Figure S1. Map of Northern England domain used in this study, taken as region of England grid north of Chester (53.19°N, 2.89°W). Local Authority boundaries (black lines), HydroBASINS Level 9 catchments (grey lines) and FOREWARNS rainfall-field sampling locations (pink circles) shown. The Ordnance Survey MiniScale® basemap contains public sector information licensed under the Open Government Licence v3.0.

S1 Supplementary details for user workshop

Workshop case study (r30, p98) FOREWARNS forecasts

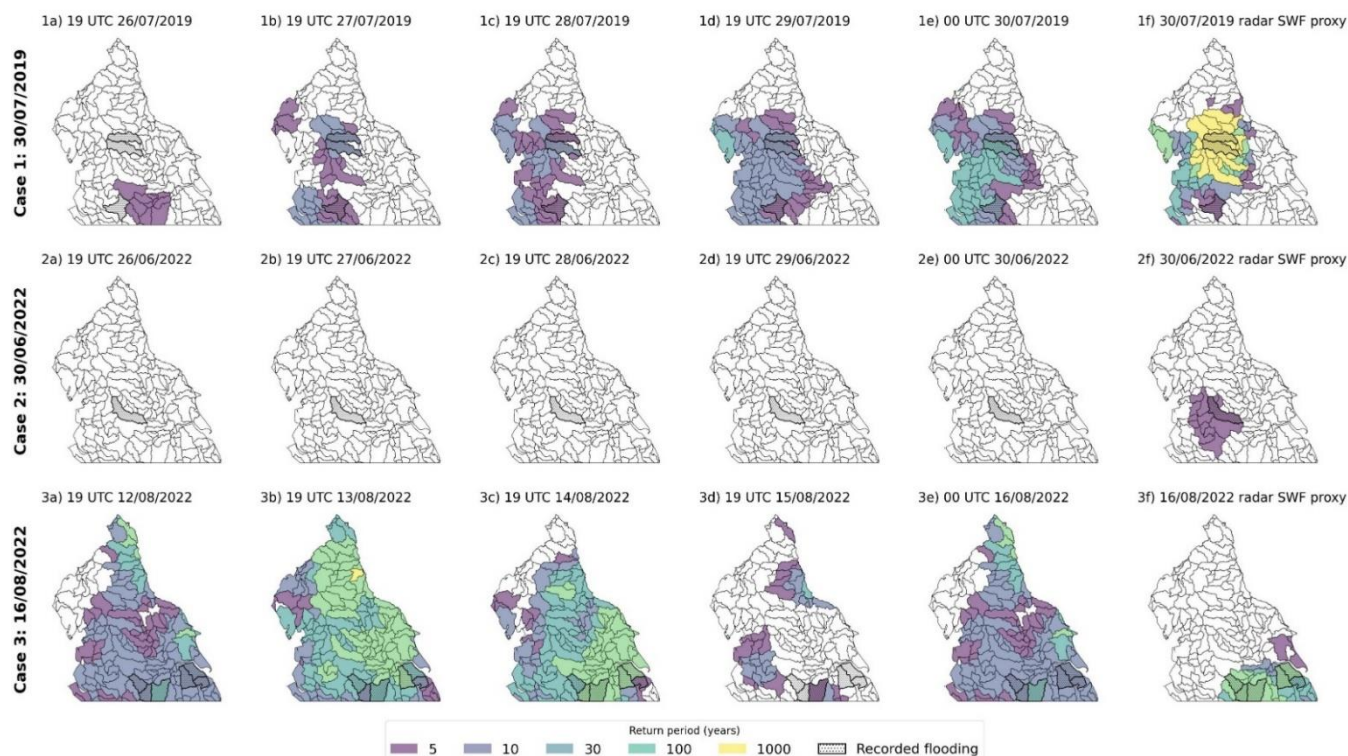



Figure S2. (r30, p98) FOREWARNS forecasts available at decreasing lead times for the three workshop case study flood events on 30/09/2019, 30/06/2022 and 16/08/2022, respectively. Forecasts valid for these dates are grouped by corresponding row number. Common lead times of 4 days, 3 days, 2 days, 1 day, and midnight the day of the event, are grouped by columns. Forecasts in columns (a) to (d) are based on the 15:00 UTC MOGREPS-UK ensemble from the date in header, where FOREWARNS would be available to users at approximately 19:00 UTC that day. Forecasts in column (e) based on 20:00 UTC MOGREPS-UK ensemble from day before event, where FOREWARNS would be available to users at approximately 00:00 UTC the same day. Column (f) shows (r30, p98) radar SWF proxy for the event. Recorded flood locations are shown by stippled catchments in forecasts and proxies (all panels).

