APPENDIX 3.1 – Occupancy Covariates Justification & Calculations

Lion site use covariates

A total of 17 covariates were hypothesised to influence lion habitat-use at the home range selection (HRS) and/or the temporary use selection within the home range scales (TUS).

Biotic

Although little data is available on lion prey preference in Ruaha-Rungwa, buffalo have been shown to be preferred prey elsewhere in Africa (e.g. Hayward & Kerley, 2005; Hayward et al., 2011), and to be a strong predictor of lion occurrence (Everatt *et al.*, 2014). To quantify the effect of preferred prey availability, I employed the replicated detection/non-detection spoor data from the spoor survey to develop a probability of use model for buffalo, at both spatial scales at which large carnivore habitat use was investigated (HRS & TUS). I assumed that the probability of site use of buffalo was representative of the encounter rate of lion with preferred prey at that site, at that scale.

In addition to buffalo, lions have also been recorded predating on a number of other ungulate species in the study area, including giraffe, zebra, sable, roan, kudu, and impala (Ruaha Carnivore Project, unpublished data). I therefore employed detection/non-detection data from the spoor surveys to develop site use models for each of these species, at both spatial scales. Not all were however included in every sampling grid: species which were considered 'preferred' (i.e. buffalo; based on Hayward et al., 2005) were included in all HRS analyses, while species which were not considered preferred, but were also consumed regularly (giraffe, zebra, sable, kudu, and impala; Hayward, 2005), were included only in the sampling grids focussing on less impacted areas (HR_PAS and HR_NAT) and at the TUS scale. These guidelines were also followed when modelling site use of the other large carnivore species investigated.

Finally, a covariate representing the probability of encounter for lion with any of the above prey species was calculated by averaging the site-use estimates of each at each site ('AllPrey'), and an alternate prey covariate ('SecPrey'), equal to the mean of the probability of use of all species except buffalo, was

calculated for each site. These were included in all HRS sampling grids. See 'Prey site use covariates' for details on the covariates employed to model site use of the different prey species.

In addition to prey availability, lions have been shown to also shape their space use based on a number of other biotic features, including vegetation type and landscape features facilitating prey catchability (e.g. Midlane et al., 2014; Hopcraft et al., 2005). As a result, the availability of riparian habitat at each site was modelled as the mean Euclidean distance of each 30 x 30 m pixel in each site to the nearest river or drainage line, using the Spatial Analyst tool in ArcMap 10.4.1 (a lower covariate value therefore indicating higher availability of riparian habitat within the site).

To quantify cover availability, and test whether lion were more associated with closed habitat or with open areas, tree cover was quantified by converting to a raster an existing 20 m resolution land cover product (ESA CCI Land Cover, http://2016africalandcover20m.esrin.esa.int/), and assigning pixel values of '1' to all pixels within the 'Trees cover areas' and 'Shrub cover areas' categories, and a value of '0' to all pixels within other land cover categories. This method permitted considerably higher (ten-fold) resolution mapping of cover than alternative, more commonly employed vegetation cover products (e.g. MOD44B, NASA LP DAAC, https://lpdaac.usgs.gov), thus being ideal for finer-scale assessments of habitat use.

During surveys, lions were recorded in both of its two major habitat (vegetation) types, miombo woodlands and A*cacia-Commiphora* bushland. However, to account for potential preferences between these, primary vegetation type was modelled by developing a 30 m resolution raster from existing layers (Olson et al., 2001; https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world). Pixels were assigned a value of '1' if in miombo woodland, and a value of '0' if in *Acacia-Commiphora*; a site-specific value of '1' therefore indicates the site is fully within *miombo* habitat, while a site-specific value of '0' that it is located solely within *Acacia commiphora* grasslands/bushlands. A positive covariate effect (+) therefore signifies association with miombo habitat, while a negative effect (-) one with *Acacia-Commiphora* habitat.

Anthropogenic & management

Protected areas provide refuge for large carnivores, both directly from human persecution and indirectly by protecting their prey base, and increased levels of protections have as a result been found to be positively associated with lion occurrence (Henschel et al., 2016). As a result, legal protection was measured as the proportion of each site within a gazetted protected area; existing PA layers (IUCN World Database on Protected Areas, WDPA) were modified based on knowledge acquired from protected area managers, and converted into a 30 m resolution raster. Values of 1 were assigned to pixels within a designated protected area, and values of 0 otherwise.

