Large-Scale Predictors for Extreme Hourly Precipitation Events in Convection-Permitting Climate Simulations

STEVEN C. CHAN

School of Engineering, Newcastle University, Newcastle upon Tyne, United Kingdom

ELIZABETH J. KENDON

Met Office Hadley Centre, Exeter, United Kingdom

NIGEL ROBERTS

MetOffice@Reading, Reading, United Kingdom

STEPHEN BLENKINSOP AND HAYLEY J. FOWLER

School of Engineering, Newcastle University, Newcastle upon Tyne, United Kingdom

(Manuscript received 12 June 2017, in final form 30 November 2017)

ABSTRACT

Midlatitude extreme precipitation events are caused by well-understood meteorological drivers, such as vertical instability and low pressure systems. In principle, dynamical weather and climate models behave in the same way, although perhaps with the sensitivities to the drivers varying between models. Unlike parameterized convection models (PCMs), convection-permitting models (CPMs) are able to realistically capture subdaily extreme precipitation. CPMs are computationally expensive; being able to diagnose the occurrence of subdaily extreme precipitation from large-scale drivers, with sufficient skill, would allow effective targeting of CPM downscaling simulations. Here the regression relationships are quantified between the occurrence of extreme hourly precipitation events and vertical stability and circulation predictors in southern United Kingdom 1.5-km CPM and 12-km PCM present- and future-climate simulations. Overall, the large-scale predictors demonstrate skill in predicting the occurrence of extreme hourly events in both the 1.5- and 12-km simulations. For the present-climate simulations, extreme occurrences in the 12-km model are less sensitive to vertical stability than in the 1.5-km model, consistent with understanding the limitations of cumulus parameterization. In the future-climate simulations, the regression relationship is more similar between the two models, which may be understood from changes to the large-scale circulation patterns and land surface climate. Overall, regression analysis offers a promising avenue for targeting CPM simulations. The authors also outline which events would be missed by adopting such a targeted approach.

1. Introduction

Extreme precipitation events from subhourly to multi-day time scales are driven by specific weather processes and drivers. Understanding future human-induced changes

in extreme precipitation, as well as changes due to natural variability, requires an understanding of these processes and drivers. Fortunately, these processes and drivers are well established for present-day observed events. For instance, convective storms are, by and large, caused by vertical instability (Johns and Doswell 1992), and most flash floods are caused by convective storms (Doswell et al. 1996). In the midlatitudes and subtropics, intense and/or slow-moving synoptic and mesoscale weather systems such as weather cyclones, fronts, and atmospheric rivers are drivers for both convective and large-scale extreme precipitation on hourly and daily time scales (Morcrette et al. 2007; Champion et al. 2015). High vertical instability and weather systems are readily

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-17-0404.s1>.

Corresponding author: Steven C. Chan, steven.chan@metoffice.gov.uk

DOI: 10.1175/JCLI-D-17-0404.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

discernible from surface and upper-air observations, and can be thought of as “ingredients” for extreme precipitation (Doswell et al. 1996; Wetzel and Martin 2001). In climate model projections, Púčik et al. (2017) show an increase of days with conditions favorable for severe weather in future-climate representative concentration pathway 4.5 and 8.5 (RCP4.5 and RCP8.5) European Coordinated Regional Climate Downscaling Experiment (EURO-CORDEX) regional climate simulations (Giorgi et al. 2009). Regional climate models cannot adequately represent severe weather, but they may be able to simulate the environments that are favorable for their occurrence (Púčik et al. 2017).

In principle, extreme precipitation in dynamical weather forecast and climate models is caused by the same processes and drivers as in reality. The different formulations of different dynamical models mean they have different precipitation responses under similar weather conditions. For instance, many dynamical models—especially ones that are lower in resolution (i.e., ≥ 10 -km gridbox size)—use cumulus parameterization schemes [parameterized convection models (PCMs)]. Cumulus parameterization schemes inhibit a good representation of the precipitation diurnal cycle and realistic evolution in convective development (Chan et al. 2014b; Clark et al. 2016). In contrast, convection-permitting models (CPMs) can be operated without cumulus parameterization schemes, and they have improved representations of the precipitation diurnal cycle, spatial variability, and hourly extreme precipitation in both NWP and climate simulations (Roberts and Lean 2008; Kendon et al. 2012; Ban et al. 2014; Chan et al. 2014b; Keller et al. 2016; Prein et al. 2015; Chan et al. 2016; Clark et al. 2016).

Given the differences between CPMs and PCMs, one expects different hourly (extreme) precipitation responses under similar model meteorological conditions. In terms of future projections, we expect larger disagreements between CPMs and PCMs during summer when convection is more prevalent. Current Met Office future-climate projections indicate that DJF has larger 1-h extreme precipitation intensification in both absolute and percentage terms than JJA, and the projected winter extreme changes are consistent between CPM and PCM simulations (Chan et al. 2014a). Unlike winter, the summer CPM and PCM projections disagree with each other; only the CPM simulations show a consistent intensification, and the intensification is moderated by large reductions in precipitation probability, leading to smaller changes in future return levels (Chan et al. 2014a).

Despite the improved realism in representing extreme precipitation in CPMs, their computational costs are high; as a consequence, their use is limited to specific regions. Such limited-area simulations are often termed

dynamical downscaling as they are driven by lower-resolution reanalysis and GCM data. Both PCMs and CPMs are extensively used in dynamical downscaling (Giorgi et al. 2009; Kendon et al. 2017).

A low-cost alternative to dynamical downscaling is statistical downscaling. However, statistical downscaling tends to underestimate extreme intensities (Fowler et al. 2007) and assumes that present-day relationships between predictors and extreme intensities remain unchanged in the future. Given CPMs’ added value in representing extreme subdaily precipitation (Kendon et al. 2012), we would argue that extreme subdaily precipitation is best simulated using CPMs. Instead of using statistical downscaling to predict intensities of hourly extremes, we predict the occurrences of extreme hourly precipitation, and hence when dynamical downscaling is needed.

The predictors for extreme hourly precipitation should be consistent with what is known a priori, namely the importance of vertical instability and synoptic weather conditions. Here, they are diagnosed from the driving simulation to predict events in the downscaling CPM simulation. In practice, the use of overly detailed measures of vertical stability and synoptic circulation (e.g., vertical stability at every model grid point for every hour) for regression analysis is self-defeating and amounts to overfitting. The goal is to build simple and elegant relationships that are predictive but not complicated ones that appear to fit well but have low predictive skill and difficult interpretation. Hence we seek general proxies and diagnostics that summarize the overall vertical stability and synoptic weather condition, which we can diagnose from low-resolution data. The proxies for instability and circulation regimes can then be regressed to extreme precipitation occurrences. Logistic regressions are regression models that fit predictors to probabilities for a binary outcome (i.e., occurrence or nonoccurrence of an extreme event). The regression probabilities inform us of the chance of extreme hourly precipitation occurring in the CPM simulation without performing the actual CPM simulation. As CPM simulations are expensive, these probabilities may inform us how to conduct CPM simulations more selectively, and hence reduce our computational cost.

The selection criteria for CPM downscaling need to balance accuracy with cost. Receiver operating characteristic (ROC; Mason and Graham 2002; Wilks 2011) analysis is designed specifically for this purpose. In ROC, we aim to find forecast thresholds that maximize the number of extreme events that we capture [the true positive rate (TPR)] with the least computer time spent on modeling times when no extremes occur [the false positive rate (FPR)]. The most efficient forecast

threshold is one that maximizes the margin between TPR and FPR.

The only way to achieve full accuracy is to conduct full continuous simulations—the very thing that we wish to avoid if we seek to reduce our computation cost. Selective CPM simulations are almost certain to miss some events [false negatives (FNs)] as it is impossible for our large-scale predictors to represent all drivers for extreme hourly precipitation. The regression models are not models of the actual physical processes. The examination of these FN events is part of our goal as it may reveal to us which drivers and processes we have missed. Such examination gives us physical insights into the inner workings of the CPM.