20 **Table S1: Questions posed to workshop participants in online debrief survey. Survey conducted at the end of workshop activity.**

Number	Question
1	What is your name? Please leave blank if you wish to remain anonymous.
2	What organisation do you primarily work or volunteer for?
3	What type of organisation do you primarily work or volunteer for? *[Options: Local authority; Water company; Forecast provider; Emergency services; Community flood group; University; Other (please describe)]*
4	What is your primary job/voluntary position title?
5	What are your SWF duties?
6	Overall, how useful did you find the National Severe Weather Warning Service (NSWWS) products for informing your decision making in the workshop case studies? *[Rated 1 to 5, where 1 is "Very unhelpful" and 5 is "Very useful"]*
7	Overall, how useful did you find the Flood Guidance Statement (FGS) products for informing your decision making in the workshop case studies? *[Rated 1 to 5, where 1 is "Very unhelpful" and 5 is "Very useful"]*
8	Do you have any comments regarding the NSWWS or FGS warnings?
9	How strongly do you agree with the following statement: "The enhanced forecast information would have made a difference to my decision making prior to the flood event in the Yorkshire Dales, July 2019"? *[Rated 1 to 5, where 1 is "Strongly disagree" and 5 is "Strongly agree"]*
10	How strongly do you agree with the following statement: "The enhanced forecast information would have made a difference to my decision making prior to the flood event in Shipley, June 2022"? *[Rated 1 to 5, where 1 is "Strongly disagree" and 5 is "Strongly agree"]*
11	How strongly do you agree with the following statement: "The enhanced forecast information would have made a difference to my decision making prior to the flood event in Sheffield, August 2022"? *[Rated 1 to 5, where 1 is "Strongly disagree" and 5 is "Strongly agree"]*
12	Overall, how useful would the enhanced forecast information be to your organisation? *[Rated 1 to 5, where 1 is "Not very useful" and 5 is "Very useful"]*
13	In your view, what added value (if any) did the enhanced forecasts provide over the operational (NSWWS, FGS, CCA) inputs?
14	From the enhanced forecasts, name one aspect that you found useful in your decision making (if any).
15	From the enhanced forecasts, name one aspect that you did not find useful in your decision making (if any).

16	Was the enhanced forecast information easy to interpret?
	[Yes; No]
17	What aspect of the enhanced forecast information was particularly difficult to interpret? How could that information be made easier to interpret?
18	What aspect of the enhanced forecast information particularly supported your interpretation?
19	How would enhanced surface water flood forecasts up to **3 days in advance** be used by your organisation?
	[Options: Would not use; Used as part of routine forecast checks; Used to monitor situation if FGS or NSWWS warning in place; Used for action planning; Other (please describe)]
20	How would enhanced surface water flood forecasts up to **1 day in advance** be used by your organisation?
	[Options: Would not use; Used as part of routine forecast checks; Used to monitor situation if FGS or NSWWS warning in place; Used for action planning; Other (please describe)]
21	Where applicable, what hypothetical level of forecast accuracy would be required to take each of the actions listed below?
	[Likert table of **actions** (listed below) vs **accuracy ratings**: <20%; 20-40%; 40-60%; 60-80%; >80%; n/a.]
22	Where applicable, what is the minimum advance warning of potential flooding required by your organisation to take the same actions listed below?
	[Likert table of **actions** (listed below) vs **lead times:** 4 days; 3 days; 2 days; 1 day; Same day; During/after the event; n/a.]
23	What kinds of evidence would you need to see for your organisation to gain confidence in using the enhanced forecast system?
24	Do you have any surface water flood records (since 2013) which we could use to aid evaluation of the enhanced forecasts over a 10 year period? If yes, how could we access this information?

Actions listed in Likert tables (following Ochoa-Rodríguez *et al.*, 2018): *Clearing trash screens; Placing street scene crews and extra resources on standby; Deployment of temporary flood defences (eg sandbags); Monitoring status of pumping stations; Ongoing event monitoring in working hours; Event monitoring outside of working hours; Notification of contractors and partners; Notification of flood wardens; Notification of public; Road closures; Closure of public locations susceptible to pluvial flooding (e.g. passages); Other (please describe below).*

