



## An enhanced Pfafstetter catchment reference system

**J.L. Stein<sup>1</sup>**

<sup>1</sup> The Fenner School of Environment and Society, Australian National University, Canberra, A.C.T., Australia.

Corresponding author: Janet Stein ([Janet.Stein@alumni.anu.edu.au](mailto:Janet.Stein@alumni.anu.edu.au))

### Key Points:

- Five modifications to the widely adopted Pfafstetter system for continent-wide catchment analysis are described and tested
- Altering the stopping rule for basin decomposition had the greatest impact on both the size and numbers of sub-basins
- Other enhancements enable coding of the topological relationships of distributary streams and endorheic basins

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2018WR023218

## Abstract

The Pfafstetter system delineates and codes hierarchically nested catchments. Its simple coding scheme denotes stream network position, enabling systematic analysis of the impacts of any activity on a river section without need for a GIS. However, widespread application has identified significant limitations. Issues raised include an inability to code complex drainage systems or large numbers of endorheic basins, the variability in both the levels of basin decomposition and sub-basin sizes and main stem identification criteria not conforming to local understanding. To address these issues, the Pfafstetter system was modified to use an independent regionalization for the initial sub-division of the continent. New procedures were developed to code distributary drainage networks and endorheic basins and alternate criteria devised to identify the main stem and to produce a more even decomposition of a catchment. These modifications were successfully applied to the task of delineating and coding hierarchically nested catchments for the Australian continent including its complex distributary and anabranching drainage networks and large numbers of endorheic basins. The independent basin grouping produced initially smaller and more evenly sized sub-basins and modelled estimates of flow identified the main stem correctly slightly more often than the original contributing area criterion. Enabling further sub-division of catchments with fewer than four tributaries produced the largest change in catchment delineations, doubling the number of sub-basins and halving their size. Achieving the vision for a comprehensive basin reference system will require further development to include anthropogenic and other hydrological features not controlled by topography.

## 1. Introduction

The Pfafstetter system was proposed as a global reference system for river basins (Verdin and Verdin, 1999). Based on the concepts articulated by a Brazilian engineer, Otto Pfafstetter, the scheme stands out among other catchment coding schemes in terms of ease of implementation and interpretation, global applicability, economy of coding and conveyance of useful information within the codes. The scheme delineates and codes hierarchically nested catchments guided by the topology of the drainage network and the size of the drainage area (Figure 1). One of the most useful properties of the Pfafstetter coding is its ability to indicate topological relationships in a catchment. Network position is inferred from the ordinal value of a digit and by whether it is odd or even. A larger digit indicates a section of river further upstream while an even digit designates a tributary off the main channel. With simple algebraic queries on the Pfafstetter codes it is possible to determine whether or not an activity is upstream and thus likely to affect a particular river section without reference to a map or flow routing functions in a GIS.

The Pfafstetter system is widely used. Recently developed global, hierarchically nested catchment datasets HydroBASINS (Lehner & Grill, 2013)(<http://hydrosheds.org/page/hydrobasins>) and the Hydrologic Derivatives for Modeling and Analysis (Verdin, 2017), were coded using the Pfafstetter system as was a predecessor, the HYDRO1k database (U.S. Geological Survey, 2001). Pfafstetter was also selected as the coding scheme for the European-wide stream and catchment database (Vogt et al., 2003; Vogt et al., 2007) and the Australian Hydrological Geospatial Fabric (Bureau of Meteorology, 2010; Stein et al., 2014) following comprehensive reviews.

Nested catchments derived using the Pfafstetter system support a variety of hydrological and environmental modelling activities. These include sub-basin decomposition for computationally efficient hydrological modelling (Beighley & He, 2009; Beighley et al.,

2015; Formetta et al., 2011; Ramadan et al., 2012; Yang & Musiaka, 2003; Yoon & Beighley, 2015) and exploration of basin scale effects on distributed hydrological model simulations (Ao et al., 2003). The system can also provide ecologically meaningful planning units for systematic conservation planning (Klein et al., 2009) with ready accounting of river network connectivity (Linke et al., 2012).

While demonstrating its utility, these applications also exposed limitations in the Pfafstetter system, particularly when applied to the diverse drainage structures apparent at continental scales. The Pfafstetter system allows for coding of just one endorheic (internally draining) basin at each level of decomposition, that being the largest, but has no method of coding smaller endorheic basins. This is particularly problematic in low relief areas with large numbers of endorheic basins. More than 50% of the Australian continent drains internally (Stein et al., 2014), with more than 20,000 recognized endorheic basins terminating in an inland lake, clay pan or other topographic depression (Hutchinson et al., 2008). Yet only the largest of these basins would receive a Pfafstetter code. While smaller endorheic basins may be combined manually with an adjacent, larger basin for coding (e.g. HydroBASINS, Lehner & Grill, 2013) there is no method to ensure consistent and repeatable treatment of endorheic basins. Coding multiple inlet streams to terminal lakes is also difficult - one inlet stream is not logically a tributary of any other and standard rules of up and downstream do not apply. Moreover, the Pfafstetter system includes no specific procedure for coding the multi-channel drainage systems that are common among large rivers (Jansen & Nanson, 2004).

The relative contributing area criteria that the scheme uses to define the tributary and main stem, a critical step for Pfafstetter code assignment, do not always conform to accepted conventions and so may reduce acceptance by the user community. For example, the Murray River is the name of the major river draining Australia's largest coastal draining river basin and is carried downstream to the river mouth. Yet at its confluence with the Darling River it drains the smaller area, and so using Verdin and Verdin's (1999) criteria, would be coded as a tributary of the Darling River.

Other reported limitations of the Pfafstetter system relate to the variable level of decomposition of basins and the uneven size of the sub-basins that it may produce (Lehner, 2014; Zhang et al., 2007). The lack of a consistent relationship between the approximate size of the units and the hierarchical level is perceived as a major disadvantage. This is both a natural consequence of variable drainage density and of the Pfafstetter coding scheme that limits the number of catchment units at each level of sub-division to ten so that each can be coded with a single digit 0 to 9. Uneven size of sub-basins can also be a consequence of the stopping rule that is invoked when there are less than four tributaries preventing the sub-division of some large tributary catchments.

To address these issues various changes to the Pfafstetter system have been proposed, including the use of an independent regionalisation (de Jager & Vogt, 2010; Stein et al., 2014) or a manual grouping of basins (Lehner, 2014) to provide the initial levels of the nested catchment hierarchy, using modelled flow, upstream channel length or stream name continuation as alternative criteria for main stem identification (Fürst & Horhan, 2009; Silva et al., 2008; Stein, 2006), increasing the numbers of tributaries coded at each level (Bertolo, 2000; Formetta et al., 2011), extending the methodology to accommodate more complex drainage structures including distributary and anabranching systems (Chen & Jin, 2011; Stein, 2006) and changing the stopping rule (Jasiewicz & Metz, 2011; Lehner, 2014; Stein, 2006).

Building on this body of work, I describe here five key modifications to the Pfafstetter system, including three not previously published. These modifications address the known limitations of the original scheme while retaining its advantages, in particular the informative and efficient coding scheme. I also systematically test these changes by applying the modified scheme to the delineation and coding of hierarchically nested catchments for the Australian continent which has extensive distributary and anabranching drainage networks and large numbers of endorheic basins.

## **2. A modified Pfafstetter system**

### **2.1 Initial division of the continent**

Verdin and Verdin's (1999) scheme initially divides each continent into ten basin and inter-basin units: -the single largest internally draining basin, the four largest coastal draining basins and five inter-basin units formed by grouping the coastal draining areas that lie between the largest coastal basins. The largest four coastal basins are assigned the even Pfafstetter codes 2,4,6 and 8 in order, travelling around the coast in a clockwise direction beginning at the basin closest to due north. In the same order, the intervening coastal inter-basins are coded with the odd digits 1 to 9 with the area lying between basins 1 and 8 divided between inter-basins 1 and 9.

The coastal inter-basin groups however, are typically much larger than the basins particularly when the largest four drainage basins are clustered geographically. This initial division thus contributes to the uneven size and level of decomposition of the nested catchments.

The modified Pfafstetter system uses, instead, an independent regionalization of drainage basins to produce a more evenly sized initial division of the continent, a method employed previously (de Jager & Vogt, 2010; Lehner, 2014; Stein et al., 2014). Subsequent levels of sub-division then follow the Verdin and Verdin scheme, delineating the four largest externally draining basins and five inter-basins within each of the initial basin regions. Starting at a boundary between basin regions and travelling in a clockwise direction around the coastline, the four largest basins are coded with the even digits and the intervening basins with the odd digits. Drainage basins on islands are coded as though they lie between the closest mainland basin and its mainland neighbour (Figure 2).

