Overview of the junction angle dataset extraction procedure

This global junction angle dataset was derived from a global channel network extracted from NASA's 30 m resolution void-filled Shuttle Radar Topography Mission Digital Elevation Model Version 3 (SRTM-DEM) (Farr et al., 2007). Initially, the SRTM-DATA was compiled into a DEM extending between 60°N and 56°S. High arctic regions and Antarctica were excluded from analysis. Computational limitations demanded tiling this DEM prior to channel extraction. To limit the subdivision of basins we used buffered basin outlines from the global HydroBASINS dataset (Lehner et al., 2013) as 'cookie-cutters' and clipped out a 30 m resolution SRTM-DEM tile for every individual basin in the HydroBASINS dataset larger than 10 km².

Based on centroid coordinates each basin tile was projected into the relevant Universal Transverse Mercator (UTM) zone prior to hydrological processing. The largest of Earth’s drainage basins span several UTM zones, resulting in the distortion of drainage areas at the longitudinal extremities of the largest basins. To avoid this problem we subdivided any basins with a drainage area greater than 200,000 km². This was achieved by resampling the basin DEM tile to 250 m resolution (to facilitate efficient processing) and using the open-source topographic analysis software LSDTopoTools (https://github.com/LSDtopotools) to extract all complete sub-basins with drainage areas between 10 and 200,000 km². The output sub-basins were then used as 'cookie-cutters' to clip 30 m DEM basin tiles. These tiles were then buffered and processed alongside the input HydroBASINS basin outlines.

To prepare for hydrological flow-routing, each basin DEM tile was pre-processed using the hybrid filling and carving algorithm of Lindsay et al., (2016). To avoid treating large internally-drained basins as topographic depressions when 'filling' the DEM, a hydrological sink was burned into the DEM at the location of the lowest point in every substantial internally-drained basin. The sink locations, taken from a global dataset of endorheic basins (Lee et al., 2013), were originally derived from the HydroSHEDS global channel network dataset (Lehner et al., 2013), with each sink manually verified.

Hydrological flow routing and channel extraction were then performed on every basin tile using LSDTopoTools (https://github.com/LSDtopotools). When extracting channel networks, channels were initiated at a threshold drainage area of 1 km².

Once a channel network was generated, LSDTopoTools was used to automate the measurement of every junction angle and the drainage area, length and slope of every network link. The LSDTopoTools junction angle measurement tool works by performing orthogonal linear regression through all the points that together form a channel link. Junction angles are then measured as the angles formed between the regression lines fitted to each of the three links meeting at any junction. We remove junctions where any of the three segments (e.g., two tributaries and a receiver segment) have fewer than 8 pixels.

Artefacts in junction angle geometry can arise from the application of flow routing algorithms across relatively flat, low-gradient landscapes. To remove these artefacts from the dataset, any junctions composed of a channel segment with a gradient less than 0.0001 (1 m elevation loss in every 10 km) were omitted from analysis.

A global dataset of aridity (Trabucco et al., 2019) was used to assign every junction an aridity index (AI) value based its location. The aridity index is the ratio of mean annual precipitation to mean annual evapotranspiration, with lower values indicating progressively more arid climates. To avoid
the analysis of junctions where perhaps no real channels exist due to lack of any streamflow accumulation, all junctions located within 'hyper-arid' regions (aridity index < 0.03) were removed from the dataset. These regions are predominantly located in the deserts of North Africa, the Arabian Peninsula and Central Asia (Trabucco et al., 2019).

For a detailed summary of our junction angle dataset extraction methods please refer to the Fig. 1. The software package we used to perform this analysis is freely available at https://github.com/LSDtopotools. The following sections contain further details, discussion and justification of our extraction parameters choices.