This paper is structured as follow: A basic description of the regional climate simulations is provided in [section 2](#), and forecast verification and statistical downscaling techniques are described in [section 3](#). Results from the regression analysis of the large-scale predictors and extreme hourly precipitation are presented in [section 4](#). We then examine the properties of the events that large-scale predictors have failed to identify ([section 5](#)). A discussion of the results and the main conclusions are presented in [section 6](#), where we also outline the significance of these results in the context of strategies and cost efficiencies for CPM dynamical downscaling.

2. Regional climate model data

The model simulations analyzed here have been documented extensively in the past (e.g., [Kendon et al. 2012](#)). The 13-yr southern United Kingdom regional climate model (RCM) simulations¹ are sets of one-way-nested downscaling simulations from noncoupled HadGEM3 present-day and RCP8.5 end-of-twenty-first-century atmospheric GCM simulations ([Meinshausen et al. 2011](#); [Mizieliński et al. 2014](#)). The HadGEM3 simulations are first downscaled with the 12-km European RCM, and the 12-km European RCM simulations are then downscaled with the 1.5-km southern United Kingdom (SUK) CPM ([Kendon et al. 2012](#)). The 1.5-km simulations do not use cumulus parameterization, whereas the 12-km simulations use that of [Gregory and Rowntree \(1990\)](#). A common land surface parameterization is used by both simulations ([Best et al. 2011](#)).

Previous analyses of the 1.5- and 12-km ERA-Interim ([Dee et al. 2011](#)) hindcast simulations demonstrate greater confidence in the representation of the diurnal cycle and extreme events by the 1.5-km model, but its

simulation of mean precipitation shows considerable positive biases ([Kendon et al. 2012](#); [Chan et al. 2013](#)). The 12-km simulations have lower mean precipitation biases but poorer representation of the diurnal cycle and extreme subdaily precipitation events ([Kendon et al. 2012](#); [Chan et al. 2013, 2014b](#)). The 1.5-km HadGEM3-driven simulations project a future intensification of summer extreme subdaily precipitation, but the 12-km HadGEM3-driven simulations show no such change ([Chan et al. 2014b](#); [Kendon et al. 2014](#)). Both simulations project intensifications of winter extremes ([Chan et al. 2014b](#)).

Atmospheric circulation and its future projected change in the 12-km simulations are constrained by the driving HadGEM3 simulations. Analysis of the driving HadGEM3 simulations is ongoing, but initial results indicate improved North Atlantic blocking relative to lower-resolution GCM simulations ([Schiemann et al. 2017](#)). Circulation changes are similar to CMIP5 projections with flow becoming more anticyclonic and cyclones less frequent near to the British Isles because of the poleward shift of the summer storm track ([Zappa et al. 2013b](#); [Belleflamme et al. 2015](#); [Chan et al. 2016](#)).

We do not regrid any model data in our analysis as the upscaling of high-resolution data may blur out critical information that we wish to retain. In particular, the United Kingdom is known to experience damaging localized extreme precipitation ([Golding et al. 2005](#)), and we are interested in identifying whether the occurrence of such events is predictable from large-scale precursors.

3. Methods

The analysis here uses summary diagnostics of extreme precipitation and large-scale conditions for the southern U.K. domain. For precipitation, we choose the maximum overland 1-h precipitation intensity for each day. Similar procedures are applied to the predictors ([section 3d](#)).

a. ROC

The ROC score ([Swets 1973](#); [Mason and Graham 2002](#); [Stephenson 2000](#); [Wilks 2011](#)) is a skill metric that compares the true-positive event detection rate (i.e., the TPR) with the false alarm rate (i.e., the FPR) across a range of detection thresholds. A prediction is skillful if correct detections [i.e., true positives (TPs)] exceed false alarms or false positives (FPs). The most efficient detection threshold is one that maximizes the margin between the TPR and FPR, meaning that one gets the maximum number of true predictions with the least number of false alarms.

¹The 1.5-km-simulation domain is illustrated in the supplemental material.

Here we use ROC scores to diagnose the relationship between large-scale atmospheric conditions and extreme hourly precipitation in the climate model world. A detailed description of ROC scores is provided in [appendix A](#). In summary, useful prediction skill require ROC curves to be to the left of the $y = x$ diagonal (i.e., the ROC curve is in the upper-left half of the diagram, and $\text{TPR} > \text{FPR}$ for most detection thresholds).

The ROC curve is often summarized by the area-under-curve (AUC) metric. This literally named metric is the area between the ROC curve and $y = 0$. Skillful predictions have $\text{AUC} > 0.5$. In technical terms, AUC represents the probability of positive events scoring higher in the forecast diagnostic than nonevents ([Fawcett 2006](#)). For instance, if 850-hPa relative vorticity ξ_{850} is a skillful predictor for precipitation extremes, one expects that an extreme precipitation day is more likely (i.e., $\text{AUC} > 0.5$) to have higher ξ_{850} than a nonevent day; otherwise, ξ_{850} is useless in predicting precipitation extremes as it cannot distinguish between event and nonevent days. In layman's terms, warnings are only credible if the warnings contain more truth than lies (as measured by AUC), and the individual points of the ROC curve represent the different warning thresholds.

b. Definition of extremes and basis of CPM added value

Before examining the large-scale predictors, we begin by asking a simpler question: can the existence of an "extreme" hourly precipitation event in the coarser-resolution 12-km PCM data be used to predict an extreme hourly event in the downscaled 1.5-km CPM simulation? This serves as a baseline level of skill for our analysis as extremes in the 12-km PCM and 1.5-km CPM likely depend on the same large-scale predictors at a different sensitivity. Hence, one would expect a 12-km precipitation predictor variable to demonstrate some skill. However, it is unclear how one should define an appropriate extreme threshold for the two models since the driving 12-km model has a lower resolution than the CPM, and so the extreme threshold for the driving simulation has to be lower than the CPM threshold as a result of gridpoint averaging.

We chose a range of potential JJA and DJF extreme thresholds for the 12- and 1.5-km simulations based on previous extreme precipitation analysis ([Chan et al. 2014b](#)). ROCs can then be used to guide which thresholds are optimal. An example is in [Fig. 1](#) showing the ROC diagram for the detection of JJA 20 mm h^{-1} events in the 1.5-km simulation using different present-climate simulation thresholds for the 12-km RCM. We find that 1.25, 2.5, 3.75, 5.0, and 6.25 mm h^{-1} are all suitable predictive thresholds for the 12-km present-climate

simulation, and all demonstrate comparably useful skill in predicting 20 mm h^{-1} events in the 1.5-km simulation. For the future simulation (see supplementary Fig. 1 in the supplemental material), ROC favors a similar range of thresholds for the 12-km simulation, with 1.25 mm h^{-1} being the optimum threshold. As a compromise between future- and present-climate simulations and to avoid thresholds that are subextreme or computationally inefficient,² we therefore use a common threshold of 2.5 mm h^{-1} for JJA for both present- and future-climate 12-km simulations. Previous extreme value analysis for the 12-km simulations showed that 2.5 mm h^{-1} is higher than the average 95th percentile of JJA "wet" values at each grid point ($\sim 1.8 \text{ mm h}^{-1}$; [Chan et al. 2014b](#)).

To choose 1.5-km JJA simulation thresholds, sensitivity tests show that 15, 20, 25, and 30 mm h^{-1} thresholds all give similar ROC analysis results (not shown). A common 20 mm h^{-1} JJA threshold is therefore chosen for the 1.5-km simulations. Any fixed threshold carries caveats; by our definition, clear weather days, days with long periods of moderate precipitation, and days with maximum precipitation of up to 20 mm h^{-1} are all excluded.

The ROC curve for the present-climate simulation ([Fig. 1](#)) lies above and is essentially parallel to the diagonal of no skill for thresholds between 1.25 and 6.25 mm h^{-1} . This is in contrast with a more curved ROC curve for the future-climate simulation (see supplementary Fig. 1). The lack of curvature in the present-climate simulation ROC means that subextreme thresholds are as skillful as higher thresholds, which results in a lower AUC for the present-climate simulation than the future simulation. By design, cumulus parameterization schemes respond to and remove instability forcing. If enough instability is in place to trigger the 12-km model cumulus parameterization, some convection and precipitation arise but not necessarily at the right vigor, as the cumulus parameterization scheme dissipates instability more rapidly than CPM simulations ([Clark et al. 2016](#)). The 12-km convective precipitation response under strong instability is therefore likely to be too weak, and this leads to a poor relationship between 12- and 1.5-km precipitation intensities in JJA, as is evident in [Fig. 1](#).