S2 Supplementary forecast verification materials

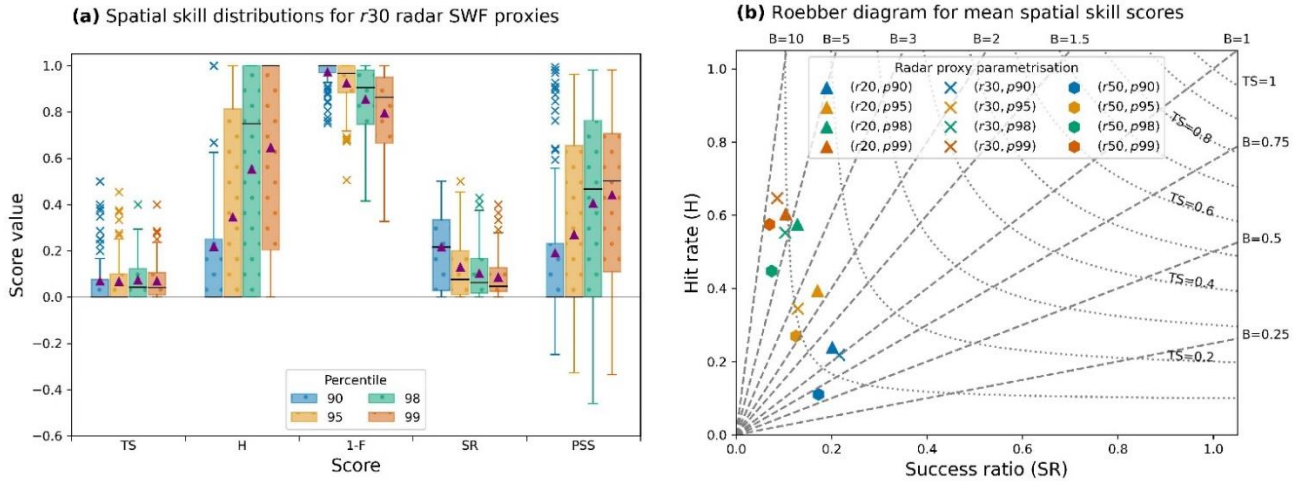



Figure S3. Spatial skill scores for radar SWF proxies computed (for all SWF return periods) against recorded flood locations for 82 days from May–October, 2013–2022. (*a*) Distributions of *TS, H, 1-F, SR* and *PSS* for *r*30 proxies, grouped by increasing percentile (indicated by shading). Mean (median) values are shown by purple triangles (black horizontal lines). All measures are equitable (worst score 0, perfect score 1) except PSS, for which random forecasts score 0 and the worst score is -1. (*b*) Roebber performance diagram showing mean spatial scores for 12 radar proxy parametrisations. Marker style indicates common radius *r*, colour shading indicates common percentile *p*.

Verification requires a common observational dataset against which all forecasts are benchmarked. To identify a single radar SWF proxy parameterisation we compute spatial skill measures for different proxies by calculating contingency tables against catchment-level locations of SWF for the 82 days with recorded flooding. Figure S3a shows the distributions of spatial skill scores for *r*30 proxies with increasing percentile *p*. Both *TS* and *SR* show generally poor skill distribution, with medians and mean values always close to zero. This reflects the intrinsically high rate of spatial false alarms, which is to be expected when comparing a RWCRS against a lower bound on flood occurrence. For a proxy SWF observation set intended as the upper bound, capturing known instances of flooding is more desirable than minimisation of spatial false alarms. The greatly improved skill distributions for *H* offer reassurance that this is achieved. The clear improvement at higher percentiles is reflected by the greatly improved distributions of *PSS*, with the small improvement from *p*98 to *p*99 reflecting the deterioration in *F*.

Figure 3b displays mean score values for all parametrisations, plotted on a Roebber diagram. Similarly to Fig. 5, increasing percentile *p* has a stronger effect on the hit rate *H* than the radius, *r*. Although highest *TS* values are here shown at low percentiles, these are accompanied by low hit rates. Given the need for high *H*, and acceptability of accompanying high spatial false alarms in this context, we adopt (*r*30, *p*98) as the standard radar SWF proxy. We necessarily expect the radar proxy to significantly overestimate SWF occurrence: the recorded flood locations against which we conduct proxy verification represent the absolute minimum extent.

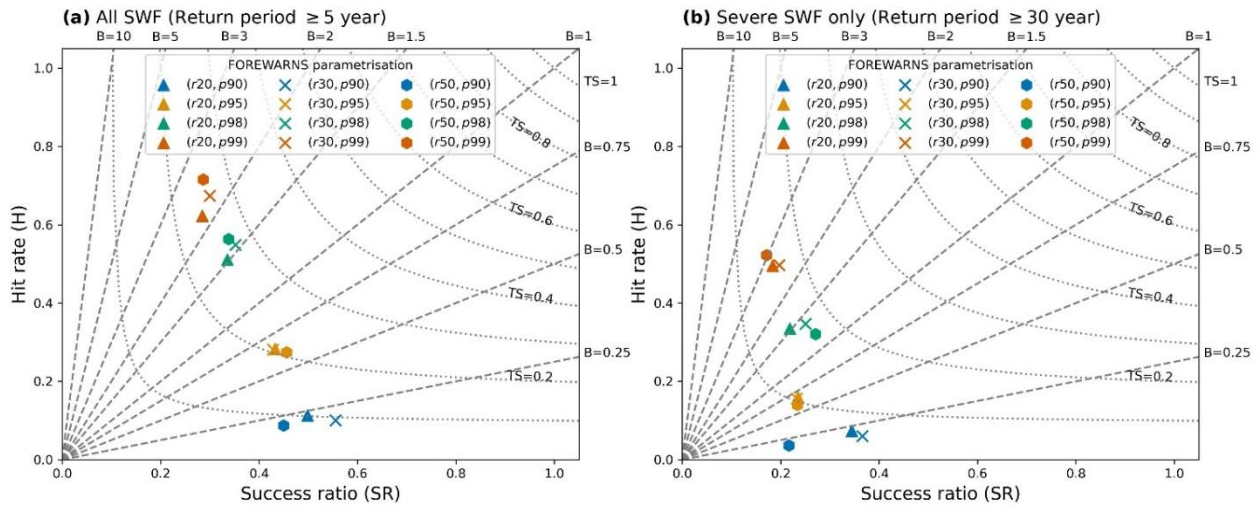



Figure S4. Roebber performance diagrams of mean spatial scores for 12 FOREWARNS forecast parametrisations, for 82 days with recorded flooding from May–October, 2013–2022. (*a*) Skill scores for all SWF return periods, computed against radar SWF proxy. Marker style indicates common radius parameter *r*, colour shading indicates common percentile *p*. All forecasts based on 15:00 UTC MOGREPS-UK ensemble issued the day before an event. (*b*) Repeated for severe SWF return periods (30 years or higher) only.

