INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **35**: 481–505 (2015)
Published online 4 April 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.3996

RMetS
Royal Meteorological Society

An analysis of the synoptic and climatological applicability of circulation type classifications for Ireland

Ciaran Broderick* and Rowan Fealy
Department of Geography, NUI Maynooth, Co. Kildare, Ireland

ABSTRACT: Circulation type classifications compiled as part of the COST733 Action, 'Harmonisation and Application of Weather Type Classifications for European Regions', were evaluated based on their ability to describe variations in surface temperature (maximum and minimum) and precipitation across the Irish landmass. In all 16 different classification schemes, representative of four general approaches in synoptic typing (leader algorithm, optimization scheme, predefined types, eigenvector analysis) were considered. Several statistical measures variously quantifying performance in arranging daily observations into clearly defined homogenous groups were employed. Based on the results it was not possible to identify a single optimum classification or general approach in synoptic typing. This is related to inconsistencies in performance with respect to the specific target variable and statistical measures used; the results were also shown to be conditional on the number of circulation types (CTs) as well as spatiotemporal dependencies in performance. However, the study did indicate that those typing schemes based on predefined thresholds (Litynski, GrossWetterTypes, Lamb Weather Type) – along with the Kruizinga and Lund classifications – were better able to resolve surface temperature. With respect to precipitation those classifications derived using some optimization procedure (simulated annealing, Self Organizing Maps, *k*-means clustering) were consistently among the best-performing schemes. In capturing the relationship between synoptic-scale circulation and precipitation the importance of incorporating some measure of vorticity was highlighted; in contrast the inclusion of discrete directional patterns was shown to be important for resolving variations in local temperature. The classifications generally performed best for winter, reflecting the closer coupling between circulation and surface conditions during this period. Spatial patterns in the synoptic–climatological relationship were more apparent for precipitation. In this case those more westerly/south-westerly stations open to zonal airflow exhibited a stronger response to circulation variability.

KEY WORDS circulation classification; circulation types; synoptic climatology; atmospheric circulation; COST733; Ireland; Western Europe; evaluation of circulation classifications

Received 14 June 2012; Revised 18 November 2013; Accepted 3 March 2014

1. Introduction

The systematic categorization of atmospheric behaviour into a series of discrete archetypal circulation patterns, provides a means for studying the dynamics of large-scale circulation and its relationship with variability in surface weather and climate (Barry and Perry, 1973; Sweeney and O'Hare, 1992; Buishand and Brandsma, 1997; Jacobeit *et al.*, 2003; Ustrnul, 2006; Beck *et al.*, 2007; Esteban *et al.*, 2009). Previously synoptic typing has been used to investigate the link between distinct modes of circulation and a multitude of 'exotic' or environmental variables including: human mortality rates (Kassomenos *et al.*, 2001); hydrological drought (Stahl and Siegfreid, 1999; Vicente-Serrano and Lopez-Moreno, 2006; Fleig *et al.*, 2010) and flooding (Samaniego and Bárdossy, 2007; Pattison and Lane, 2011; Prudhomme and Genevier, 2011); viticulture (Jones and Davis, 2000); urban heat islands (Unger, 1996; Mihalakakou *et al.*, 2002; Wilby, 2003); wildfire occurrence (Kassomenos, 2009); and a range of air quality parameters (Comrie and Yarnal, 1992; McGregor and Bamzelis, 1995; Demuzere *et al.*, 2009; Leśniok *et al.*, 2010; Dayan *et al*, 2012). In addition classification schemes provide a means for exploring long-term trends and variability in large-scale circulation – in terms of both the frequency of occurrence and changing properties (e.g. persistence, within-group characteristics) of particular weather types (Werner *et al.*, 2000; Kysely and Domonkos, 2006; Kysely and Huth 2006; Beck *et al.,* 2007). Using model simulated data their application has also been extended to studying changes in climate and atmospheric dynamics under different forcing scenarios (Huth, 1997, 2000; Yarnal *et al.*, 2001; Huth *et al.*, 2008).

Their ability to describe surface climate and environmental phenomena is a key attribute of circulation type classifications (CTCs), and has contributed to the emergence of a diverse array of methods in synoptic typing – ranging from those which are purely subjectively based to those which utilize some quantitative or objective criteria. This presents difficulties when seeking to select the most appropriate scheme for a given application (Yarnal *et al.*, 2001; Huth *et al.*, 2008). Such

* Correspondence to: C. Broderick, Department of Geography, NUI Maynooth, Co. Kildare, Ireland. E-mail: Ciaran.Broderick@nuim.ie

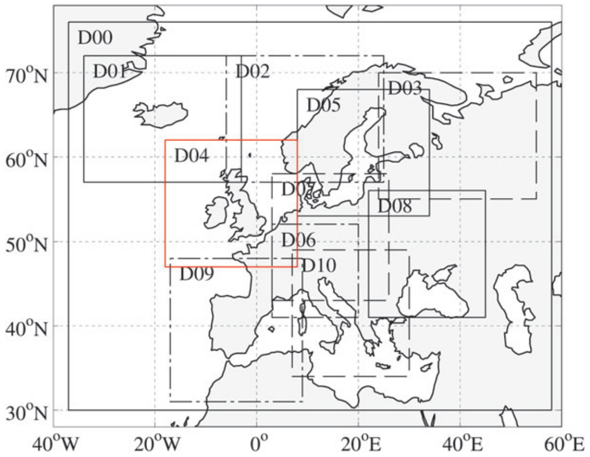

Figure 1. Spatial extent of the 11 domains for which the COST733 classifications are compiled; this study considers circulation type classifications for the D04 domain.

difficulties are compounded by inconsistencies in the development – including differences in the number of circulation types (CTs), input data, spatial extent and classifying area used – and intended application of individual schemes, both of which limits the degree to which an equitable comparison between methods can be undertaken. This provided the impetus for establishing a European CoOperation in Science and Technology (COST) Action entitled, 'Harmonisation and Application of Weather Type Classifications for European Regions' (Hereafter referred to as COST733), under which a co-ordinated evaluation and inter-comparison of CTCs could be conducted. Central to this was developing a catalogue (COST733CAT) of the most commonly employed methods in synoptic typing (Philipp *et al.*, 2010). This database archives classification series compiled independently for 12 fixed European domains (Figure 1) using a common set of input variables (ERA-40 reanalysis). It represents the core material used under the framework of the COST733 Action. By allowing a number of confounding factors to be removed or controlled, the catalogue provides an ideal basis for undertaking a systematic assessment of different methods. It must however be acknowledged that research into the development and application circulation classifications, particularly with respect to synoptic climatological and environmental analysis, has a history far beyond the COST733 Action, and exists on a much broader scale (Barry and Perry, 1973; Kalkstein *et al.*, 1987; Yarnal, 1993; Kidson, 1994a, 1994b; Bardossy *et al.*, 1995; Sumner *et al.*, 1995; Buishand and Brandsma, 1997; Romero *et al.*, 1999; Linderson, 2001; Dayan *et al.*, 2012).

Previous inter-comparison studies have focused on the performance of classifications when applied to a common task, for which the same dataset (e.g. gridded pressure fields, surface climate variables) and evaluation criteria were employed (Huth, 1996; Buishand and Brandsma, 1997; Philipp, 2009; Lupikasza, 2010; Schiemann and Frei, 2010). Others have undertaken an assessment of their statistical attributes or examined their identified CTs (Jones *et al.*, 1993; Stehlik and Bardossy, 2003). In investigating the application of circulation classifications to climatological analysis, a number of studies have considered their ability to arrange objects (e.g. daily precipitation) into well defined groups – which ideally feature maximum internal homogeneity whilst simultaneously exhibiting maximum inter-group dissimilarity (Beck and Philipp, 2010; Casado *et al.*, 2010).

To date a number of studies exploring various aspects of the COST733 catalogue have been undertaken. Beck and Philipp (2010) employed several statistical measures to quantify the performance of classifications when applied to discretize gridded mean sea level pressure (MSLP), 2-m temperature and precipitation from the ERA-40 reanalysis. Consistent with previous inter-comparison studies, the performance of individual schemes was found to vary according to both the season and climate parameter considered; their ability to resolve surface conditions was also shown to vary spatially – both within and between the COST733 domains (Figure 1). The study generally indicated that classifications performed better for winter, and exhibited a closer link with MSLP as opposed to surface temperature or precipitation; in addition better performance scores were returned for the smaller and more westerly located domains. Whilst this study represents a comprehensive overview of the catalogue – in terms of the number of classifications, target variables/domains and statistical measures it considered – by using reanalysis data it is limited in its investigation of performance in relation to sub-grid scale processes and local climate variability. The ability of classifications to resolve conditions at this scale is important for determining their potential applicability to environmental prediction and analysis. In addition this study is limited in providing an in-depth assessment of performance for each specific region. With respect to this several studies of the catalogue which focus on individual domains or specific geographic regions within them have been conducted.

Huth (2010) assessed the COST733 classifications based on their ability to stratify daily maximum and minimum temperature records from a network of 97 observational stations located across Europe. Similar to Beck and Philipp (2010), it was found that their performance varied depending on the season and spatial domain considered. In general, the schemes which provided the best results were those derived using some threshold based criterion (e.g. GrossWetterTypes, Jenkinson Collison Types and Litynski). It is noted that the original and objectivized version of the Hess and Brezowsky classification performed well over both the larger European (D00) and smaller Central European (D07) domains (Figure 1). Both classifications were also found to perform well for the winter and summer seasons respectively. In contrast the correlation-based methods (e.g. Lund and Kirchhofer), along with both the neural network and T-mode PCA-based schemes were among the poorest performing members of the catalogue.

Tveito (2010) and Casado *et al.* (2010) investigated the COST733 classifications based on their proficiency in resolving winter precipitation over Norway and Spain respectively. Casado *et al.* (2010) found that their performance was spatially dependent; in the Northern and Atlantic regions the optimal methods were those employing an optimization procedure (e.g. *k*-means clustering). In contrast over the Mediterranean region classifications based on predefined thresholds returned the best results. Those schemes which were found to perform worst included the subjective types and those employing a neural network approach. Similarly Tveito (2010) indicated that the neural network type classifications returned the least favourable results; this study also found that, for those classifications which possess a relatively small number of CTs, the threshold-based schemes generally produced the best results.

Schiemann and Frei (2010) assessed the COST733 classifications with respect to their performance in describing spatial variations in daily precipitation for the Alpine region. Using the Brier score the authors found that classifications were better able to resolve low to moderately intense precipitation (as opposed to heavy precipitation events). In this study, the Brier score was calculated for different precipitation thresholds related to exceedance probabilities. It was found that automatic methods were superior to manual schemes; this was with the exception of the SCHÜEPP (Schüepp, 1957, 1979) and ZAMG (Baur, 1948; Lauscher, 1985) classifications during summer – both of which have been developed specifically for the Alpine region. Performance levels were also found to be both spatially and temporally dependent, with the best results being returned for winter over regions to the west and north of the Alps. One of the findings common to previous inter-comparison studies is that no one optimal scheme or approach to synoptic typing could generally be identified. Essentially this suggests that the best, or most suitable scheme may be determined not only by its intended application, but also by the temporal and spatial parameters of the study. On the basis of previous research it is also clear that the performance of typing methods is dependent on the particular evaluation criteria and quantitative measures employed.

In the present study classification series from the COST733CAT catalogue (Philipp *et al.*, 2010) are evaluated based on their ability to resolve daily precipitation and temperature (minimum and maximum) at 14 meteorological stations distributed across the Irish landmass. The main objective of this research is to identify those CTCs which are most appropriate to studying the relationship between large-scale circulation and surface climate for this region. Different aspects of the classifications are considered including: their sensitivity to the number of CTs, the explanatory power of individual schemes; and the existence of spatiotemporal dependencies in performance. The article also discusses why certain classifications may be more applicable to describing surface temperature and precipitation respectively. The island's maritime climate and exposure to mid-latitude weather systems means it presents a unique set of conditions for undertaking an assessment of the catalogue. Thus the study is important for advancing our knowledge of circulation classifications and their application to synoptic–climatological analysis over the North-West Atlantic. The study also adds to knowledge already accumulated under the framework of the COST733 Action, and mirrors similar regionally specific studies previously undertaken. Whilst Beck and Philipp (2010) did include this landmass in their investigation of the COST733 classifications – albeit as part of the larger British Isles domain – it is the first study to conduct an in-depth inter-comparison of typing schemes specifically for Ireland. Furthermore, by using point scale records the article provides an indication of their descriptive power at the scales most relevant to local/regional climate and environmental analysis - and by extension their applicability to tasks including: statistical downscaling, trend analysis and environmental modelling.

2. Data

The study considered CTCs from version 2 of the COST733CAT catalogue compiled for the D04 domain (47°N–62°N, 18°W–8°E; Figure 1). This catalogue differs from earlier versions in that the objective schemes have been recalculated using the classification software developed as part of the COST733 Action. Given its inception as a typing scheme for the British Isles, the objectivized version of the original Lamb weather type (LWTo; James, 2006) – produced independently of the COST733 software – was also considered. It represents a modified version of the Jenkinson and Collison classification (Jenkinson and Collison, 1977), which in turn is the objectivized representation of the subjective Lamb weather type (Lamb, 1950). To ensure an equitable comparison between methods, only those series compiled using MSLP from the ECMWF (European Centre for Medium Range Weather Forecasts) ERA-40 reanalysis (Uppala *et al.*, 2005) were included. In the case of the D04 domain the input fields used exist at a $1° \times 1°$ resolution and cover the period 1 September 1957 to 31 August 2002. As the pressure data relates to daily conditions at 12:00 hours UTC, the resultant series are only considered representative of large-scale circulation at this particular instant in time.

In total, 16 different objective classification schemes were considered, each of which can be categorized into one of four groups depending on the general method in circulation typing they employ. Included in this is a group for classifications (GWT, JCT, LWTo and LIT) which use thresholds to assign daily patterns according to distinct CTs. The principal circulation patterns are pre-defined in that they are not identified through any prior analysis of the input fields. In contrast the three remaining groups comprise classifications whose principal types are identified using the circulation information contained in the input data. Included in this is a category for T- (temporal) or S (spatial)-mode principal component analysis (KRX, PXE, PCT and PTT); a category for classifications employing a leader algorithm (LND, KIR and ERP); and finally a category for classifications which use an optimization procedure or some variant of non-hierarchical cluster

Table 1. Details of the COST733 classifications (COST733CAT Version 2) used in the study (Philipp *et al.*, 2010).