The 1.5-km simulations have almost no 20 mm h^{-1} events during DJF (not shown), and hence a lower 10 mm h^{-1} extreme threshold is chosen for the 1.5-km simulations. The present-climate simulation ROC curve for 12-km event thresholds is shown in [Fig. 1](#). As in JJA,

² A higher but equally good threshold is preferred.

FIG. 1. The ROC curves for the detection of JJA 20 mm h^{-1} (blue solid line) and DJF 10 mm h^{-1} (blue dashed line) extreme events over land (France and Ireland excluded) in the dynamically downscaled 1.5-km present-climate CPM simulation using exceedance events of various precipitation thresholds in the driving 12-km simulation as predictor. The number near each point indicates the 12-km precipitation threshold for that point.

the detection of 12-km precipitation events exceeding the threshold is a skillful predictor, and the optimal thresholds are 2.5 and 3.0 mm h^{-1} . The same values are also the optimal thresholds for the future simulation (see supplementary Fig. 1). Given the similarity with JJA results, we use the same 2.5 mm h^{-1} threshold for DJF. This threshold exceeds the spatial average for the 95th percentile of all DJF wet values ($\sim 2.2 \text{ mm h}^{-1}$; Chan et al. 2014b).

c. Logistic models

A logistic model is a statistical model that regresses binary outcomes (1 for positive and 0 for negative) to continuous or categorical predictors, and gives estimates to the probability of a positive outcome (i.e., the tick marks across the top in Figs. 2a–c). A full description of the logistic model is provided in appendix B. The ROC and the Akaike information criterion (AIC; Akaike 1974; Stone 1977) are used to compare the usefulness of different regression predictors.

The AIC is a regression model selection measure. It compares the relative goodness of fit and predictive skill

between regression models for the same dependent variable. The regression model with the lowest AIC is preferred.

d. Extreme precipitation predictors

Based on forecasting and statistical downscaling experience, we chose the following predictors:

- 1) Daily averaged 850-hPa relative vorticity [$\xi_{850} = (\nabla \times \mathbf{v}_{850}) \cdot \mathbf{k}$] as a synoptic circulation metric (Cavazos and Hewitson 2005), computed from model horizontal winds $\mathbf{v}_{850}$, using the areal average for the southern U.K. domain. Relative vorticity quantifies weather variations and circulation regimes; fair weather days tend to have negative vorticity, and stormy days tend to have positive vorticity. Areal averaging acts as a filter that only retains information at the CPM domain scale.
- 2) Daily averaged mean sea level pressure (MSLP), areally averaged for the entire southern U.K. domain. This is not used together with ξ_{850} as both are strongly correlated. MSLP is a popular and effective choice for statistical downscaling of daily precipitation

JJA

FIG. 2. Single predictor (x axis), (a) MSLP, (b) ξ_{850} , and (c) stability, logistic model regressions for the probability of 1.5-km-model 20 mm h^{-1} and 12-km-model 2.5 mm h^{-1} event over land in JJA for the present- and future-climate simulations. Blue and orange represent the regression probabilities for 1.5- and 12-km simulations respectively; solid lines and dashes are for present- and future-climate simulations. The actual model-simulated event occurrences ($=1$; top ticks and crosses, with ticks for present-climate and crosses for future-climate simulations) and nonoccurrences ($=0$; bottom ticks and crosses) are marked with the same colors as the regression probabilities.

TABLE 1. Seasonal means and standard deviations of predictors. Future projected changes are shown as well.

Simulation	Stability (K)			ξ_{850} ($10^{-0.5} \text{ s}^{-1}$)			MSLP (hPa)		
	Mean	Median	σ	Mean	Median	σ	Mean	Median	σ
JJA present	2.45	2.30	1.57	−0.24	−0.43	1.44	1018.1	1018.8	6.5
JJA future	2.50	2.34	1.77	−0.62	−0.75	1.32	1019.0	1019.5	5.3
JJA change	0.05	0.04	0.20	−0.38	−0.32	−0.12	0.9	0.7	−1.2
DJF present	3.93	3.63	2.44	−0.27	−0.44	1.95	1015.7	1017.1	12.6
DJF future	4.01	3.78	2.48	−0.55	−0.62	1.80	1018.7	1020.2	12.2
DJF change	0.08	0.15	0.04	−0.28	−0.18	−0.15	3.0	3.1	−0.4

(Fowler et al. 2007). The differences between MSLP and ξ_{850} are explored in this manuscript.

- 3) Daily averaged vertical moist static stability is diagnosed as

$$\text{stability} = 0.5[\theta_s(500 \text{ hPa}) + \theta_s(700 \text{ hPa}) - 2\theta_w(925 \text{ hPa})], \quad (1)$$

where θ_s and θ_w are saturated wet-bulb (Morcrette et al. 2007; Clark et al. 2014) and wet-bulb potential temperatures (Hewson 1937) respectively,³ and they have been used in forecasting and storm analysis (Hewson 1937; Clark et al. 2014). It is similar to the lifted and K indices (Galway 1956; George 1960). It has units of kelvins. Higher values of stability indicate a more stable troposphere and an environment that is more unfavorable to convection. As the spatial variability of instability occurs on the meso-scale, only a small part of the domain needs to have enough instability to trigger convection; hence, we use the 10th spatial percentile as a proxy for instability pockets within our domain.

Convective available potential energy (CAPE) is an alternative measure of instability. However, CAPE is difficult to compute with climate model data that have a limited number of vertical levels (Púčik et al. 2017). In principle, one could add 12-km model precipitation as a predictor; however, our focus is in the use of direct thermodynamic and circulation diagnostics as predictors, but we do explore 12-km model precipitation as a regression predictor in the supplemental material.

Unlike precipitation, the areal averaging and percentiles for the predictors include nonland points. Stability can be thought of as a thermodynamic driver for extreme hourly precipitation, while ξ_{850} and MSLP are linked to the synoptic variability and circulation drivers

for extreme hourly precipitation. However, thermodynamic and circulation drivers are not independent of each other.⁴

The regressions are fitted separately for the present- and future-climate simulations. The means and variances of the predictors change between the future and present-climate simulation (Table 1), caused by circulation and humidity differences in the driving GCM simulations.

e. Limitations of predictors

Our goal is to choose predictors that are easy to diagnose from lower-resolution PCM simulations, are consistent with our meteorological knowledge, and have predictive skill. Yet, it is critical to understand their limitations.

The use of daily averaged ξ_{850} smooths out subdaily synoptic variability. To fully account for subdaily vorticity variability, one must include subdaily variations, as pressure systems may move at different speeds. A rapid change from an anticyclonic environment to a cyclonic environment and vice versa is likely to be smeared out by daily averaging. Hence, 6-h vorticity data should, in theory, be more representative. However, sensitivity tests (not shown) for prediction using 6-h vorticity data produced similar results to those using daily vorticity data. There is no straightforward way to link the timing of vorticity changes with extreme precipitation events. For instance, the bulk of midlatitude cyclone precipitation is along fronts; fronts usually lead the cyclone center, but post- and prefrontal precipitation are common (Hobbs 1978). The fronts often extend hundreds of kilometers away from the cyclone center, and carry their own vorticity signatures. Hourly precipitation intensities are also often controlled by the steering speed of the pressure systems and convective storms. Slow-moving systems are more likely to exceed extreme thresholds.

³ Note that θ_s takes the parcel down to 1000 hPa via moist adiabatic without any lifting.

⁴ The Spearman correlations between stability and ξ_{850} are ~ -0.5 and ~ -0.4 for the present- and future-climate simulations, respectively.

Given the above caveats, we have chosen the daily averaged ξ_{850} , fully aware of its limitations.