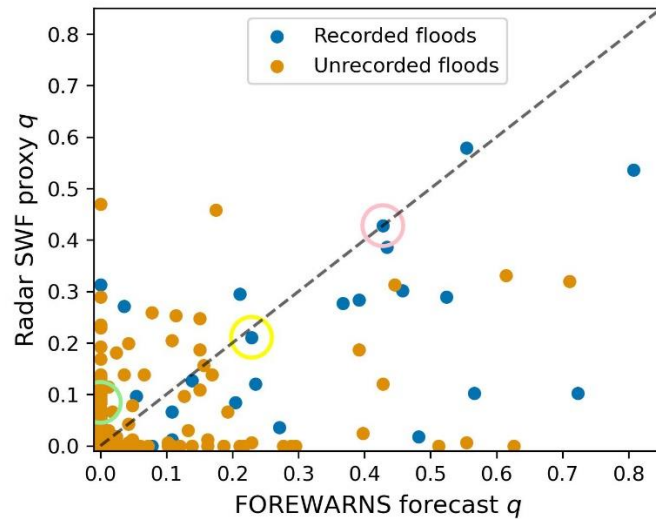



Figure S5. Scatter plot of observed versus forecast *q*, where *q* is the SWF coverage of Northern England catchments (proportion highlighted as showing SWF) on a given day, for May–October, 2019–2022. Observations based on radar SWF proxy, while (*r*30, *p*98) FOREWARNS forecasts are based on 15:00 UTC MOGREPS-UK ensemble and would be available to users at approximately 19:00 UTC, valid next day. Points for 41 days with recorded flooding are coloured blue. User workshop case study days circled: case 1 (30/07/2019), pink; case 2 (30/06/2022), green; case 3 (16/08/2022), yellow.