Name	Code	Number of types (D04)	Key reference	Description
Threshold-based methods				
GrossWetterTypes	GWT	8, 18, 27	Beck *et al.* (2007)	Correlation with prototype circulation patterns
Jenkinson-Collinson -Types	JCT	9, 18, 27	Jenkinson and Collison (1977), James (2006)	Objectification of Lamb weather type (Cost733software)
Objectivized Lamb Catalogue	LWTo	10, 18, 26	Jenkinson and Collison (1977), James (2006)	Objectification of Lamb weather type (original)
Litynski	LIT	9, 18, 27	Litynski (1969)	Based on the direction of advection and vorticity of air masses
PCA-based methods				
Kruizinga	KRZ	9, 18, 27	Kruizinga (1979), Buishand and Brandsma (1997)	Using S-mode PCA scores
PCA extreme scores (without optimization)	PXE	9, 17, 18	Esteban *et al.* (2005)	Using extreme S-mode PCA scores (reassigned by Euclidean distance)
PCA obliquely rotated	PCT	9, 18, 27	Richman (1986), Huth (2000)	PCA obliquely rotated using T-mode
PCA orthogonally rotated	PTT	9, 18, 21	Huth (1993)	PCA orthogonally rotated using T-mode
Methods using Leader algorithms				
Lund	LND	9, 18, 27	Lund (1963)	Based on correlations between daily patterns
Kirchhofer	KIR	9, 18, 27	Kirchhofer (1973), Blair (1998)	Considers the spatial correlation of daily patterns along with overall correlations
Erpicum	ERP	9, 18, 27	Erpicum *et al.* (2008)	Employs a proximity index to determine the similarity of daily patterns
Methods using optimization algorithms				
k-Means clustering	CKM	9, 18, 27	Enke and Spekat (1997)	Non-hierarchal cluster analysis employing *k*-means algorithm
self-organizing maps	SOM	9, 18, 27	Michaelides *et al.* (2007)	Artificial Neural Network - Kohonen map using 1-D topology
Cluster analysis of principal components	CAP	9, 18, 27	Yarnal (1993)	*k*-Means of low-pass filtered PCA scores
PCA extreme scores (with optimization)	PXK	9, 17,18	Esteban *et al.* (2006)	Using extreme S-mode PCA scores (reassigned by *k*-means)
Simulated annealing clustering	SAN	9, 18, 27	Philipp *et al.* (2007)	Non-hierarchical cluster analysis employing simulated annealing

analysis (CKM, SOM, CAP, PXK and SAN). To explore the affect that varying the number of CTs has on the classifications, in compiling the catalogue the number of types was fixed (where possible) at 9, 18 and 27. The COST733 catalogue contains each scheme calculated using the full annual series of daily values as input data; it also holds the same classifications compiled independently for each season. In this study classifications developed using the annual series only were considered. In all, 48 individual series from the catalogue were included; Table 1 provides a description of their associated typing scheme, it also lists their assigned abbreviation and number of CTs. For a more comprehensive description of the COST733 catalogue refer Philipp *et al.* (2010).

Observational records of daily precipitation and temperature (minimum and maximum) were provided by the Irish meteorological service Met Éireann; data for Northern Ireland was obtained from the Armagh Observatory (Figure 2). Where possible the station records used cover the same period as that of the ERA-40 reanalysis. A temporal offset exists between the classifications (12:00 hours UTC) and the time for which meteorological data are reported. It was anticipated that this would not have an undue influence on the study findings. The geographical spread of the network ensured that regional variations in climate, and the associated regionally specific response to large-scale circulation were adequately represented. It is acknowledged that, given the low density of the observational network, employing a gridded dataset may have constituted a more appropriate approach; however, the smoothed nature of reanalysis data limits the assessment of classifications in resolving sub-grid

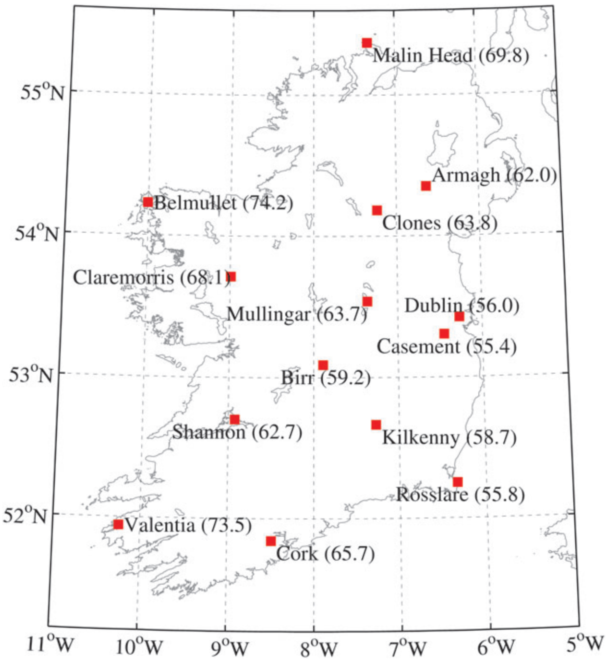

Figure 2. Location of the 14 synoptic stations whose daily records for temperature and precipitation are used in the study. Also shown is the percentage wet days (≥0.2 mm) for each station record.

scale processes and higher quantile events. At least one study has found differences between reanalysis and station records for a location in Europe (Mooney *et al.*, 2011). Furthermore, due to the gridding procedure associated with reanalysis products, and the requirement to maintain the covariance structure, using it in the current context may bias the results. Given that it was same data as was used to compile the classifications, this applies particularly in the case of the ERA-40 reanalysis.

3. Methods

The proficiency of classifications in describing precipitation and temperature (maximum and minimum) for the target area was assessed based on their ability to group daily values for each of the three variables according to CT. Based to the evaluation criteria employed, better performing schemes possess circulation patterns which correspond to a distinct regime or unique set of surface conditions. With respect to this each of the statistical measures outlined below were variously employed to assess two key attributes of the data once arranged according to CT – that of the within-group similarity and between-group separability. These criteria are generally employed to measure the quality of data partitioning in conventional cluster analysis (Gerstengarbe, 1997); however, they have been adopted as key parameters of performance in studies exploring the descriptive power of classifications (Beck and Philipp, 2010; Casado *et al.*, 2010). To account for uncertainty in the evaluation criteria, and given that no optimum approach to quantifying the quality of cluster separation exists, it was considered more pragmatic to employ several different performance measures. For certain schemes (most notably the predefined types) the surface regime specific to a given circulation pattern varies depending on the season. In order to address this, the study was conducted on a seasonal basis only; spatial patterns in performance were explored by analysing variations in the strength of the relationship between the classifications and each meteorological station.

Both of the above criteria were quantified differently by the respective performance measures used (e.g. statistical parameter, distribution function, geometric distance). While each essentially assesses the extent to which, once arranged according to CT, a defined structure in the data exists, each statistic has its own limitations, and the means by which the quality of data partitioning is quantified are methodologically different. The accuracy of each is not solely dependent on their formulation, or whether they consider the between-group separability or within-group similarity – or as in the case of the silhouette index some measure of both – but is also conditional on the challenges which the data itself presents (e.g. noisy, unbalanced group sizes, sub-clusters, uneven number of classes). The individual statistics used were selected on the basis that they have been employed in previous studies of the COST733 catalogue (Beck and Philipp, 2010; Casado *et al.*, 2010; Huth, 2010; Ustrnul *et al.*, 2010). This allows for an equitable comparison between studies, and ensures that this article can contribute to the larger body of work associated with the catalogue. It must be noted however that the performance measures were also selected on the basis that they have a history of use outside the COST733 Action (Buishand and Brandsma, 1997).

Daily precipitation has a discrete-continuous character consisting of the initial occurrence and ensuing intensity of a given event. In recognition that these processes may be affected differently by synoptic conditions it was necessary to consider each independently. In the case of rainfall occurrence a threshold value of 0.2 mm was applied to recode each time series into a binary sequence signifying wet/dry days. The SD-Occ statistic (described below) is the only measure which was applied to precipitation in this transformed state. To ensure that the relationship between each circulation pattern and surface climate was not obscured by a seasonal cycle (which may have favoured some classifications), daily anomalies calculated for temperature (both maximum and minimum temperature respectively) and precipitation respectively were used. As the study is conducted on a seasonal basis, the anomalies were calculated by subtracting the long-term mean for each season from the corresponding daily values. To account for uncertainty in the performance estimates each of the measures outlined below were calculated for 1000 bootstrapped samples obtained from each station record. For this, daily values from each time series were randomly sampled with replacement, thus the same observations could be included in the analysis multiple times. In this case the sub-samples used were of the same length as the original time series. The probability distributions generated using the output from this procedure allowed some level of statistical confidence to be associated with the results.

The first measure employed was the explained variation (*EV*) or coefficient of determination; it quantifies the proportion of variance present in the dependent variable which can be accounted for by group membership (Buishand and Brandsma, 1997; Beck and Philipp, 2010; Casado *et al.*, 2010). It is estimated by comparing the within-type variance to the total variance of the dependent variable, and can be written as:

$$\mathrm{EV} = 1 - \frac{\mathrm{WSS}}{\mathrm{TSS}}$$

here WSS refers the within-type sum of squares and is calculated as the sum of the variances specific to K groups (CTs) – estimated as the squared deviation of each observation (n) from its group mean $\left(\overline{y}_1, \overline{y}_2, \overline{y}_3, \ldots, \overline{y}_K\right)$. TSS represents the total sum of squares and is calculated as the squared deviation of each observation from the overall mean $(\overline{\overline{y}})$, summed across all values.

$$\mathrm{WSS} = \sum_{j=1}^{K}\sum_{i=1}^{n}\left(y_{ji} - \overline{y}_j\right)^2$$

$$\mathrm{TSS} = \sum_{i=1}^{n}\left(y_i - \overline{\overline{y}}\right)^2$$

The statistic assumes a value between 0 and 1, with a higher value signifying a classification possessing greater discriminatory power. The pseudo-*F* statistic (*PF*) (also known as the variance ratio score) is estimated by finding the ratio of the between group sum of squares and the within-group sum of squares, each divided by their respective degrees of freedom (Calinski and Harabsz, 1974; Milligan and Cooper, 1985; Beck and Philipp, 2010; Casado *et al.*, 2010). According to this statistic a classification is better able to resolve surface conditions if its resultant groupings return a low value for the within-group similarity, whilst simultaneously returning a high value for the between-group separability. The measure is equivalent to the F statistic in the standard analysis of variance (ANOVA), and can be written as:

$$\mathrm{PF} = \frac{\mathrm{BSS}/(K-1)}{\mathrm{WSS}/(n-K)}$$

here K denotes the number of CTs and n the number of observations; BSS represents the between-group sum of squares, and is calculated as the sum of the squared differences between the mean value for each group $\left(\overline{y}_K\right)$ and the grand mean $\left(\overline{\overline{y}}\right)$.

$$\mathrm{BSS} = \sum_{j=1}^{\mathrm{K}}\left(\overline{y}_j - \overline{\overline{y}}\right)^2$$

A large PF value is evidence that there is a greater difference between than within-groups, thus suggesting a better separation between CTs. As both the EV and PF statistics are based on determining the effect which the imposed groupings have by proportioning variance, they produce similar results; however, one of the weaknesses associated with this statistic is its dependence on the number of classes (i.e. as the number of groups increases so too does the explained variance), which precludes an equitable comparison between classifications possessing a dissimilar number of types. In contrast the inclusion of the normalization term $(n-K)/(K-1)$ prevents the PF value being biased towards larger values of K. In the context of conventional cluster analysis the optimum number of partitions is defined as the value which maximizes the PF criterion.

The third performance measure used was the within-type variability (WSD). This is estimated as a weighted mean of the variances for K groups, wherein the observations are grouped according to their associated CT. In this case the weighting is calculated based on the frequency of occurrence for a given type, with a higher occurrence frequency receiving a greater weighting in the calculation. The final value is taken as the root of the weighted mean variance. It is referred to as the pooled standard deviation by Beck and Philipp (2010), and is a modification of the formula given by Kalkstein *et al.* (1987). This statistic explicitly considers both the number of CTs (K) and number of observations (n); it is written as:

$$\mathrm{WSD} = \sqrt{\frac{\sum_{j=1}^{K}\left(n_j - 1\right)\cdot s_j^2}{\sum_{j=1}^{K}\left(n_j - 1\right)}}$$

here s represents the standard deviation and is estimated for each of K groups; n_j denotes the number of observations for a given CT (j). As this statistic quantifies the deviation of each group from its respective mean, a smaller value is indicative of a classification with more compact groupings (i.e. smaller within-type deviation). The statistic is limited to providing an estimate of the within-group similarity and does not have a defined range – it is thus considered in absolute terms.

The SIL index (SIL) – initially proposed by Rousseeuw (1987) as a geometrical measure for assessing the quality of cluster separation – was employed to quantify classification performance in grouping daily observations. The index has been employed previously in a similar context by Beck and Philipp (2010). As it is estimated based on the distance of a given observation (y_i), to other observations in its own group, as well as to observations in the closest neighbouring group, the index is a combined measure of both the between-group separability and within-group similarity. For the statistic a distance measure is used to quantify the proximity of observations, in this case the Euclidean distance was employed. The SIL index (SIL) is defined as:

$$\mathrm{SIL} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{SIL}_i$$

where

$$\mathrm{SIL}_i = \frac{(b_i - a_i)}{\max(a_i, b_i)}$$

here a_i is the average distance of the observation y_i to other observations in its own group; b_i is estimated as the average distance from y_i to all other observations in the nearest neighbouring group. The difference between b_i and a_i is divided by a_i or b_i depending on which is largest ($\max(a_i, b_i)$). A small value for a_i indicates that y_i is well matched (i.e. in close proximity) to observations in its own group; conversely a small value for b_i indicates that y_i is closer to observations in the nearest neighbouring group. This group is identified based on the minimum average distance from a given observation to each observation in all groups to which it is not a member. The SIL index is taken as the average silhouette value (SIL_i) for all observations; the resulting index ranges from −1 and +1, with a value close to +1 indicating the existence of well separated and compact groupings. In contrast a value closer to −1 indicates that the data are inappropriately grouped – or in the context of cluster analysis misclassified – and lacks a discernible structure. Given the requirement to consider the distance to all points outside an objects own group, the index may be considered more rigorous when compared with the preceding measures. This is with particular reference to the PF statistic, which only considers differences between the mean of each type when quantifying the between-group separability.

Two performance measures based on the nonparametric Kolmogorov–Smirnov (KS) test for the homogeneity of distribution functions were also used (Huth, 2010; Tveito, 2010). In both cases the test was applied to quantify the between-group separability. The statistic was firstly employed to determine whether each classification, when applied to discretize the climate series produced a type specific distribution which was distinguishable from the probability distribution of the unclassified data – this measure is referred to as KS type I in the following sections. The rejection of the null hypothesis, signifying that two distributions are significantly different indicates that the values were sampled from a specific region of the unclassified data in a structured rather than random manner. The final value was calculated as the percentage of actual rejections when the number of all possible rejections were considered. Higher values indicated that individual CTs correspond to surface conditions which were dissimilar to the non-discretized data. The KS test was also used to perform a comparison between the associated distribution functions of each CT for a given classification – this is referred to as KS type II in the following sections. In this instance, once the surface data were grouped according to CT, each type-specific distribution was compared using the KS test. The proportion of rejections for each classification was calculated and reported as a percentage of all possible rejections. This statistic thus examines the ability of CTs identified under each classification to characterize a statistically unique set of surface conditions.