In the present analysis, daily values and averages are defined from 0000 to 2400 UTC, and effects resulting from the diurnal cycle and the subdaily timing of events are not considered. However, the diurnal cycle modifies vertical stability and localized processes like sea breezes. The maximum diurnal effect for convective storms is expected to be near local late afternoon (~ 1500 UTC). U.K. extreme hourly precipitation events cluster in the late afternoon (Blenkinsop et al. 2017); however, overnight extreme hourly precipitation cannot be ruled out, with the passage of synoptic systems not affected by the diurnal cycle. The convention of 0000–2400 UTC time averaging is likely to introduce errors for overnight synoptic systems as a single vorticity signature is averaged out over two separate days. Spatial averaging introduces errors as well; centers of synoptic systems may be at the edges or even outside of our analysis domain. Hence, the sampled vorticity and MSLP are likely to underestimate the true intensities of synoptic systems, while stability is likely to be less affected by time averaging and spatial location because of its tendency to synchronize with the diurnal cycle and for maximum instability to occur in close proximity to convection. We note that coarse-resolution models remove vertical instabilities at each call to the cumulus parameterization scheme. Hence, the time-averaged model data used here may underestimate maximum instability.

The above discussion focuses on limitations in individual predictors, but the predictors may also correlate with each other (i.e., multicollinearity). Vertical static stability is closely linked with moisture availability, and moisture availability is connected to both synoptic variability and circulation regime. Many statistical downscaling studies use not only MSLP but also horizontal wind speeds at different pressure levels as predictors (Fowler et al. 2007) despite both being closely related to each other. Moisture recycling may also be a problem. Higher recycling and soil moisture leads to higher moisture availability in the absence of synoptic drivers. Recycling ratios and soil moisture content may change for the future-climate simulation, and this affects the effectiveness and multicollinearity of the synoptic predictors.

JJA and DJF summary statistics for the predictors are given in Table 1. In the future simulation, the atmosphere tends to become more stable (stability increased for both JJA and DJF; right column of Table 1), and circulation becomes more anticyclonic (increased mean MSLP and reduced ξ_{850} for both JJA and DJF; center and right columns of Table 1). However, the future change in stability is only a few percent of its

standard deviation, whereas future changes in ξ_{850} and MSLP are about 15%–30% of their standard deviation (not shown explicitly). Hence, predictor changes between the present- and future-climate simulations are dominated by circulation changes. This raises the question of how much of the extreme precipitation changes are driven by circulation changes. We do note that this circulation change is based on a single GCM projection, and internal variability cannot be disentangled from the forced change. To address this, the Met Office is currently conducting the first ensemble of CPM simulations driven by a perturbed physics GCM ensemble for the next set of U.K. future climate projections (UKCP 2017).

The more anticyclonic environment in the future-climate simulation is also associated with reductions in the standard deviations for MSLP and ξ_{850} , indicating an overall decrease of transient synoptic activity in both seasons. Changes in transient activity and regional circulation are coupled (Karoly 1990; DeWeaver and Nigam 2000) and cannot be considered as separate change signals.

4. Logistic regressions for the probability of subdaily extremes

a. Summer results

The JJA logistic regressions for the present- and future-climate simulations are shown in Fig. 2. For both 12- and 1.5-km simulations, 1-h precipitation event probabilities increase with decreasing stability and MSLP (Figs. 2a,c) and increasing ξ_{850} (Fig. 2b). Present-climate simulation event probabilities exceed 0.7 for both models when $\text{MSLP} \lesssim 1000$ hPa, $\xi_{850} \gtrsim 2 \times 10^{-5} \text{ s}^{-1}$, and stability $\lesssim 0.5$ K. For the future simulations, the same probabilities can be found at higher MSLP and lower ξ_{850} (i.e., higher event probabilities under a more anticyclonic environment); event probabilities exceed 0.7 for both models when $\text{MSLP} \lesssim 1010$ hPa and $\xi_{850} \gtrsim 1 \times 10^{-5} \text{ s}^{-1}$; there are statistically significant changes (<0.001 level) to the regression coefficients for MSLP and ξ_{850} between the present- and future-climate simulations (not shown). For the present-climate simulations (solid lines), 1-h precipitation in the 1.5-km CPM simulation is more sensitive to stability and ξ_{850} changes than the 12-km simulation PCM simulation, and the differences in regression coefficients between the 1.5- and 12-km simulations are significant at levels beyond 0.001 (not shown); regression probability changes across the sampled ranges for stability and ξ_{850} are larger for the 1.5-km CPM simulation than the 12-km simulation.

TABLE 2. AICs of JJA logistic regressions. The AIC of the “1” null model is given on the first row. The best single-predictor regressions for each RCM simulation are marked with asterisks.

Predictor	Present		Future	
	1.5 km	12 km	1.5 km	12 km
1	1507.1	1609.7	1410.9	1512.2
MSLP	1430.0	1534.0	1203.8	1254.2
ξ_{850}	1314.5	1505.2*	1145.1	1216.2*
Stability	1188.0*	1543.6	1063.7*	1273.9
Stability + ξ_{850}	1146.2	1491.0	959.8	1120.6

For both the present- and future-climate simulations, the regression probabilities with MSLP and ξ_{850} as predictors are higher for the 12-km PCM than for the 1.5-km CPM. The differences between the two model resolutions become narrower for the future-climate simulation for all three predictors. The slopes of MSLP and ξ_{850} regressions become steeper, meaning much higher sensitivities and potential predictability in both 12- and 1.5-km future-climate simulations. The 12-km future-climate simulation also becomes more sensitive to stability changes than the present-climate simulation.

The relative goodness of fit and prediction skill of the regressions are given by their AICs, and they are shown in Table 2. For both present- and future-climate simulations, ξ_{850} is the best predictor ($AIC_{\xi_{850}} < AIC_{MSLP} < AIC_{stability}$) for the 12-km simulations, but stability is the best predictor for the 1.5-km simulations ($AIC_{stability} < AIC_{\xi_{850}} < AIC_{MSLP}$). For both CPM and PCM, ξ_{850} is a better predictor than MSLP.

Given that stability and ξ_{850} have higher skill and independence, we compute the regressions using both together.⁵ AICs (Table 2, bottom row) have all decreased with respect to the one-predictor regressions, indicating overall improvements using both predictors together. Hence, we use the two-variable stability + ξ_{850} regression as the basis to examine the limitations of our method (section 5).

The ROC curves for the logistic regressions are shown in Fig. 3. All ROC curves are to the left of the diagonal with AUC scores above 0.5, indicating positive skill for all regressions. For both present- and future-climate simulations, the stability + ξ_{850} regression is the most skillful regression in terms of AUC. With an AUC exceeding 0.8, more than four out of five event days have a lower stability and ξ_{850} than non-event days. The 0.4 probability threshold is close to optimal; therefore, it is used as the prediction threshold

for comparing the differences between TP events and FN events (section 5).

For the 1.5-km CPM regressions, AUCs increase in the same order as AICs decrease: $AUC_{stability+\xi_{850}} > AUC_{stability} > AUC_{\xi_{850}} > AUC_{MSLP}$ and $AIC_{stability+\xi_{850}} < AIC_{stability} < AIC_{\xi_{850}} < AIC_{MSLP}$. For the present-climate simulation ROCs, the AUC differences between different regression models and the diagonal are significant at the 1% level using the Mann–Whitney U test (Mason and Graham 2002). For the future-climate simulation ROCs, the AUC differences between different regression models and the diagonal are significant at the 5% level.

A more important comparison is with Fig. 1, which compares the skill of the large-scale predictors with 12-km PCM precipitation thresholds. For the present-climate simulation, only MSLP has a lower AUC than 12-km PCM precipitation thresholds, meaning MSLP has less skill than 12-km PCM precipitation in predicting 20 mm h^{-1} events in the 1.5-km CPM simulation. The AUC difference between ξ_{850} and 12-km PCM precipitation thresholds is not statistically significant at the 10% level. The AUCs from the stability and stability + ξ_{850} regressions (0.801 and 0.815, respectively) are higher than the AUC from 12-km PCM precipitation thresholds (0.721; Fig. 1). The differences are significant at the 0.1% level; hence, these regression predictors are more skillful than simply using 12-km PCM precipitation.