In addition, strong effective protection was quantified as the proportion of each site within areas of the landscape which received regular protection anti-poaching patrols. Data on this was gathered from surveys and from protected area managers during the survey, and converted to a raster at 250-m resolution. Values of 1 were assigned to pixels within this area, and values of 0 otherwise.

Finally, distance to nearest ranger post was included as a covariate, as it was reasoned that areas in closer proximity to posts would also benefit from increased patrol protection. This was measured as the mean Euclidean distance of each 30 x 30-m pixel in each site to the nearest ranger post using the Spatial Analyst tool in ArcMap 10.4.1.

I also hypothesised that lions in our landscape would avoid agricultural and developed areas, due to direct persecution and lack of favourable habitat (Abade et al., 2019; Schuette *et al.,* 2013). A measurement of natural habitat conversion to agriculture (i.e. cropland) was therefore obtained through Google Earth Grids (Jacobson et al., 2015), a hand-digitised 1-km resolution raster of cropland (a ground-truth exercise revealed this to be more accurate than alternative finer scale products available).

As a result of the high incidence of smaller hamlets, livestock enclosures, and small-scale farms close to PA boundaries in the study area, it was reasoned that distance to PA boundary may be more representative of development-related anthropogenic disturbance outside of PAs than distance to human settlements. Site-specific distances to PA boundary were therefore obtained by converting existing protected area vector layers (IUCN World Database on Protected Areas, WDPA), modified based on

knowledge acquired from protected area managers, into a 30 m resolution raster. The distance of each pixel to the nearest PA boundary was then calculated using the Spatial Analyst tool in ArcMap 10.4.1. Because it was reasoned that the strength of any anthropogenic effect would decrease with distance, the natural logarithm (ln) of distance to boundary was used.

However, I reasoned that employing this covariate in the analyses encompassing the complete landscape (HRS-ALL) would not be ideal, given that levels of development of non-protected areas of similar distance to PA boundaries could vary considerably based on their distances to highly-developed areas. As a result, in the HRS-ALL analysis the effect of anthropogenic disturbance outside of PAs was quantified based on distance to human settlements (villages) instead. Villages and other settlements in the study area were hand-digitised in QGIS 3.6.3 based on GoogleEarth satellite imagery and OpenStreetMap layers, and a 30 m resolution raster of distance to nearest human settlement was generated using the Spatial Analyst tool in ArcMap 10.4.1. As in the case of distance to PA boundary, the natural logarithm (ln) of distance to settlements was used for all analyses, as it was reasoned that the strength of any effect would decrease with distance

In impacted habitat, lions avoid bushmeat poachers and other potentially detrimental human presence within protected areas (Everatt et al., 2015). I therefore decided to model risk of encounter with humans (excluding rangers, tourists, and hunters) in PAs, by developing an illegal human use occupancy model, to obtain site-specific probabilities of illegal human activity. This was restricted to sites located within protected areas, and was carried out at both spatial scales investigated. See 'Illegal human use site covariates' for more details on the predictive variables employed for this.

Finally, trophy hunting has been shown to have the potential to negatively impact lion populations if carried out unsustainably, or if additive to other sources of anthropogenic mortality (e.g. Packer et al., 2011). To test whether trophy hunting affected lion occurrence and persistence in Ruaha-Rungwa, I employed two related covariates. The first was the proportion of a site located within a legally designated trophy hunting area, regardless of whether or not trophy hunting was taking place at the time of surveying. Existing protected area vector layers were converted into 30 m resolution rasters, and a value of '1' was assigned to pixels within legally designated trophy hunting areas (GRs, GCA, and OA),

and a value of ‘0’ to pixels located within areas which did not allow trophy hunting (NP, WMAs, non-PAs). Then, I also calculated the proportion of each site located within a hunting area which was being actively managed (i.e. hunted) by hunting operators at the time f surveying. Of the 14 individual hunting blocks in the landscape, six were not occupied by hunting operators, meaning no trophy hunting took place during the survey period. The same methodology was followed, except that a value of ‘1’ was assigned to pixels within an actively-hunted hunting block, and a value of ‘0’ to pixels located within either non-hunting or vacant hunting areas.