For the KS test the D_s statistic, which represents the largest absolute difference between the distribution functions of two datasets (Y_1 and Y_2) is calculated using:

$$D_s = \sup_x \left| F_n(Y_1) - F_m(Y_2) \right|$$

here F_n and F_n represent the empirical cumulative distribution function of each data set. The null hypothesis regarding the similarity of both distributions – that both samples are drawn from the same population – is rejected when D_s exceeds a critical value (α), which is dependent on the sample size and selected significance level; in the present study a significance level of 0.01 was used. As the number of observations per group has the potential to bias the KS test, steps were taken to ensure it was applied in a robust manner. The statistic is considered reasonably accurate for sample sizes n_1 and n_2, such that $(n_1 \times n_2)/(n_1 + n_2) \geq 4$; in this case n_1 and n_2 refer to the size of the respective data sets being compared. However, to ensure the test produced valid results it was only conducted in cases where both groups included 30 or more observations. Whereas the preceding measures are focused on determining the spread/proximity of the data with respect to a given datum, or are limited to considering a single statistical parameter, the KS test has the advantage of comparing the entire probability distribution; however, one of its shortcomings is its potential to be biased by group size (Huth, 2010).

The final performance measure (SD-Occ) used was employed to determine whether the classifications could differentiate between a wet (≥0.2 mm) and dry day (Tveito, 2010). Before calculating this measure the precipitation series from each station was recoded into a binary (0, 1) sequence signifying the occurrence of a wet/dry day. To apply the statistic the standard deviation in the frequency of precipitation occurrence across CTs was calculated using the following:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{K} (p_i - p)^2}{K - 1}}$$

here p_i is the relative frequency of wet days per CT, and p is the frequency of wet day occurrence for the full series. The mean frequency of occurrence for the full series was taken as the point about which dispersion is measured. A high SD-Occ value is indicative of a classification which is better able to distinguish between the occurrence and non-occurrence of a wet day.

4. Results

In total 48 classification series from the COST733 catalogue were assessed based on their ability to describe daily temperature (minimum and maximum) and precipitation across the Irish landmass. To this end the statistical measures outlined above were applied to 1000 bootstrapped samples taken from each of the 14 station records used.

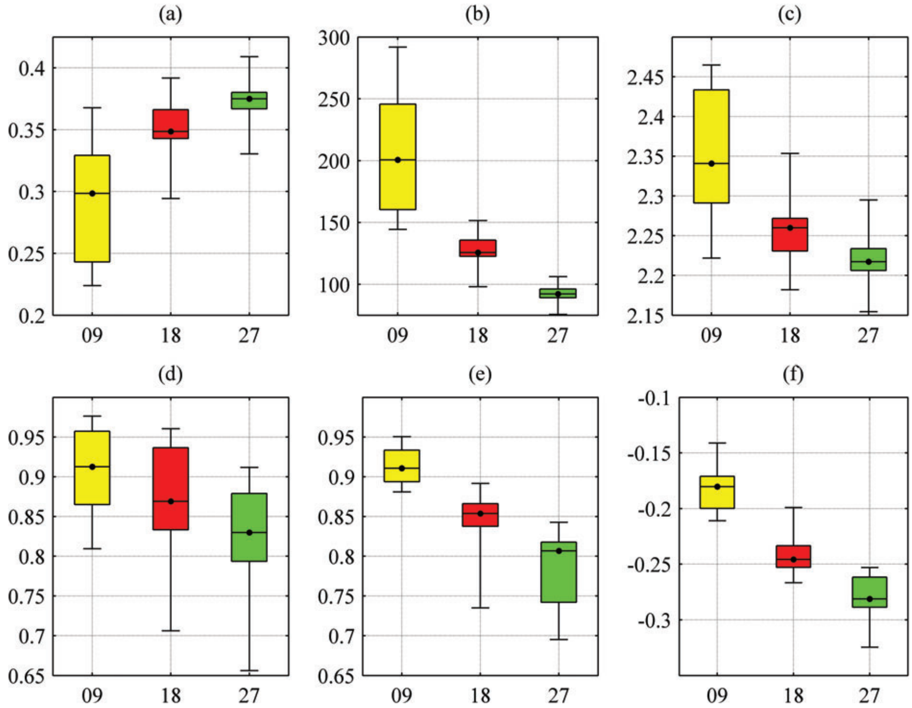

Figure 3. Classification performance for daily winter (DJF) temperature (maximum) assessed using six evaluation measures: EV (a), PF (b), WSD (c; °C, *y*-axis), KS type I (d), KS type II (e) and SIL (f). The plots are constructed using the bootstrapped test results for 36 classification series grouped according to the number of circulation types (9, 18 and 27). On each plot the central mark denotes the median, whilst the edges represent the 25th and 75th percentiles respectively; the whiskers denote the 90% interval.

Several difficulties arose during the study for which adjustments were made. Firstly, although the classifications may equate roughly to the three sub-groups possessing approximately 9, 18 and 27 CTs, in some cases the occurrence of a particular type was found to be seasonally dependent which, conditional on the period under assessment, reduced the overall number of types present. In addition several of the classifications possess circulation patterns which occur relatively infrequently (<1% probability of occurrence), the result of which is smaller group sizes. Generally these issues are exacerbated as the number of CTs is increased, and are more so associated with summer as opposed to other seasons. In order to avoid both these factors biasing the results it was necessary to exclude certain classifications (PXE, PTT, ERP and PXK) from the analysis. Given that it corresponds to the period when large-scale circulation typically has the most perceptible influence on local climate - and thus subject to the least amount of noise – the study focussed primarily on classification performance for the winter (DJF) season (Beck and Philipp, 2010; Schiemann and Frei, 2010).

4.1. Sensitivity to the number of CTs

The number of CTs is an important parameter in determining whether key synoptic states are identified. It also affects the degree to which circulation variability is captured by the classification series. Figures 3 and 4 show the aggregated bootstrapped test results for each statistic used to assess performance in describing daily winter temperature (maximum) and precipitation respectively. The boxplots correspond to a set number of types (9, 18 and 27), under which the series for each scheme was compiled; only those classifications which hold a number of types approximating to this were included. A high degree of correspondence between the results for each variable is evident, with all of the evaluation measures exhibiting a clear dependency on the number of classes. Both Figures indicate that, with an increase in the number of types, a greater proportion of the variance is captured (EV), the ability of classifications to differentiate between wet and dry conditions is improved (SD-Occ), and a closer grouping of objects is achieved (WSD); conversely, with a greater number of types the discriminatory power of classifications is also reduced, leading to circulation patterns whose associated groupings possess similar characteristics - as is demonstrated by the reduction in both KS statistics. An increase in the number of types also results in an incremental decrease in the PF statistic. The exclusively negative values returned for the SIL index indicate that the data was misclassified, with a large proportion of values being closer to those in the nearest neighbouring group as opposed to their assigned grouping. Both the SIL index and PF statistic are used in conventional cluster analysis to identify the optimum number of partitions in a dataset. Their applicability to this is due to both statistics incorporating mechanisms (such as the normalization constant in the PF statistic) which prevent the efficiency score being biased towards an increase in the number of clusters – which is one of the limitations associated with the EV statistic. Offsetting the number of groups against

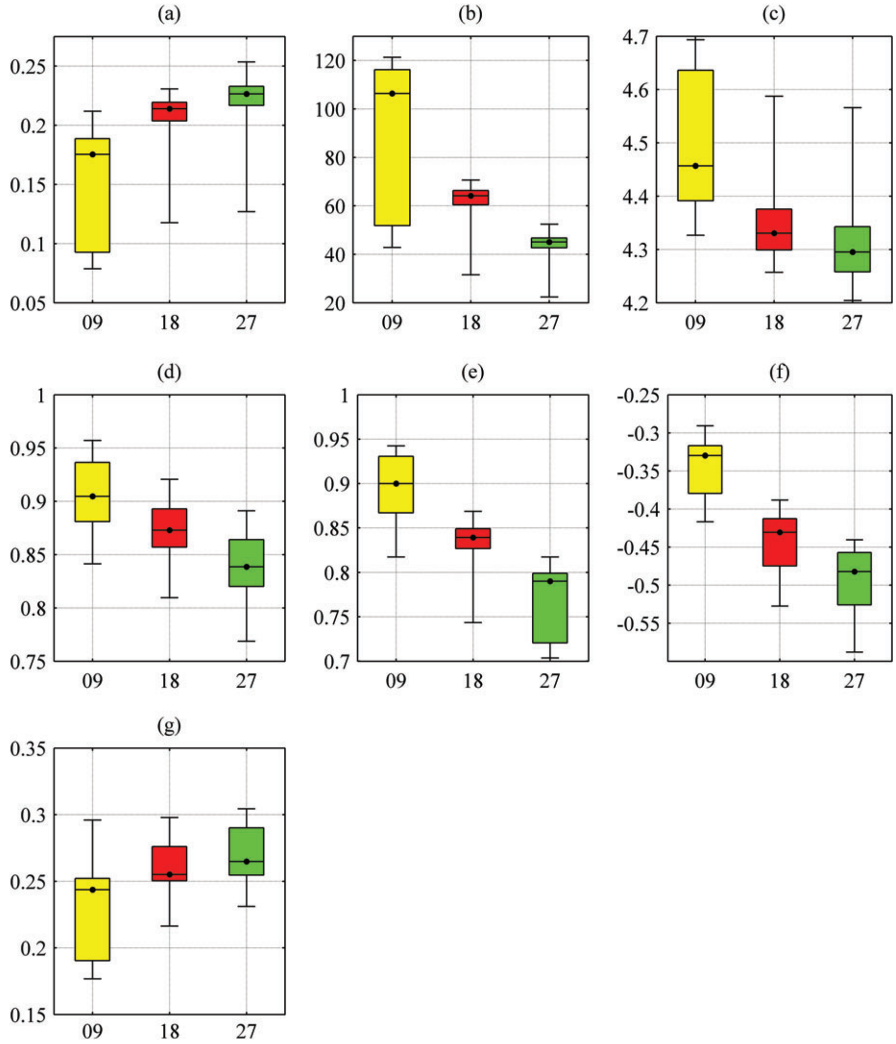

Figure 4. Classification performance for daily winter (DJF) precipitation assessed using seven evaluation measures: EV (a), PF (b), WSD (c; mm day^{-1}, *y*-axis), KS type I (d), KS type II (e), SIL (f) and SD-OCC (g). The plots are constructed using the bootstrapped test results for 36 classifications grouped according to the number of circulation types (9, 18 and 27). On each plot the central mark denotes the median, whilst the edges represent the 25th and 75th percentiles respectively; the whiskers denote the 90% interval.

the associated increase in explained variance allows for a more equitable comparison between classifications possessing a dissimilar number of types. On the basis of these statistics, it would suggest that classifications possessing 18 or 27 CTs have little added value over those incorporating 9 types; however, it is recognized that the EV and WSD statistics indicated that an increase in the number of classes above 9 resulted in a better overall performance.

4.2. Evaluation and inter-comparison of CTCs

Given the sensitivity of each statistical measure to the number of classes, an equitable comparison could only be realized if this factor was removed; hence classifications were assessed by focusing on three sub-groups [Group 1 (8–10 types); Group 2 (17–18 types); Group 3 (26–27 types)] each possessing approximately the same number of types. In the following sections classification performance with respect to each target variable is examined. Also considered is the consistency in performance across sub-groups and evaluation criteria. In assessing performance the classifications were ranked based on the aggregated bootstrapped results from all synoptic stations; thus spatial variations in performance were not considered. Whilst the study focused on the winter and to a lesser extent summer (JJA) seasons, for comparison purposes the results for spring (MAM) and autumn (SON) are also provided.

4.2.1. *Evaluation of CTCs for temperature*

Figure 5 shows the proportion of variance in maximum daily temperature which is accounted for by each series assessed on a seasonal basis. For winter the classifications captured approximately 35% of the observed variance (dependent on the number of CTs), with the best performing schemes accounting for over 40%. The lowest EV values were returned for spring and autumn respectively, for which less than 20% of the variability present was resolved. Classification performance is noted as being most variable during summer, for which EV values ranging between 0.05 and 0.35 were returned.

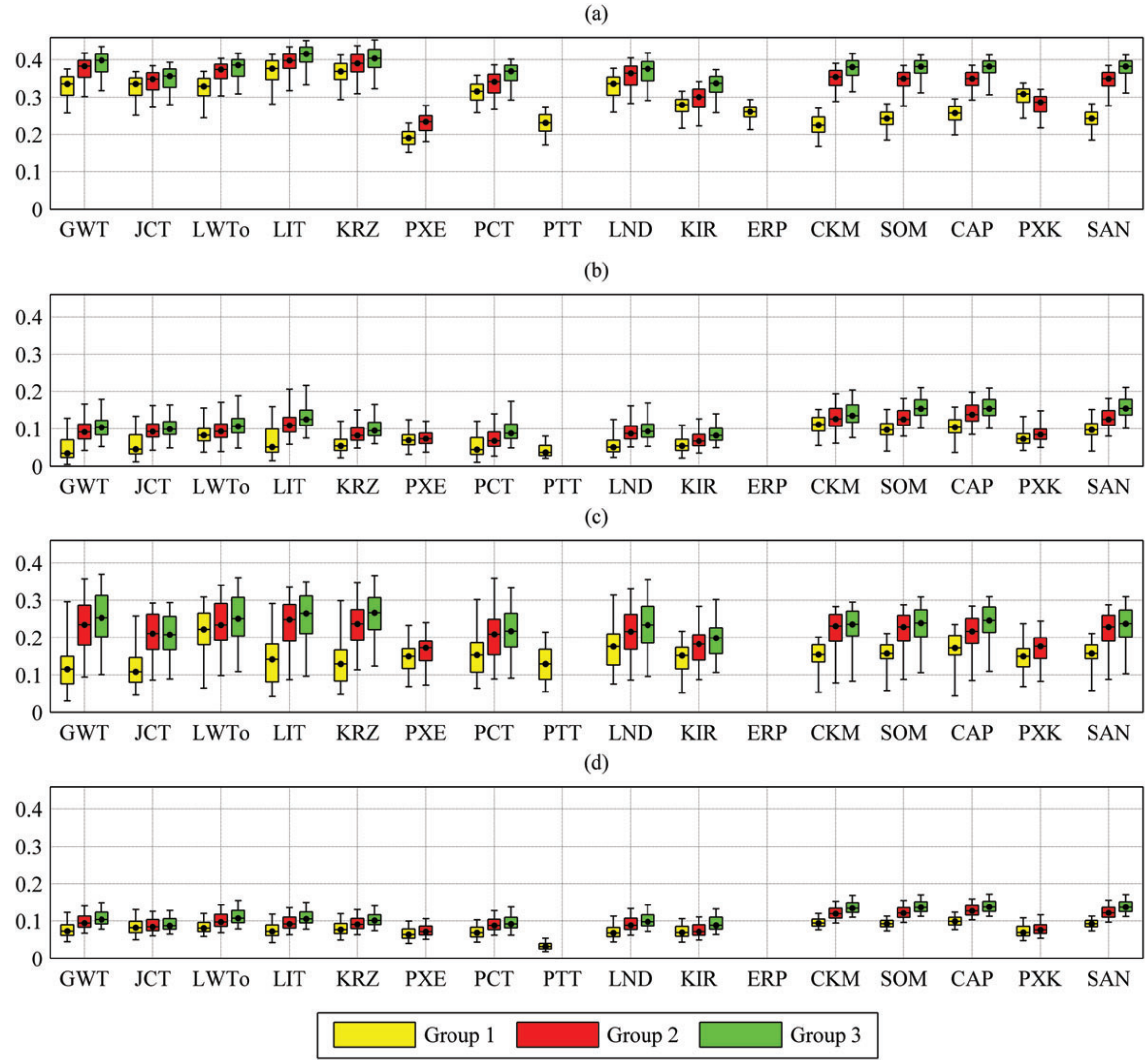

Figure 5. Classification performance for maximum daily temperature assessed seasonally, DJF (a), MAM (b), JJA (c), SON (d), based on the proportion of explained variance (EV). The plots are constructed using the results from 1000 bootstrapped samples taken from each station record. On each plot the central mark denotes the median, whilst the edges represent the 25th and 75th percentiles respectively; the whiskers denote the 90% interval. The groupings signify the number of circulation types (Group 1: 8–10 CTs; Group 2: 17–19 CTs; Group 3: 26–28 CTs).