The AUC for future 12-km PCM precipitation thresholds is ~ 0.848 (see supplementary Fig. 1). Nearly all large-scale predictor regressions (except stability + ξ_{850}) have AUCs lower than 0.848, indicating lower relative skill compared to the use of 12-km precipitation thresholds. Although the 0.859 AUC for stability + ξ_{850} is greater than 0.848, the difference between the two is not statistically significant at the 10% level.

b. Winter results

The DJF regressions for the present- and future-climate simulations are shown in Fig. 4. ROC analysis for the regressions shows the large-scale predictors are skillful (see supplementary Fig. 4 in the supplemental material). The responses to each predictor are similar to JJA with the probabilities of a 10 mm h^{-1} event in the 1.5-km simulations increasing with decreasing stability and/or MSLP and increasing ξ_{850} . For the present-climate simulations, both models' event probability exceeds 0.7 when $MSLP \leq 1002 \text{ hPa}$, $\xi_{850} \geq 2 \times 10^{-5} \text{ s}^{-1}$, and stability $\leq 1 \text{ K}$. Unlike summer results, the order of AICs is different (Table 3). According to the AIC ordering, MSLP is a better predictor than ξ_{850} for DJF for both 12-km and 1.5-km present- and future-climate simulations. Like JJA, the 1.5-km CPM simulations

⁵ The actual regressions are shown in supplementary Figs. 2 and 3 of the supplemental material.

FIG. 3. JJA ROC curves for the synoptic and mesoscale predictor logistic models with different regression probability thresholds for the present-climate (solid lines) and future-climate (dashed lines) simulations. Blue and orange curves are for the 1.5-km and 12-km simulations, respectively. The respective AUCs are given in the title of each panel.

have stability as the best predictor, but the 12-km PCM simulations have MSLP as the best predictor in contrast with JJA, when ξ_{850} is superior.

When DJF (Fig. 4) is compared with JJA (Fig. 2), DJF samples a wider range of MSLP and ξ_{850} than JJA. DJF also has a greater number of points in the cyclonic range ($\xi_{850} \geq 0.6 \times 10^{-5} \text{ s}^{-1}$ and $\text{MSLP} \leq 1000 \text{ hPa}$). This is consistent with increased midlatitude synoptic storminess during DJF.

As in JJA, stability is the best predictor for the 1.5-km simulations. The average DJF stability is higher than JJA. Values below 0 K are rarer in DJF than in JJA for both present- and future-climate simulations, and JJA has lower extreme values ($\leq -1.0 \text{ K}$) for stability than DJF, consistent with higher JJA temperatures and specific humidities. Relative to DJF in the present-climate 1.5-km CPM simulation, the same amount of instability is more likely to produce extreme

DJF

FIG. 4. As in Fig. 2, but for DJF.

events in the future-climate simulation as the future regression probabilities are shifted to the right as one can see in Fig. 4. This shift is possibly related to warming-driven tropospheric moistening (Trenberth

et al. 2003), but detailed analysis is beyond the scope of this study.

It is worth remembering that the DJF 10 mm h^{-1} extreme threshold for the 1.5-km simulation is lower than

TABLE 3. As in Table 2, but for DJF.

Predictor	Present		Future	
	1.5 km	12 km	1.5 km	12 km
1	1430.0	1496.3	1498.0	1495.6
MSLP	1138.3	1271.5*	1118.2	1158.6*
ξ_{850}	1148.7	1280.4	1119.1	1216.9
Stability	1096.6*	1351.2	1099.8*	1271.3
Stability + ξ_{850}	1042.8	1266.5	1018.7	1186.2

the 20 mm h^{-1} JJA threshold. In contrast with winter, projected summer 1-h extreme precipitation changes differ greatly between the two simulations; only the 1.5-km simulations show an intensification in 1-h extreme precipitation (Chan et al. 2014a). Hence in the next section, we focus on characterizing summer events in detail.

5. Character of excluded summer events

Rather than starting with the TP events that are correctly predicted, we explore the FN events (i.e., the events that were missed) as they may reveal accuracy limitations of our methodology. Accuracy is distinct from skill as skill accounts for FPs (i.e., type-I errors), but accuracy only involves TPs and FNs (i.e., type-II errors). FN events are JJA 20 mm h^{-1} events with sub-0.4 probabilities for the stability + ξ_{850} regressions (section 4a). If there are common features in the excluded events, important processes and predictors may have been missing from the regression analysis. It is possible that these excluded events share no common features; that would suggest that the FN events occur randomly and are caused by details specific to each of the event. We would not want to quantify all these details as that may amount to overfitting.

The use of fixed probability thresholds implies cases just below the 0.4 threshold are excluded in a “black and white” manner. Our threshold choices are based on ROC analysis. ROC analysis is not intended to find the most “accurate” threshold but the most “efficient” threshold that accounts for cost of false positives. The only possible way to have no FN events is to apply no threshold and examine all data.

a. Convective fraction and actual hourly accumulations

The FN events, by definition, relate more poorly with our predictors than TP events. Hence, their characterizations have to be based on other metrics. We use a gradient-based separation of convective-like and stratiform-like precipitation (section 3e in Kendon et al. 2012). Hourly precipitation are first regridded to 5-km grid boxes. The

FIG. 5. The violin plot for JJA convective fraction (y axis, from 0 to 1) of TP and excluded FN events modeled by the stability + ξ_{850} regression with a 0.4 event probability threshold. Results from the present- and future-climate simulation are on the left and right, respectively. The violin plot combines the traditional box plot (as indicated by the black lines) with the probability distribution (the colored “violin” widths).

separation between high-gradient (convective like) and low-gradient (stratiform like) precipitation is computed hourly by comparing individual gridpoint precipitation intensities with their neighbors; if the difference exceeds 2 mm h^{-1} , the gradient is deemed to be high-gradient, or convective-like. We define the daily convective fraction to be the total overland convective-like precipitation divided by the total convective-plus-stratiform precipitation. The differences in daily convective fraction between FN events and TP events are shown in Fig. 5.

The TP events, for both future and present-climate simulation, have higher convective fraction than the excluded cases. Apart from the present-climate FN cases, convective fractions appear bimodal. A lower convective fraction for the FN events suggests the FN events tend to be large-scale-like. This could be as a result of localized maxima embedded within large-scale orographic and frontal precipitation. Forced frontal uplifts require less instability to occur, so frontal events may have higher stability than convective events and are missed by our regression analysis. Detection of fronts from the model data is not straightforward as they have localized vorticity signatures, which are smeared out by our spatial and temporal averaging.

One might expect that weaker predictor forcing is likely to produce weaker precipitation intensities, but undiagnosed physical processes and event-specific processes can push intensities above the 20 mm h^{-1} extreme threshold. In those cases, regression probabilities may

FIG. 6. As in Fig. 5 but for 1.5-km model simulated 1-h precipitation intensities (y axis, mm h^{-1}) that are above 20 mm h^{-1} .

be around or just below 0.4 , and hourly precipitation intensities are above the 20 mm h^{-1} extreme threshold. The probability distributions of the underlying 1.5-km precipitation hourly intensities are shown in Fig. 6. The FN events have lower intensities and skewness than TP events for both present- and future-climate simulations. Most FN events have intensities just above the 20 mm h^{-1} threshold. This suggests that marginal events under weaker forcing are harder to predict. Comparing the future- and present-climate simulation intensities, both TP and FN event intensities are higher in the future-climate simulation, consistent with previous analysis of precipitation intensity changes (Kendon et al. 2014); however, the intensity increase is higher for TP events.

b. Spatial distribution of missed events

If FN events are related to large-scale orographic precipitation, one may expect a spatial pattern for them (i.e., clustering of FN events over orography). However, spatial distribution analyses are inconclusive (see the supplemental material). There is no clear evidence supporting orography as a cause for FN events. However, this does not rule out other surface processes that our predictors cannot represent.

