AI for Hydrologic Feature Extraction

Questions and Answers

1. **What licenses are required for the purpose of today's webinar? Is image analyst sufficient?**

Spatial Analyst and Pro Advanced.

2. **Is it possible to use arc hydro with an ArcGIS university account?**

Yes. Arc Hydro is a free add-on for anyone with access to ArcGIS Pro. We recommend using Pro 3+ and the accompanying Arc Hydro version https://www.esri.com/en-us/industries/water-resources/arc-hydro/downloads#arc-hydro-for-arcgis-pro

3. **Do you stretch continuous indices such as TWI and TPI to unsigned 8-bit 0-255? Do you perform any transformations on the categorical data, namely Geomorphons?**

We do not perform any transformations for the predictor variables prior to model training with the UNET, ConnecNet, or MTRE model architectures. According to our source code, inputs are automatically normalized for these model architectures. However, this is not true for all models that esri offers. Users should confirm whether their model requires normalized inputs. If so, this step should be completed before Export Training Data for Deep Learning.

I also set transforms = False when running the Python API for ArcGIS prepare_data() function. Since my predictor variables represent physical geomorphology, I do not want the model to change pixels values to augement training data in the same way it would change pixel values for satellite imagery (e.g., brightness).

4. **Can you elaborate how the geomorphon tool is different from just deriving the slope of a given area?**

The Geomoprhon Landforms tool uses a robust algorithm that considers both slopes and visibility concepts. Slope alone provides elevation differences between immediate neighborhoods. Geomorphon Landforms determines a neighborhood by creating a viewshed based on a specified number of proximate cells to look past and a search distance that dictates how far outwards to look. Once elevation differences are calculated within the neighborhoods, Geomorphon Landforms also classifies the information into discrete landform types. This results in recognizable categories rather than a raw index.

5. **In the slide that showed testing streams vs. predicted streams, what would be the best approach to improve the accuracy of the predicted streams?**

Providing a model with more training data may improve accuracy. More training data can provide additional examples of how streams may appear, and therefore give your model a more robust understanding of streams. You can also reconsider your predictor variables. If you have reliable landcover data, you could include that raster as another predictor variable, for example. It is also possible that your model would perform better if it trained for more epochs, which you can specify during model.fit().

6. **How would you compare the efficacy of deep learning for stream mapping with using flow accumulation thresholding as in ArcGIS Pro's Derive Stream As Line tool?**

Answer

7. **For the pre-trained models do the inputs have to be from the same data source that the model was trained on?**

Yes. Users must match the data source that was used to train pre-trained models. This information can be found in the Item Details page of any of the Living Atlas models.

8. **For the pre trained land cover model is that provided across the US including Hawaii?**

The pre-trained land cover model should work with any Sentinel-2 image, which includes global coverage. The Item Details page for this model specifies that it is expected to work well in the US and Europe. Since the landscape in Hawaii is vastly different from CONUS, incorrect predictions can occur. I would recommend creating a small set of your own training data in your area of interest and using that to fine-tune (i.e., further train) the Living Atlas model.

9. **Do you have any AI extraction tools of this sort, specifically for Planet Scope Imagery?**

There may be pre-trained models that are well-suited to PlanetScope imagery. To verify, check the Item Details page for a model's expected number and type of imagery bands and resolution. There are models that are trained on high-resolution imagery.

10. **Are there any publications, technical reports or other published assessments of stream network delineation methods and accuracy metrics?**

Dr. Passalacqua is a prominent researcher in this field. She has authored multiple papers on methods to extract streams from geomorphology. I recommend reviewing her work and digging into researchers that are citing her work to support machine learning approaches to stream extraction.