Two additional covariates hypothesised to affect lion habitat-use were initially considered: habitat degradation as a result of over-grazing (based on decreases in land-productivity over time; UNCCD, 2017), and presence of livestock & pastoralists (to be measured by a probability of use model based on detection/non-detection data). However, both were ultimately excluded from all analyses as they were were very strongly correlated with other covariates believed to be more meaningful at the scales researched, such as distance to boundary/village, legal protection, and habitat conversion to cropland.

Leopard & cheetah site use covariates

Most covariates employed to model lion habitat use were also used to model leopard and cheetah site use in Ruaha-Rungwa (see Tables 3.2, 3.3 in the main text).

However, as both species preferentially prey on small-medium ungulates, such as impala (Hayward et al., 2006a,b), buffalo was replaced with impala as a proxy of preferred prey. The alternate prey covariate employed in the lion models (‘SecPrey’) was also excluded, while the covariate for mean ungulate availability (‘AllPrey’) was retained. The selection of which prey species to include in the different grids followed the methodology detailed for lion; as a result, for both leopard and cheetah, kudu and sable were also included in the HRS-PAS, HRS-NAT and TUS analyses (Hayward et al., 2006a,b).

Research has shown that cheetahs attempt to minimise the potential for detrimental interactions with larger sympatric carnivores (e.g. Durant, 2000). Because in the landscape lion densities appear highest along the main rivers (Chapter 5), a 30 m resolution distance to large rivers (defined as all rivers with a

river size score of 1 or 2; see Appendix 3.1.1) matrix was generated using the Spatial Analyst tool in ArcMap 10.4.1.

African wild dog site use covariates

Most covariates utilised to model lion occurrence were retained when modelling African wild dog occurrence across the landscape (Table 3.5 in the main text).

However, impala and greater kudu replaced buffalo as two covariates modelling the probability of encounter with preferred prey (Hayward et al., 2006c). The alternate prey covariate employed in the lion models ('SecPrey') was excluded, while the covariate quantifying mean ungulate availability ('AllPrey') was retained. Additional prey species were included following the methodology detailed for lion site use covariates; sable site use was therefore also included in the HRS-PAS, HRS-NAT, and TUS analyses (Hayward et al., 2006c)

Furthermore, it has been suggested that wild dogs avoid areas with high risk of mortality from lions (e.g. Creel & Creel, 2002); to test this, I included the same distance to large rivers covariate used to model cheetah site use.

Spotted hyaena site use covariates

Most covariates utilised to model lion site use were retained when modelling spotted hyaena site use (Table 3.6 in the main text).

However, as a result of the width dietary breadth of the species (Hayward et al., 2006d), no individual prey species was included as a preferred prey covariate at the HRS scale. Instead, only the covariate representing mean encounter probability with all prey species ('AllPrey') was retained. All prey species were included at the TUS, as they are all at least occasionally hunted (Hayward, 2006).

Prey site use covariates

I used the data collected in this study to develop site-use models for six ungulate prey species (buffalo, zebra, giraffe, roan & sable, kudu, and impala), to be employed as covariates in the large carnivore

models. Since detections for sable and roan antelope were pooled, probability of use by sable should be interpreted as the probability of use by sable *or* roan. For all prey species modelling, the same assumptions were made as for the large carnivores, including that of interpreting the estimator (Ψ) as probability of site use.

Although detection data for eland and elephant were also collected, site use models were not developed for either of these species. Detections of eland were not sufficient, and, while elephants have been known to be predated upon by lions in Ruaha-Rungwa (Ruaha Carnivore Project, unpublished data), it is unlikely that large carnivores would make habitat use decisions at the scales investigated based on elephant occurrence. As a result, they were both excluded from the analysis.

Because ungulate space use is mainly determined by the nutritional value (nitrogen levels) of vegetation, as well as water availability, predation risk, and avoidance of anthropogenic risks (Winne *et al.*, 2008), the following covariates were employed to model heterogeneity in prey site use: an index of primary vegetation productivity (Normalised Difference Vegetation Index, NDVI), water availability, vegetation cover, habitat (vegetation) type, natural habitat conversion to cropland, and distance to human settlement. Although some of these were expected to impact habitat use of some species less than that of others, all were tested for all species, as I expected habitat use to be driven by similar mechanisms. Nevertheless, the hypothesised direction of the effect of some covariates was different to that of the other modelled prey species.