6. Discussion and conclusions

a. Key results from regression and event analysis

The occurrences of hourly precipitation extremes in the U.K. CPM simulations can be skillfully predicted by large-scale predictors, encompassing stability and circulation, from the lower-resolution-driving PCM data. The AUCs

from the statistical downscaling with ξ_{850} and vertical stability as predictors indicate that statistical downscaling can distinguish about $80\%+$ of the 1.5-km model $20+ \text{ mm h}^{-1}$ event days from nonevent days. Among the chosen diagnostics, stability is the most sensitive and skillful predictor for both summer and winter months for the 1.5-km CPM simulations. Circulation-based predictors such as vorticity and MSLP demonstrate skill but to a lesser degree than stability. The same predictors can also be applied to diagnose the large-scale conditions for precipitation extremes in the 12-km PCM driving data. Extremes in the 12-km PCM simulations are also sensitive to the same predictors but to a lesser degree than the 1.5-km CPM simulations.

The predictors are chosen based on meteorological experience: convective storms require instability to develop, and extreme precipitation tends to occur during favorable synoptic conditions: lower MSLP and higher vorticity are both proxies for synoptic low pressure systems and fronts. It is not clear if vorticity is a better predictor than MSLP. For summer, vorticity appears to be better, but MSLP appears to be better for winter. We note that vorticity is a more direct measure of the synoptic circulation regime than MSLP as the latter is derived from winds. The two have a high negative correlation (~ -0.75). The spatial coincidence of the two applies not just to centers of blocking and synoptic lows but also to weather fronts.

Hourly extremes that are of lower intensity and more large-scale in nature are harder to predict than more intense and convective-like events. This is consistent with weaker thermodynamic forcing, the black-and-white nature of our threshold-based analysis methodologies, and localized vorticity signatures that have been smeared by spatial averaging. Comparing the total number of TP and FN events (see supplementary Table 2, right column, in the supplemental material), about 30% of the events are missed. Large-scale precipitation events are often associated with orographic precipitation, but there is no clear evidence that suggests FN events are associated with orography.

b. Summer changes in response to extreme precipitation drivers

The 1.5- and 12-km simulations predictor responses converge in the future climate simulations. The convergence is associated with increased extreme event predictability for both 1.5- and 12-km models as indicated by increased AUCs in the future-climate simulations (Fig. 3).

The future-climate simulations have higher MSLP over the northern Europe (Chan et al. 2016). The average synoptic conditions in the future simulations are more

unfavorable to precipitation. As well as an overall increase of MSLP in the future simulation, there is also a reduction in MSLP variability in the future (Table 1 and Fig. 2).

Summer soil moisture conditions are much drier in the future simulation (see supplementary Figs. 5 and 6 in the supplemental material). The simulated future drier surface conditions are more likely to be a consequence of decreased mean rainfall and increased surface temperature. Preexisting dry conditions resulting from dry soils and warm temperature act as local rainfall suppressors as recycling is reduced. Dry conditions are likely to persist without influx of moisture from synoptic systems. Hence, the triggering of extremes in the future simulation becomes more synoptically driven for both the 12- and 1.5-km simulations. Yet, it is the lack of synoptic systems that keeps the soil dry and relative humidity low. As the synoptic variabilities for both simulations are the same, prediction skills converge when synoptic variability become the main driver in a moisture-limited climate.

c. Why the 1.5-km CPM responds more strongly to stability

For both summer and winter, the 1.5-km simulations are more sensitive to stability changes than the 12-km simulations. Within the 12-km PCM, instability triggers an immediate convective response by the cumulus parameterization scheme. Most cumulus parameterization schemes relax the instability back to the “closure” stable state. The relaxation is tuned to represent the average convective cloud (Arakawa 2004) and not to extremes. Hence, the response to instability is dampened by design, which leads to lower extreme thresholds (Chan et al. 2014b) and weaker precipitation response to instability forcing.

The above is a major reason why parameterized convection models perform poorly in representing extreme precipitation and the diurnal cycle. Cumulus parameterization is not designed to have a proper buildup of vertical instability and convective storms, which are essential for extreme precipitation.

d. Dynamic downscaling applications of the analysis

Multiensemble CPM climate simulations are currently too expensive even for institutes that have state-of-the-art computational facilities. For the first time, the Met Office (UKMO) is planning to conduct an ensemble of CPM climate simulations, as part of the next set of U.K. climate projections (UKCP 2017). Hence, the UKMO is currently exploring different options for conducting these new simulations that balance quality and limited computational resources.

Most computer time is spent on simulating days without extreme precipitation. If extreme precipitation

is the main added value of interest, the simulation cost is reduced if driving data are prescreened for the probability of subdaily extremes. In that case, the results here offer a promising avenue to target CPM simulations. However, CPMs have other added value, for instance the representation of snow, land surface feedbacks, orographic precipitation, and urban climate (Prein et al. 2015), and hence targeting just extreme precipitation events may not be appropriate.

The current analysis gives physically based guidance for targeting simulations. The prescreening strategy is intended to maximize efficiency instead of accuracy. The only sure way to achieve best accuracy is to do a complete full downscaling simulation. The prescreening diagnostics used here are on daily time scales. At longer time scales, probabilities of extremes and synoptic patterns are likely to be linked to dominant climate modes over Europe and North Atlantic, for instance the North Atlantic Oscillation (Hurrell 1995). The simulation of such climate modes remains a challenge for coupled GCMs (Davini and Cagnazzo 2014), and fundamental properties of the storm track and synoptic transients are often not well simulated by coupled GCMs (Zappa et al. 2013a). For “large” (say continental-scale) simulation domains, it is almost certain that extreme events will occur somewhere in the simulation domain; in that case, a prescreening strategy is unlikely to help.

Other important caveats also exist. Targeted simulations focus on periods with increased precipitation extremes, but the extreme risks outside of the targeted period are unknown. The 30% FN event estimate (section 6a) is diagnosed with hindsight of the full simulation, and it is not known if that 30% estimate can be applied to other CPM simulations and to other regions. The CPM data from targeted simulations may capture the “best” snapshots and profiles of extreme precipitation events, but they do not generate continuous multiyear data time series. Such time series are necessary to give accurate estimates for future extreme precipitation changes. Seasonal or snapshot simulations cannot simulate land surface feedbacks properly as the land surface climate requires at least a few months to spin up correctly.

e. Conclusions

We have demonstrated that model-simulated extreme hourly precipitation can be skillfully predicted by examining physically based large-scale variables. The skill is dependent on both the type of model (CPM or PCM) and the underlying climate regime. Not all events are predictable, and local processes (like sea breezes) are potentially important. As our predictors are based on fundamental meteorology, one may expect skillful prediction

for other extratropical regions and models; however, the sensitivity to the predictors may differ due to differences in model physics and regional climate.

As CPMs and PCMs are physically based models, large-scale proxies for the drivers for hourly precipitation extremes are expected to have skill in predicting extreme events in the model world. However, one would expect different sensitivities to the drivers as CPMs and PCMs have substantial differences in their representation of atmospheric convection. Climate regime dependency in our regressions indicates that the responses to the large-scale drivers are modified by complex feedbacks and interactions between different components of the model climate world.

Targeted simulations reduce simulation costs, but they come with many caveats and limitations. Intelligent use of high-performance computing can lower the computational cost (Leutwyler et al. 2016; Váña et al. 2017). Faster computation resources will still be needed; will the rapid increase of computational power in the past (Moore 2006) continue in the future to allow even higher resolution and more realistic climate models? The future of climate modeling is a scientific challenge as well as a technological challenge.

Acknowledgments. This research is part of the projects CONVEX and INTENSE and the UKMO Hadley Centre research program that are supported by the United Kingdom NERC Changing Water Cycle program (Grant NE/I0066-80/1), European Research Council (Grant ERC-2013-CoG-617329), and the Joint Department of Energy and Climate Change and Department for Environment Food and Rural Affairs (Grant GA01101), respectively. Hayley J. Fowler is also funded by the Wolfson Foundation and the Royal Society as a Royal Society Wolfson Research Merit Award (WM140025) holder. Large portions of the analysis have been carried out with the free-and-open-source software R and Python.