Two publications to start with:

https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2009WR008812

https://www.sciencedirect.com/science/article/abs/pii/S1364815216301219

11. **It is possible to include infiltration values for soil or the process only includes pervious, impervious and water categories?**

Yes. Mechanically, any raster representation of the physical landscape can be used as a predictor variable, as long as its cell size and extents match the model configuration. However, it is important to assess inputs, at least visually, before including them in your model. Widely-available soil data is often too coarse to be a useful predictor. Visually, this may look like a single soil type covering and entire watershed. An alternative is to use proxies for soil characteristics, which TWI and DTW are for near-surface soil moisture. In general, evaluate whether other inputs are capturing your prediction target adequately.

12. **If you are familiar with the work of Mainali et al (2023), have you tried reproducing their results using NAIP and seasonal composites from Sentinel-2? Their research suggests that good results can be obtained without lidar data, elevation models, or derived indices.**

Yes, we are experimenting with reproducing their workflow. They show promising results. It would be interesting to see how their method and accuracy scales to other areas where NWI data may not be reasonable for training and testing.

13. **Is there a water body size limit to what is able to be detected? Curious if this remote sensing technique could pick up smaller / <1 acre inland lakes.**

The minimum output mapping unit depends on the input data resolution. For model configurations that use predictor variables derived from 3m x 3m cells, the smallest feature that can be detected will be a single cell (3m x 3m). If you are targeting relatively small features, it is important to use input data with sufficient horizontal accuracy.

14. **Can this be used with a time series to determine wetland resiliency?**

By introducing satellite data from varying years and seasons, you could create temporally fluctuating wetland predictions. Additional analyses would be needed to translate that to wetland resiliency. Esri offers space-time cube capabilities that may be beneficial https://spatialstats-analysis-1.hub.arcgis.com/.

15. **What is the resolution of the DEM used in the WIM example?**

We used 3m DEMs for the WIM demo.

16. **What spatial resolution does geomorphon produce its results at?**

The geomorphons layer showed in our demo was derived from a 3m DEM. This DEM was also smoothed using the Perona-Malik method before deriving geomorohpons. This smoothing method has been shown to reduce microtopographic noise while preserving geomorphon boundaries.

17. **Is WIM able to train/predict different wetland classes, e.g., bog, marsh, etc.**

WIM is mechanically equipped to train for and predict multiple classes, rather than binary classes alone. However, predicting wetland classes is a more complicated task compared to wetland vs. nonwetland. Users would likely need imagery and vegetation data at a high resolution to be able to differentiate between wetland classes. The target classes would also need to be included in the training dataset.

18. **What resolution DEMs do you recommend for WIM and the deep learning method for wetland identification?**

We have found that wetland predictions derived from 3m DEMs performed better than those derived from 1m DEMs. The DEM resolution dictates the scale for which hydrology is being modeled. The best resolution may vary based on the prediction target.

19. **How can one use Archydro to automatically create multiple pourpoints as ArcSWAT does?**

Basic Arc Hydro terrain processing generates catchments based on the specific stream threshold limit. During the process, a set of “drainage points” can be created that represent the outflow points from each catchment (these are catchment pour points). Arc Hydro document “Arc Hydro – Support for Hydrologic Modeling” has additional discussion on building foundation for hydrologic modeling support using Arc Hydro tools.

20. **On average, how many training samples should you use for random forest machine learning versus deep learning?**

More training data will always be better than less training data and figuring out the best trade-off training configuration is part of the model development process. I recommend 1 sq. Km of known wetland and nonwetland data per unique landscape for Random Forests. Deep Learning would perform better with three times that. It really depends on how representative any training data is of variants of your target. It is hard to give a definitive answer for all applications. Hopefully, these initial numbers can get you started!

AI for Hydrologic Feature Extraction

21. **Can you apply the stream prediction model for polar regions? Specifically on ice sheets?**

Areas with glacial influence will have a more complex subsurface than is assumed for Topographic Wetness Index. It is best to approach that use-case with through the lens of a hydrologists and determine what descriptors of streams are valid in that region, and how you can derive those. Once fitting predictor variables are available, the rest of the model mechanics are the same.