**Basin outline buffering procedure**

We used HydroBASINS basin outlines as 'cookie-cutters' to clip out DEM tiles to hydrologically process from the Shuttle Radar Topography Mission Digital Elevation Model (SRTM-DEM). The HydroBASINS dataset was derived from a resampled (90 m resolution) version of the 30 m SRTM-DEM. As such we expect basin boundaries extracted from the 30 m SRTM-DEM to be generally aligned with the HydroBASINS basin boundaries. However, as HydroBASINS outlines are provided at a relatively coarse 15s (~450 m at the equator) resolution, there are spatial discrepancies between basin boundaries extracted from 30 m SRTM-DEM data and the HydroBASINS basin boundaries at local scales. To account for these discrepancies and to allow for the extraction of fully intact basins, we buffered the edges of each basin tile prior to hydrological processing. For tiles with areas less than 100 km² this buffer was fixed at 2 km. For tiles with areas between 100 km² and 15,625 km² the buffer radius (R) dependent on the area of the tile (A) as according to:

\[ R = 0.25 \sqrt{A} \]

For tiles with areas greater than 15,625 km² R was fixed at 25 km. This buffering ensures that when flow routing was performed on the 30 m SRTM DEM basin tiles the great majority of basins have no truncation of basin headwaters at the tile margins. There are rare instances where spatial discrepancies between our extracted basin boundaries and the HydroBASINS basin boundaries result from differences are significant enough to extend beyond the buffered zone. Given these disagreements highlight significant discrepancies in flow routing paths, all junctions located on pixels draining from such areas were rejected from the dataset as a precaution against measuring incorrect drainage areas.

**Selecting a channel extraction threshold**

When extracting channel networks from topographic data the location of channel heads strongly controls the density and topology of the resulting network. While it is possible to apply algorithms to extract channel heads from topographic data (Clubb et al., 2014) they require high pixel resolution and considerable processing time - constraints which made this approach unfeasible in our analysis. Instead we opted to use a simplistic threshold approach, whereby channels are initiated at a threshold drainage area.

What constitutes an appropriate drainage area threshold for channel initiation? Using a threshold that is too low may result in the extraction of false junctions where no channels exist. Using a threshold that is too high would result in fewer extracted junctions, sacrificing statistical confidence.
The scale at which junction angles are measured is also controlled by the extraction threshold as channel links will be longer when this threshold is higher.

In 21 field studies of channel heads (Wohl et al., 2018) the largest recorded initiation threshold area for a channel was 0.8 km$^2$ with the great majority of channels initiated below a threshold of 0.2 km$^2$. We opted to use a relatively conservative threshold for channel extraction of 1 km$^2$.

**Hydrological flow-routing**

For surface flow-routing algorithms to properly extract drainage networks input topographic data should be free of voids and topographic depressions - these features behave as hydrological sinks, terminating extracted channels and falsely depriving downstream channels of drainage area. To avoid voids we used SRTM-DEM Version 3 (Farr et al., 2007) which is void-filled, having been patched with alternative data where voids in the original data existed. Although topographic depressions are not artefacts, it is standard practice when preparing DEM’s for hydrological applications to assume flow into the depression has a hydrological connection to the surrounding watershed. DEM processing to maintain hydrologic connectivity can be accomplished by either filling depressions, breaching obstructions, or some combination of the two. For this study we used the hybrid breaching and filling algorithm of Lindsay et al. (2016) with the slope of filled surfaces set to 0.0001. Employed on the void-filled SRTM dataset, this algorithm results in a processed DEM wherein each pixel has a continuous downstream flow-path that terminates at the DEM margin. A weakness of this approach is that anywhere depressions are filled, planar low-gradient surfaces are artificially generated. This results in an artificial flow network pattern through filled pixels. We mitigate this effect in by explicitly accounting for large internally drained basins.