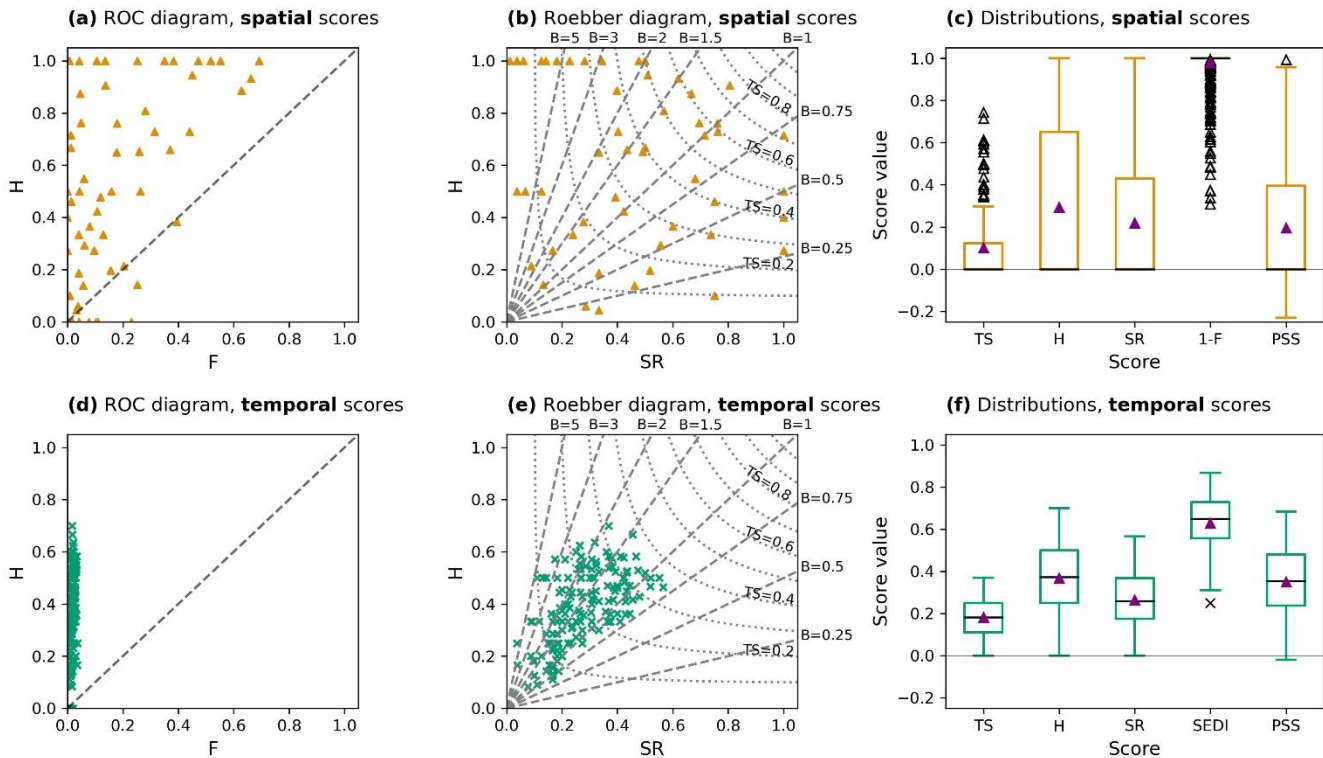



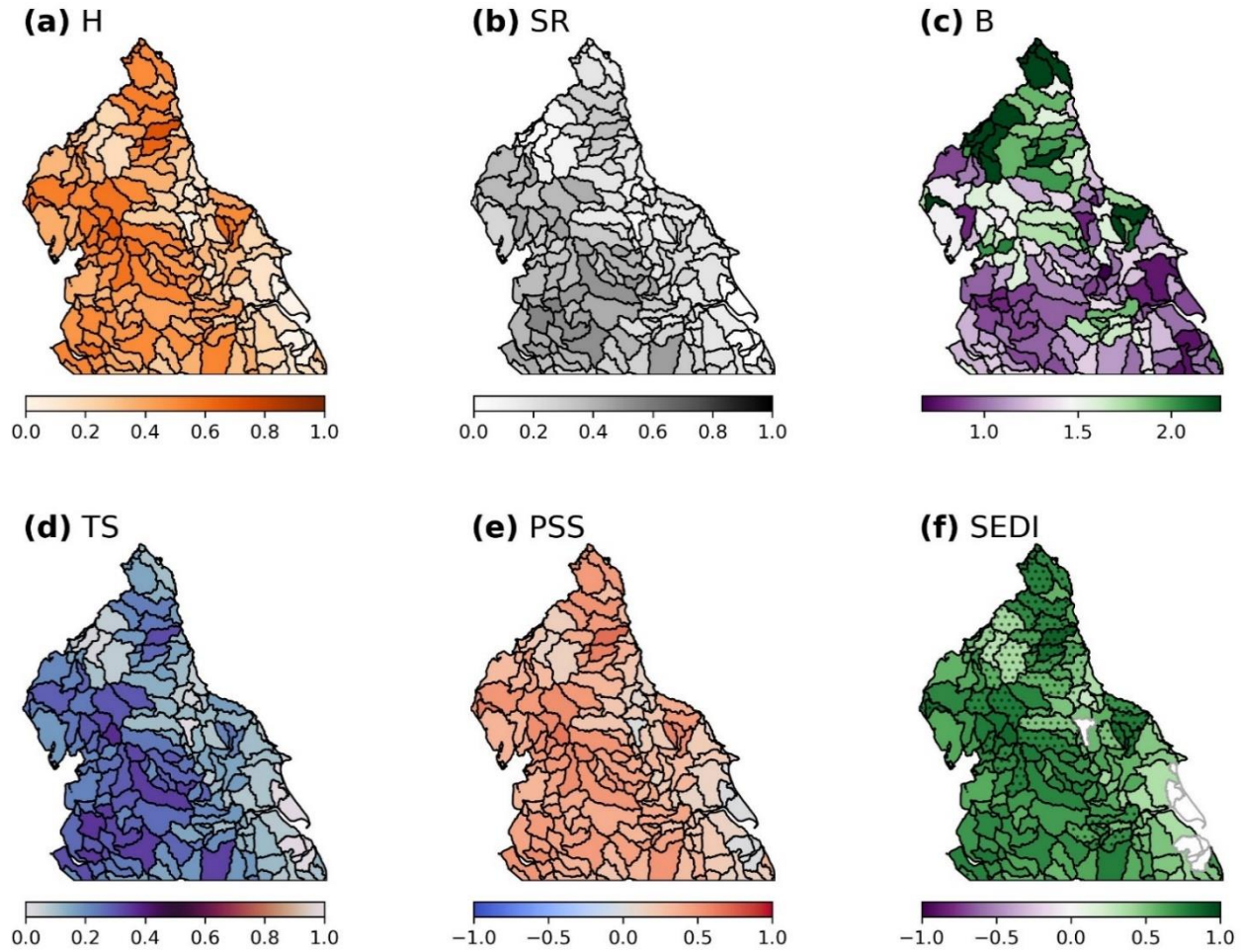
Figure S6. Skill diagrams for spatial (top row) and temporal (bottom row) skill scores for May–October, 2019–2022 (*r*30, *p*98) FOREWARNS forecasts evaluated against radar SWF proxy. All forecasts based on 15:00 UTC MOGREPS-UK ensemble and would be available to users at approximately 19:00 UTC, valid next day. (*a*) ROC diagram of spatial (*F*, *H*) values for the 725 individual forecasts – 570 values at origin due to no catchment-level events. Perfect forecasts lie towards top left of diagram. (*b*) Roebber diagram of spatial skill for same sample. Here perfect forecasts lie towards top right of diagram. (*c*) Skill distributions for spatial *TS*, *H*, *SR*, 1-*F* and *PSS* for same sample. Mean (median) values indicated by purple triangles (black lines) – trivial median values reflect 570 days with no catchment-level events. All measures are equitable (worst score zero, perfect score one) except *PSS*, for which random forecasts score zero and the worst score is -1 (*d*) ROC diagram of temporal (*F*, *H*) values for 166 catchments, calculated over full forecast period [repeat of Figure 8a]. (*e*) Roebber diagram for same sample [repeat of Figure 8b]. (*f*) Skill distributions for temporal *TS*, *H*, *SR*, *SEDI* and *PSS* scores for sample. *SEDI* distribution takes same value range as *PSS* but excludes catchments where *B*<0.67 or *B*>1.5.

