Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

HYDROLOGICAL OUTLOOK

Brief description of seasonal river flow forecasts using persistence and historical analogy

2013

NERC SCIENCE OF THE ENVIRONMENT

Funding

The research described in this publication has been funded by the Natural Environment Research Council.

Author

Cecilia Svensson, Centre for Ecology & Hydrology

Publication Address

Centre for Ecology & Hydrology
Maclean Building
Benson Lane
Crowmarsh Gifford
Wallingford
Oxfordshire
OX10 8BB
UK

General and business enquiries: +44 (0)1491 692562
E-mail: enquiries@ceh.ac.uk

Revision date: 24 March 2014

Contents

1 Foreword 4

2 General methodology 4

3 Technical details 4

4 Interpretation of the forecast map 6

5 Example monthly information sheets 12

1 Foreword

This document describes one method of generating a seasonal river flow forecast that is used to inform the UK Hydrological Outlook.

2 General methodology

The historical flow analogues approach to seasonal river flow forecasting is based on selecting the previously observed sequences of flows that are the most similar to the recently observed past. The assumption is that this similarity will carry on in the coming few months. New one- and three-month forecasts are made each month using monthly river flows at 93 stations in the National Hydrological Monitoring Programme. These stations have at least 30 years of data in the period from January 1883 onwards. The locations of the stations are shown on the sample forecast map in Figure 1. The bulk of the forecasts are persistence forecasts, which are made when these outperform the historical analogues approach. They are particularly useful for slowly responding catchments with large underground water storage in aquifers.

3 Technical details

Monthly river flows vary much from one month to the next. Firstly, they vary in terms of the magnitude of the average (climatological mean) flow, with flows tending to be higher in the winter months than in the summer months. But they also vary from one calendar month to the next in terms of their range, with generally higher variability in the winter months than in the summer months (compare, for example, with the seasonal variation of the “normal” flow range in Figure 2). In order to put equal emphasis on the flows from one month to the next when selecting the analogues, it is necessary to transform the data. Standardized (mean = 0, standard deviation = 1) anomalies of log-transformed monthly mean river flows were therefore calculated. The log-transform means that less emphasis is put on the very highest river flows, and makes the distribution around the climatological mean more symmetrical.

When a forecast is made, the monthly anomalies of the past six or nine months, for the one- and three-month forecast respectively, are compared with all possible historical sequences of anomalies covering the same months of the year. Since the calendar months in the recent past has to map onto the same calendar months in the historical record, only one possible analogue is available per year. The five historical analogues most similar to the recent past are selected, based on the root mean square error. Three forecast methods can then be applied to generate the forecast for the next month(s): (i) a weighted mean of the five analogues, (ii) a shifted weighted mean, taking into account the difference between the observed flow in the most recent month, and the weighted

mean of the analogues for the same month, and (iii) persistence of the flow observed in the most recent month. The use of the shifted forecast and the persistence forecast means that forecasts can be made that exceed the historical envelope of observations.