### **2.2 Main stem and tributary identification**

Application of the Pfafstetter system requires that the main stem be identified at each stream confluence (Figure 1). In Verdin and Verdin's scheme this is always the stream with the larger contributing area, even where this is inconsistent with accepted naming conventions and local understanding. The modified Pfafstetter system tests an alternative rule: that the main stem is that stream with the greater flow. To apply this rule a measured, or modelled estimate of flow is required for each stream at every confluence.

### **2.3 Endorheic basins**

The treatment of endorheic or internally draining basins initially follows the Verdin and Verdin scheme. Thus, at each level of sub-division, the largest internally draining basin in each higher-level basin group is identified and assigned a zero code. To attribute a Pfafstetter code to smaller endorheic basins however, a new procedure is applied that assigns each of the smaller endorheic basins the Pfafstetter code of the lower catchment into which they would spill were they to overflow the lowest point on the drainage divide (Figure 3). If

an internally draining basin spills to another un-coded internally draining basin, the Pfafstetter code is that of the basin into which the lower basin would spill and so on. If the receiving basin had been further sub-divided, the Pfafstetter code assigned is that of the individual tributary or main stem sub-catchment adjacent to the spill point on the drainage divide.

Where there are multiple inlet streams to a terminal lake, a secondary “Pfafstetter basin” is delineated for each to enable Pfafstetter tributary – main stem decomposition. These Pfafstetter basins delineate the catchment area drained by the inlet stream to its terminus (sink) within the lake (Figure 3b). Areas draining directly into the lake are included within the Pfafstetter basin of the inlet stream with the greatest upstream contributing area.

Pfafstetter basins are assigned the Pfafstetter code of the parent basin (the entire area draining to the lake) until individually coded as the largest internally draining basin in the higher-level catchment unit, so ensuring that all lake Pfafstetter basins are initially coded together.

#### 2.4 Multi-channel stream networks

The Pfafstetter system of Verdin and Verdin was also extended to code the tributary and main stem relationships of multi-channel drainage structures. The method employed depends on whether the distributary channels re-join downstream or terminate without re-joining the main channel. Thus, branching streams that re-join, including anabranching, anastomosing and braided streams, are treated as though a single channel for Pfafstetter tributary and main stem coding at all levels of basin sub-division (Figure 4). In contrast, the tributary / main stem relationships of terminal distributaries are coded independently of the parent stream once the distributary catchment is recognized as either the largest endorheic basin or if externally draining, one of the four largest coastal draining basins within the higher-level catchment unit.

This is achieved by delineating another Pfafstetter basin for terminal distributaries, being the catchment area draining from the point at which the terminal distributary offtakes the main stem to the distributary terminus – either an inland sink or the sea (Figure 4). These Pfafstetter basins, like those delineated for the multiple inlet streams to terminal lakes, are combined with all other basins draining to a single outlet, and together, provide the initial catchment units for Pfafstetter grouping and sub-division. Until independently coded, the Pfafstetter basins of distributaries are attributed with the Pfafstetter code of the main stem segment from which they offtake. To facilitate network tracing using the Pfafstetter codes, the identity of the main stem stream segment upstream of the distributary and in the case of the main stem, the identity of the offtake channel, are stored in additional fields in the Pfafstetter table.

#### 2.5 Level of decomposition

Verdin and Verdin’s scheme stops the Pfafstetter sub-division of a catchment whenever there are less than four tributaries. The modified Pfafstetter system employs an alternative stopping rule (Jasiewicz & Metz, 2011; Lehner, 2014; Stein, 2006) to ensure that basins are sub-divided as many times as is supported by the resolution of the stream network, so that sub-division continues as long as there is at least one tributary within a catchment. Tributary catchments are still numbered with the even digit codes 2,4,6,8 as in the original scheme though the higher value codes may not always be allocated. When no tributaries remain, a Pfafstetter code of “1” is applied to the catchment unit for the current and subsequent divisions.

## 2.6 Testing the modified Pfafstetter system

To compare elements of the modified and original Pfafstetter systems five test runs were undertaken (Table 1). Each run produced a 45-level nested catchment hierarchy by first grouping then successively sub-dividing drainage basins based on the topology of the stream network. Updated versions of the 9-second resolution drainage basins, stream networks and associated sub-catchments that supplied the foundation data layers (Stein *et al.*, 2014) for the Australian Hydrological Geospatial Fabric version 2 (“Geofabric”) (Bureau of Meteorology, 2010) were used. These data were produced using the ANUCatchment custom tools developed for Geoscience Australia to assist the production of the 1-second resolution version 3 Geofabric products. Applied to the 9-second resolution DEM used for version 2, these tools delineated 63837 drainage basins of which 42439 were exorheic and 21398 endorheic, and nearly 1.4M stream segments and segment sub-catchments.

Levels 1 and 2 of the nested catchment hierarchy were delineated either by an independent grouping of drainage basins (Stein *et al.*, 2014) based on the Australian Water Resources Council Drainage Divisions and River Basins (AUSLIG, 1997) or by application of the original Pfafstetter scheme as described by Verdin and Verdin (1999). The independent basin groupings were those derived directly from the 9 second DEM prior to any manual readjustment of boundaries (Stein *et al.*, 2014).

The main stem was identified as in the original scheme, as that with the greater contributing area, or for the modified Pfafstetter system, as that with the greater flow. Contributing area was calculated at 9 second resolution (Stein *et al.*, 2014) and for an anabranch, included the total contributing area of the main stem from which it offtakes. While the contributing area of the main stem is not divided at distributary points it is also not double counted when an anabranch re-joins the main stem.

Modelled estimates of annual mean flow had been calculated for each network stream segment by accumulating upstream gridded estimates of runoff volume. Two independent sources of modelled runoff were employed to test the sensitivity of the main stem identification to model choice and time period: 1) the eMast monthly runoff estimates for the period 1970 to 2012 (Hutchinson & Stein, 2014) and 2) the Australian Water Availability Project (AWAP) annual runoff estimates for the period 1900 to 2007 (Raupach *et al.*, 2008; 2012). The 0.01-degree eMast and 0.05-degree AWAP gridded runoff estimates were resampled to 9 second grid resolution (0.0025 degree) consistent with the resolution of the flow direction grid that defined the surface flow pathways used to accumulate the upstream flow.

## 3. Results

Analysis of the five test runs (Table 1) showed the effect of the modifications to the Pfafstetter system.

The method used for the initial continental division impacted the size and numbers of sub-basins but only at higher levels in the nested catchment hierarchy. Forming levels 1 and 2 by grouping drainage basins using the independent regionalisation, rather than the Pfafstetter system, produced generally smaller and less variably sized sub-basins up to level 9 (Figure 5a). The peak number of basin sub-divisions also occurred earlier (level 5 compared to level 7) when starting with the independent basin grouping. Irrespective of the starting configuration, the number of sub-basins increased rapidly, then plateaued after about 10 levels of sub-division as increasingly more drainage basins were sub-divided as far as

possible for the given stream network (Figure 5b). After 45 levels, there was little difference between the runs in either the average size or the number of sub-basins delineated (Table 2).

Relaxing the stopping rule to enable sub-division of basins with less than four tributaries, had a much greater impact on both the numbers and the size of sub-basins. Under the new stopping rule, sub-basins were smaller and their size less variable (Figure 6a, Table 2). After 45 levels of sub-division, the average size of the sub-basins produced by run L1\_A\_0 employing the new stopping rule was about half that of run L1-A-4 that used the original Pfafstetter system rule. Moreover, there were about twice the number of sub-basins (Figure 6b, Table 2) reflecting the greater number of times basins and sub-basins were sub-divided (Figure 6c). The new stopping rule also ensured that all basins were decomposed as far as possible. In contrast, at the completion of the run employing the original rule requiring at least four tributaries, more than 150,000 sub-basins were undivided despite the presence of at least one stream confluence (Figure 6d).

The extended Pfafstetter system attributed a Pfafstetter code to all endorheic basins, the larger basins a zero code as the single largest in each higher-level basin grouping, the smaller basins the code of the basin or sub-basin to which they would overflow. However, it required many levels of division before most of the endorheic basins were individually coded, many more than was required for exorheic basins (Figure 7a). Accordingly, endorheic basins were fully sub-divided later than those that drain externally (Figure 7b). At least 36 levels of division were needed to ensure 98% of the endorheic basins with at least one stream confluence were individually coded and sub-divided as far as possible. In contrast, just 11 levels of division were required to achieve this for exorheic basins (runs L1-A-0 and L3-A-0). Not all endorheic basins were individually coded even after 45 levels. The pattern was similar for run L1-A-4 though the numbers of basins of all types was lower as many smaller basins with less than four tributaries were never sub-divided (Figure 7b).