80

Figure S7. Spatial distribution of temporal skill scores for (*r30, p98*) FOREWARNS forecasts across Northern England; for distribution of base rate *s*, see Fig. 3. All forecasts based on 15:00 UTC MOGREPS-UK ensemble, where FOREWARNS would be available to users at 19:00 UTC, and valid next day. Scores computed against radar SWF proxy, for all return periods. (*a*) Distribution of values for *H*. Scores are
85 **equitable, with perfect forecasts scoring one and worst score zero. (*b*) Values of *SR*, also equitable. (*c*) Bias scores for forecast sample. Catchments shaded purple have *B*<1.5, with darker colours then indicating lower bias values. Perfect unbiased forecasts have *B*=1. Catchments shaded green have *B*>1.5, indicating a strong tendency towards overforecasting SWF. Here colourbar maximum value corresponds to 90th percentile of bias values. (*d*) Equitable values for *TS*. (*e*) Values for *PSS*, where perfect score is one, random forecasts**
90 **score zero and worst score is -1. (*f*) Values for *SEDI*, with same scale as *PSS*. Stippling indicates where *B*<0.67 or *B*>1.5 and values should hence be treated with caution. Grey bordered empty catchments (where *H*=0) indicate degenerate values due to logarithmic dependence of *SEDI* on *H*.**

Participant	Hits	False alarms	Misses	Correct rej.	*TS*	*H*	*F*	*SR*	*PSS*	*SEDI*	*B*
1	38	40	51	596	0.29	0.43	0.06	0.49	0.36	0.57	0.88
2	52	34	45	594	0.40	0.54	0.05	0.60	0.48	0.69	0.89
3	47	29	47	602	0.38	0.50	0.05	0.62	0.45	0.67	0.81
4	30	39	57	599	0.24	0.34	0.06	0.43	0.28	0.48	0.79
5	26	39	60	600	0.21	0.30	0.06	0.40	0.24	0.43	0.76
6	47	38	46	594	0.36	0.51	0.06	0.55	0.45	0.65	0.91
7	37	29	35	624	0.37	0.51	0.04	0.56	0.47	0.69	0.92
8	6	31	62	626	0.06	0.09	0.05	0.16	0.04	n/a	0.54
9	21	41	63	600	0.17	0.25	0.06	0.34	0.19	0.35	0.74
10	46	27	36	616	0.42	0.56	0.04	0.63	0.52	0.73	0.89
Column means	35.0 ±4.5	34.7 ±1.7	50.2 ±3.2	605.1 ±3.9	0.29 ±0.04	0.40 ±0.05	0.05 ±0.00	0.48 ±0.05	0.35 ±0.05	0.58 ±0.04	0.81 ±0.04

Table S2: Contingency table category totals and dependent skill scores from subjective assessment of May–October, 2019–2022 (*r*30, *p*98) FOREWARNS vs radar SWF proxy pairs. Assessment conducted on 155 forecast-proxy pairs which did not exclusively show correct rejections by group of meteorologists comprising 6 of the authors and 4 additional practitioners. Further 570 default correct rejections included in final calculations. All forecasts based on 15:00 UTC MOGREPS-UK ensemble, where FOREWARNS would be available to users at approximately 19:00 UTC, and valid on day of radar SWF proxy. Single contingency table category allocated per day.

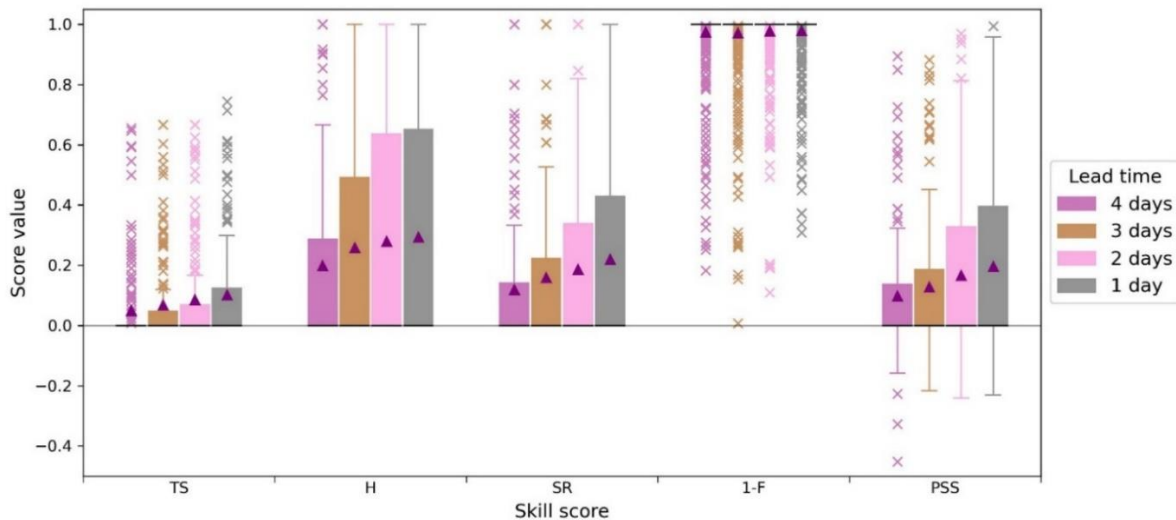



Figure S8. Distributions of spatial *TS, H, SR,* 1-*F* and *PSS* scores for daily (*r*30, *p*98) FOREWARNS forecasts. Scores computed from 166 catchments, for all return periods, against radar SWF proxy on forecast validity date. Distributions calculated from forecasts for 725 days, May–October 2019–2022, with mean (median) values shown by purple triangles (black horizontal lines). Shading indicates lead time in days, with scores for forecasts issued at one day's lead time shaded grey. All forecasts are based on 15:00 UTC MOGREPS-UK ensemble, where FOREWARNS would be available to users at approximately 19:00 UTC. All measures are equitable (worst score zero, perfect score one) except PSS, for which random forecasts score zero and the worst score is -1.