22. **Is it possible to make a model builder to streamline the train/predict steps?**

Yes. All of our geoprocessing tools can be brought into a model builder. We do recommend automating through a python script instead for added flexibility.

23. **What accuracy estimator for a model based on ML/ Deep learning is the best for hydrological workflows?**

We rely on precision, recall, and the precision-recall curve to assess model accuracy. These metrics are more representative of model performance for anomaly detection. This is important given that wetlands and streams typically occur less frequently in a landscape.

24. **Do the pre-trained layers work for places outside of the continental US? Like in Alaska/Hawaii?**

See answer to Question 8.

25. **Is there an option to generate a probobalistic output rather than a descrete output? It might be more informative to see a continuous probobalistic raster.**

Yes. The Random Trees algorithm implemented in WIM will output class probabilities in addition to the discrete classification.

26. **I've struggled with training my own models, could you send some documentation/help on that?**

Samples like the one found here provide steps for executing an end-to-end application.

27. **Small headwater streams represent most of the network in the landscape. Are you developing any tools specifically for headwater analysis and prediction? Such as algorithms for the prediction of channel initiation and flow permanence.**

We are not explicitly investigating this. Perhaps a model could differentiate flow permanence given flow accumulation as a predictor variable.

28. **Could stream permanence be estimated with these tools?**

See answer to Question 27.

29. **Is the stream output layer, polyline or polygon, if it is polyline, is there a way to use this tool to identify streams as polygon that can detect the boundary between inbank and outbank?**

The stream extraction workflow will output raster pixels that are predicted to belong to the "stream" class. Postprocessing could be added to shrink or grow groupings of these predicted pixels based on geomorphology. The "lidar" python package could be a good option to "snap" predictions to hydrologic contours. Keep in mind that postprocessing will be limited by obscured bathymetry in DEMs.

30. **Your examples seemed to be watershed boundary bound. Your inputs seem to be covering all of a previously known watershed area. Does the model work better when you work on a watershed bases?**

We recommend setting processing extents for inputs to watershed boundaries. We typically do so for HUC12 boundaries, but the smallest watershed that encompasses your training data will work. Deriving hydrologic indices in watershed units ensures that hydrologic connectivity in maintained.

31. **How do you measure accuracy with each options?**

See answer to Question 23.

32. **Can this tool help identify stream and riparian widths?**

See answer to Question 29.

33. **I'm having trouble finding pre-trained models in Living Atlas/ArcGIS Online. Can you share a link?**

https://livingatlas.arcgis.com/en/browse/?q=dlpk#d=2&q=dlpk

34. **Can the Arc Hydro AI tool be used to determine, for example, some similarity index between watershed characteristics (land use, soil characteristics, etc) of different say, HUC12s?**

I think there are indirect ways to do this. Differences in modeling accuracy between different HUCs may indicate different watershed characteristics.

35. **What will be the accuracy in result of AI calculations to predict or prepare the hydrogeological model of a particular site??**

Accuracy is highly dependent on the quality of training data and how well predictor variables represent a landscape for your specific application. Our goal in modeling is not to achieve perfect results (if the model is perfect you probably don't need a model to find that feature), but rather to provide a useful screening or guidance to subsequent data refinement. Our applications have reached wetland recall of 80+% but with some degree of overestimation (wetland precision of 50%). This is not to say that other model configuration cannot perform better or worse.

36. **Are there good AI models/methods being researched for identifying barriers to stream networks (and connecting to known water features)?**

We are actively researching how to incorporate built drainage into our hydro feature extraction workflows. Since features like culverts are not modeled in a DEM, flow accumulation data can overestimate pooling in areas. No answers here yet, but we're working on it!

37. **Can I add my own data into a pretrained deep learning model and retrain it?**

Yes, and that is what I find to be the most exciting application of our pretrained models. You can consider the pretrained models to be starting points for your model development, and fine-tune them with training data specific to your application and area of interest. See this blog for tips on how to get started.