APPENDIX A

Basics of the Receiver Operating Characteristic

The receiver operating characteristic (ROC; Mason and Graham 2002; Wilks 2011) is a skill metric that compares rates of correct positive forecast [i.e., accuracy; true positive (TP)] against incorrect positive forecast [false positive (FP)] rates. The metric is designed to check if accuracy is worth the cost for the number of incorrectly predicted events, assuming the costs from each false alarms is equal to the benefits from each correct positive forecasts.

A positive forecast is issued whenever a forecast threshold is met, and a negative forecast is issued otherwise. We construct contingency tables that keep track of correct and incorrect forecast for each observed positive and negative events for various thresholds. The ROC curve is formed with points using different signal thresholds with the x axis representing the FP rate and y axis representing the TP rate:

$$\begin{aligned} \text{TPR} &= y \\ &= \frac{\text{No. of true positives}}{\text{No. of true positives} + \text{No. of false negatives}} \end{aligned} \quad (\text{A1})$$

and

$$\begin{aligned} \text{FPR} &= x \\ &= \frac{\text{No. of false positives}}{\text{No. of false positives} + \text{No. of true negatives}}. \end{aligned} \quad (\text{A2})$$

The “perfect” diagnostic has points at $x = 0, y = 1$ that detects all TPs with no FPs. The best and most efficient threshold for an imperfect diagnostic that balances between good TPR and low FPR is the one closest to $x = 0, y = 1$. A “useless” diagnostic that cannot distinguish between real events and false alarms has ROC points close to the $y = x$ line. The overall skill of the diagnostic across N number of reasonable thresholds is represented by the integral area below the ROC curve [area under curve (AUC)]:

$$\text{AUC} = \sum_{n=1}^N \text{TPR}_n \Delta(\text{FPR})_n. \quad (\text{A3})$$

AUC represents the probability of a TP to score higher in the forecast diagnostic than an FP, and measures the predictive skill of that forecast diagnostic (Fawcett 2006). A useless diagnostic has an AUC of approximately 0.5, and an “useful” diagnostic has AUC significantly higher than 0.5. Statistical significance of AUC can be computed based on the equivalence of AUC and the nonparametric Mann–Whitney U statistic (Mason and Graham 2002).

APPENDIX B

Logistic Regression and the Akaike Information Criterion

Logistic models are extensions of linear regression that model binary states (McCullagh and Nelder 1989). They linearly relate the natural log of the odd ratios

$\text{Pr}/(1 - \text{Pr})$ (or logit; Pr is the event probability) for a 1-or-0 binary category with i independent variable(s) for j observations x_{ij} :

$$\ln\left(\frac{\text{Pr}}{1 - \text{Pr}}\right) = \text{logit}(\text{Pr}) = y_0 + \sum_i \beta_i x_{ij}. \quad (\text{B1})$$

Observed values take on the value of either 0 or 1, and logistic regression estimates the probability of scoring 1. Regression coefficient differences between group samples can be tested with the method described in Clogg et al. (1995).

The Akaike information criterion (AIC; Akaike 1974; Stone 1977) is a standard metric for regression model selection that balances goodness of fit with model complexity. AIC is defined by the number of predictors k and the maximum likelihood L of the statistical model ($\text{AIC} = 2k - 2\ln L$). The statistical model with the lowest AIC is the most preferred. AIC comparisons are specific to same set of dependent variables, and cannot be compared between different sets of training data.

Logistic regression can be linked with ROC in which an advisory against a binary outcome is issued whenever a threshold probability is exceeded. Forecasts from each threshold can then be validated and be used to construct a ROC curve.

REFERENCES

- Akaike, H., 1974: A new look at the statistical model identification. *IEEE Trans. Autom. Control*, **19**, 716–723, <https://doi.org/10.1109/TAC.1974.1100705>.
- Arakawa, A., 2004: The cumulus parameterization problem: Past, present, and future. *J. Climate*, **17**, 2493–2525, [https://doi.org/10.1175/1520-0442\(2004\)017<2493:RATCPP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2).
- Ban, N., J. Schmidli, and C. Schär, 2014: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *J. Geophys. Res. Atmos.*, **119**, 7889–7907, <https://doi.org/10.1002/2014JD021478>.
- Belleflamme, A., X. Fettweis, and M. Erpicum, 2015: Do global warming-induced circulation pattern changes affect temperature and precipitation over Europe during summer? *Int. J. Climatol.*, **35**, 1484–1499, <https://doi.org/10.1002/joc.4070>.
- Best, M. J., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description—Part 1: Energy and water fluxes. *Geosci. Model Devel.*, **4**, 677–699, <https://doi.org/10.5194/gmd-4-677-2011>.
- Blenkinsop, S., E. Lewis, S. C. Chan, and H. J. Fowler, 2017: Quality control of an hourly rainfall dataset and climatology of extremes for the UK. *Int. J. Climatol.*, **37**, 722–740, <https://doi.org/10.1002/joc.4735>.
- Cavazos, T., and B. C. Hewitson, 2005: Performance of NCEP–NCAR reanalysis variables in statistical downscaling of daily precipitation. *Climate Res.*, **28**, 95–107, <https://doi.org/10.3354/cr028095>.
- Champion, A. J., R. P. Allan, and D. A. Lavers, 2015: Atmospheric rivers do not explain UK summer extreme rainfall. *J. Geophys. Res. Atmos.*, **120**, 6731–6741, <https://doi.org/10.1002/2014JD022863>.
- Chan, S. C., E. J. Kendon, H. J. Fowler, S. Blenkinsop, C. A. T. Ferro, and D. B. Stephenson, 2013: Does increasing the spatial resolution of a regional climate model improve the simulated daily precipitation? *Climate Dyn.*, **41**, 1475–1495, <https://doi.org/10.1007/s00382-012-1568-9>.
- , —, —, —, and N. M. Roberts, 2014a: Projected increases in summer and winter UK sub-daily precipitation extremes from high resolution regional climate models. *Environ. Res. Lett.*, **9**, 084019, <https://doi.org/10.1088/1748-9326/9/8/084019>.
- , —, —, —, —, and C. A. T. Ferro, 2014b: The value of high-resolution Met Office regional climate models in the simulation of multihourly precipitation extremes. *J. Climate*, **27**, 6155–6174, <https://doi.org/10.1175/JCLI-D-13-00723.1>.
- , —, N. M. Roberts, H. J. Fowler, and S. Blenkinsop, 2016: Downturn in scaling of UK extreme rainfall with temperature for future hottest days. *Nat. Geosci.*, **9**, 24–28, <https://doi.org/10.1038/ngeo2596>.
- Clark, P. A., K. A. Browning, C. J. Morcrette, A. M. Blyth, R. M. Forbes, B. Brooks, and F. Perry, 2014: The evolution of an MCS over southern England. Part 1: Observations. *Quart. J. Roy. Meteor. Soc.*, **140**, 439–457, <https://doi.org/10.1002/qj.2138>.
- , N. Roberts, H. Lean, S. P. Ballard, and C. Charlton-Perez, 2016: Convection-permitting models: A step-change in rainfall forecasting. *Meteor. Appl.*, **23**, 165–181, <https://doi.org/10.1002/met.1538>.
- Clogg, C. C., E. Petkova, and A. Haritou, 1995: Statistical methods for comparing regression coefficients between models. *Amer. J. Sociol.*, **100**, 1261–1293, <https://doi.org/10.1086/230638>.
- Davini, P., and C. Cagnazzo, 2014: On the misinterpretation of the North Atlantic Oscillation in CMIP5 models. *Climate Dyn.*, **43**, 1497–1511, <https://doi.org/10.1007/s00382-013-1970-y>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- DeWeaver, E., and S. Nigam, 2000: Do stationary waves drive the zonal-mean jet anomalies of the northern winter? *J. Climate*, **13**, 2160–2176, [https://doi.org/10.1175/1520-0442\(2000\)013<2160:DSWDTZ>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<2160:DSWDTZ>2.0.CO;2).
- Doswell, C. A., H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, **11**, 560–581, [https://doi.org/10.1175/1520-0434\(1996\)011<0560:FFFAIB>2.0.CO;2](https://doi.org/10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2).
- Fawcett, T., 2006: An introduction to ROC analysis. *Pattern Recognit. Lett.*, **27**, 861–874, <https://doi.org/10.1016/j.patrec.2005.10.010>.
- Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.*, **27**, 1547–1578, <https://doi.org/10.1002/joc.1556>.
- Galway, J. G., 1956: The lifted index as a predictor of latent instability. *Bull. Amer. Meteor. Soc.*, **43**, 528–529.
- George, J. J., 1960: *Weather Forecasting for Aeronautics*. Academic Press, 673 pp.
- Giorgi, F., C. Jones, and G. R. Asrar, 2009: Addressing climate information needs at the regional level: The CORDEX framework. *WMO Bull.*, **58**, 175–183.
- Golding, B., P. Clark, and B. May, 2005: The Boscawen flood: Meteorological analysis of the conditions leading to flooding on 16 August 2004. *Weather*, **60**, 230–235, <https://doi.org/10.1256/wea.71.05>.
- Gregory, D., and P. R. Rowntree, 1990: A mass flux convection scheme with representation of cloud ensemble characteristics