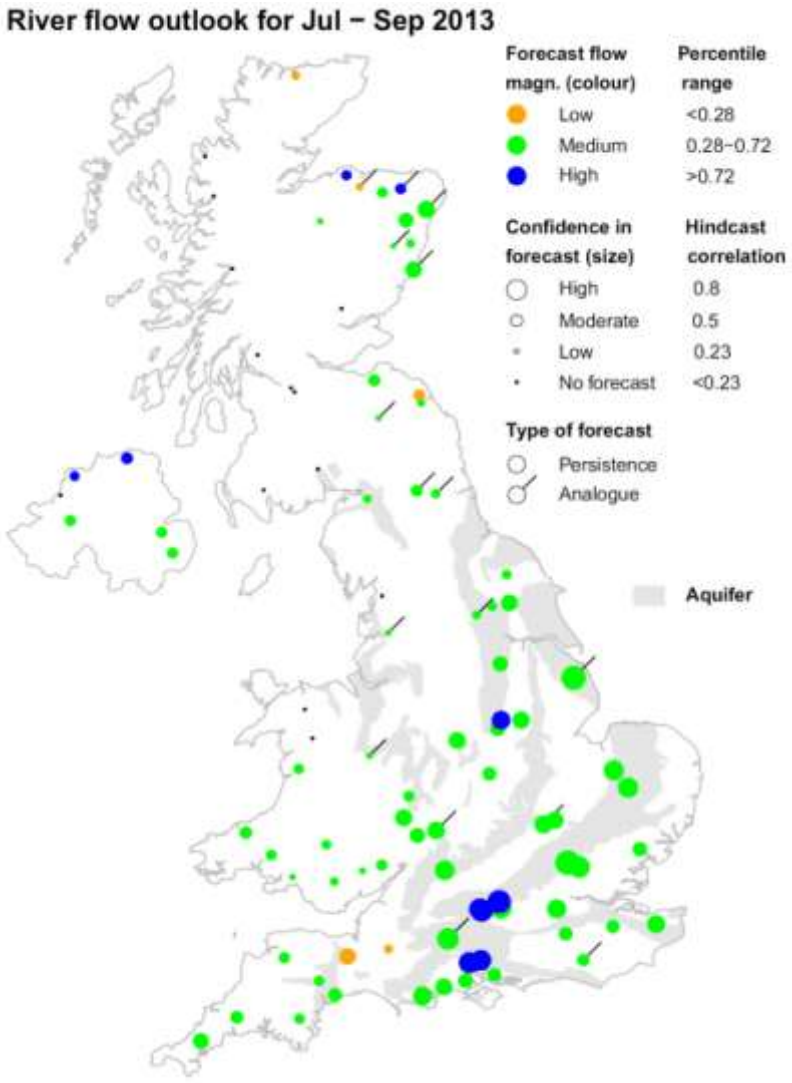


Figure 1. *Sample forecast map: Forecast for July to September 2013. No forecast is available for many of the catchments in the north and west, due to poor performance of the hindcasts (small black dots). Flows are generally forecast to be in or above the normal range in slowly responding catchments on the aquifer outcrops in the southeast, and there is a moderate to high degree of confidence in these forecasts (medium to large green and blue dots, with correlations between the hindcast and observed flows of about 0.5-0.9).*

For each calendar month, the method that has shown the best performance for that month in the past is selected for making the monthly forecast for the coming one or three months (the forecast period). Performance is based on “jack-knife” hindcasts for the entire period of observations

available at each station (30-130 years). This involves dropping one year at a time from the record, and forecasting the flows in this missing year using the remaining years in the record. In this way, a hindcast will be made for each year which can be compared with the flow that was actually observed in that year. For each method, the correlation between its hindcast record and the observed record is calculated for each month, and is used as a measure of its performance. The selected method for each station and month is listed in Table 1. Overall, hindcasts with correlations significant at the 10% level can be made for 81% (70%) of the station-months for the one-month (three-month) forecast period. Of these, 16% (19%) are historical analogue forecasts.

The averaging of the river flow anomalies, over the five analogues and over the three forecast months, that occurs in methods (i) and (ii) results in a reduction of the variance of the forecast flows compared with the observed flows. The hindcast series therefore need re-standardising to obtain series with mean = 0 and standard deviation = 1. The mean and standard deviation from the hindcast series are used to re-scale the forecast when it is converted back to flow (in m^3/s) (Figure 2). However, the general uncertainty in the flow forecast means that this type of time series presentation implies a precision of the forecast which is not justified. Instead, the presentation is restricted to the flows occurring within a low, medium or high interval (Figure 1). It is the empirical distribution of the hindcast series that is used to define the limits between these intervals, which correspond to the 28^{th} and 72^{nd} percentiles. Note that the ranges of these intervals vary with the season (Figure 2).

4 Interpretation of the forecast map

A sample forecast map is shown in Figure 1, with the forecast shown at each river gauging station as a dot of a particular colour (representing flow magnitude) and size (representing confidence in the forecast). There is large uncertainty in the precision of the forecast, and the forecast flow is therefore presented as occurring within one of three intervals: high flows (comprising the highest 28% of flows), medium flows (comprising the middle 44% of flows), and low flows (comprising the lowest 28% of flows). The middle interval spans the same percentiles as the middle interval (the “normal range”) in the Hydrological Summary, whereas the various intervals for increasingly higher/lower flows in the Summary have been collapsed into single intervals for high and low flows, respectively.