The initial division of the continent had a relatively small effect on the coding of endorheic basins under the extended Pfafstetter system. While there were fewer endorheic basins not individually coded when Pfafstetter sub-division was applied from level 1 (Run L1-A-0), these 679 basins were about twice as large on average as the 2156 basins not individually coded when the independent basin grouping supplied the level 1 and 2 sub-division (Run L3-A-0) (mean area of basins not individually coded was 26.3km<sup>2</sup> and 11.9km<sup>2</sup> respectively). Most of the basins not coded individually occurred within the very arid South-western Plateau Drainage Division and for Run L3-A-0, also within the Pilbara-Gascoyne Drainage Division.

The extended Pfafstetter system also attributed a standard Pfafstetter code to all multi-channel and lake inlet streams. Pfafstetter sub-basins were delineated for 2450 terminal lake inlet streams and 772 terminal distributaries. These Pfafstetter basins were largely endorheic and followed a similar pattern of sub-division to that of other endorheic basins (Figure 7). Many more of the multi-channel stream networks rejoined downstream to flow to a single outlet and so were coded as though a single channel.

The pattern of sub-division and final number of sub-basins differed little when modelled flow replaced contributing area as the criterion for tributary-main stem identification (Table 2). In most cases, the modelled flow and contributing area criteria agreed on the main stem selection. The modelled eMast and AWAP flow selected a stream segment with a lower contributing area at just 2.3% and 2.9%, respectively, of the 512,848 confluences. Both measures of modelled flow correctly selected the main stem accepted by local convention, as indicated by upstream name continuation, at almost 95% of the

confluences of named streams. This result is only slightly better than 94.4% achieved by the contributing area criterion.

While the figures overall are very similar, there were important differences in main stem selection and hence the Pfafstetter coding of individual basins. A notable example is the Murray Darling Basin. Both estimates of modelled flow selected the Murray River as the main stem and the Darling River as the tributary at their confluence, even though the Darling River catchment of 511,304 km<sup>2</sup> is more than 300,000 km<sup>2</sup> larger than that of the Murray River upstream of their confluence (198,517km<sup>2</sup>). This accords with both stream name continuation and accepted local convention. Similarly, the Burdekin River that flows into the Coral Sea in north Queensland drains a catchment some 33,000km<sup>2</sup> smaller than that of the Suttor River upstream of their confluence, though in this case only the eMast modelled runoff correctly selected the Burdekin as the main stem.

The alternate estimates of modelled flow selected a different main stem at a total of 8,639 confluences or 1.7% of the total number of confluences, of which about 1,300 were the confluence of named streams. However, when a different stream was selected it was almost always one with the same Strahler stream order or differing by just one. At just 185 confluences, about 2% of the confluences, where the main stem selection differed depending on the choice of modelled runoff, tributary stream orders differed by 2 or more.

#### **4. Discussion**

Modifications and extensions to the Pfafstetter system of Verdin and Verdin have been described and applied to delineate and code nested catchments for the Australian continent.

Deriving the initial basin grouping based on an established regionalisation in widespread use is likely to gain greater acceptance by the user community and facilitates linkages with other information organized according to this regionalisation. The Australian regionalisation, for example, is based on the Australian Water Resources Council River Basins that provided the framework for the collection of water information for many decades. Nonetheless, while using an independent grouping of basins to initiate the continental sub-division produced smaller and less variably sized sub-basin units for the early sub-divisions, it made little difference to either the total number of sub-basins or their size subsequently. The variation in the size of the sub-basins reflects the natural variability in drainage density and thus is largely unavoidable if confluences are to be the basis for sub-division. If more homogeneously sized units are required there are simple work-arounds such as extracting sub-basins at basin specific levels (Stein, 2006) or as implemented in HydroBASINS, skipping a level of sub-division when the sub-basin area falls below a level-dependent range (Lehner, 2014).

An independent regionalisation is also less sensitive to the start position of the coastal numbering system that determines the initial Pfafstetter basin and inter-basin coding. For example, starting the basin numbering system at the most north-easterly, rather than the most northerly, of the basins, produces a more variable sized Level 1 Pfafstetter delineation leaving more than half of the continental landmass coded as coastal inter-basin 7 (Figure 8, Stein, 2006). In contrast, this large area of over 4.1M km<sup>2</sup>, is split between inter-basins 1 and 9 when coding the most northerly of the four largest basins with a Pfafstetter code of 2. However, unless the independent regionalisation has 10 or fewer regions, catchment codes will be longer and thus less efficient. The Australian code, for example, was extended by 7 characters to accommodate the 12 Drainage Divisions that formed Level 1, the 191 Level 2 basin regions, and the dashes that were used to separate the region and Pfafstetter codes.

The modelled estimates of flow correctly identified the main stem as indicated by continuation of the stream name slightly more often than did the contributing area criteria employed by the original Pfafstetter system. They are, therefore, more likely to produce results that conform to local understanding and thus be more acceptable to users. Using flow to identify the main stem requires a measure or modelled estimate of flow for every stream link. As shown here, suitable estimates of flow can be derived using information readily available at monthly or annual time steps. The new HydroAtlas database (<http://wp.geog.mcgill.ca/hydrolab/hydroatlas/>) for example, will supply global estimates of natural annual discharge for each river reach of HydroSHEDS.

While it was expected that the results would be sensitive to the runoff model chosen and the time period over which flow was estimated, this occurred only rarely. Where the two estimates of modelled flow differed in their main stem selection, it was almost always at the confluence of streams of a similar size. To reduce uncertainty, additional criteria could be introduced such as stream name continuation or the selective use of contributing area. For example, the main stem of a confluence of named streams would be that which continues the same stream name of the downstream stream segment. For smaller streams, where the difference in flow is small and the modelled estimates of flow might be a less reliable indicator, the main stem selection could revert to being that with the greater contributing area.

Many levels of division were required before most of the numerous endorheic basins were individually coded as the single largest internally draining basin. The extensions to the Pfafstetter system ensured, however, that these smaller endorheic basins were supplied a Pfafstetter code that also reflected the higher-level topographic relationships among basins. Fewer levels will likely be required for other continents where endorheic basins occupy a substantially smaller proportion of the landmass (Hammer, 1986).

A new Pfafstetter basin was delineated for the secondary stream systems draining into a terminal waterbody, supporting a more logical application of the Pfafstetter tributary – main stem coding. Another option would be to treat the waterbody edge as though it were a coastline and number basins and inter-basins in a clockwise direction around the edge in a manner akin to the initial numbering of basins around the coastline. In this case, the basins to be coded with the even digits would be the four largest catchments of the waterbody inlets and the inter-basins, coded with the odd digits, are the remaining areas that drain into the lake directly. This method, however, is only logically employed for very large lakes such as Lake Eyre (Stein, 2006).

The Pfafstetter system was also extended to delineate and code Pfafstetter basins for terminal distributaries enabling the topological relationships of distributary streams to be analyzed using the standard Pfafstetter coding. To do so however, requires additional fields recording the main stem segment from which the terminal distributary off takes and for the main stem, the identity of the distributary segment. Multi-channel braided and anastomosing systems occupy a single floodplain (Riley, 1975) and accordingly, are coded as though a single channel.

In contrast, the extended Pfafstetter coding scheme of Chen and Jin (2011) uses a combination of prefixes and suffixes to represent the topological relationships of individual stream segments in these complex channel types. Thus, an anabranch is assigned the main stem code together with the suffix “B” while an anabranch that is also coded as a tributary at a downstream confluence, receives both the tributary code and the anabranch code prefixed with an “A”. While this has the advantage of attributing an individual code to each stream segment, the codes potentially become very long and unwieldy, for example, where channels repeatedly bifurcate and re-join as in extensively anastomosed stream systems.

Altering the stopping rule produced the most substantial change. Under the original Pfafstetter system of Verdin and Verdin, catchment sub-division was halted prematurely in large numbers of catchments with less than four tributaries, even where the smaller catchments nested within them had four or more tributaries that could otherwise be further sub-divided. Accordingly, the final nested hierarchy contained about half as many sub-basins as did those where catchments with less than four tributaries were sub-divided, and the sub-basins were more variable in size, and on average, larger.

The change to the stopping rule does not impact on the ability to infer stream network topological relationships from the Pfafstetter codes. Thus, for example, if just two tributaries are mapped in a catchment only the codes one to five will be used, but the odd-even and ordinal nature of the codes is retained so that main-stem and tributary relationships and up – downstream linkages are readily inferred. Differing also from the original Verdin and Verdin numbering scheme is the application of a code “1” rather than zero after the stopping rule is invoked. This however, conforms with the logic of the Pfafstetter numbering system where a zero code indicates an internally draining basin and a one the most downstream of the main stem stream segments in a catchment.