Filling errors are most problematic in internally drained (endorheic) basins. When fill algorithms are applied, large areas of completely artificial channel networks are generated. Due to the filling process these channels effectively spill over into neighbouring basins causing drainage area errors to cascade downstream. Endorheic basins constitute approximately 18 percent of the Earth’s land surface. To address the endorheic basin problem we generated an artificial hydrological sink at the location of the lowest elevation point within every significant internally drained basin, preventing the basin from being algorithmically filled. This was achieved using a global dataset of endorheic basins (Lee et al., 2018) derived from the HydroBASINS dataset (Lehner et al., 2013).

In lowland regions DEM-based extraction of river networks may result in poor matches between the true location of channels and DEM-based channel locations due to filling and errors in flow routing in low relief terrain. The data can be filtered based on a slope threshold to remove junctions with considerable flow-routing uncertainties.

Using the SRTM-DEM data for hydrological processing presents various other problems related to how the DEM was derived and what the data is capable of representing. SRTM-DEM is a digital surface model (DSM) which does not distinguish between the real land surface and other natural/artificial surface features. Features such as urban developments, dense forests, and hydro-power/reservoir developments are represented as topography in the SRTM-DEM and these structures can result in the misrepresentation of real hydrology. Similarly, poor-quality flow routing can occur where rivers are confined in gorges narrower than the DEM pixel resolution (30 m) or where drainage does not occur at the land surface (e.g. karstic landscapes with subterranean rivers). No attempt was made to mitigate these effects in the channel network extraction and poorly represented channels may impact the location and geometry of some extracted junctions. The very
large number of junction angles in the dataset (25,913,054) are assumed to statistically mitigate for any locally misrepresented hydrology.

**Junction angle measurement**

Once the DEM is preprocessed, including steps to avoid artefacts caused by endorheic basins (see above) and continuous flow paths are computed using the hybrid carving and filling algorithm \cite{lindsay_efficient_2016}, we use the resulting flow network to extract the junction network. Junctions each have a receiver junction (base level junctions' are their own receivers) and two or more donor junctions. We reject any junction that has more than two donor channels (this is extremely rare), and base level junctions. For each junction, we then add the location of every pixel along the channel pathway from the two donor junctions to the junction in question, and from that junction to the receiver junction. Base level junctions are ignored. The bearing of each of these three segments is calculated based on orthogonal regression (that is, for the set of channel pixels in each of the three segments, a line is fitted that minimises the sum of squared perpendicular distances from the data points to the regression line). These bearings are then used to calculate the junction and bending angles using standard vector calculations. The spacing between junctions is determined by the channel extraction threshold: smaller thresholds will have greater network density and thus more closely spaced junctions and shorter channel segments. We use channel segments between junctions rather than segments made up only of channel pixels near the junctions because we are interested in capturing the structure of how channels drain the entire landscape rather than focusing on the hydraulics near the confluence of tributaries. Any junction with a channel segment composed of less than 8 pixels is rejected. This avoids the inclusion of junctions dominated by local hydraulics and, more importantly, provides sufficient degrees of freedom for angular calculations from the linear regression to avoid the production of systematic values.

**Dataset information**

Dataset headers: **latitude** (decimal degrees), **longitude** (decimal degrees), **donor1_stream_order** (Horton-Strahler stream order of first tributary), **donor2_stream_order** (Horton-Strahler stream order of second tributary), **receiver_stream_order** (Horton-Strahler stream order of the channel formed at the junction), **donor1_drainage_area** (m$^2$), **donor2_drainage_area** (m$^2$), **this_junction_drainage_area** (m$^2$), **donors_junction_angle** (°), **donor1_receiver_junction_angle** (°), **donor2_receiver_junction_angle** (°), **gradient_donor1**, **gradient_donor2**, **gradient_receiver**, **ai** (aridity index value from Trabucco et al., (2019)), **AR** (ratio of tributary drainage areas), **AI_class** (as according to Trabucco et al., (2019))

**Acknowledgements**

Funding for this work was provided by NERC grant NE/L002558/1.

**Authors**

Callum M. Strong and Simon M. Mudd, The University of Edinburgh
References