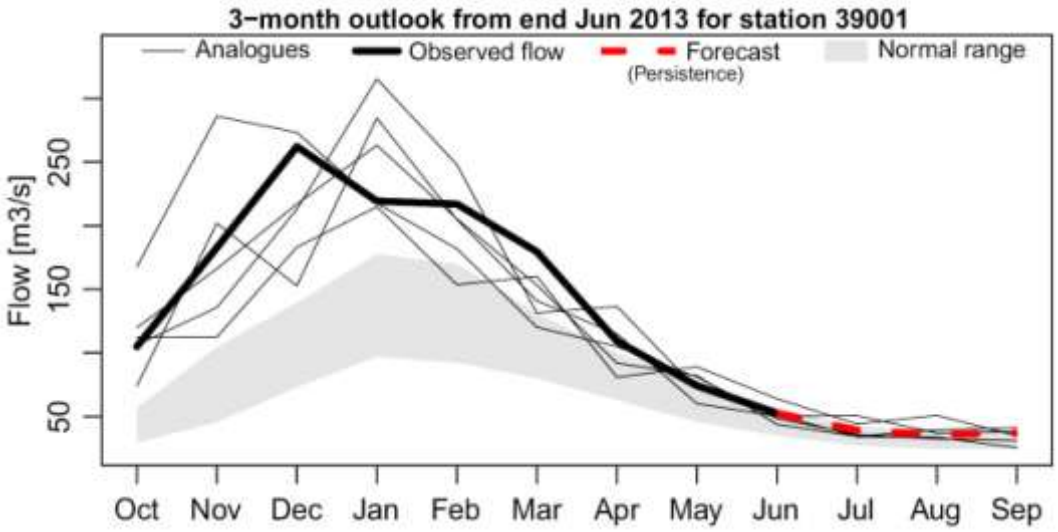


Figure 2. *Forecast for July to September 2013 for the Thames at Kingston (station number 39001). The medium ("normal") range (the middle 44 % of the observed distribution for the recent past, and the middle 44% of the hindcast distribution for the forecast period) is shaded grey. The five closest analogues are shown, although the forecast method used is Persistence.*

Forecasts are presented if the correlation between the hindcasts and the observations exceeds 0.23 and has a significance level of at most 10% (a correlation of 0.23 corresponds to a 10% significance level for a 51-year record, a typical record length). For correlations just above 0.23, there is still a large proportion of the hindcast flows which does not occur in the correct interval. Instead, they occur in the neighbouring interval, and to a lesser extent in the "opposite" interval (for example, a high flow is predicted when, in fact, a low flow is observed). As correlations increase, the risk of predicting the opposite extreme decreases. Contingency tables cross-reference the hindcast flows with the actual observed flows, and some examples are shown in Figures 3 to 5. If all hindcasts were perfect, then all counts would be in the boxes along the diagonals, i.e. the number of hindcast low (medium/high) flows would be the same as the number of observed low (medium/high) flows. The contingency tables complement the correlations in providing a measure of how well the methods perform.

The correlations between hindcasts and observations are generally high in the slowly responding catchments on the aquifer outcrop areas in southeast England, in the order of 0.6 to above 0.9. Here, the best forecast method is generally the persistence of the flow. Conversely, correlations are often poor in the north and west, particularly for spring (Table 1). In these areas, the lack of underground water storage means that the flow in one month does not greatly influence the flow in the coming months. Forecasts are often not possible for several months of the year, and the correlations are generally lower than for the southeast of the country.

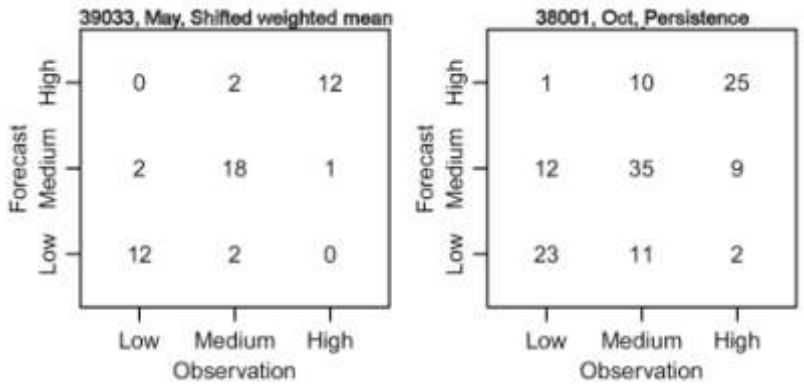


Figure 3. *Examples of contingency tables for hindcasts and observed river flows for which the correlations are high. The middle column/row comprises 44% of the occurrences, and the outer columns/rows (low and high intervals) comprise 28% each. In these particular cases, the correlations are 0.93 (left) and 0.69 (right). Note that there is no strict relationship between the correlation and the values in the contingency tables, these will vary from case to case.*

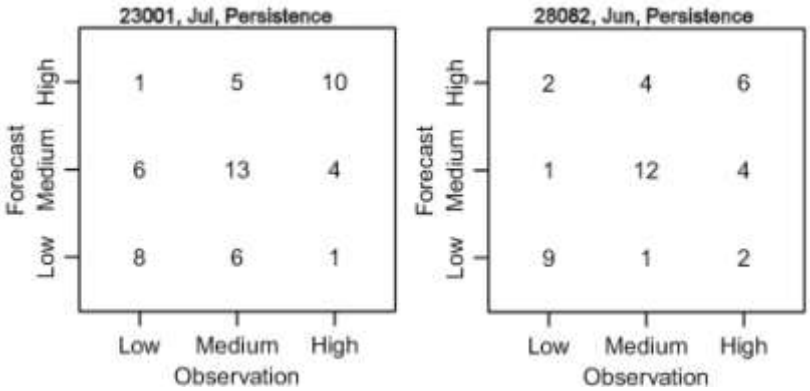


Figure 4. *As in Figure 3, but for hindcasts and observed river flows for which the correlations are moderately high. In these particular cases, the correlations are 0.49 (left) and 0.54 (right).*