Emerging methods based on a graph theoretic approach also facilitate network traversing and connectivity analysis (Chen et al., 2018; Daltio and Medeiros, 2016; Cui et al., 2011). However, application to distributary drainage networks are still developing (Tejedor et al., 2015) and, unlike the enhanced Pfafstetter system described here, there is no apparent method for endorheic basins or that would produce a hierarchical decomposition of basins.

## 5. Conclusions

Pfafstetter is an efficient and effective catchment reference system. Most importantly, its simple coding is readily accessible to those without sophisticated analytical skills. Pfafstetter continues to be the system of choice for national and global databases.

The extensions and modifications described here address limitations of the Pfafstetter system. New procedures were introduced to code complex distributary and anabranching drainage networks and to more logically code multiple stream inlets to waterbodies. Another novel extension is the coding method for endorheic basins that is informed by the higher-level topographic relationships of basins. With this extension, all endorheic basins are logically and consistently coded. Application to delineate and code hierarchically nested catchments for the Australian continent demonstrated the positive effect of these changes. Relaxing the stopping criterion had the greatest overall impact, substantially increasing the level of basin decomposition. More significant locally, was the effect of switching the criterion for tributary and main stem identification to an estimate of modelled flow that more often agreed with local understanding. Importantly, these results will assist users understand the effect of Pfafstetter system choices on main stem selection and key catchment properties including numbers, size and variability.

While a number of issues have been addressed, others remain that limit its applicability as a “fundamental spatial framework “ (Verdin & Verdin, 1999 p.12) for water information. Designed to code the topological relationships of a connected stream network, the Pfafstetter system is, unsurprisingly, not well suited to coding other components of the hydrological system including groundwater connections, non-riverine surface water features such as palustrine wetlands or water infrastructure that are not dependent upon topographically driven flow. The Pfafstetter code of an associated stream segment might be attributed but would likely need to be amended if it were to correctly reflect the topological

relationships among features. In the case of an inter-basin transfer pipeline, for example, it would require consideration of the codes of both the offtake and outflows segments. Nevertheless, with further development and testing the Pfafstetter system might truly meet Verdin and Verdin's vision for a "comprehensive reference system for the Earth's river basins" (1999, p.1).

## Acknowledgements

This research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. Michael Hutchinson and John Stein are thanked for constructive comments that improved an earlier version of this manuscript. The manuscript also benefited from insightful comments and helpful suggestions from John Gallant, Riccardo Rigon and a third anonymous reviewer.

A program implementing the modified Pfafstetter system is available in the ANUCatchment toolkit. The toolkit, together with the data used to generate the results presented here, will be submitted to the ANU's research repository. While this manuscript is under review, the data is available for download from <https://drive.google.com/drive/folders/1ldMEDw5mvsygJU6Ji2De6rbQFTM8kJ2L?usp=sharing>

## References

- Ao, T., J. Yoshitani, K. Takeuchi, K. Fukami, T. Mutsuura, & Ishidaira, H. (2003), Effects of sub-basin scale on runoff simulation in distributed hydrological model: BTOPMC in *Weather Radar Information and Distributed Hydrological Modelling (Proceedings of symposium HS03 held during IUGG2003 at Sapporo, July 2003)*, IAHS Publication no.282, pp. 1-7.
- AUSLIG (1997), River Basins of Australia, Australian Survey and Land Information Group, Canberra.
- Beighley, R. E., & He, Y. (2009), Predicting model uncertainty at river junctions due to drainage network structure, *Journal of Hydrologic Engineering*, 14(5), 499-507.
- Beighley, R. E., Eggert, K., Wilson, C. J., Rowland, J. C. & Lee, H. (2015), A hydrologic routing model suitable for climate-scale simulations of arctic rivers: application to the Mackenzie River Basin, *Hydrological Processes*, 29(12), 2751-2768.
- Bertolo, F. (2000), Catchment Delineation and Characterisation. A Review. EuroLandscape - Catchment Characterisation and Modelling (CCM), Report No. 1, 36 pp, Space Applications Institute, Joint Research Centre, Ispra(Va), Italy.
- Bureau of Meteorology (2010), Australian Hydrological Geospatial Fabric (Geofabric) Product Guide. Version 1.0, edited, Bureau of Meteorology, Canberra.. Retrieved from [http://www.bom.gov.au/water/geofabric/documents/product\\_guide.pdf](http://www.bom.gov.au/water/geofabric/documents/product_guide.pdf)
- Chen, X., & Jin, F. (2011), Coding Method and Application for Complicated River Network Based on Surveyed River Network, *Fourth International Symposium on Computational Intelligence and Design (ISCID)*, pp. 47-50, Hangzhou, China. <https://10.1109/ISCID.2011.113>
- Chen, J., Xiao, C. & Chen, D. (2018), Connectivity evaluation and planning of a river-lake system in East China based on graph theory, *Mathematical Problems in Engineering*, 2018, 1361867.

- Cui, Z., Koren, V., Neftali, C., Voellmy, A. & Moreda, F. (2011), Hydroinformatics advances for operational river forecasting: using graphs for drainage network descriptions." *Journal of Hydroinformatics* 13(2): 181-197.
- Daltio, J. & Medeiros, C. B. (2016), Hydrograph: Exploring geographic data in graph databases. *Brazilian Journal of Cartography*, 68(6):1181-1189.
- de Jager, A. L., & Vogt, J. V. (2010), Development and demonstration of a structured hydrological feature coding system for Europe, *Hydrological Sciences Journal*, 55(5), 661-675.
- Formetta, G., Mantilla, R., Franceschi, S., Antonello, A. & Rigon, R. (2011), The JGrass-NewAge system for forecasting and managing the hydrological budgets at the basin scale: models of flow generation and propagation/routing, *Geoscientific Model Development*, 4, 943-955.
- Fürst, J., & Horhan T. (2009), Coding of watershed and river hierarchy to support GIS-based hydrological analyses at different scales, *Computers & Geosciences*, 35(3), 688-696.
- Hammer, U. T. (1986), Saline Lake Ecosystems of the World, Volume 59 of *Monographiae Biologicae*, Dr. W. Junk, Dordrecht
- Hutchinson, M.F., & Stein, J. L. (2014), Monthly runoff: ANUclimate 1.0, 0.01 degree, Australian Coverage, 1970-2012, Terrestrial Ecosystem Research Network's (TERN) Ecosystem Modelling and Scaling Infrastructure (eMAST). Retrieved from <http://portal.tern.org.au/monthly-runoff-anuclimate-1970-2012/20323>.
- Hutchinson, M. F., Stein, J. A., Stein, J. L., Anderson, H. & Tickle, P. (2008), GEODATA 9 Second DEM and D8. Digital Elevation Model Version 3 and Flow Direction Grid User Guide, Geoscience Australia, Canberra. Retrieved from <https://data.gov.au/dataset/geodata-9-second-dem-and-d8-digital-elevation-model-version-3-and-flow-direction-grid-2008>
- Jansen, J. D., & Nanson, G. C. (2004), Anabranching and maximum flow efficiency in Magela Creek, northern Australia, *Water Resources Research*, 40(4), W04503.
- Jasiewicz, J., & Metz, M. (2011), A new GRASS GIS toolkit for Hortonian analysis of drainage networks, *Computers & Geosciences*, 37(8), 1162-1173.
- Klein, C., Wilson, K., Watts, M., Stein, J., Berry, S., Carwardine, J., et al (2009), Incorporating ecological and evolutionary processes into large-scale conservation planning, *Ecological Applications*, 19(1), 206-217.
- Lehner, B. (2014), HydroBASINS Global watershed boundaries and sub-basin delineations derived from HydroSHEDS data at 15 second resolution. Technical Documentation Version 1.c (with and without inserted lakes).retrieved from [http://hydrosheds.org/images/inpages/HydroBASINS\\_TechDoc\\_v1c.pdf](http://hydrosheds.org/images/inpages/HydroBASINS_TechDoc_v1c.pdf)
- Lehner, B., & Grill, G. (2013), Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, *Hydrological Processes*, 27(15), 2171-2186.
- Linke, S., Kennard, M. J., Hermoso, V., Olden, J. D., Stein, J., & Pusey, B. J. (2012), Merging connectivity rules and large-scale condition assessment improves conservation adequacy in river systems, *Journal of Applied Ecology*, 49(5), 1036-1045.