Figure 6 shows the results for each classification aggregated based on the principal method in synoptic typing they employ. In this case, performance when applied to discretize maximum daily temperature for winter was considered. The best results were generally returned by the predefined type method, an outcome which is particularly apparent for the first sub-group according to the PF, EV and KS type I statistics. For the second and third sub-groups a greater degree of consistency between methods is evident; also notable for these sub-groups is the relative improvement in performance of the optimization-based schemes – this is most apparent under the EV, PF, WSD and KS type II statistics.

Tables 2 and 3 show the classification series ranked (lowest to highest) according to their performance (best to worst) when applied to daily maximum temperature for the winter and summer seasons respectively. The rankings were determined independently for each sub-group (approximately 9, 18 and 27 CTs) based on the bootstrapped test results. In this case the rankings were deemed statistically significant if they remained constant across 95% of samples. These tables are supplemented by the boxplots shown in Figures 5, 7 and 8, which provide a graphic of the aggregated results for each series. With respect to the second and third sub-groups, the LIT, KRZ and GWT classifications generally attained the lowest rankings – with four (EV, PF, WSD, KS type II) of the six statistics applied indicating that the LIT scheme was the optimal classification. According to the EV statistic it explained over 40% of the variability present in maximum daily temperature for this season (Figure 5(a)). It is less clear which classification performed best for the first sub-group; however, the LIT scheme was ranked lowest according to four of the six statistical measures used. Also among the consistently lowest ranked classifications for this group (8–10 CTs) were the KRZ and LND schemes.

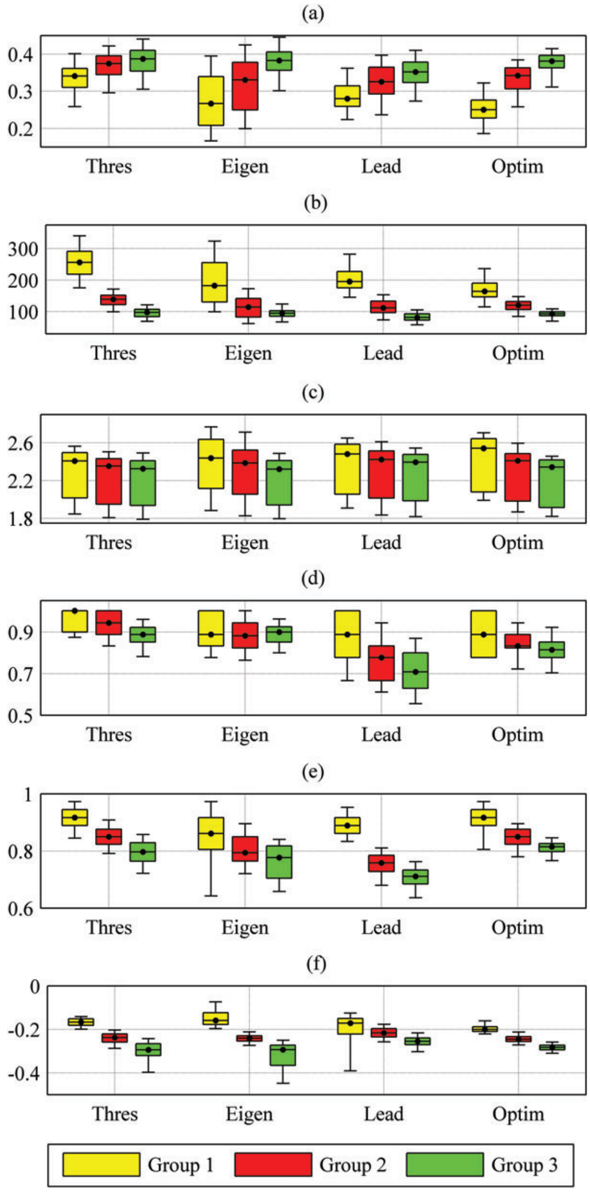

Figure 6. The performance of each principal typing method (pre-defined threshold, eigenvector based, leader algorithm, optimization procedure) for maximum daily winter (DJF) temperature assessed using six evaluation measures: EV (a), PF (b), WSD (c; °C, *y*-axis), KS type I (d), KS type II (e) and SIL (f). On each plot the central mark denotes the median, whilst the edges represent the 25th and 75th percentiles respectively; the whiskers denote the 90% interval. The groupings signify the number of circulation types (Group 1: 8–10 CTs; Group 2: 17–19 CTs; Group 3: 26–28 CTs).

Whilst the ranking of most classifications remained relatively stable across sub-groups, particularly with respect to the best/worst performing schemes, the descriptive ability of the optimization type classifications (e.g. SOM, SAN and CKM) was found to noticeably improve as the number of types increased (relative to other series). This is illustrated by the EV values for the CKM scheme (Figure 5; a) which increased from ~0.2 (Group 1) to ~0.4 (Group 3). In contrast, for the KS type II statistic there was a noted reduction in performance for the PCT, LND and KIR schemes as the number of CTs was expanded from 9 to 18. Although the results for minimum daily winter temperature are not shown, the LIT, KRZ and GWT series were also found to be among the lowest ranked classifications for this variable. Also consistent with maximum temperature was the consistently poor performance of the PXE, PXK and KIR schemes.

There appears to be a high degree of correspondence between measures regarding which classifications performed best for winter (Table 2). Given that the statistics variously assess the between-group separability and/or within-group similarity, it suggests that classification performance was consistent with respect to both evaluation criteria. Variations in the EV, PF and WSD statistics can be attributed to slight differences in the number of CTs, and the means by which this is accounted for by each statistical measure. However, despite a high level of agreement between statistics, their results were generally found to be inconsistent with those of the SIL index – for which the PTT, LND and KIR schemes were ranked lowest for the first, second and third sub-groups respectively. The divergence between this and other statistics is best illustrated by the LIT classification, whose poor performance under the SIL index was in complete disagreement with each of other measures used. In addition all series returned negative values for the index, indicating that the data was misclassified irrespective of the imposed groupings.

Generally the results for winter were not replicated for the summer season, where a greater divergence across sub-groups and evaluation criteria is evident (Table 3). Due to the greater variability in results - as shown by the fewer number of statistically significant rankings - it is more difficult to identify a single best or worst performing scheme for this season. In the case of the first sub-group, the LWTo, LND, CAP and SOM schemes were among the lowest ranked according to the majority of statistics. In contrast, the JCT, PTT and PCT classifications were found to be consistently amongst the poorest performing schemes. The disagreement noted between individual measures was most apparent for the second and third sub-groups. In this case the optimization based schemes (particularly SOM and SAN) generally attained the lowest rankings for both KS statistics. In contrast the pre-defined types (notably the GWT, LWTo and LIT), along with the KRZ scheme performed best according to the EV, PF and WSD measures. In keeping with the results for winter, the PXE, PXK and PTT schemes were found to be among the worst performing classifications. Also of note for summer was the high ranking of the JCT classification (consistent across sub-groups), which is generally at odds with the other threshold based methods, despite the fact that they all possess similar circulation patterns (for example each include eight clearly defined directional types). In relation to the ranking of individual schemes, the results for summer maximum temperature mirror those for minimum temperature; however, the classifications typically exhibited a lesser ability to resolve this variable.

The low SIL values for classifications which performed well according to other statistical measures (e.g. EV,

Table 2. Circulation type classifications ranked (lowest to highest) according to their performance (best to worst) when applied to discretize maximum daily temperature (°C) for the winter season (DJF)

	Group 1 (8–10 types)						Group 2 (17–19 types)						Group 1 (26–28 types)					
	EV	PF	WSD	KS I	KS II	SIL	EV	PF	WSD	KS I	KS II	SIL	EV	PF	WSD	KS I	KS II	SIL
GWT	3.9	**3.0**	4.0	1.4	6.0	**4.0**	**3.0**	**3.0**	**3.0**	2.5	**9.0**	**2.0**	**3.0**	**3.0**	**3.0**	1.2	7.9	10.2
JCT	5.0	4.9	4.9	4.2	10.3	**5.0**	9.8	9.8	9.8	1.6	7.6	10.4	**11.0**	**11.0**	**11.0**	7.0	**9.0**	10.8
LWTo	5.6	7.8	5.6	3.5	9.5	10.8	**4.0**	**4.0**	**4.0**	2.6	7.3	4.1	5.9	**4.0**	6.1	2.1	5.8	2.5
LIT	1.2	1.2	1.2	4.3	1.4	8.7	**1.0**	**1.0**	**1.0**	6.5	**1.0**	**14.0**	**1.0**	**1.0**	**1.0**	5.1	**1.0**	9.2
KRZ	1.8	1.8	1.8	2.1	1.6	7.7	**2.0**	**2.0**	**2.0**	3.2	2.8	8.3	**2.0**	**2.0**	**2.0**	3.4	3.3	4.9
PXE	**16.0**	**16.0**	**16.0**	14.6	14.9	10.2	**14.0**	**14.0**	**14.0**	11.9	11.3	7.7	–	–	–	–	–	–
PCT	**7.0**	**6.0**	**7.0**	10.8	12.3	**2.0**	10.9	10.9	**11.0**	6.4	12.9	7.6	10.0	9.9	10.0	3.7	11.1	11.6
PTT	14.1	14.1	14.1	7.8	**16.0**	**1.0**	–	–	–	–	–	–	–	–	–	–	–	–
LND	3.5	4.1	3.5	6.3	11.3	**3.0**	**5.0**	**5.0**	**5.0**	11.4	**14.0**	**1.0**	**9.0**	9.1	**9.0**	10.4	11.8	1.8
KIR	**9.0**	**9.0**	**9.0**	15.7	10.5	7.5	**12.0**	12.3	**12.0**	**14.0**	11.8	5.2	12.0	12.0	12.0	12.0	10.1	1.8
ERP	10.2	**10.0**	10.1	13.4	7.4	**16.0**	–	–	–	–	–	–	–	–	–	–	–	–
CKM	14.9	14.9	14.9	13.2	5.9	**13.0**	6.2	6.2	6.2	5.2	3.5	6.3	**6.8**	7.2	**6.6**	5.6	6.9	7.2
SOM	**13.0**	**13.0**	**12.0**	10.6	4.2	**15.0**	8.9	8.9	8.9	10.3	5.6	11.8	**5.8**	**6.4**	**5.8**	**8.4**	2.9	7.3
CAP	10.9	**11.0**	10.9	11.7	7.4	**12.0**	7.2	7.3	7.2	9.0	2.9	12.7	5.9	6.4	6.0	10.3	5.3	4.4
PXK	**8.0**	7.2	**8.0**	6.9	14.0	**6.0**	**13.0**	12.7	**13.0**	11.2	**10.0**	3.2	–	–	–	–	–	–
SAN	**12.0**	**12.0**	**13.0**	9.6	3.2	**14.0**	7.9	7.9	7.9	9.2	5.4	10.9	5.5	6.1	5.5	8.7	2.9	6.4

The average ranking for each classification, estimated using the results from 1000 bootstrapped samples taken from each station record is shown. The individual rankings are considered statistically significant (shown in bold) if they remain constant across 95% of samples. Classifications are grouped according to the number of circulation types (~ 9, 18 and 27), and are shaded based on their principal typing method (pre-defined threshold, eigenvector based, leader algorithm, optimization procedure).

Table 3. Circulation type classifications ranked (lowest to highest) according to their performance (best to worst) when applied to discretize maximum daily temperature (°C) for the summer season (JJA)

	Group 1 (8–10 types)						Group 2 (17–19 types)						Group 1 (26–28 types)					
	EV	PF	WSD	KS I	KS II	SIL	EV	PF	WSD	KS I	KS II	SIL	EV	PF	WSD	KS I	KS II	SIL
GWT	13.7	8.0	13.8	7.5	7.8	**1.0**	3.0	2.2	3.0	6.4	6.8	6.4	3.0	1.3	3.0	6.6	**8.0**	2.3
JCT	**15.0**	14.9	**15.0**	10.3	11.7	**2.0**	10.6	10.9	10.6	10.7	12.7	3.5	**11.0**	**11.0**	**11.0**	10.5	**12.0**	7.3
LWTo	**1.0**	**1.0**	**1.0**	5.4	3.2	7.7	2.5	2.7	2.5	6.6	3.9	2.5	3.8	2.7	3.8	5.5	5.1	1.8
LIT	9.9	8.9	10.1	13.5	7.1	6.7	1.6	1.9	1.4	7.4	4.6	8.4	2.2	4.3	2.1	6.5	6.2	4.2
KRZ	11.9	11.9	**12.0**	7.3	9.3	10.7	3.0	3.8	3.1	9.8	6.7	**10.0**	**1.0**	2.4	1.1	8.8	6.5	6.7
PXE	9.7	11.5	9.7	13.6	8.9	5.7	**14.0**	13.9	**14.0**	11.7	11.7	5.2	–	–	–	–	–	–
PCT	4.1	4.0	4.4	7.7	14.0	8.9	10.2	10.1	10.3	13.9	13.9	4.5	**10.0**	**10.0**	**10.0**	10.8	10.8	8.5
PTT	13.3	14.1	13.2	14.0	14.9	9.5	–	–	–	–	–	–	–	–	–	–	–	–
LND	**2.0**	**2.0**	2.1	9.5	5.2	3.8	7.5	8.6	7.8	6.1	9.8	8.6	6.2	8.9	6.4	8.4	10.1	5.7
KIR	10.2	12.0	9.9	12.1	5.8	6.5	**12.0**	12.7	**12.0**	13.0	11.4	4.6	**12.0**	**12.0**	**12.0**	10.7	**9.0**	3.9
CKM	7.2	7.5	7.0	5.3	7.3	14.2	6.3	4.4	5.9	1.9	7.0	**14.0**	**9.0**	4.3	8.8	4.1	4.2	6.4
SOM	6.0	6.2	4.9	2.3	2.1	14.4	6.3	7.5	5.8	4.4	2.6	12.5	6.8	6.8	6.8	1.5	1.5	10.0
CAP	**3.0**	3.1	2.9	2.9	9.6	**12.1**	9.1	7.2	9.0	2.5	2.8	11.2	5.1	6.4	5.1	2.6	3.0	10.8
PXK	8.2	9.8	8.2	7.6	12.1	3.5	**13.0**	12.4	13.0	7.1	9.5	1.2	–	–	–	–	–	–
SAN	5.0	5.2	5.9	1.2	1.1	13.4	6.1	6.7	6.7	3.4	1.6	12.3	7.9	7.8	7.9	2.0	1.6	10.4

The average ranking for each classification, estimated using the results from 1000 bootstrapped samples taken from each station record is shown. The individual rankings are considered statistically significant (shown in bold) if they remain constant across 95% of samples. Classifications are grouped according to the number of circulation types (~ 9, 18 and 27), and are shaded based on their principal typing method (pre-defined threshold, eigenvector based, leader algorithm, optimization procedure). The ERP classification is excluded for this season.