S2.1 Technical Note on *SEDI* calculation

The bias distribution shown in Figure S7f is an important counterpart to results given for the symmetrical extreme dependence index (*SEDI*) score in this study. As highlighted in the paper body, *SEDI* is designed for use with unbiased forecasts where $B\sim1$ – this is to ensure that the score shows the correct limiting behaviour when *s* is small (Ferro and Stephenson, 2011). Typically, bias is removed by forecast calibration: for categorical forecasts examining threshold exceedance, Ferro and Stephenson, 2011, recommend that for forecasts of events with base rate *s*, thresholds should be modified and instead chosen as the upper *s* quantiles of forecasted and observed variables, respectively. Alternative calibration methods for probabilistic forecasts (Magnusson et al., 2014) or specific quantiles of continuous variables (Sharpe et al., 2018) have also been implemented. Although FOREWARNS does rely on threshold exceedance to indicate SWF risk, three accumulation periods are compared at each sampling point, of which there may be multiple in a given catchment. The suggested recalibration of the underlying thresholds is therefore not tractable. The proxy observational record used here also inevitably makes true forecast bias measures uncertain.

We have therefore followed North et al., 2013, and chosen to use uncalibrated forecasts displayed with bias indications. Given the very high bias values shown in some catchments (north of domain especially), we also only use catchments with $0.67<B<1.5$ for calculations of *SEDI* distributions, so as to exclude misleading values.

Table S3: Date and event locations of recorded flood events across Northern England, May–October 2013–2022. There are 82 days in total. The 28 days identified as having especially significant SWF impacts are highlighted with bold font.

Date	Regions	Locations
2013-07-28	West Yorkshire	Chapeltown (Sheffield)
2013-07-29	**West Yorkshire**	**Upper Calderdale**
2014-07-08	**East York**shire	**Cottingham (Hull); Halifax**
2014-07-19	South Yorkshire	Rotherham
2014-07-20	**East Yorkshire**	**Market Weighton and South Cave**
2014-08-08	West Yorkshire	West Yorks - widespread, Castleford, Shipley and Bingley in particular
2014-08-10	East Yorkshire	Hull
2015-08-22	**West Yorkshire**	**Garforth and surrounding area**
2015-09-01	Lancashire	Carnforth
2015-09-02	Wirral, Cheshire	Bebington, Heswell, Thornton Hough; Chester, Neston
2016-06-05	Merseyside	Southport; Birkdale
2016-06-07	**West York**shire	**Baildon**
2016-06-08	**Greater Manchester; Merseyside**	**Manchester, Oldham, Rochdale**
2016-06-10	**South Yorks; Derbyshire; Greater Manchester**	**Barnsley area - Darton, Staincross, Grimethorpe; Sheffield; Manchester, Oldham, Rochdale; Liverpool, Cressington, Aigburth**
2016-06-11	**Greater Manchester**	**Stockport, Hazel Grove, Offteron; Whaley bridge; Poynton, Disley**
2016-06-12	Cheshire	Northwich
2016-06-13	Lancashire; Cumbria; Lincolnshire	Lancaster, Morecambe; Barrow; Lincoln?
2016-06-14	**Cheshire; Merseyside**	**Lymm, Wrexham, Chester, Warrington; Liverpool**
2016-06-15	South Yorkshire; Derbyshire; Lincolnshire; Cheshire; Merseyside	Sheffield; Cherry Willingham, Tealby; Chesterfield; Frodsham; Liverpool
2016-06-16	South Yorkshire; Cheshire; Merseyside	Sheffield, Liverpool
2016-06-24	Northumberland	Alnwick
2016-08-22	**North Yorkshire; Lancashire; Cumbria**	**Ingleton area; Lancaster; Clitheroe; Churchtown; Millom; Bootle; Dent**
2016-09-13	**Greater Manchester**	**Salford, Manchester**
2017-05-27	West Yorkshire; Cumbria	Leeds/Otley area; Windemere, Kendal, Ullverston
2017-06-06	Teeside; North Yorkshire	Middlesborough, South Bank, Grangetown, Guisborough
2017-07-06	**North Yorkshire**	**Ryedale**
2017-07-19	Lancashire	Lancaster, Fylde coast, Morecambe, Carnforth Blackpool
2017-08-08	East Yorkshire; NE Lincolnshire	Humberside and East coast towns; Grimsby
2017-08-18	Cumbria	Carlisle