- Ramadan, H. H., Beighley, R. E., & Ramamurthy, A. S. (2012), Modelling streamflow trends for a watershed with limited data: case of the Litani basin, Lebanon, *Hydrological Sciences Journal*, 57(8), 1516-1529.
- Raupach, M. R., Briggs, P. R., Haverd, V., King, E. A., Paget, M., & Trudinger, C. M. (2008), Australian Water Availability Project, CSIRO Marine and Atmospheric Research, Canberra, Australia.
- Raupach, M. R., Briggs, P. R., Haverd, V., King, E. A., Paget, M., & Trudinger, C. M. (2012), Australian Water Availability Project (AWAP): CSIRO Marine and Atmospheric Research Component: Final Report for Phase 3, *CAWRC Technical Report No. 013*, CSIRO Marine and Atmospheric Research, Canberra, Australia. Retrieved from <http://www.csiro.au/awap>
- Riley, S. J. (1975), Some differences between distributing and braiding channels, *Journal of hydrology, New Zealand*, 14(1), 1-8.
- Silva, N., Ribeiro, C., Barroso, W., Ribeiro, P., Soares, V., & Silva, E. (2008), An improved stream network addressing system: The modified Pfafstetter coding scheme, *Revista Arvore*, 32(5), 891-897.
- Stein, J. L. (2006), A continental landscape framework for systematic conservation planning for Australian rivers and streams, PhD thesis, Australian National University, Canberra. Retrieved from ANU Open Research Repository (<http://hdl.handle.net/1885/49406>).
- Stein, J. L., Hutchinson, M. F., & Stein, J. A. (2014), A new stream and nested catchment framework for Australia, *Hydrology and Earth System Sciences*, 18(5), 1917-1933.
- Tejedor, A., Longjas, A., Zaliapin, I. & Foufoula-Georgiou, E. (2015), Delta channel networks: 1. A graph-theoretic approach for studying connectivity and steady state transport on deltaic surfaces, *Water Resources Research*, 51: 3998–4018.
- U.S. Geological Survey (2001), HYDRO1K Elevation Derivative Database, U.S. Geological Survey EROS Data Center. <https://lta.cr.usgs.gov/HYDRO1K>
- Verdin, K.L., 2017, Hydrologic Derivatives for Modeling and Analysis—A new global high-resolution database: U.S. Geological Survey Data Series 1053, <https://doi.org/10.3133/ds1053>.
- Verdin, K. L., & Verdin, J. P. (1999), A topological system for delineation and codification of the Earth's river basins, *Journal of Hydrology*, 218(1-2), 1-12.
- Vogt, J., Colombo, R., Paracchini, L., de Jager, A., & Soille, P. (2003), CCM River and Catchment Database. Version 1.0, 31 pp, Institute for Environment and Sustainability, EC Joint Research Centre, Ispra (Varese), Italy.
- Vogt, J., Soille, P., de Jager, A., Rimavičiūtė, E., Mehl, W., Foisneau, S., et al. (2007), A Pan-European River and Catchment Database, *JRC Rep. 40291*, 124 pp, European Commission, Joint Research Center, Institute for Environment and Sustainability, Ispra, Italy. Retrieved from [http://ccm.jrc.ec.europa.eu/documents/CCM2-Report\\_EUR-22920-EN\\_2007\\_STD.pdf](http://ccm.jrc.ec.europa.eu/documents/CCM2-Report_EUR-22920-EN_2007_STD.pdf)
- Yang, D., & Musiak, K. (2003), A continental scale hydrological model using the distributed approach and its application to Asia, *Hydrological Processes*, 17(14), 2855-2869.

Yoon, Y., & Beighley, E. (2015), Simulating streamflow on regulated rivers using characteristic reservoir storage patterns derived from synthetic remote sensing data, *Hydrological Processes*, 29(8), 2014-2026.

Zhang, L., Wang, G. Q., Dai, B. X., & Li, T. J. (2007), Classification and codification methods of stream networks in a river basin: A review, *Environmental Informatics Archives*, 5, 364-372.

Accepted Article

## Tables

Table 1 Pfafstetter coding runs contrasting elements of the original and modified Pfafstetter system. For the listed element "original" runs adhere to the Pfafstetter system of Verdin and Verdin (1999), "modified" that described here. All runs employ the modified procedures to code multi-channel stream networks and endorheic basins

Element	Original RunID	Modified RunID	Description
Initial division	L1-A-4	L3-A-4	Pfafstetter division is applied from level 1 (original) or level 3 (modified- an independent regionalisation delineates the first 2 levels of the nested catchment hierarchy)
Stopping rule	L1-A-4	L1-A-0	Decomposition of a catchment ceases whenever there are less than 4 remaining tributaries (original) or no tributaries remaining (modified)
Main stem identification criteria	L1-A-0	L1-R1-0 L1-R2-0	Main stems are those with the greater contributing area (original) or that with the greater modelled flow (modified). Modelled flow source eMast runoff (R1) or AWAP (R2) (see text for details)

Table 2 Numbers and area of sub-basins after 45 levels of sub-division

	Run ID				
	L1-A-4	L1-A-0	L3-A-4	L1-R1-0	L1-R2-0
<b>Area (km<sup>2</sup>): Minimum</b>	< 1	< 1	< 1	< 1	< 1
<i>Mean</i>	14.9	7.4	14.9	7.4	7.4
<i>Median</i>	4.9	2.8	4.9	2.8	2.8
<i>Maximum</i>	13,730	13,730	13,325	13,730	13,730
<i>Standard deviation</i>	71.7	40.8	72.8	40.8	40.8
<b>Number of sub-basins</b>	517,868	1,047,317	516,436	1,046,385	1,046,441

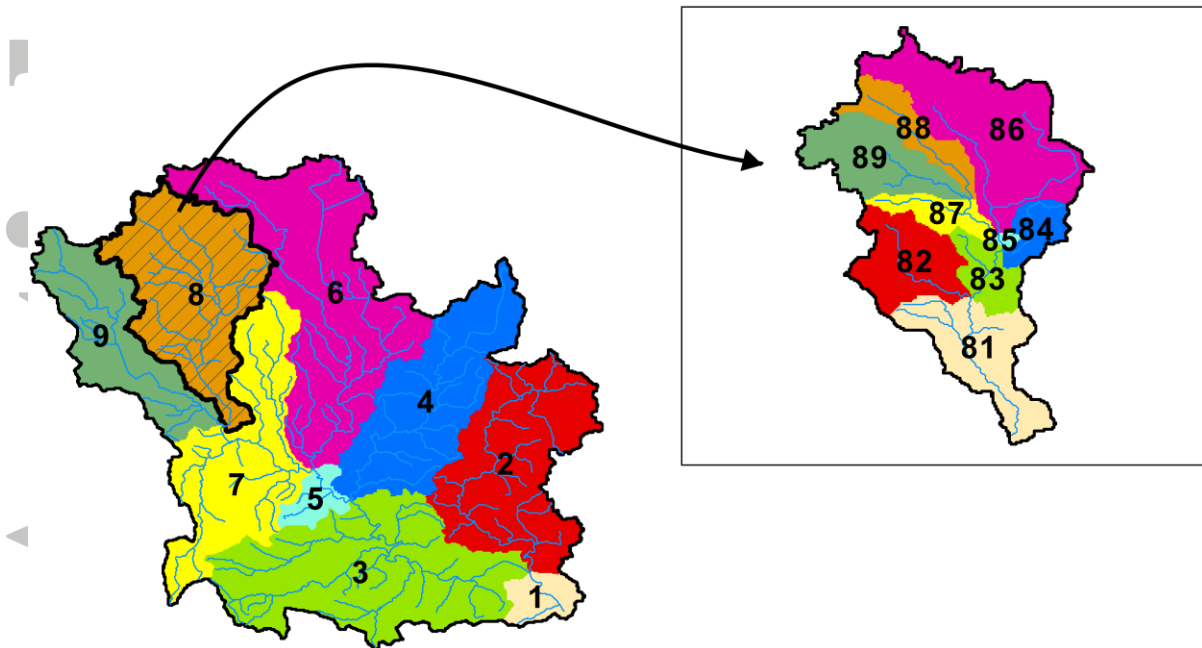


Figure 1 Pfafstetter decomposition of a basin. The four largest tributary catchments are coded with the even digits 2, 4, 6 and 8 in order upstream from the basin outlet. The intervening areas draining to the main stem are coded with the odd digits 1 to 9. Sub-basins are successively sub-divided using the same system as long as four tributaries remain. Inset – Pfafstetter sub-division and coding of sub-basin 8