PF and KS type II) is due to the within-group similarity - or in the context of the SIL index the distance between observations assigned to the same group (measured using a_i in the SIL formula above) - being outweighed by deficiencies in the between-group separability. Related to this is how the SIL index quantifies the difference between groups. In this case the average linkage is used - measured as the mean proximity of a given value to all other values in the nearest neighbouring group (quantified using b_i in the SIL formula above). This is best illustrated by the PTT scheme, which is ranked lowest according to the SIL index for classifications incorporating 8-10 CTs. Using maximum daily winter temperature from the Valentia station as a test case, Figure 9(a) shows that the PTT classification returned the second highest value for the sum of a_i, indicating that its groups were poorly defined and lacked internal cohesion. Generally the sum of a_i was in agreement with the EV, PF and WSD statistics, all of which indicate that the LIT classification was the best performing scheme. In contrast, the PTT classification returned the highest value for the sum of b_i (Figure 9(b)), indicating

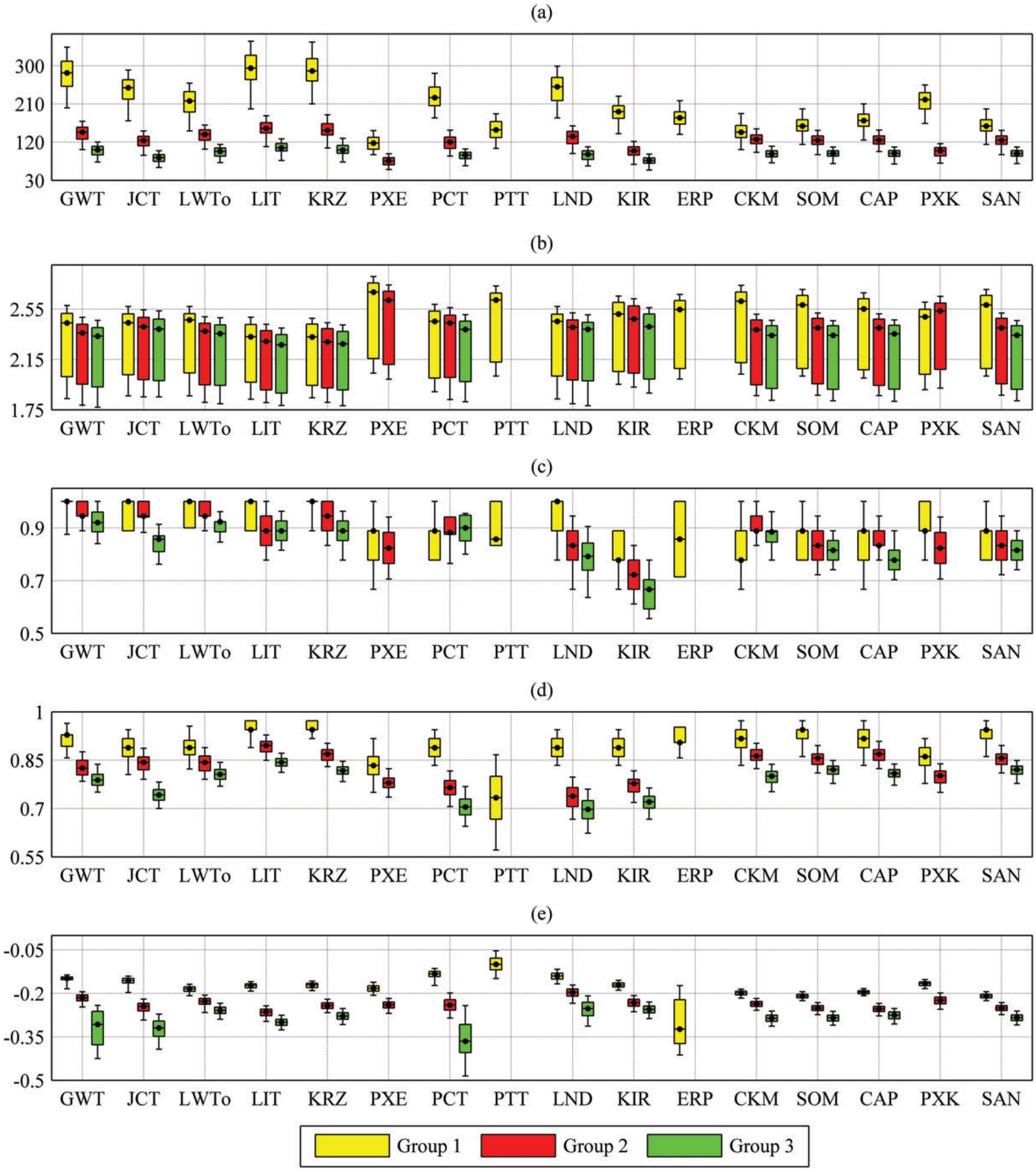

Figure 7. Classification performance for maximum daily winter (DJF) temperature assessed using five evaluation measures: PF (a), WSD (b; °C, *y*-axis), KS type I (c), KS type II (d) and SIL (e). The plots are constructed using the results from 1000 bootstrapped samples taken from each station record. On each plot the central mark denotes the median, whilst the edges represent the 25th and 75th percentiles respectively; the whiskers denote the 90% interval. The groupings signify the number of circulation types (Group 1: 8–10 CTs; Group 2: 17–19 CTs; Group 3: 26–28 CTs).

that the observations were closer to values inside their own group as opposed to the nearest neighbouring group. In this case a high value for the sum of b_i resulted in the PTT classification attaining a better overall index value (Figure 9(c)) – irrespective of its high value for the sum of a_i. Essentially this indicates that observations belonging to a particular group appear well separated from the nearest neighbouring group, simply because the groups are so poorly defined. Despite the prominence of the between-group separability in determining the low ranking of the PTT scheme, the same classification did not perform to a similar level under either KS statistics. Therefore differences in how this criterion is quantified by each measure (Euclidean distance as opposed to probability distributions) are important in determining the overall performance. An additional assessment conducted using the so called 'faster silhouette index' (Beck and Philipp, 2010) – for which the distance between data points and the centroid of the nearest neighbouring group is used – returned the same (negative values) results as the conventional index, albeit with slight variations in the individual rankings.

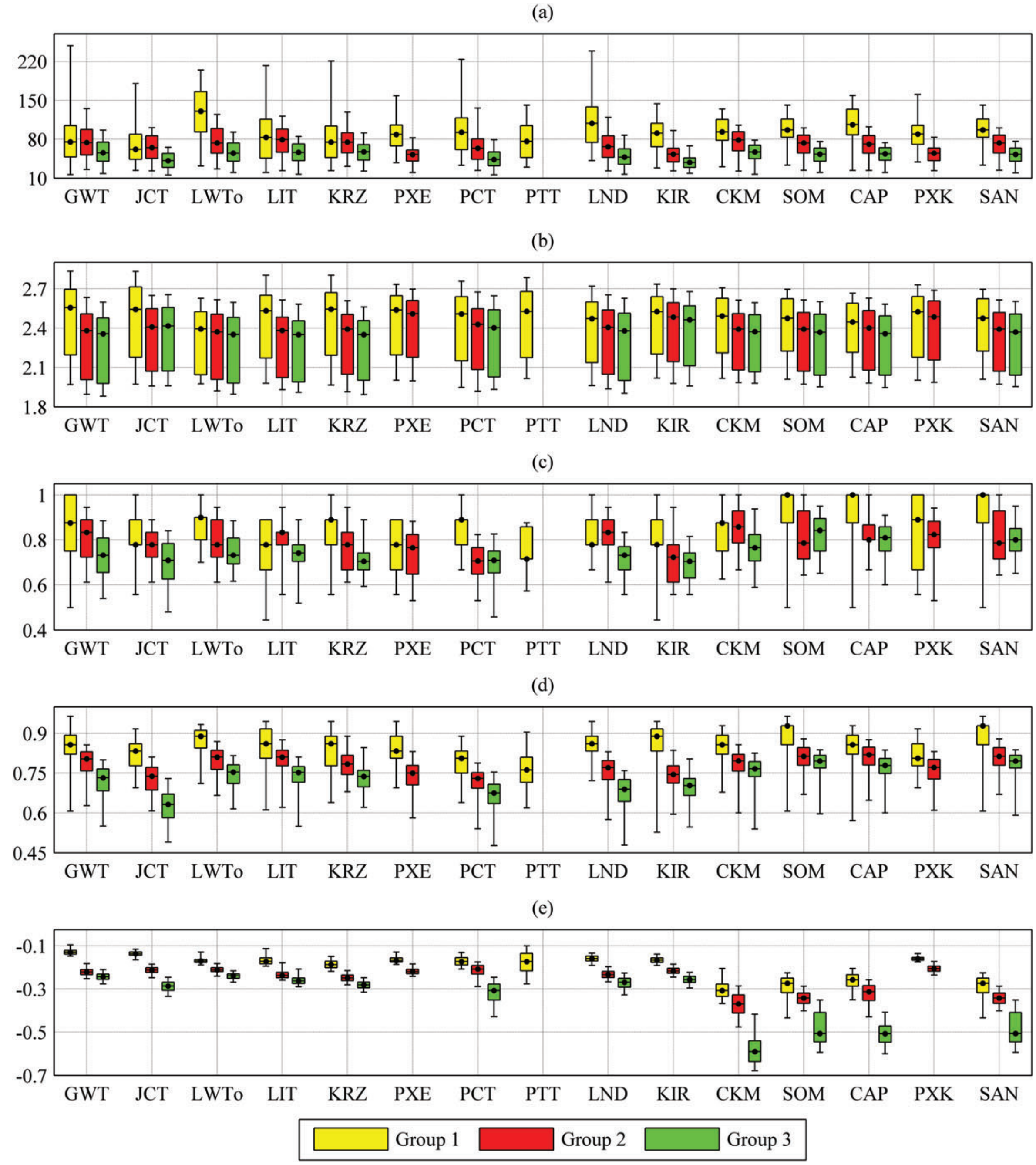

Figure 8. Classification performance for maximum daily summer (JJA) temperature – assessed using five evaluation measures: PF (a), WSD (b; °C, *y*-axis), KS type I (c), KS type II (d) and SIL (e). The plots are constructed using the results from 1000 bootstrapped samples taken from each station record. On each plot the central mark denotes the median, whilst the edges represent the 25th and 75th percentiles respectively; the whiskers denote the 90% interval. The groupings signify the number of circulation types (Group 1: 8–10 CTs; Group 2: 17–19 CTs; Group 3: 26–28 CTs). The ERP classification is excluded for this season.

4.2.2. *Evaluation of CTCs for precipitation*

Figure 10 shows the proportion of variance (EV) in daily precipitation which is explained by the COST733 classifications estimated for each season. According to this measure the classifications generally performed best for winter and worst for summer respectively. In relation to the former, the individual series explained between ~7 (PTT; 9 CTs) and ~28 (KRX; 27 CTs) per cent of the variability in receipts. In contrast, for summer the classifications captured between ~3% (LIT; 9 CTs) and ~18% (JCT; 27 CTs) of the variance present. Differences in the transition (i.e. spring and autumn) seasons were less apparent, however marginally higher EV values were returned for spring. As shown in Figure 10 the performance of individual series (relative to one another) remained relatively constant across each season and sub-grouping.

Figure 11 shows the bootstrapped test results for winter aggregated according to each general approach in synoptic typing. With respect to classifications possessing 8-10

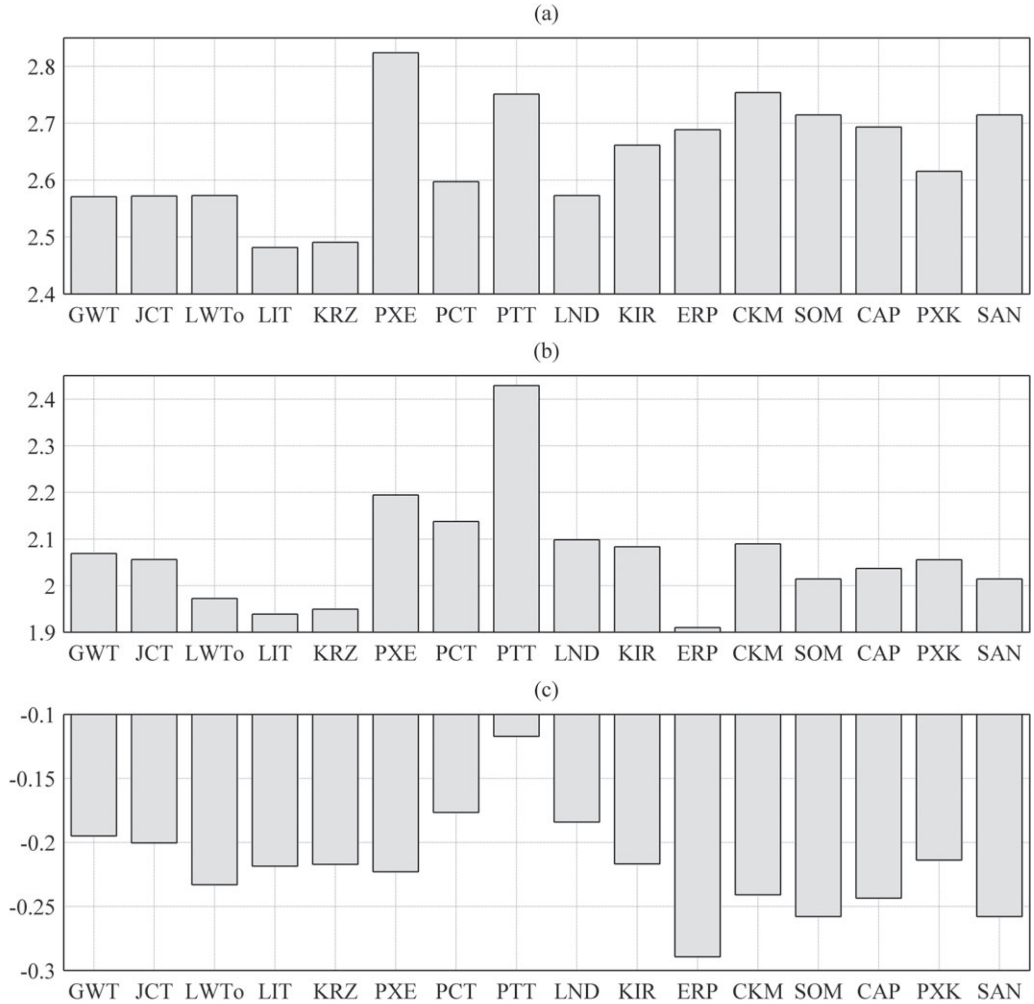

Figure 9. Deconstruction of the SIL index showing the sum of the distance between observations assigned to the same group (a; a_i); the sum of the distance from a given observation to all observations in the nearest neighbouring group (b; b_i); and the silhouette values (c; SIL). The statistic is applied to daily winter (DJF) temperature (maximum) from the Valentia station.