Since both higher primary productivity (mean NDVI) and greater variation in vegetation composition and growth (high standard deviation in NDVI) positively correlate with ungulate occurrence (Oindo and Skidmore, 2002), I generated raster layers for mean and standard deviation of NDVI in the study area. This was done at 250-m resolution, using the MOD13Q1 product from NASA Earthdata (NASA LP DAAC, https://lpdaac.usgs.gov), to obtain 16-day NDVI imagery for both survey periods (July 1-November 30th 2017; July 1-December 16th 2018). These were then averaged, to obtain a single mean and standard deviation NDVI value for each pixel, and a mean site-specific value was then obtained for each sampled site. NDVI was selected over alternative vegetation indices (including: Enhanced Vegetation Index, EVI; Lear Area Index, LAI; and Fraction of Absorbed Photosynthetically Active

Radiation, Fraction of Absorbed Photosynthetically Active Radiation, FPAR) following preliminary analyses, and based on knowledge from ground surveys and satellite imagery.

Availability of cover and habitat closeness/openness, as well as main habitat type (miombo or *Acacia-Commiphora*) were modelled as per the methodology detailed in 'Lion site use covariates'.

Many ungulates species are water dependent, and must drink daily (Estes, 1991). Dry season surface water availability was therefore estimated based on a probability of dry season water availability raster, developed for this study from existing river & drainage lines, hand-digitisation of satellite imagery, and ground surveys. Appendix 3.1.1 (below) provides detailed information on how this was obtained.

Finally, protected areas provide a refuge for ungulates from direct and indirect persecution by humans, and agricultural and developed areas are likely to be avoided. As a result, legal protection, strong effective protection, habitat conversion to agriculture, distance to PA boundary, and distance to village were all quantified using the same procedure defined in 'Lion site use covariates',

Illegal human activity site use covariates

As for prey, site-specific probability of illegal human use estimates were extracted at both scales of interest (HRS & TUS), which are treated as two scales of undefined biological relevance, although the nomenclature is retained for clarity. A single analysis is carried out at the HRS scale (using the same sampling grid employed for the HRS-PA large carnivore analyses), and a single one at the TUS scale (employing all data within PAs), employing the same covariates.

A number of site covariates were hypothesised to influence illegal human activity across the PA complex. Areas farther from PA boundary were reasoned to be more difficult to access, and distance to PA boundary was therefore included as a covariate. This was kept linear (rather than converted to its natural logarithm, as for large carnivores), as it was reasoned the effect would not be asymptotic. Conversely, it was reasoned that areas closer to ranger posts and more strongly protected areas would be avoided, so both strong on-the-ground law enforcement and distance to ranger post were modelled as covariates, measured as detailed above.

I also hypothesised that more densely forested areas would be more attractive for logging, areas with higher prey availability to bushmeat poachers, and increased probability of dry season surface water to both bushmeat poachers and fishers, all of which are present in the study area (TAWIRI, 2019). All three were therefore included as covariates, as described above.

APPENDIX 3.1.1 – Water Availability Digitisation & Calculations

Water availability was initially measured as the mean Euclidean distance of each 30 x 30-m pixel in each site to the nearest river or drainage line using the Spatial Analyst tool in ArcGIS, using the best available rivers & drainage lines layer for the region (HydroRIVERS; https://www.hydrosheds.org/page/hydrorivers; Fig. 3.1.1).