Accept

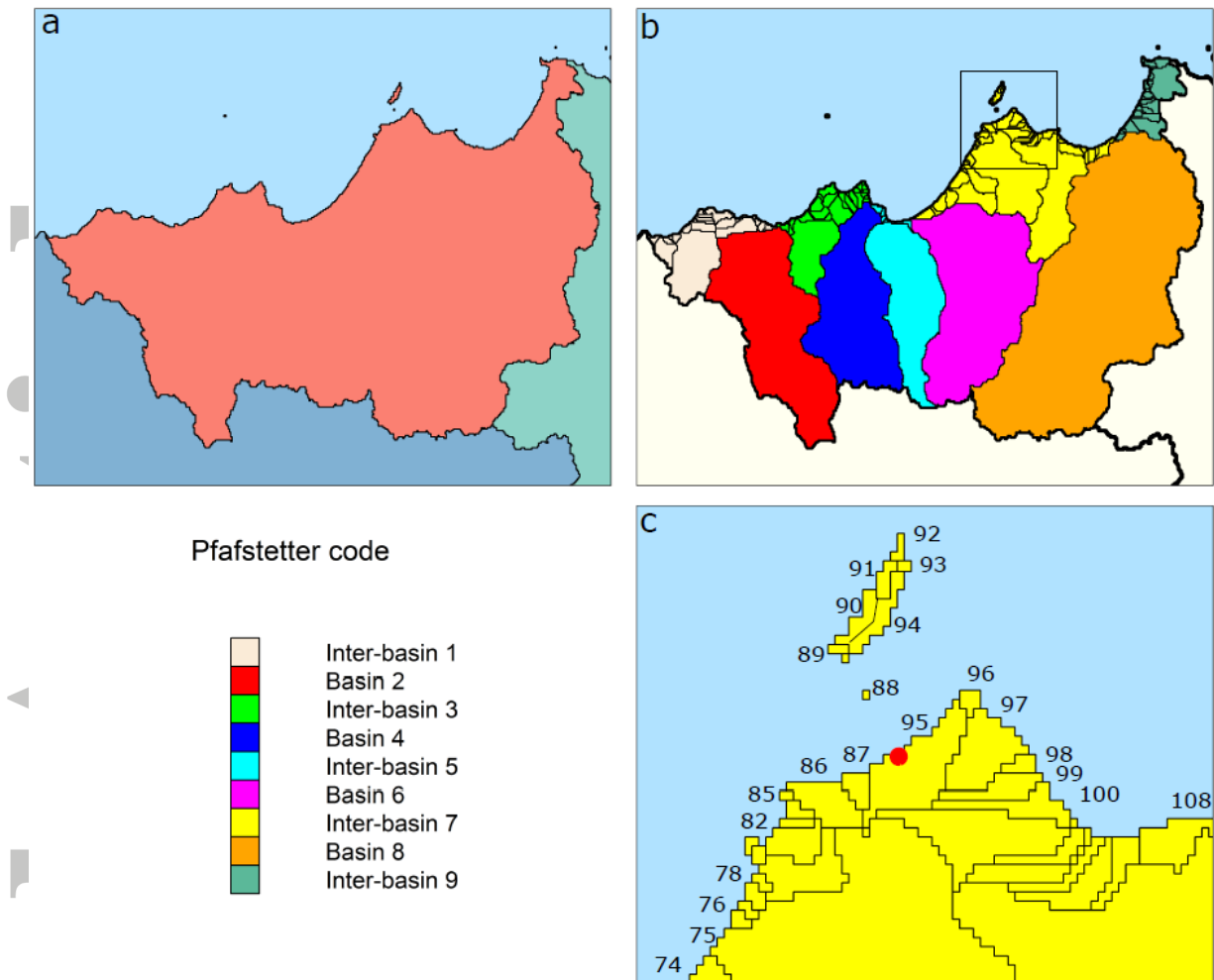


Figure 2. Coastal and island basin coding. a) Regionalization delineating the initial grouping of basins b) First level Pfafstetter sub-division - the four largest coastal draining basins within the region are coded with the even digits 2 to 8, the intervening coastal draining basins are grouped and coded with the odd digits 1 to 9 in clockwise sequence around the coast, c) Inset area in b) showing sequential numbering of basin outlets. Island basins are ordered in the sequence immediately before the closest basin outlet on the mainland, in this example, before basin 95.

Accepted

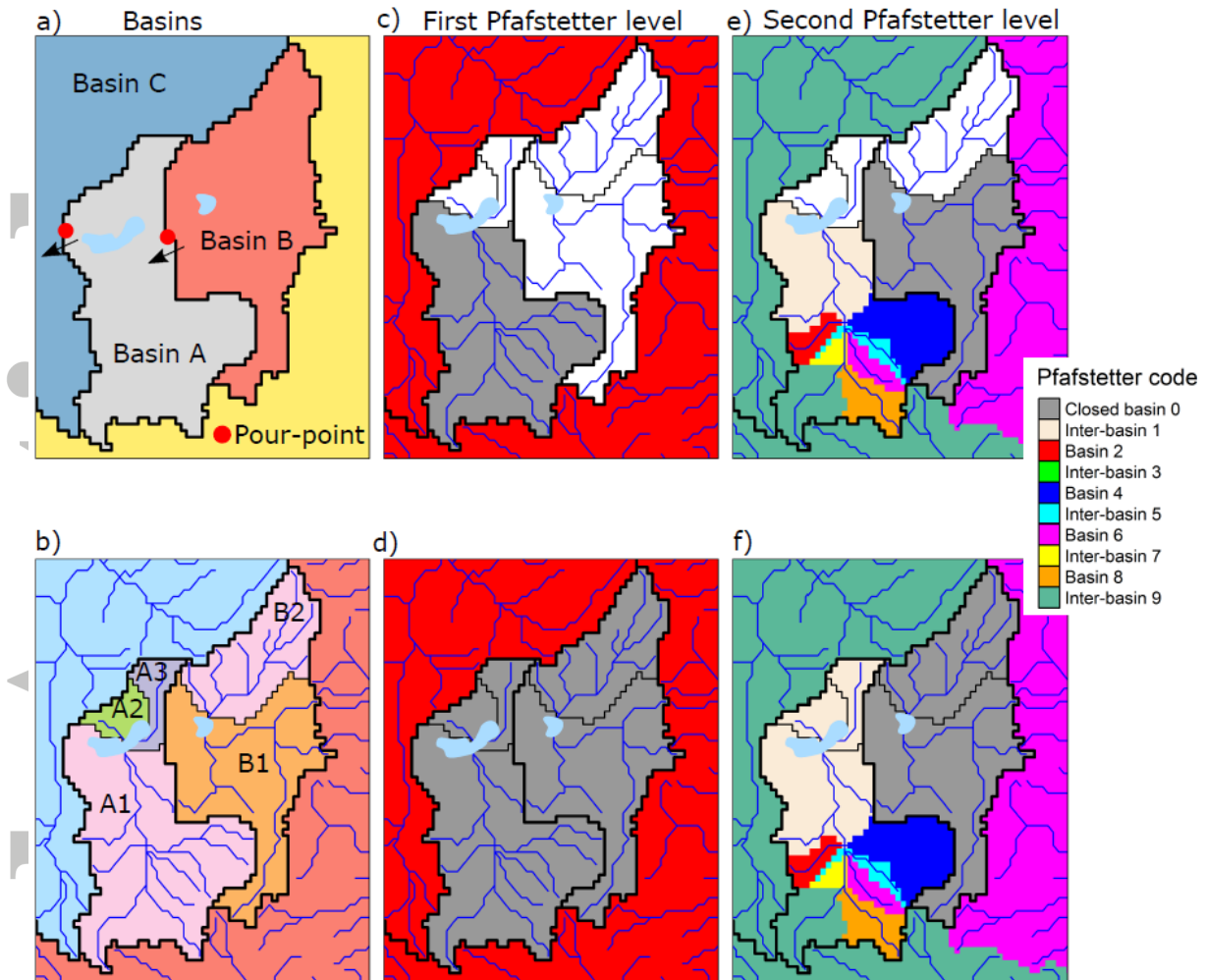


Figure 3. Pfafstetter coding of endorheic basins. a) Until independently coded, endorheic basins are attributed with the code of the neighbouring basin into which they would spill were they to overflow the pour-points on the drainage divide. In this example, Basin B would first overflow into Basin A which in turn would overflow to Basin C. b) “Pfafstetter basins” are delineated for each stream inlet of a waterbody. c) The largest endorheic basin (A1) is coded. d) Smaller endorheic basins receive the code of the parent basin (A2 and A3 from A1) or that into which they would spill (B1 and B2 from A1). e) Next level of coding - B1 as the largest endorheic basin, A1 sub-divided into tributary catchments and main-stem sub-basins. f) Smaller basins A2 and A3 receive inter-basin 1 from the waterbody inlet stream of A1, B2 receives the zero code from B1