CTs, the optimization based schemes generally produced the best results. This is most apparent under the EV, PF, KS type II and SD-Occ statistics. The first sub-grouping is also notable for the poor performance of the predefined type schemes; however, there is a relative improvement in their performance as the number of CTs increased from 9 to 18/27. For the second and third sub-group the leader algorithm based schemes performed relatively poorly, as is evidenced in the results of the EV, PF, KS type I and KS type II statistics.

Table 4 shows the classification series ranked (lowest to highest) based on to their ability (best to worst) to resolve daily precipitation for the winter season. The average rankings were estimated based on the bootstrapped test results. This data was also employed to construct the boxplots shown in Figures 10 and 12. The EV, PF and WSD statistics all returned similar results, as did both KS statistics. In accordance with the findings for daily temperature, the SIL index was generally found not to be in agreement with each of the other measures used. Also consistent with the results for temperature were the exclusively negative values returned by the index. Whilst there was a degree of consistency across measures as to the ranking of individual series, there was a general lack of agreement between sub-groups as to which was the best/worst performing methods.

For classifications possessing 8–10 CTs the optimization-based schemes (with the exception of the PXK scheme) generally performed best. In contrast, for this sub-group the predefined type schemes returned consistently poor results. This however is with the notable exception of the LWTo series, which is ranked lowest under five of the seven statistical measures used. Results for the SD-Occ statistic indicated that the LWTo, LND, PCT, SOM and GWT classifications were the most proficient in differentiating between wet and dry conditions. The results for this statistic were in general agreement with the EV, PF and WSD measures, each of which also explicitly considers the within group-similarity. For this sub-group the PTT classification generally returned the least favourable results. For classifications possessing 17–19 CTs there was a noted improvement (relative to other series) in performance for the threshold based schemes. Those series derived using an eigenvector

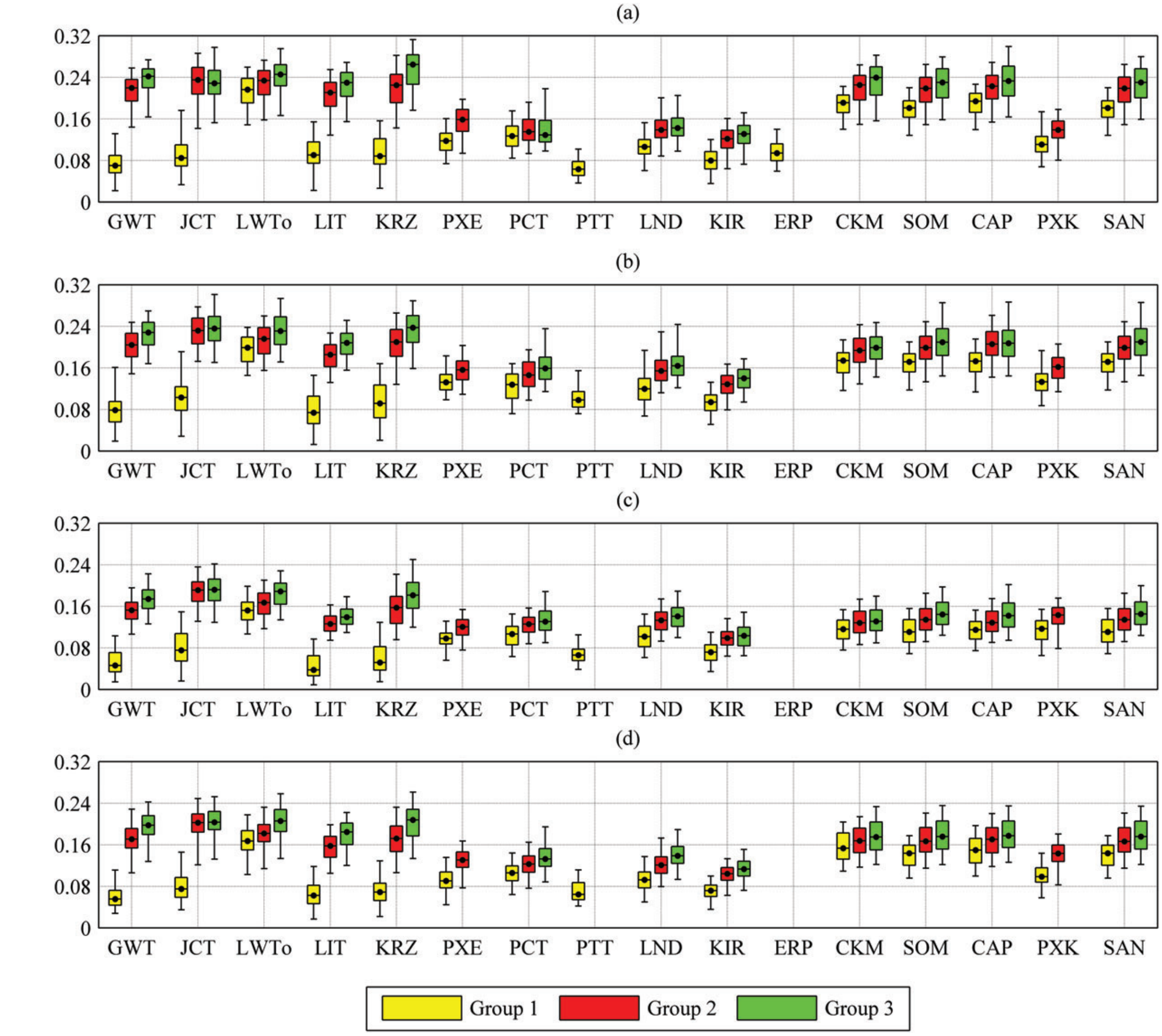

Figure 10. Classification performance for daily precipitation assessed seasonally, DJF (a), MAM (b), JJA (c), SON (d), based on the proportion of explained variance (EV). The plots are constructed using the results from 1000 bootstrapped samples taken from each station record. On each plot the central mark denotes the median, whilst the edges represent the 25th and 75th percentiles respectively; the whiskers denote the 90% interval. The groupings signify the number of circulation types (Group 1: 8–10 CTs; Group 2: 17–19 CTs; Group 3: 26–28 CTs).

technique or some variant of a leader algorithm were generally found to perform poorly, with the LND, KIR, PXK and PTT schemes in particular producing unfavourable results. As shown in Table 4 the LWTo scheme was among the lowest ranked classifications for this sub-group, as was the JCT, CKM and KRZ schemes. Patterns evident in the performance of classifications for the second sub-group were largely repeated for the third sub-group. In this case those series derived through some optimization procedure, or which are based on predefined thresholds attained the lowest rankings, with the LWTo, CKM, SAN, SOM and GWT classifications in particular returning favourable results. In contrast the KIR, LND and PCT schemes were generally ranked as the poorest performing schemes for this sub-group.

In general the LWTo classification, along with the optimization based schemes (excluding PXK) are noted as performing well according to each evaluation criteria and across all sub-groups. In contrast the PTT, ERP and KIR were consistently found to be among the poorest performing schemes. With respect to the ranking of individual series, the findings for winter were generally repeated for other seasons – as is evidenced by the EV values shown in Figure 10 – and thus are not discussed. This comes with the caveat that performance scores were lower, and the differentiation between classifications in terms of their ranking was less well defined.

4.3. Spatiotemporal variations in classification performance

To fully investigate the suitability of the COST733 classifications to synoptic–climatological analysis of the study area, it was necessary to explore spatiotemporal variations in their performance. To this end the proficiency of the classifications in resolving daily precipitation and temperature at individual synoptic stations was investigated.

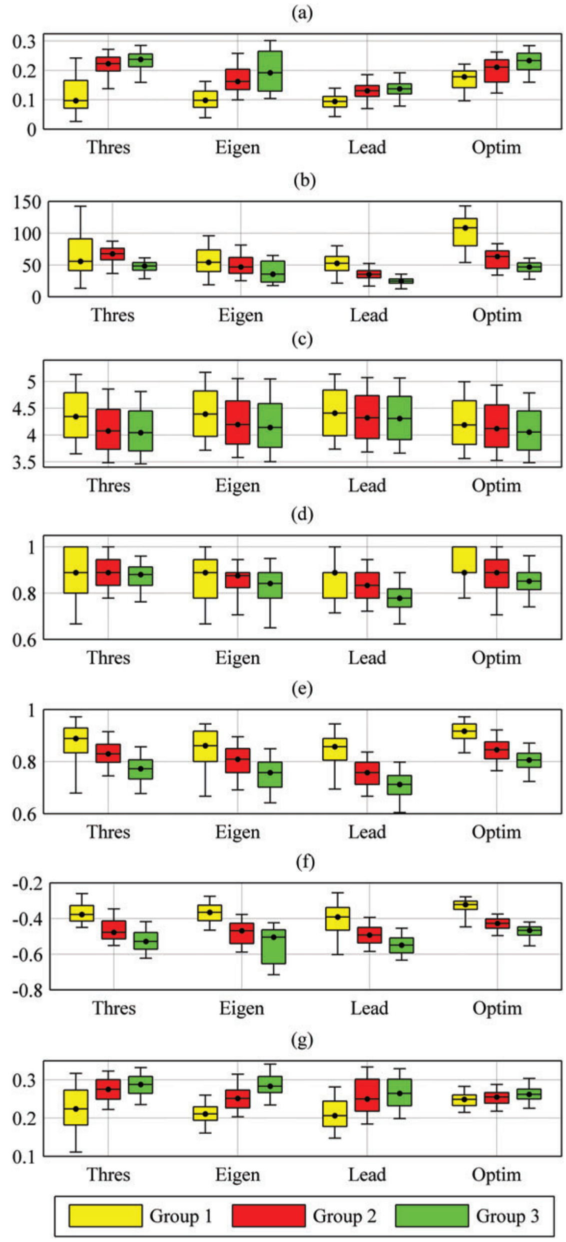

Figure 11. The performance of each principal typing method (pre-defined threshold, eigenvector based, leader algorithm, optimization procedure) for daily winter (DJF) precipitation assessed using seven evaluation measures: EV (a), PF (b), WSD (c; mm day^{-1}, *y*-axis), KS type I (d), KS type II (e), SIL (f), SD-Occ (g). The plots are constructed using the results for each scheme aggregated based on their principal classifying method. On each plot the central mark denotes the median, whilst the edges represent the 25th and 75th percentiles respectively; the whiskers denote the 90% interval. The groupings signify the number of circulation types (Group 1: 8–10 CTs; Group 2: 17–19 CTs; Group 3: 26–28 CTs).

Previous studies undertaking such an examination have focused on comparing performance over distinct climatic regions (e.g. Casado *et al.*, 2010); however, owing to its location on the western edge of Europe, the Atlantic Ocean exerts a moderating influence on climate, meaning Ireland does not possess contrasting climatic zones, nor does its climate generally exhibit extremes of conditions. As a result the proficiency of classifications in capturing conditions across highly dissimilar climate regimes/regions could not be assessed. The scope of the study was therefore limited to examining whether any spatial biases in performance exist.

In order to explore this the results from two performance measures (PF and KS type II) – standardized using the results from all observational records for each season (DJF, MAM, JJA, SON) – were plotted against one another. These criteria were chosen on the basis they both consider different aspects of performance. The PF statistic describes the ratio of the within type-variance to the between type-variance, with large values indicating the existence of well defined groups. With respect to the KS type II statistic, it measures the separability between groups based on the similarity of their statistical distributions. In the case of both statistics, their results were standardized by taking the seasonal results for each station (4×14) and calculating their respective z-score (i.e. subtracting the overall mean from each value and dividing by the standard deviation). Each point on a given plot represents the value returned for an identified synoptic station. To construct the plots two classification series were considered, both of which possess 18 CTs and sample different methodological approaches. The LIT and KIR schemes were selected to examine maximum temperature, whilst the JCT and KIR series were chosen for assessing daily precipitation. Each classification was selected on the basis that it represented among the best/worst performing schemes for that variable.

Figures 13 and 14 show seasonal scatter plots of the standardized scores returned by the LIT and KIR series when applied to records of maximum daily temperature. Despite differences in the overall performance of both series, there are a number of commonalities in the plots for each. Both classifications performed best for winter, signified by the tight clustering of points in the northeast quadrant of each plot (Figures 13(a) and 14(a)). As the stations appear to be randomly distributed within this, there is no obvious spatial dimension to performance. This is in contrast to summer, for which the scatter plots suggest the existence of a strong spatial component, illustrated by the structured spread of points across each plot (Figures 13(c) and 14(c)). It is clear that the lowest values were returned for those stations located in southerly coastal areas (Cork, Rosslare). Those synoptic stations situated further inland are generally clustered around the centre of each plot. The most favourable results were returned by stations located in more western and northerly coastal locations (Malin Head, Belmullet and Valentia). The disparity in performance across stations is reflected in the more variable EV values returned for this season (Figure 5). The classifications generally performed better for winter as opposed to spring or autumn, both of which exhibited no clear spatial pattern in the strength of the synoptic–climatological relationship. However, for spring both Malin Head and Rosslare sit outside the general grouping, as does Dublin and Casement Aerodrome, whose position on the plots reflects their geographical proximity (Figures 13(b) and 14(b)).