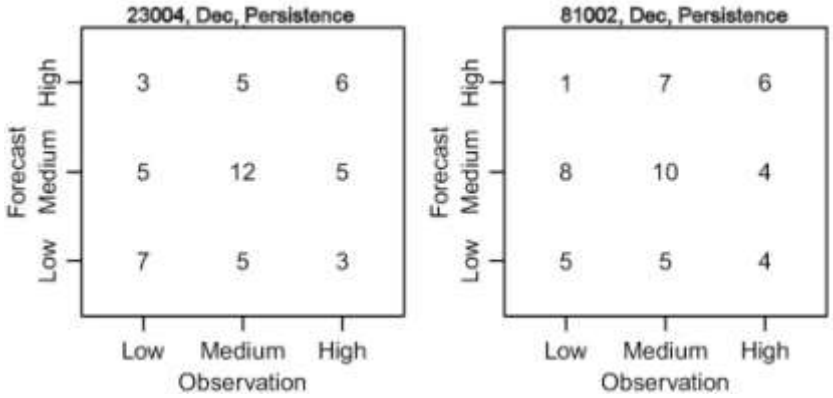


Figure 5. *As in Figure 3, but for hindcasts and observed river flows for which the correlations are low. In these particular cases, the correlations are 0.23 (left) and 0.24 (right).*

Table 1. *Codes for the best forecast method for each river flow station and month, based on jack-knife hindcasts made in July 2013. The codes for the methods are 1= weighted mean, 2= shifted weighted mean, 3= persistence, and 5= no forecast possible. The month is the last month of the recent past (i.e. the forecast period is for the following one or three months). The six-digit station numbers are for Northern Ireland. National Grid coordinates are in 100m units, with Northern Ireland stations in the Irish Grid coordinate system.*

Station	Name	East (hm)	North (hm)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
7003	Lossie at Sheriffmills	3194	8626	3	3	5	3	3	3	3	3	3	3	1	1
8005	Spey at Boat of Garten	2947	8192	3	5	3	2	5	3	3	3	3	5	5	5
8006	Spey at Boat o Brig	3319	8518	5	5	5	2	1	1	3	3	3	2	5	5
9001	Deveron at Avochie	3532	8464	3	2	5	3	2	3	3	3	3	2	1	3
9002	Deveron at Muiresk	3706	8498	3	2	5	3	2	2	3	3	3	3	1	3
10003	Ythan at Ellon	3947	8304	3	2	5	3	3	2	3	3	3	3	3	3
11002	Don at Haughton	3757	8202	2	3	5	3	3	3	3	3	3	3	2	3
12001	Dee at Woodend	3635	7956	5	5	5	5	5	2	3	5	3	5	5	5
12002	Dee at Park	3797	7983	5	5	5	5	5	3	3	5	3	5	5	5
13001	Bervie at Inverbervie	3826	7734	3	5	1	2	5	2	3	3	3	3	5	3
15006	Tay at Ballathie	3148	7367	3	5	3	5	3	5	2	3	3	5	5	5
20003	Tyne at Spilmersford	3456	6689	2	5	5	3	1	3	3	3	3	3	3	3
21006	Tweed at Boleside	3498	6334	5	5	5	5	5	2	3	3	3	5	5	5
21009	Tweed at Norham	3898	6477	5	5	5	5	3	3	3	3	3	3	5	5
21022	Whiteadder Water at Hutton Castle	3881	6550	2	2	5	3	3	3	2	3	3	3	3	3
23001	Tyne at Bywell	4039	5617	5	5	5	3	3	2	3	3	3	5	5	2
23004	South Tyne at Haydon Bridge	3857	5647	5	5	5	5	5	2	3	2	3	5	5	3
27002	Wharfe at Flint Mill Weir	4422	4473	5	5	2	3	3	2	3	3	3	2	5	3
27009	Ouse at Skelton	4568	4554	5	5	2	2	5	3	3	2	3	5	5	5
27041	Derwent at Buttercrambe	4731	4587	2	1	2	3	3	3	3	3	3	1	3	3
27042	Dove at Kirkby Mills	4705	4855	2	1	2	3	3	3	3	2	3	3	3	3
28009	Trent at Colwick	4620	3400	2	1	3	3	3	3	2	3	3	3	3	3
28018	Dove at Marston on Dove	4237	3288	2	1	3	3	2	3	3	3	3	3	5	3
28050	Torne at Auckley	4646	4012	3	3	2	3	3	3	3	3	2	3	3	2
28060	Dover Beck at Lowdham	4653	3480	3	3	3	3	3	3	3	3	3	3	3	3
28082	Soar at Littlethorpe	4542	2973	3	3	3	3	3	3	3	3	2	3	2	3
29003	Lud at Louth	5337	3879	3	3	3	3	3	2	2	3	3	3	2	3
30001	Witham at Claypole Mill	4842	3480	3	3	3	3	3	3	3	3	3	3	3	3
33002	Bedford Ouse at Bedford	5054	2496	3	1	2	2	2	2	2	2	3	3	3	3

