various model settings were tested to support optimal learning. At the end, the model with the best performance was selected and applied to 1,050 reservoirs from a reference dataset and 16 reservoirs in New York with field measurements. The model produced reasonable estimates for reservoirs worldwide and showed similar performance when tested on independent reservoirs with field measurements. The results indicate that the model successfully learned relationships between landscape characteristics and reservoir volume, although uncertainties remain. Some of the uncertainty may originate from the training data itself, as many of the reservoir volumes used to build the model were estimates rather than direct field measurements. It can be expected that the model accuracy could be improved with a higher quality of the volume data on which the model was created.

Moreover, it was analyzed whether the model performance varies with reservoir characteristics. The deviation between the estimate and the reference volume in the dataset tended to increase with increasing volume. Apart from that, no clear relationship was visible.

Lastly, the machine learning model was compared to a model based on the geometrical approximation of a bowl shape. The machine learning model was closer to the reference volumes and showed, particularly for larger reservoirs, more accurate volume estimates.

Overall, the results demonstrate that satellite data and machine learning can be used to estimate maximum reservoir volume worldwide, providing a practical alternative where direct measurements are unavailable or difficult to obtain.

*Original title: **Estimation of maximum water storage in reservoirs using satellite data and machine learning methods**