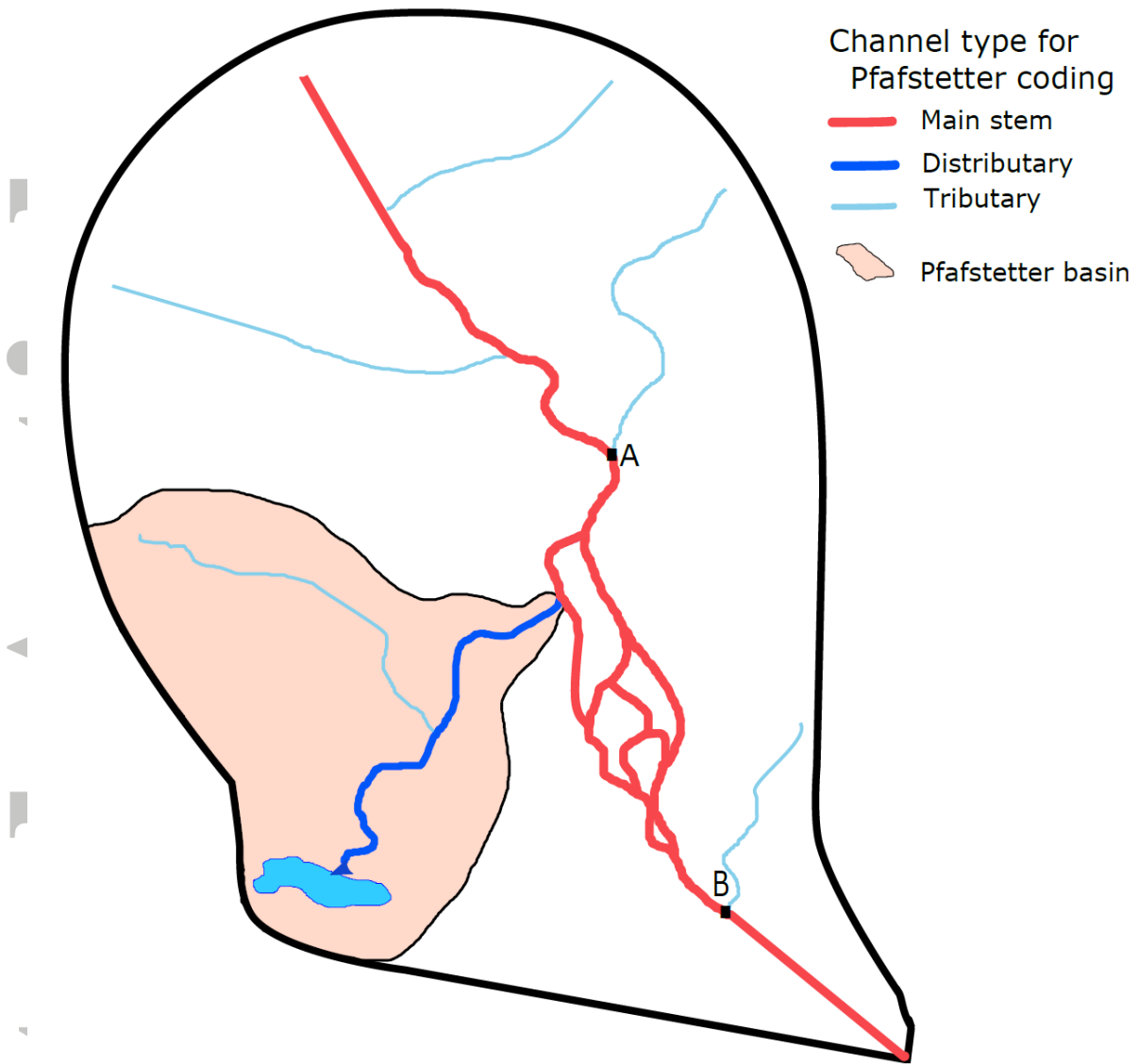


Figure 4. Pfafstetter coding of multi-channel streams. The area draining to the terminal point of a distributary channel (the “Pfafstetter basin”) is coded independently of the main stem channel from which the distributary offtakes, once recognized as the largest endorheic basin. Sub-basin decomposition of the Pfafstetter basin then follows the standard method (Figure 1). Multiple channels that rejoin such as those between stream junctions A and B are coded together as if a single channel – in this example, as a main stem, and so with the same odd digit.

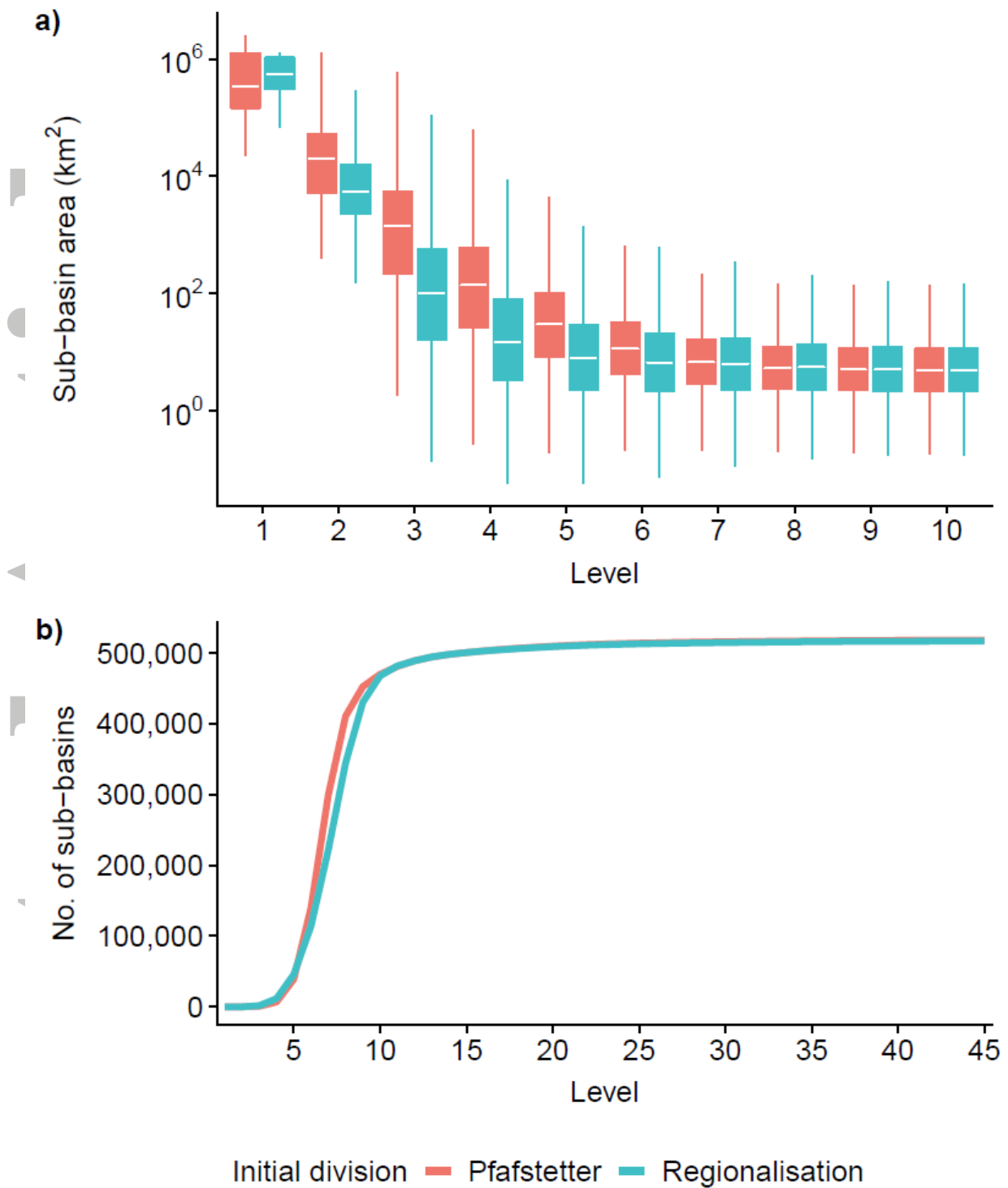


Figure 5 Effect of the initial continental division on: a) sub-basins area and b) the number of sub-basins. Note that the distribution of area values changes little after the first 10 levels of sub-division. Boxes show the inter-quartile range and median value while the whiskers are drawn to a length 1.5 times the inter-quartile range