Station	Name	East (hm)	North (hm)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
33029	Stringside at Whitebridge	5716	3006	3	3	3	3	3	3	3	3	3	3	3	3
33034	Little Ouse at Abbey Heath	5851	2844	3	1	3	3	2	3	3	3	3	3	3	3
33039	Bedford Ouse at Roxton	5160	2535	3	3	3	3	2	3	3	2	3	3	3	3
37005	Colne at Lexden	5962	2261	1	2	2	3	2	3	3	3	3	3	3	3
38001	Lee at Feildes Weir	5391	2092	1	1	3	3	3	3	3	3	3	3	3	3
38003	Mimram at Panshanger Park	5283	2133	3	3	3	3	3	3	3	3	3	2	3	3
39001	Thames at Kingston	5178	1699	3	1	3	3	3	3	3	3	3	3	3	3
39016	Kennet at Theale	4650	1708	3	3	3	3	3	3	3	3	3	3	3	3
39019	Lambourn at Shaw	4470	1682	3	3	3	3	3	3	3	3	3	3	3	3
39020	Coln at Bibury	4121	2062	3	3	3	3	3	3	3	3	3	3	3	3
39027	Pang at Pangbourne	4635	1766	3	3	3	3	3	3	3	2	3	3	3	3
39033	Winterbourne St at Bagnor	4453	1695	3	3	3	2	2	3	3	3	3	3	3	3
39069	Mole at Kinnersley Manor	5262	1462	3	5	5	3	5	3	2	5	3	3	2	3
40003	Medway at Teston / East Farleigh	5709	1530	3	5	3	3	3	3	3	3	3	3	3	3
40011	Great Stour at Horton	6115	1554	3	1	3	3	3	3	3	3	3	3	3	3
41005	Ouse at Gold Bridge	5429	1214	3	5	2	3	2	2	2	3	3	3	2	3
42001	Wallington at North Fareham	4587	1075	3	3	3	3	3	3	3	3	3	3	3	3
42003	Lymington at Brockenhurst	4318	1019	3	5	2	3	3	3	3	5	3	3	3	3
42004	Test at Broadlands	4354	1189	3	3	1	3	3	3	3	3	3	3	3	3
42010	Itchen at Highbridge & Allbrook Total	4461	1211	3	3	3	3	3	3	3	3	3	3	3	3
43005	Avon at Amesbury	4151	1414	2	2	3	3	3	2	3	3	3	3	3	3
43007	Stour at Throop	4112	960	3	3	3	3	3	3	3	3	3	3	3	3
44002	Piddle at Baggs Mill	3913	876	3	3	3	3	3	3	3	3	3	3	3	3
45001	Exe at Thorverton	2936	1016	5	5	5	5	3	3	3	5	3	2	5	3
45005	Otter at Dotton	3087	884	2	3	5	3	3	3	3	1	3	3	2	3
46003	Dart at Austins Bridge	2751	658	5	5	2	5	3	3	3	3	3	5	5	3
48004	Warleggan at Trengoffe	2160	674	3	5	2	3	3	3	3	3	3	3	5	3
48005	Kenwyn at Truro	1820	450	3	5	5	3	2	3	3	3	3	3	5	3
50001	Taw at Umberleigh	2608	1237	5	5	5	5	3	3	3	3	3	2	1	3
52005	Tone at Bishops Hull	3206	1250	3	3	5	3	3	3	3	3	3	3	3	3
52010	Brue at Lovington	3590	1318	3	5	3	3	3	3	3	3	3	3	3	3