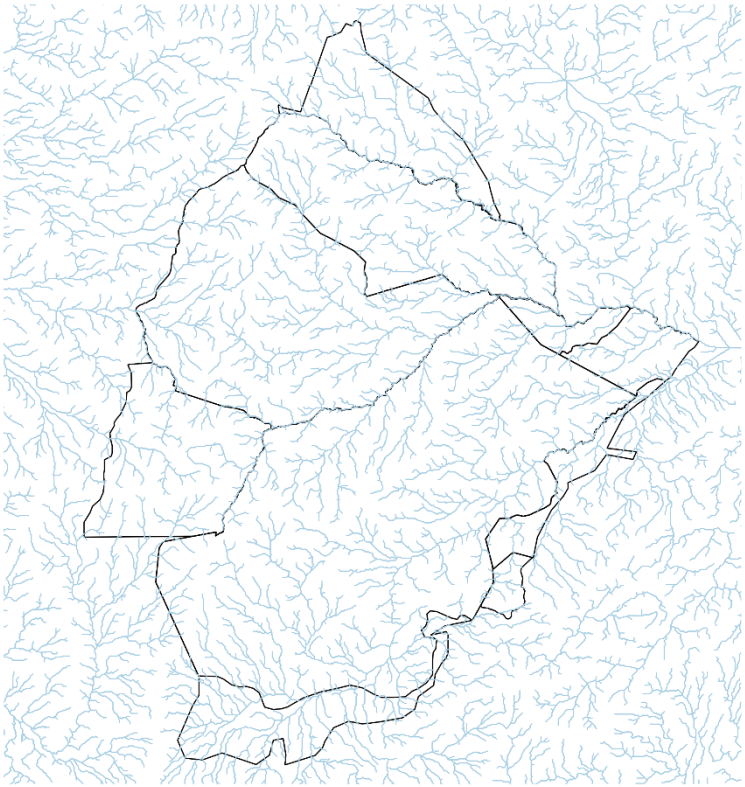

Fig. S3.1.1 Rivers and drainage lines, HydroRIVERS

However, following ground surveys it became apparent that this was not an adequate proxy for water availability, as key sources of dry-season surface water, such as *mbugas* (drainage lines in the *miombo* woodlands), were often excluded. As a result, I modified existing river & drainage line layers by hand-digitising Google Earth satellite imagery in ArcMap 10.4.1. This was achieved by systematically surveying satellite imagery of the study area at an appropriate scale (1:20,000) in ArcMap 10.4.1, and mapping all observed rivers and drainage lines (Fig. 3.1.2). In order to reflect the fact that differently-sized rivers & drainage lines would have different likelihoods of being dry season surface water reservoirs, I assigned each 1-km segment of river or drainage line an arbitrary score based on their

width: 1 = 50+ metres; 2 = 25 to 50 m; 3 = 10 to 25 m; 4 = 0 to 10 m in areas with more than 130 mm of average annual rainfall, 5 = 0 to 10 m in areas with less than 130 mm of average annual rainfall (WorldClim 2.0; Fick & Hijmans, 2017). The scores acted as a proxy for the likelihood to hold dry-season surface water (in descending order), based on water presence records from our on-the-ground surveys.

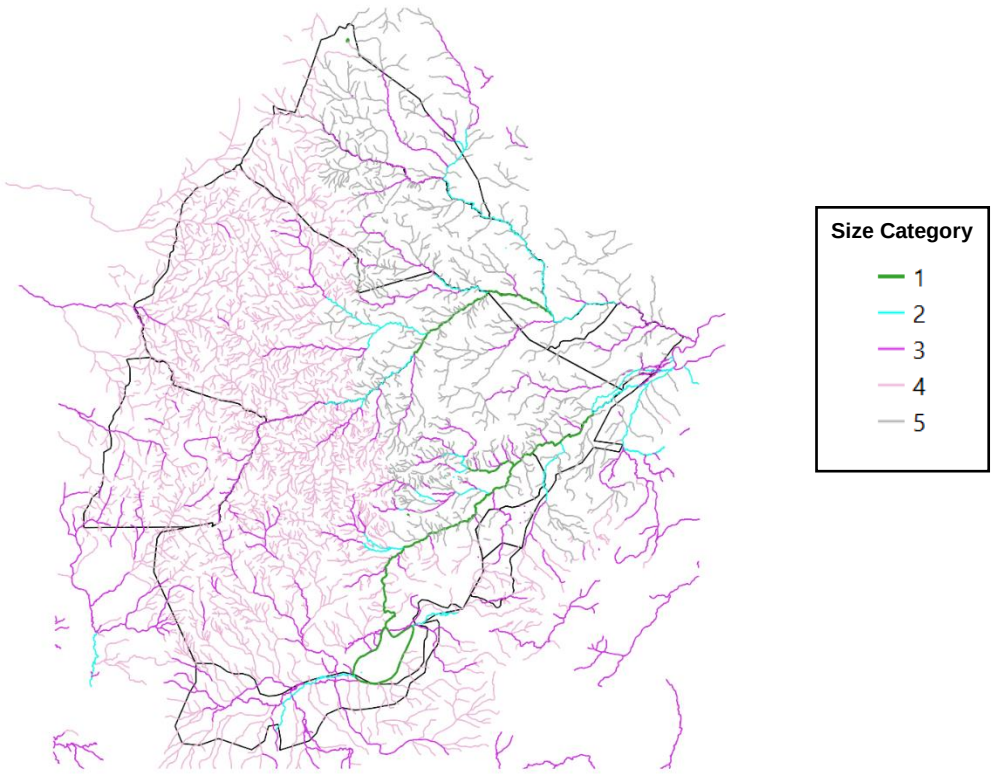

Fig. S3.1.2. Digitised rivers and drainage lines (all size categories)