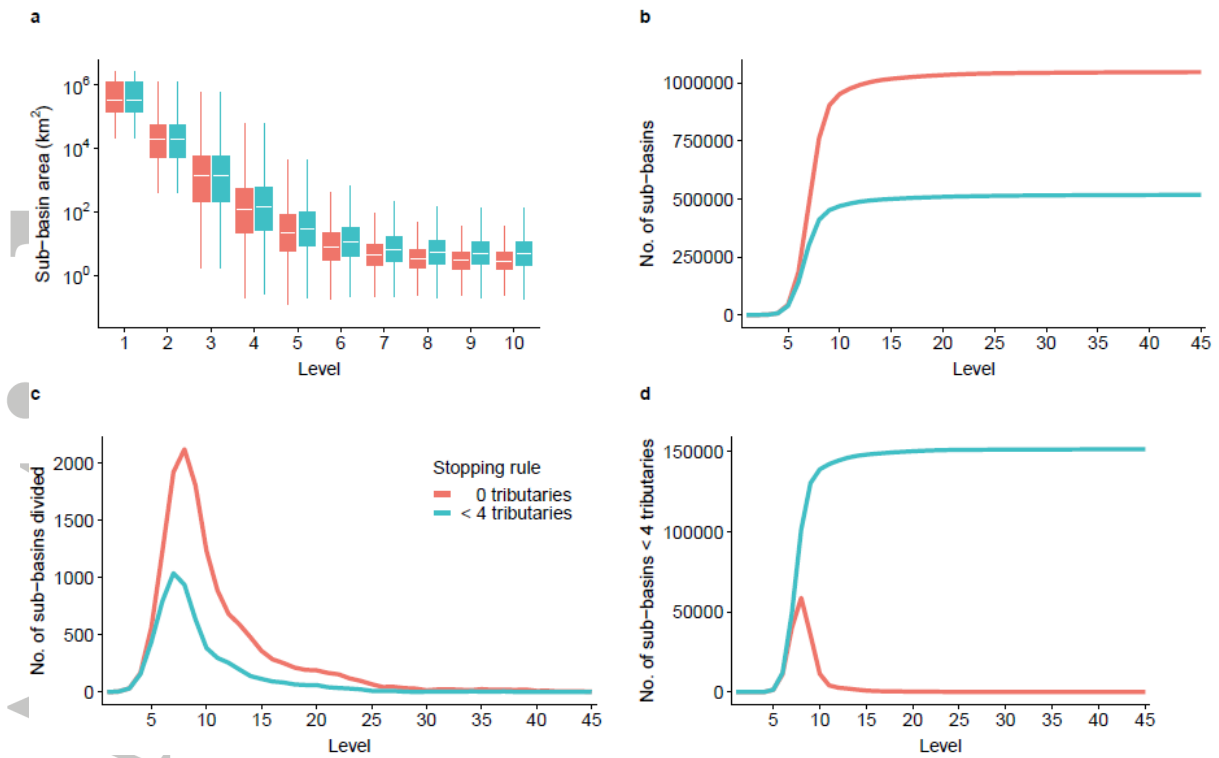
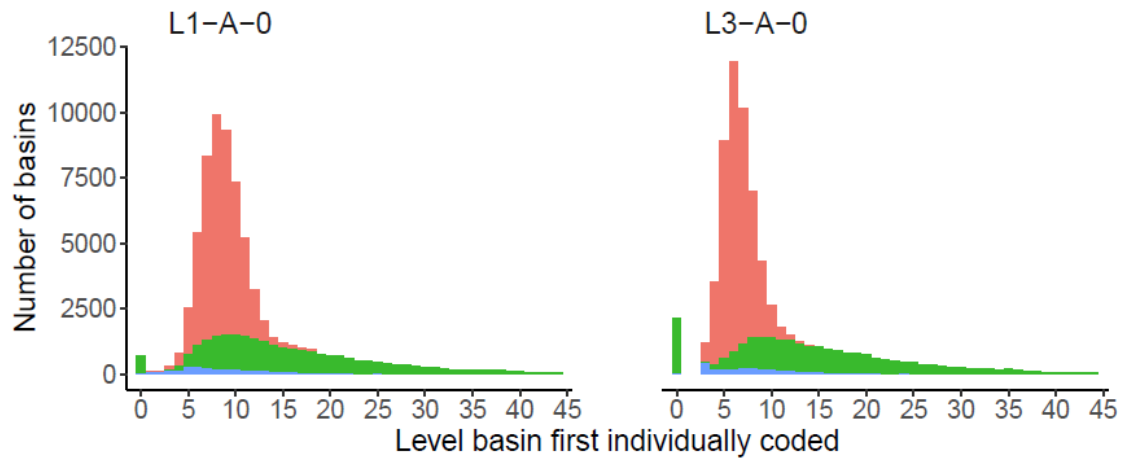


Figure 6 Effect of the stopping rule on: a) sub-basins area, b) the cumulative numbers of sub-basins in the nested catchment hierarchy, c) numbers of sub-basins divided at each level and d) the numbers of sub-basins with at least one but less than four tributaries i.e. those that would not be further sub-divided under the original stopping rule

Accepted

a)



b)

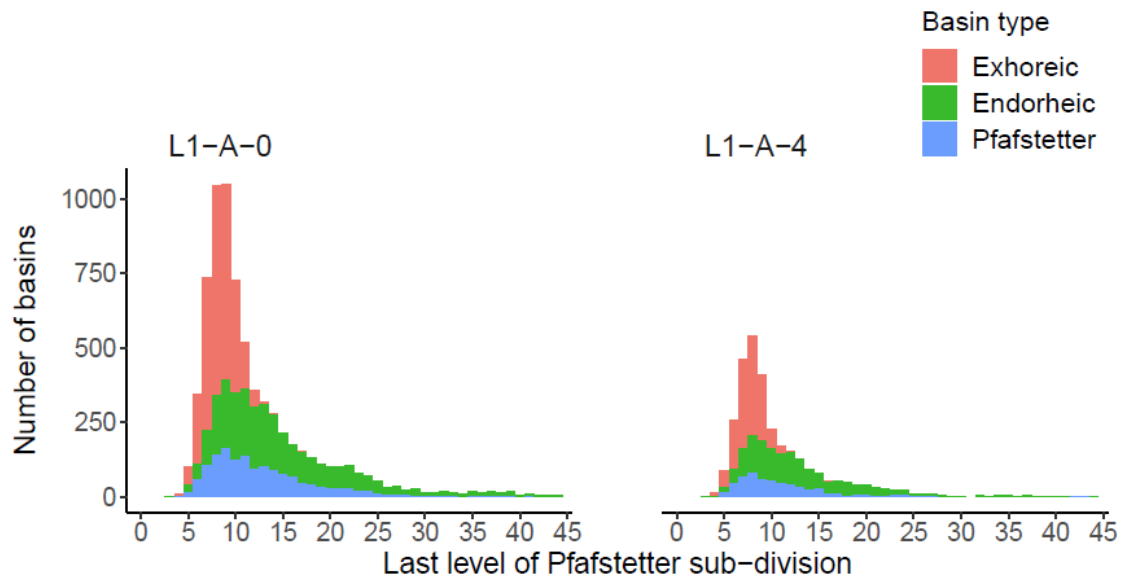


Figure 7 a) Numbers of basins first independently coded as either one of the four largest exhoreic basins or the largest endorheic basin at each level of sub-division after applying the Pfafstetter system (left) and the regionalisation (right) as the initial continental division. Basins never independently coded are plotted at level zero. b) Numbers of basins for which level is the last that the basin was sub-divided depending on stopping rule: zero (left) and less than 4 tributaries (right). Test run IDs are shown above each figure. “Pfafstetter basins” include the additional sub-basins delineated for secondary lake inlet streams (Section 2.3) and terminal distributaries (Section 2.4)

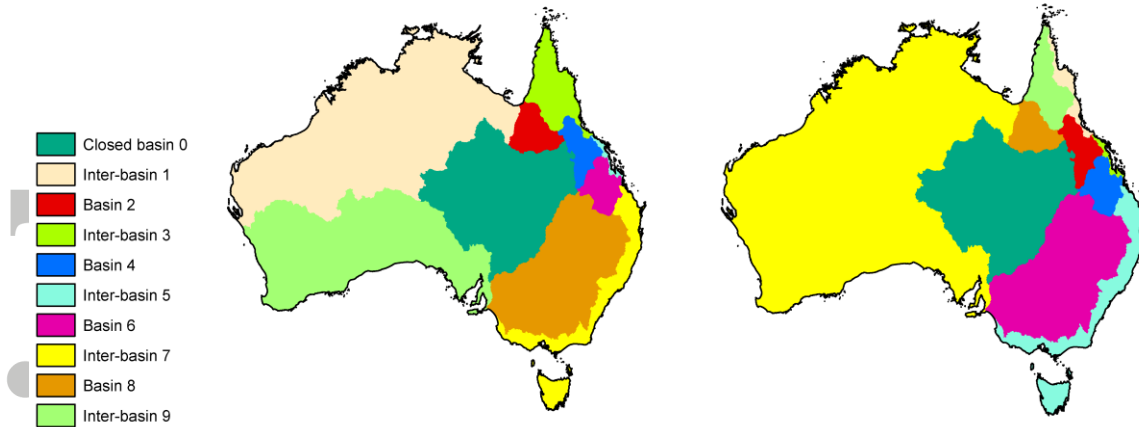


Figure 8 Effect of the start position for the coastal basin numbering system on the first level Pfafstetter sub-division: coding Basin 2 as the most northerly basin (left, this manuscript) and the most north-easterly basin (right, after Stein, 2006)

Accepted Article