Date	Region	Locations
2017-08-23	**North and West Yorkshire**	**Scarborough; Malton/Ryedale area; Wyke Beck (Leeds); Wetherby, Garforth**
2017-09-05	Lancashire	Southport
2017-09-11	Greater Manchester; Tamesdale	Manchester; Mossley; Micklehurst
2017-09-30	**Cumbria**	**Millom, Windermere**
2017-10-11	Cumbria	Borrowdale; Tebay; Millom; Plumbsland; Workington
2017-10-21	Lancashire	Rawtenstall; Accrington; Great Harwood; Euxton
2018-07-27	Lincolnshire	Scunthorpe, Broughton, Marblethorpe, Faldingworth
2018-08-12	Greater Manchester	Manchester
2018-08-13	**North Yorkshire**	**York**
2018-09-20	South Yorkshire	Sheffield - but also more widespread
2018-10-13	Cumbria	Barrow; Windermere; Blawith; Grange-over-Sands
2019-06-02	Greater Manchester	Swinton
2019-06-11	Lincolnshire	Louth, Partney, Horncastle, Skegness
2019-06-12	Merseyside/Wirral	Liverpool, Birkenhead, Ellesmere Port, Hooton, Northwich
2019-07-11	Greater Manchester; North East	Stockport, Heaton Chapel; Sunderland
2019-07-28	Lancashire; Rochdale; Greater Manchester	Manchester, Preston, Smithy Bridge (Rochdale)
2019-07-30	**North Yorkshire; Greater Manchester/Cheshire**	**Swaledale, Arkengarthdale, Wensleydale; Whaley Bridge, areas of south Manchester and E Cheshire, eg Alderley**
2019-07-31	Derbyshire; E Cheshire	Edale and wider Hope Valley; Buxton; Congleton
2019-08-04	South Yorkshire	Heeley (Sheffield)
2019-08-09	Lancashire; Merseyside	Carnforth, Lancaster; Liverpool
2019-08-10	Lancashire; Cumbria; Northumberland	Carlisle; Preston, Ribblesdale (Blackburn - Clitheroe); Otterburn area, Haltwhistle, Rothbury, Alnwick
2019-08-16	Greater Manchester	Manchester, Levenshulme, Fallowfield
2019-09-24	West Yorkshire; Manchester; Merseyside	Leeds and surrounding area; Liverpool, Knowsley, Sefton
2019-09-27	Lancashire; Cumbria	Carnforth
2019-09-28	Greater Manchester	Adlington; Bentham; Blackley
2019-09-29	**North and West Yorkshire; Calderdale; E Cheshire; Derbyshire**	**Wigan; Poynton; Newton le Willows; Saddleworth; Edale and wider Hope Valley; M62 Halifax/Huddersfield; Leyburn, Hawes, Grinton, roads between Harrogate and Ripon; Yeadon; Fylde; Oldham**
2019-09-30	Cumbria; County Durham	Carleton, Holme, Beetham, Scaleby, Carlisle; Willington
2019-10-11	Lancashire; North Yorkshire	Preston, Kirkham, Wesham
2019-10-15	Rochdale	Middleton Street/Lane
2019-10-25	Yorkshire; Manchester	Harrogate; Gargrave; Scarborough; Calderdale; Sheffield

2019-10-26	W and E Yorkshire; Lancashire; Lincolnshire	Leeds; Walkington ,Beverley, Leconfield. Brough, Warter, Hull; Doncaster-Scunthorpe train line; Knutsford, Astley, Alderley Edge; Wilmslow; Lincoln, Brigg
2020-06-15	**Greater Manchester; Rochdale**	**Manchester, Stretford, Milnrow**
2020-06-16	**Merseyside**	**Speke/Garston (Liverpool)**
2020-06-26	**South Yorkshire**	**Sheffield**
2020-08-23	Cumbria	Carlisle
2020-08-25	Cumbria; Rochdale	Carlisle; Rochdale-Todmorden line
2020-10-03	Cumbria; Chester	Waverton; Westlinton; Dalton; Chester
2020-10-06	Lancashire; Manchester	Bury; Manchester; Buckley; Clitheroe; Wigan
2020-10-26	Greater Manchester	Stockport
2020-10-28	N Yorkshire	Wennington
2020-10-29	Lancashire	Preston; Middleforth; Woodplumpton
2021-05-16	**Lancashire**	**Thornton; Chorley; Bolton**
2021-07-04	South Yorkshire; Greater Manchester	Sheffield; Whalley Range; Burnage
2021-07-06	Greater Manchester	Heywood; Stockport
2021-07-28	**Warrington**	**Warrington**
2021-08-07	Greater Manchester	Manchester Airport
2021-09-09	**Wirral**	**Birkenhead, Ellesmere Port**
2021-09-30	**Greater Manchester**	**Failsworth; Middleton; Heywood; Prestwich**
2021-10-05	**Newcastle**	**Newcastle**
2022-06-28	Lancashire	Blackpool; Cleveleys
2022-06-30	West Yorkshire	Shipley
2022-08-16	**South Yorkshire; Nottinghamshire; N Lincolnshire**	**Sheffield and surrounding area; Worksop; Market Raisen**
2022-09-30	**Cumbria**	**Borrowdale**

145 **References**

Ferro, C. A. T. and Stephenson, D. B.: Extremal dependence indices: Improved Verification measures for deterministic forecasts of rare binary events, Weather Forecast., 26, 699–713, https://doi.org/10.1175/WAF-D-10-05030.1, 2011.

Magnusson, L., Haiden, T., Ferro, C., and Stephenson, D.: Verification of extreme weather events: Discrete predictands, ECMWF Tech. Memo., 731, 699–713, 2014.

150 North, R., Trueman, M., Mittermaier, M., and Rodwell, M. J.: An assessment of the SEEPS and SEDI metrics for the verification of 6h forecast precipitation accumulations, Meteorol. Appl., 20, 164–175, https://doi.org/10.1002/met.1405, 2013.

Ochoa-Rodríguez, S., Wang, L. P., Thraves, L., Johnston, A., and Onof, C.: Surface water flood warnings in England:

overview, assessment and recommendations based on survey responses and workshops, J. Flood Risk Manag., 11, S211–S221, https://doi.org/10.1111/jfr3.12195, 2018.

155 Sharpe, M. A., Bysouth, C. E., and Stretton, R. L.: How well do Met Office post-processed site-specific probabilistic forecasts predict relative-extreme events?, Meteorol. Appl., 25, 23–32, https://doi.org/10.1002/met.1665, 2018.