Station	Name	East (hm)	North (hm)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
54001	Severn at Bewdley	3782	2762	5	5	2	5	3	3	3	3	3	3	5	3
54002	Avon at Evesham	4040	2437	3	1	3	3	3	2	2	3	3	3	1	3
54005	Severn at Montford	3412	3145	5	5	5	5	5	2	3	3	2	2	5	3
54029	Teme at Knightsford Bridge	3735	2557	3	5	5	3	3	3	3	3	3	3	3	3
54032	Severn at Saxons Lode	3863	2390	3	5	5	3	3	3	3	3	3	3	5	3
55023	Wye at Redbrook	3528	2111	5	5	2	3	3	3	3	3	2	2	5	3
56001	Usk at Chain Bridge	3346	2056	5	5	5	5	3	3	3	5	2	5	5	3
56013	Yscir at Pontaryscir	3003	2304	5	5	5	5	5	3	3	2	5	5	5	3
57004	Cynon at Abercynon	3079	1957	5	5	5	5	3	3	3	5	3	5	5	3
59001	Tawe at Ynystanglws	2686	1998	5	1	5	3	5	3	5	3	5	2	5	3
60010	Tywi at Nantgaredig	2485	2206	5	5	5	5	3	3	3	5	3	2	5	3
62001	Teifi at Glan Teifi	2244	2417	5	5	5	5	2	3	3	3	2	2	5	3
64001	Dyfi at Dyfi Bridge	2745	3020	5	5	5	3	5	3	3	5	5	5	5	2
66011	Conwy at Cwmlanerch	2801	3581	5	5	5	3	5	5	3	3	1	1	5	3
67018	Dee at New Inn	2874	3308	5	5	5	3	5	5	5	5	2	1	1	5
71001	Ribble at Samlesbury	3589	4304	5	5	5	5	5	2	2	3	3	1	5	3
72004	Lune at Caton	3529	4653	5	5	5	3	5	5	3	2	3	5	5	3
76007	Eden at Sheepmount	3390	5571	5	5	5	5	5	3	3	2	3	1	5	5
79002	Nith at Friars Carse	2923	5851	5	5	5	5	5	5	3	3	3	5	5	5
81002	Cree at Newton Stewart	2413	5653	5	5	5	5	5	5	2	3	5	1	5	3
84005	Clyde at Blairston	2704	6579	3	5	5	2	5	5	3	3	3	5	5	3
84013	Clyde at Daldowie	2672	6617	3	5	5	2	5	5	3	3	3	5	5	3
85004	Luss Water at Luss	2356	6929	2	5	5	5	5	5	5	5	3	5	5	5
90003	Nevis at Claggan	2116	7743	3	5	5	5	5	5	5	5	1	5	5	5
93001	Carron at New Kelso	1942	8429	3	5	5	5	5	5	5	5	2	5	5	3
94001	Ewe at Poolewe	1860	8803	3	5	5	5	5	5	5	5	2	5	5	5
96002	Naver at Apigill	2714	9568	5	5	5	5	5	3	5	5	3	2	5	2
201005	Camowen at Camowen Terrace	2461	3730	5	5	5	5	3	3	5	3	3	5	5	5
201010	Mourne at Drumnabuoy House	2348	3960	5	5	5	5	5	5	5	5	5	1	5	5
202002	Faughan at Drumahoe	2464	4151	5	5	5	5	5	3	3	3	3	5	5	5
204001	Bush at Seneirl Bridge	2942	4363	5	5	5	5	3	3	3	3	3	5	5	5
205004	Lagan at Newforge	3329	3693	5	5	5	5	3	3	3	5	3	5	5	5
205011	Annacloy at Kilmore Bridge	3449	3508	5	5	5	5	5	3	3	5	3	1	5	5

5 Example monthly information sheets

The information generated is summarised on a set of three information sheets. The first is a summary sheet containing two figures as described in Section 3, one for the one month ahead outlook and the other for a three month period. Key points are note are marked on the annotated figure below.