This was then used to obtain a probability of water availability raster for the study area. This was achieved by first employing the 'Points along geometry' tool in QGIS 3.6.3 to place regular points along the mapped rivers & drainage line, at different densities depending on their size category:

Table S3.1.1. Point density for different categories of river & drainage lines size

Size Category	Point density (distance between points)
1	25m
2	50m
3	100m
4	200m
5	500m

I then used the ‘Heatmap (Kernel Density Estimation)’ tool in QGIS 3.6.3 to produce, from this points layer, a map of probability of water availability across the landscape (radius = 2500m, resolution = 100m; Fig. 3.1.3). Each pixel’s value therefore represents its ‘water value’, based on distances to rivers & drainage lines of different categories of probability dry surface water availability.

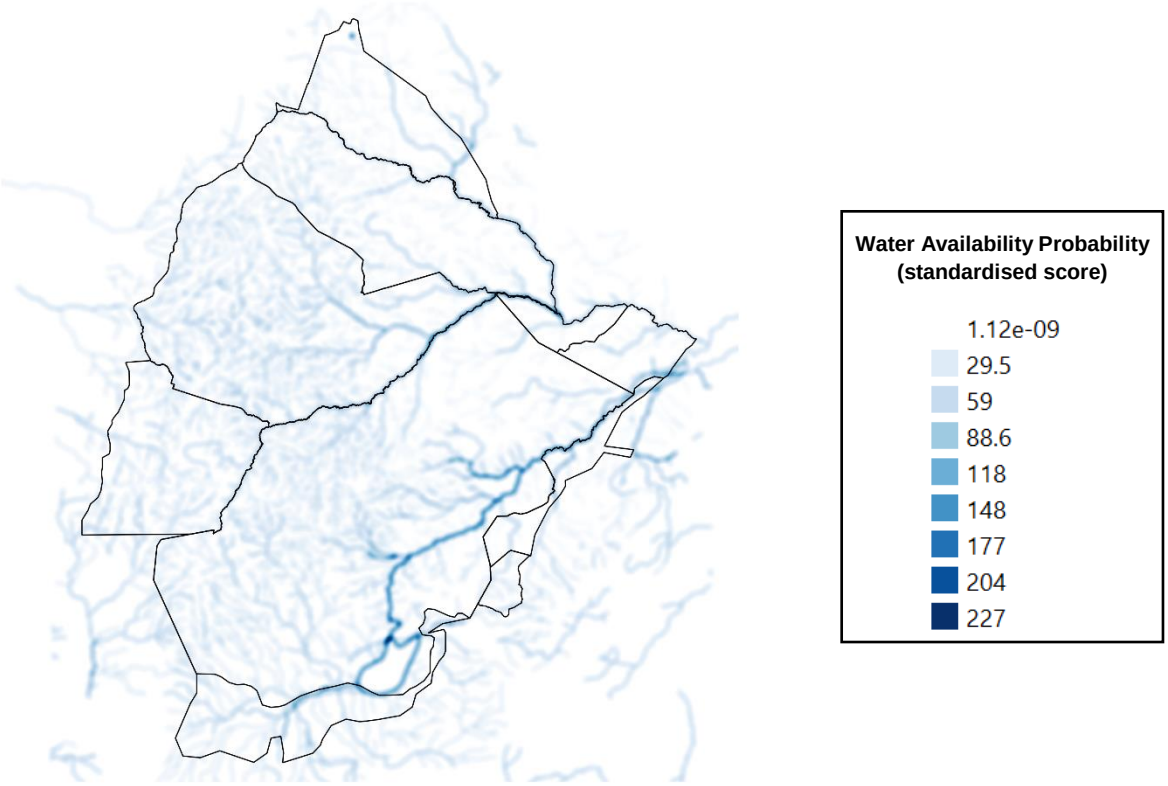

Fig. S3.1.3. Water availability probability raster developed from digitised rivers and drainage lines

Site-specific values were then extracted for each site (at both spatial scales) as the mean value of all pixels within the site, through the ‘Zonal Statistics’ tool in QGIS 3.6.3.