- and stability-dependent closure. *Mon. Wea. Rev.*, **118**, 1483–1506, [https://doi.org/10.1175/1520-0493\(1990\)118<1483:AMFCSW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1990)118<1483:AMFCSW>2.0.CO;2).
- Hewson, E. W., 1937: The application of wet-bulb potential temperature to air mass analysis. III. Rainfall in depressions. *Quart. J. Roy. Meteor. Soc.*, **63**, 323–328, <https://doi.org/10.1002/qj.49706327105>.
- Hobbs, P. V., 1978: Organization and structure of clouds and precipitation on the mesoscale and microscale in cyclonic storms. *Rev. Geophys. Space Phys.*, **16**, 741–755, <https://doi.org/10.1029/RG016i004p00741>.
- Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, **269**, 676–679, <https://doi.org/10.1126/science.269.5224.676>.
- Johns, R. H., and C. A. Doswell, 1992: Severe local storms forecasting. *Wea. Forecasting*, **7**, 588–612, [https://doi.org/10.1175/1520-0434\(1992\)007<0588:SLSF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1992)007<0588:SLSF>2.0.CO;2).
- Karoly, D. J., 1990: The role of transient eddies in low-frequency zonal variations of the Southern Hemisphere circulation. *Tellus*, **42A**, 41–50, <https://doi.org/10.3402/tellusa.v42i1.11858>.
- Keller, M., O. Fuhrer, J. Schmidli, M. Stengel, R. Stöckli, and C. Schär, 2016: Evaluation of convection-resolving models using satellite data: The diurnal cycle of summer convection over the Alps. *Meteor. Z.*, **25**, 165–179, <https://doi.org/10.1127/metz/2015/0715>.
- Kendon, E. J., N. M. Roberts, C. A. Senior, and M. J. Roberts, 2012: Realism of rainfall in a very high-resolution regional climate model. *J. Climate*, **25**, 5791–5806, <https://doi.org/10.1175/JCLI-D-11-00562.1>.
- , —, H. J. Fowler, M. J. Roberts, S. C. Chan, and C. A. Senior, 2014: Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nat. Climate Change*, **4**, 570–576, <https://doi.org/10.1038/nclimate2258>.
- , and Coauthors, 2017: Do convection-permitting regional climate models improve projections of future precipitation change? *Bull. Amer. Meteor. Soc.*, **98**, 79–93, <https://doi.org/10.1175/BAMS-D-15-0004.1>.
- Leutwyler, D., O. Fuhrer, X. Lapillonne, D. Lüthi, and C. Schär, 2016: Towards European-scale convection-resolving climate simulations. *Geosci. Model Dev.*, **9**, 3393–3412, <https://doi.org/10.5194/gmd-9-3393-2016>.
- Mason, S. J., and N. E. Graham, 2002: Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quart. J. Roy. Meteor. Soc.*, **128**, 2145–2166, <https://doi.org/10.1256/003590002320603584>.
- McCullagh, P., and J. A. Nelder, 1989: *Generalized Linear Models*. 2nd ed. Chapman and Hall/CRC, 532 pp.
- Meinshausen, M., and Coauthors, 2011: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109**, 213–241, <https://doi.org/10.1007/s10584-011-0156-z>.
- Mizieliński, M. S., and Coauthors, 2014: High-resolution global climate modelling: The UPSCALE project, a large-simulation campaign. *Geosci. Model Dev.*, **7**, 1629–1640, <https://doi.org/10.5194/gmd-7-1629-2014>.
- Moore, G. E., 2006: Cramping more components onto integrated circuits, reprinted from *Electronics*, volume 38, number 8, April 19, 1965, pp.114 ff. *IEEE Solid-State Circuits Soc. Newsl.*, **11**, 33–35, <https://doi.org/10.1109/N-SSC.2006.4785860>.
- Morcrette, C., H. Lean, K. Browning, J. Nicol, N. Roberts, P. Clark, A. Russell, and A. Blyth, 2007: Combination of mesoscale and synoptic mechanisms for triggering an isolated thunderstorm: Observational case study of CSIP IOP 1. *Mon. Wea. Rev.*, **135**, 3728–3749, <https://doi.org/10.1175/2007MWR2067.1>.
- Prein, A. F., and Coauthors, 2015: A review on regional convection-permitting climate modeling: demonstrations, prospects and challenges. *Rev. Geophys.*, **53**, 323–361, <https://doi.org/10.1002/2014RG000475>.
- Púčik, T., and Coauthors, 2017: Future changes in European severe convection environments in a regional climate model ensemble. *J. Climate*, **30**, 6771–6794, <https://doi.org/10.1175/JCLI-D-16-0777.1>.
- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, <https://doi.org/10.1175/2007MWR2123.1>.
- Schiemann, R., and Coauthors, 2017: The resolution sensitivity of Northern Hemisphere blocking in four 25-km atmospheric global circulation models. *J. Climate*, **30**, 337–358, <https://doi.org/10.1175/JCLI-D-16-0100.1>.
- Stephenson, D. B., 2000: Use of the “odds ratio” for diagnosing forecast skill. *Wea. Forecasting*, **15**, 221–232, [https://doi.org/10.1175/1520-0434\(2000\)015<0221:UOTORF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0221:UOTORF>2.0.CO;2).
- Stone, M., 1977: An asymptotic equivalence of choice of model by cross-validation and Akaike’s criterion. *J. Roy. Stat. Soc.*, **39**, 44–47.
- Swets, J. A., 1973: The relative operating characteristic in psychology. *Science*, **182**, 990–1000, <https://doi.org/10.1126/science.182.4116.990>.
- Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, **84**, 1205–1217, <https://doi.org/10.1175/BAMS-84-9-1205>.
- UKCP, 2017: UKCP18 project announcement. UK Climate Projections, accessed 15 March 2017, <http://ukclimateprojections.metoffice.gov.uk/24125>.
- Váňa, F., P. Düben, S. Lang, T. Palmer, M. Leutbecher, D. Salmond, and G. Carver, 2017: Single precision in weather forecasting models: An evaluation with the IFS. *Mon. Wea. Rev.*, **145**, 495–502, <https://doi.org/10.1175/MWR-D-16-0228.1>.
- Wetzel, S. W., and J. E. Martin, 2001: An operational ingredients-based methodology for forecasting midlatitude winter season precipitation. *Wea. Forecasting*, **16**, 156–167, [https://doi.org/10.1175/1520-0434\(2001\)016<0156:AOIBMF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2001)016<0156:AOIBMF>2.0.CO;2).
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 676 pp.
- Zappa, G., L. C. Shaffrey, and K. I. Hodges, 2013a: The ability of CMIP5 models to simulate North Atlantic extratropical cyclones. *J. Climate*, **26**, 5379–5396, <https://doi.org/10.1175/JCLI-D-12-00501.1>.
- , —, —, P. G. Sansom, and D. B. Stephenson, 2013b: A multi-model assessment of future projections of North Atlantic and European extratropical cyclones in the CMIP5 climate models. *J. Climate*, **26**, 5846–5862, <https://doi.org/10.1175/JCLI-D-12-00573.1>.