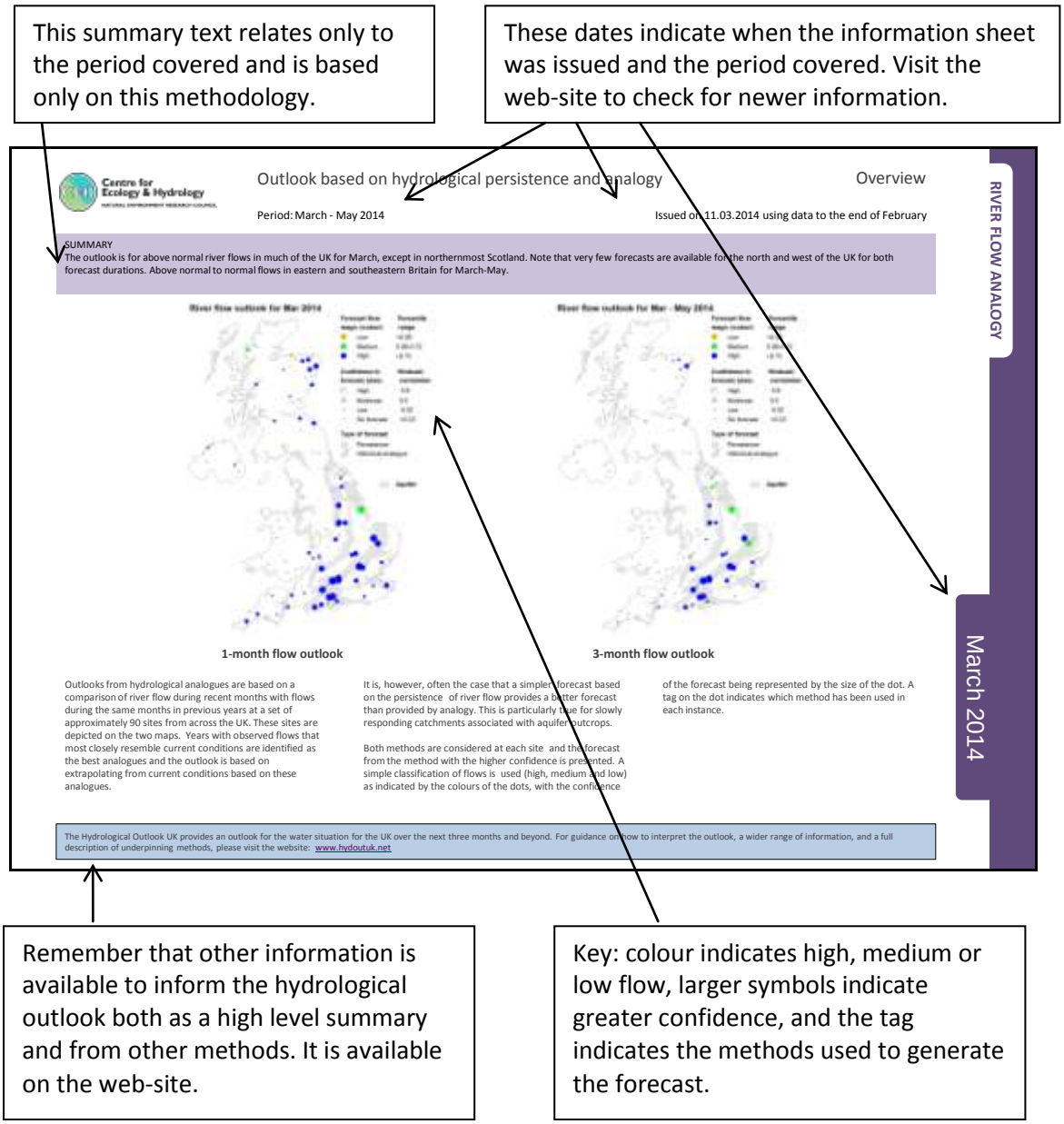
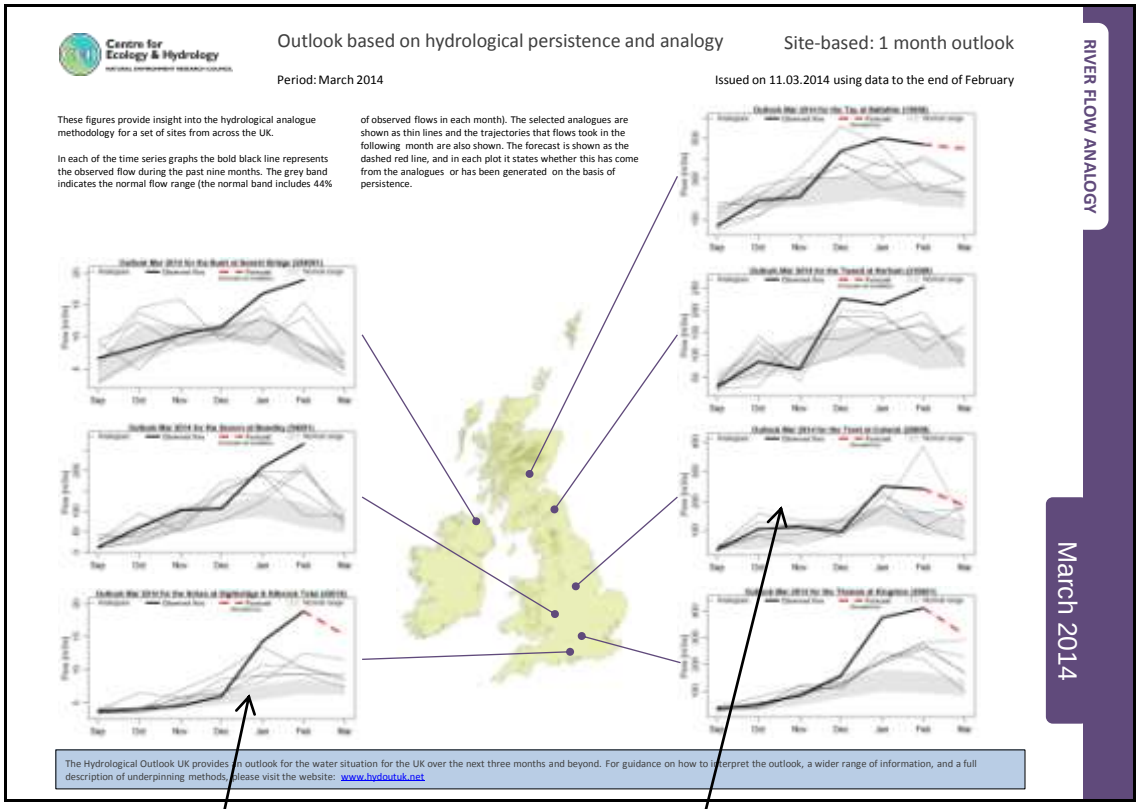