Table 4. Circulation type classifications ranked (lowest to highest) according to their performance (best to worst) when applied to discretize daily precipitation (mm) for the winter season (DJF)

	Group 1 (8–10 types)							Group 2 (17–19 types)							Group 1 (26–28 types)						
	EV	PF	WSD	KS I	KS II	SIL	SD-OCC	EV	PF	WSD	KS I	KS II	SIL	SD-OCC	EV	PF	WSD	KS I	KS II	SIL	SD-OCC
GWT	**15.0**	**14.0**	14.9	9.3	9.2	10.9	12.1	6.8	6.9	7.6	4.0	9.3	12.0	2.9	3.3	3.1	4.3	3.0	7.9	10.3	3.6
JCT	11.6	11.6	10.6	12.2	13.8	9.9	10.0	1.1	1.1	1.0	6.0	8.7	10.9	4.5	5.8	5.8	5.2	3.8	10.0	11.0	5.1
LWTo	**1.0**	1.1	**1.0**	1.6	5.8	12.1	**1.0**	1.9	1.9	**2.0**	1.8	6.1	10.2	2.1	**2.0**	**2.0**	**2.0**	1.8	4.5	8.6	1.7
LIT	12.1	12.1	11.9	13.7	9.2	**1.0**	**16.0**	**9.0**	**9.0**	**9.0**	8.3	5.3	**1.0**	**11.0**	8.8	8.9	**8.9**	5.6	5.1	2.3	11.0
KRZ	12.1	12.1	12.1	7.1	9.1	7.8	14.1	4.3	4.0	3.6	4.9	6.6	4.9	**7.0**	**1.0**	**1.0**	**1.0**	7.6	7.0	6.1	7.0
PXE	7.1	7.1	7.3	9.6	7.1	9.3	11.7	**10.0**	**10.0**	**10.0**	7.7	8.7	7.0	**13.0**	–	–	–	–	–	–	
PCT	**6.0**	**6.0**	**6.0**	7.6	13.7	4.5	8.7	12.1	12.7	12.3	13.2	13.5	**14.0**	4.7	11.0	10.9	10.9	9.0	9.9	1.3	1.5
PTT	**16.0**	**16.0**	**16.0**	15.9	15.7	15.1	8.6	–	–	–	–	–	–		–	–	–	–	–	–	
LND	**9.0**	**9.0**	**9.0**	9.5	11.0	13.7	6.3	11.5	12.2	11.2	9.1	13.5	**13.0**	**1.0**	**10.0**	10.1	10.1	11.2	10.5	8.9	3.2
KIR	**14.0**	**15.0**	14.1	7.6	11.1	6.8	14.7	**14.0**	**14.0**	**14.0**	12.7	**12.0**	8.1	**14.0**	**12.0**	**12.0**	**12.0**	11.7	11.6	7.4	**12.0**
ERP	10.3	10.2	11.4	15.0	14.8	15.7	12.1	–	–	–	–	–	–		–	–	–	–	–	–	
CKM	2.9	2.9	2.9	6.8	2.3	**2.0**	5.3	4.3	4.4	4.6	1.4	**1.0**	7.4	5.8	5.0	5.1	5.2	1.7	**1.0**	8.9	5.9
SOM	**5.0**	**5.0**	**4.0**	2.8	2.5	4.9	3.1	7.2	7.2	7.0	9.2	3.4	4.0	9.6	7.5	7.5	7.5	7.2	2.7	5.3	9.0
CAP	2.1	2.0	2.1	7.1	3.9	5.0	3.8	4.0	4.1	4.1	4.7	2.7	**2.0**	8.9	4.3	4.4	3.6	8.2	5.1	3.4	8.5
PXK	7.9	7.9	7.7	8.5	5.5	13.3	6.3	12.5	11.2	12.5	13.1	**11.0**	7.6	**12.0**	–	–	–	–	–	–	
SAN	**4.0**	**4.0**	**5.0**	1.8	1.5	3.9	2.1	6.3	6.3	6.1	8.9	3.1	3.2	8.5	7.3	7.3	7.3	7.2	2.6	4.4	9.5

The average ranking for each classification, estimated using the results from 1000 bootstrapped samples taken from each station record is shown. The individual rankings are considered statistically significant (shown in bold) if they remain constant across 95% of samples. Classifications are grouped according to the number of circulation types (~ 9, 18 and 27), and are shaded based on their principal typing method (pre-defined threshold, eigenvector based, leader algorithm, optimization procedure).

With respect to their proficiency in resolving local precipitation (Figures 15 and 16), a clear spatial dimension underlies the performance of both classifications. In this case a similar grouping of stations across the plot for each season is evident. Generally, those stations located on the western and southern Atlantic seaboard (Cork, Valentia and Belmullet) returned the highest values. In contrast those stations situated further inland – variously including Clones, Mullingar and Kilkenny – are generally clustered around the centre. The relationship was generally shown to be weakest for Dublin, Casement and Armagh respectively, signified by their location in the south-west quadrant of each plot. These patterns capture the existence of coastal and inland type regional climates across the island; they also reflect the dominant airflow trajectory and the west-to-east gradient in precipitation receipts. This gradient is demonstrated by the average annual (1961–1990 average) rainfall yield for Valentia (1430 mm year^{-1}; situated on the south western seaboard) and Dublin Airport (732 mm year^{-1}; located on the east coast) respectively. The same spatiotemporal patterns evident in Figures 15 and 16 were repeated for other series in the catalogue.

5. Discussion and conclusions

This article presented an investigation into the applicability of different methods in circulation typing to synoptic–climatological analysis of the Irish landmass. To this end CTCs compiled under the COST733 Action were assessed based on their ability to discretize daily temperature (maximum and minimum) and precipitation records (1957–2002) from 14 meteorological stations. Different aspects of the classifications were examined including: their sensitivity to the number of CTs; the performance of general methodological approaches as well as individual typing schemes; and the extent to which performance is spatiotemporally dependent. By way of the individual performance measures employed, different approaches to the quantification of key efficiency criteria (i.e. the between-group separability and within-group similarity) were adopted (e.g. distance measure, statistical parameter, homogeneity of distribution functions).

Commensurate with previous studies (Beck and Philipp, 2010; Huth, 2010) the sensitivity analysis indicated that a clear dependency between each evaluation measure and the number of CTs existed. This illustrates the need to remove the number of types as a confounding factor when comparing classifications. The findings generally indicate that an expansion in the class number leads to a reduction in the within-group similarity (EV, WSD) – suggesting greater explanatory power; however, increasing the number of types also has the effect of reducing the association between individual patterns and a set of uniquely defined surface conditions (KS statistic). These generalizations however cannot be assumed to hold for each individual scheme, nor apply equally to all climate or environmental variables. As shown by the results there was a notable improvement in the ranking of certain classifications (relative to others) when the number of types was increased from 9 to 18/27. In this case, by allowing for the inclusion of synoptic patterns which exert an important level of control, or have physically meaningful relationship with the target variable, increasing the number of types is critical for improving performance. In relation to this, the results must be interpreted with respect to the restrictions which deriving or identifying a set number of types may place on individual schemes. It is important to note that, in addition to a host of other parameters (e.g. spatial domain, input data, period of analysis, and target variable), the optimal

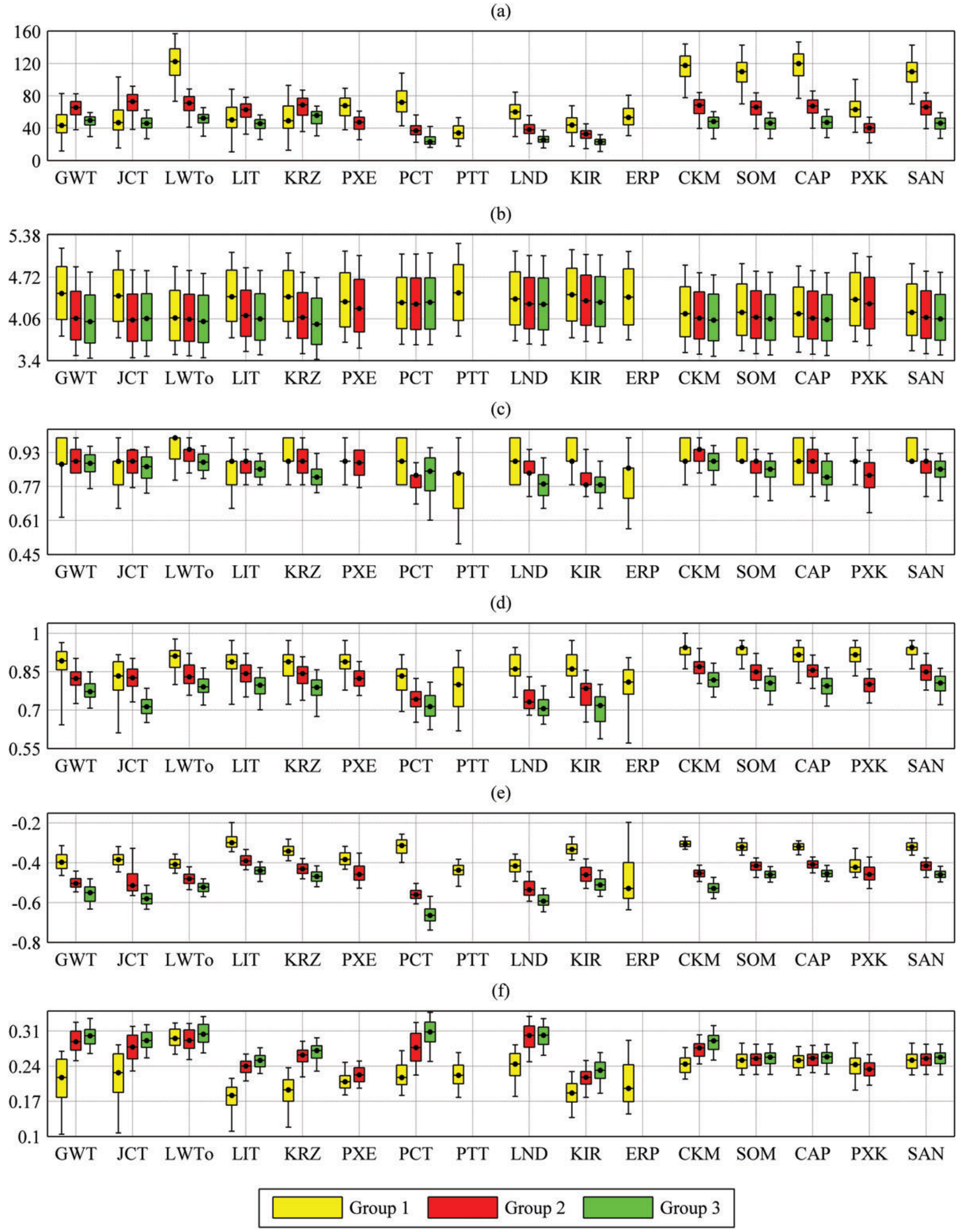

Figure 12. Classification performance for daily winter (DJF) precipitation – assessed using six evaluation measures: PF (a), WSD (b; mm day^{-1}, *y*-axis), KS type I (c), KS type II (d), SIL (e), and SD-Occ (f). The plots are constructed using the results from 1000 bootstrapped samples taken from each station record. On each plot the central mark denotes the median, whilst the edges represent the 25th and 75th percentiles respectively; the whiskers denote the 90% interval. The groupings signify the number of circulation types (Group 1: 8–10 CTs; Group 2: 17–19 CTs; Group 3: 26–28 CTs).

number of classes will always be specific to the particular typing method used. The adherence of the COST733 catalogue to a set number of types is a caveat which must be considered more generally when interpreting the study findings. In addition, as the classifications used were not independently derived for each season, but rather compiled using the annual series, this may have limited the explanatory power of some typing schemes. Further research will examine the potential added benefit of using seasonally independently series.

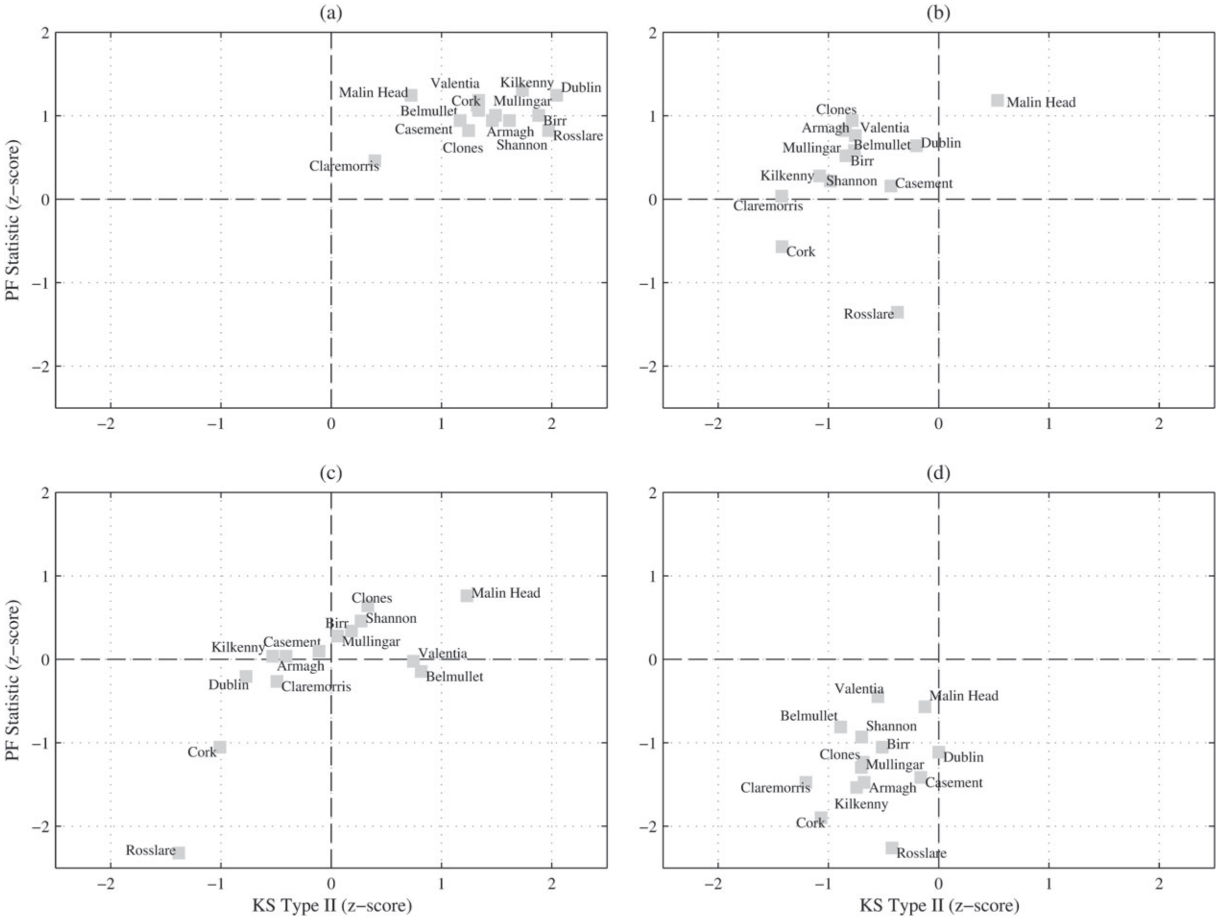

Figure 13. Seasonal scatter plots – DJF (a), MAM (b) JJA (c), SON (d) – of the standardized z-scores returned for the LIT classification (18 CTs) when applied to records of daily maximum temperature from 14 synoptic stations. The values for individual stations are standardized using the results for all stations and each season.