Figure 6. *Overview information sheet for the outlook using persistence and historical analogues.*

An example of the additional information sheet for the one month ahead outlook using the persistence and analogy methodology is shown in Figure 7.

The example flow hydrographs illustrate for the particular month the normal flow range (the grey band includes 44% of observed historical flows), the recent flow in bold, and the five selected analogues.



In this particular example the flow has been increasing steadily for the past 4 months. In September the flow was normal, but since December flow has been in the above normal range. The historical analogues look similar to the recent flow record until December, but statistically it is best to project forward on the basis of persistence.

In this example the recent flow was normal in November, above normal in January. The historical analogues are not a good match with the recent flows and how these continued into February gives a good indication of the wide range of possible future flows.

Figure 7. *Information sheet with examples of one month ahead forecasts using persistence and historical analogues.*

An example of the additional information sheet for the three month ahead outlook using the persistence and analogy methodology is shown in Figure 8. This information sheet contains the same elements as for the one month ahead outlook. The difference in the example hydrographs is that as well as providing a forecast for the next three months, the previous nine months' flow is shown as this is the period used to select the analogues.

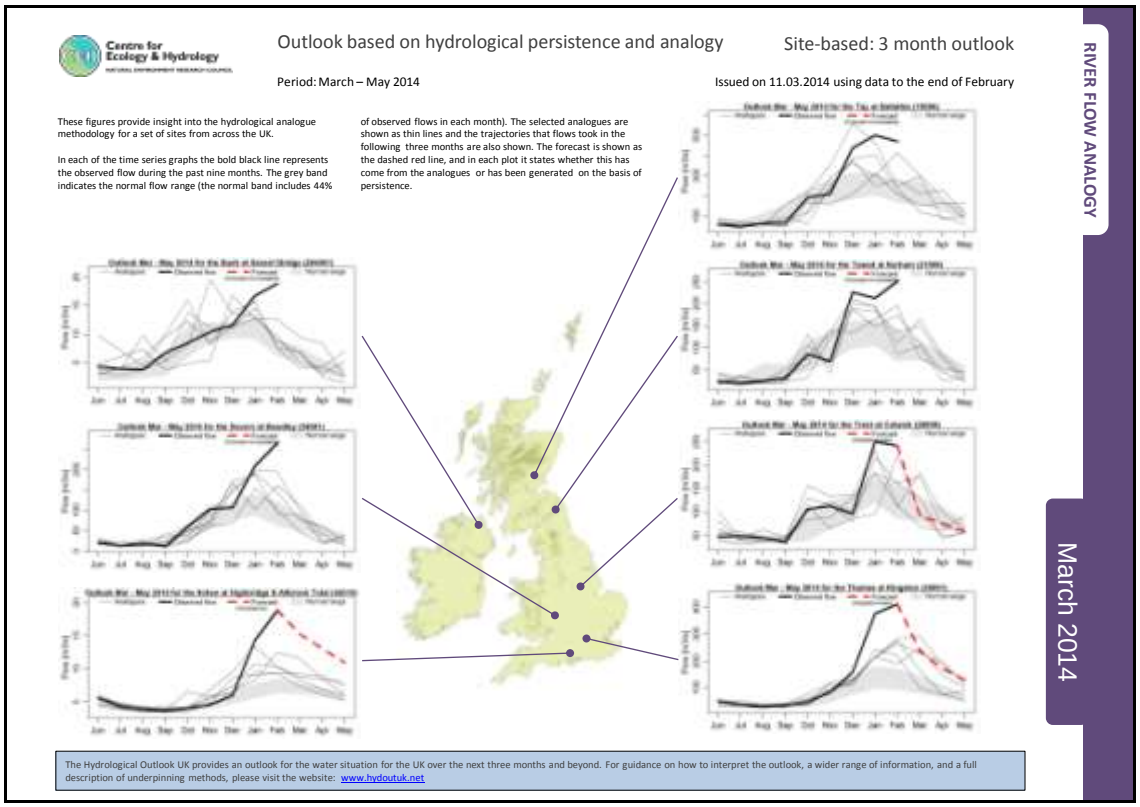


Figure 8. *Information sheet with examples of three month ahead forecasts using persistence and historical analogues.*