In addressing uncertainty the study underlines the necessity of considering multiple performance criteria, in addition it highlights the importance of adopting measures which evaluate the same criterion using different quantitative mechanisms. With respect to the agreement shown between measures, the EV, WSD, PF and both KS statistics generally returned similar results. In contrast the SIL index was often found to be in complete disagreement with each of the other measures used. In addition exclusively negative values were returned for the index, indicating that in all cases the data was inappropriately partitioned. The relationship between large-scale circulation and local climate is subject to a high degree of noise, which given how the index quantifies inter-group dissimilarity, heightens its sensitivity to this criterion. As a result classifications which cannot arrange observations into well defined clusters are unduly rewarded. Its inability to deal with the complexities of the data, coupled with its lack of agreement with other measures indicates that the index may be inappropriate for use in the current context; however, in cases where the relationship is subject to less noise - for example with gridded observational or reanalysis data - it may be more robust.

In keeping with previous studies it was not possible to identify a single optimal classification or approach to circulation typing (Beck and Philipp, 2010; Huth, 1996, 2010). However the performance assessment indicated that those typing schemes based on predefined thresholds (Litynski, GrossWetterTypes, and original Lamb Weather Type) – along with the Kruizinga and Lund classifications – generally were most proficient in resolving surface temperature. These methods were the most consistent across performance measures and for groups possessing a greater/lesser number of CTs. The PXE and PXK classifications – both of which are derived using PCA extreme scores – were consistently among the worst performing schemes for this variable. The classifications exhibited a greater ability to resolve maximum as opposed to minimum daily temperature, an outcome which may be related to a temporal offset in the occurrence of daily minima with respect to the time for which the classifications were compiled. With the notable exception of the PXK scheme, the optimization type classifications (e.g. simulated annealing, Self-Organizing Maps, k-Means clustering) generally performed best for precipitation; they were also found to be the most consistent across sub-groups and evaluation criteria. Those classifications derived using the leader algorithm approach generally performed poorly for this variable; as did both S-mode PCA schemes (PCT and PTT).

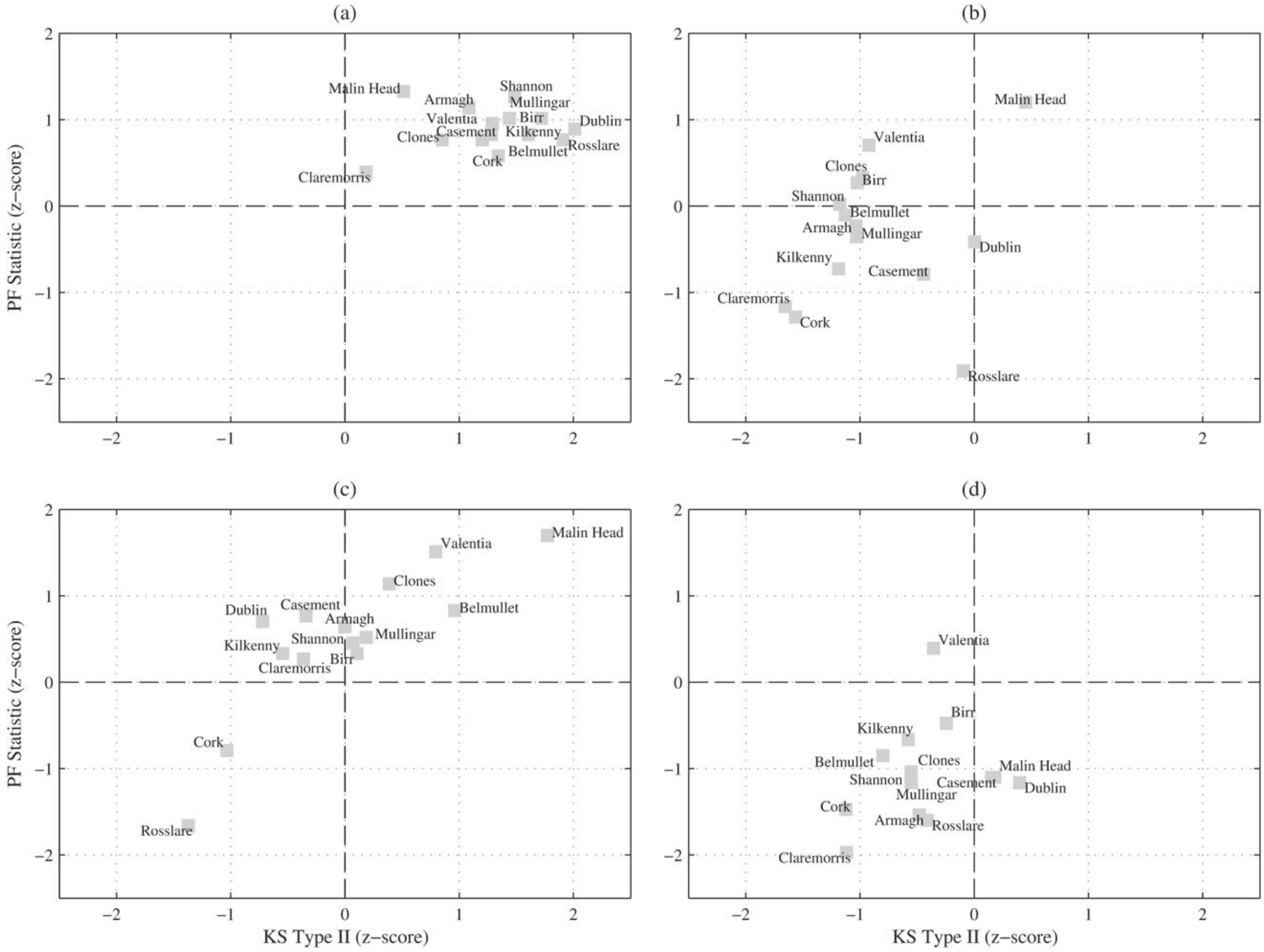

Figure 14. Seasonal scatter plots – DJF (a), MAM (b) JJA (c), SON (d) – of the standardized z-scores returned for the KIR classification (18 CTs) when applied to records of daily maximum temperature from 14 synoptic stations. The values for individual stations are standardized using the results for all stations and each season.

The strong performance of the threshold based classifications (incorporating 8–10 CTs) for temperature is due to their explicit inclusion of eight purely directional types (N, NW, W, SW, S, SE, E and NE), each of which can be related to a particular trajectory of airflow across the island bringing a distinct set of surface conditions (Hulme and Barrow, 1997). This is in contrast to the relatively poor performance of the optimization based schemes (possessing approximately 9 CTs), for which there is no prerequisite to identify or discretize daily patterns based on these directional types. Conversely, for precipitation it may be the adherence to these patterns which resulted in the relatively poor performance of the predefined type schemes (possessing approximately 9 CTs). The importance of vorticity in controlling the local response in this variable means it is essential that the identified synoptic patterns explicitly reflect this aspect of circulation behaviour (Buishand and Brandsma, 1997). This was illustrated by the relative improvement in the threshold based schemes when they were expanded (from 9 to 18/27 CTs) to include the 'hybrid' type patterns – which reflect both airflow direction and vorticity. In relation to this, the superior performance of the LWTo scheme with respect to other threshold based methods, (incorporating 8–10 CTs) can be attributed to its inclusion of hybrid as well as purely cyclonic/anticyclonic patterns.

Notable in the results for both variables was the performance of the correlation based classifications. The Kirchhofer scheme, which can be considered an advancement of the methodological blueprint set out by Lund (1963), was consistently outperformed by its predecessor (El-Kadi and Smithson, 1992). In addition the performance of the Kruizinga scheme was generally shown to be at odds with other eigenvector based methods; however, as its methodology is more closely related to that of the predefined types, it may be more appropriate to consider in this grouping (Beck and Philipp, 2010). In assessing performance it is important to highlight the relative infrequency (approximately 1% occurrence) with which a number of types occurred in the time series of several classifications (e.g. PTT, ERP and PCT). Despite their rarity, the temperature conditions associated with these patterns were generally indistinguishable from those associated with other types. This is in contrast to precipitation for which the infrequent patterns in some cases (particularly the ERP classification) were associated with high/low receipts. This suggests that classifications whose general descriptive ability is relatively poor, may be more appropriate for

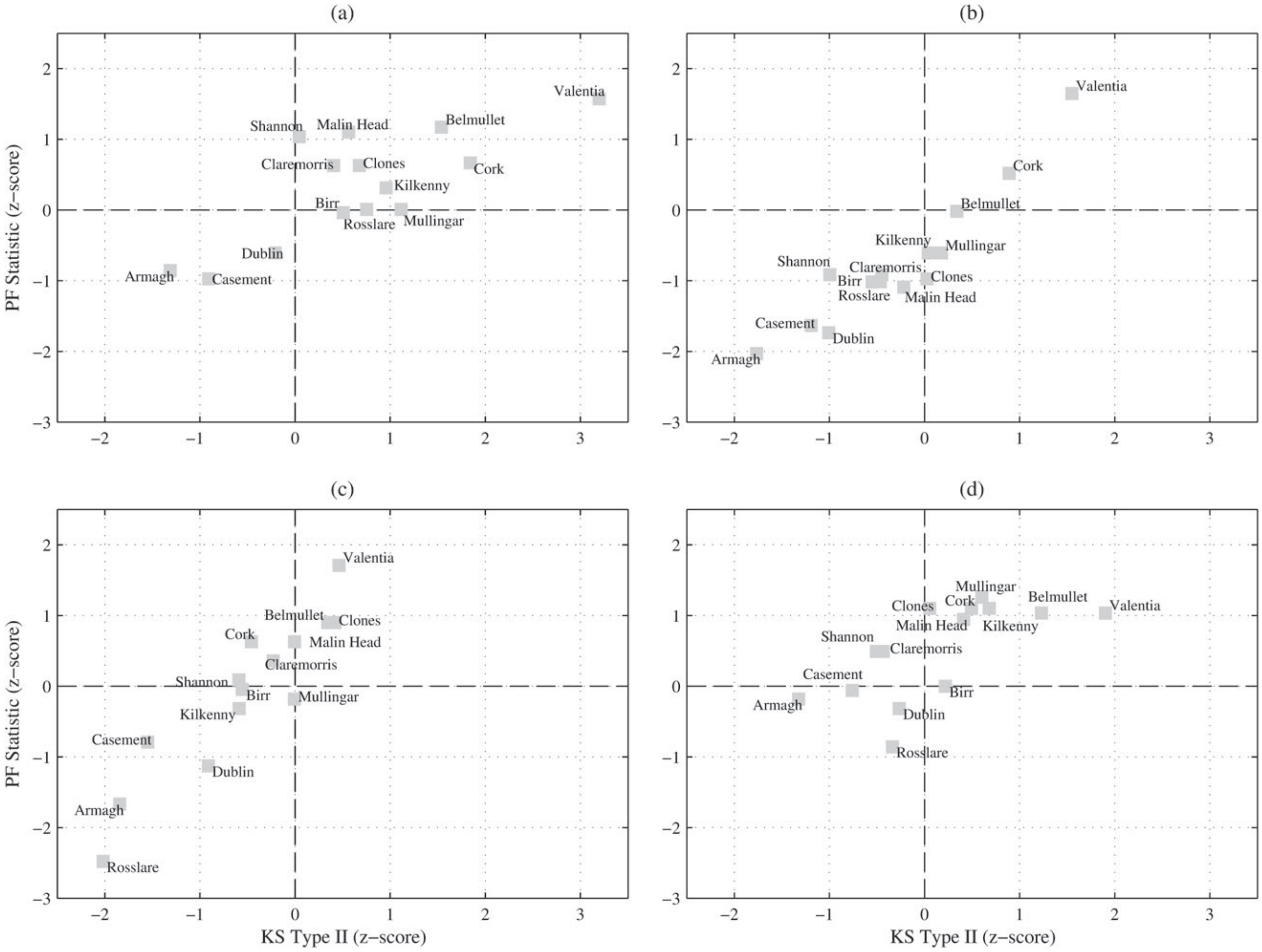

Figure 15. Seasonal scatter plots – DJF (a), MAM (b) JJA (c), SON (d) – of the standardized *z*-scores returned for the JCT (18 CTs) classification when applied to records of daily precipitation from 14 synoptic stations. The values for individual stations are standardized using the results for all stations and each season.

capturing extreme events; this however would require further investigation.

Two classification series each with 18 CTs were selected to explore the existence of spatiotemporal patterns in the relationship between large-scale circulation and local climate. This connection was found to be strongest during winter, reflecting the strengthening of zonal airflow and the occurrence of more well defined flow patterns during this season. With respect to precipitation this also corresponds to the period when frontal depressions are at their most frequent and intense. For winter the best performing classifications captured ~40% of the variance in daily temperature (maximum), and only ~26% of the variance in precipitation. With the exception of summer, spatial patterns in performance for temperature were not evident. This is in contrast to precipitation, for which a clear spatial component in the relationship was apparent across all seasons. Generally those more westerly and southerly coastal stations, whose zonal airflow is less perturbed by land surface processes, exhibited a closer connection with large-scale circulation. These patterns mirror the presence of coastal and inland type regional climates, but also reflect the west-to-east gradient in precipitation receipts and the path of westerly airflow across the island. In general the patterns found were common to each of the classification series archived in the catalogue; thus, variations in the strength of the relationship are not specific to a particular typing scheme, but are rather due to alterations in the level of control exerted by large-scale circulation, and the affect which unresolved sub-synoptic scale processes (e.g. convection, orographic uplift, smaller scale circulation components) have on the local response.

The analysis conducted here used records from a small number of synoptic stations which, although being well separated geographically, may fail to fully capture the relationship between large-scale variability and local climate. To examine more closely the explanatory power of classifications, further studies may consider a larger set of climatological stations. Additional research may also focus on a sub-set of classifications shown to be proficient in characterizing surface climate for the study area. Determining exactly why these particular schemes outperformed others would aid in the further development of typing schemes for this region. Finally, given the relative size of the Irish landmass in relation to the D04 domain (Figure 1), it may be beneficial to investigate the affect which altering the

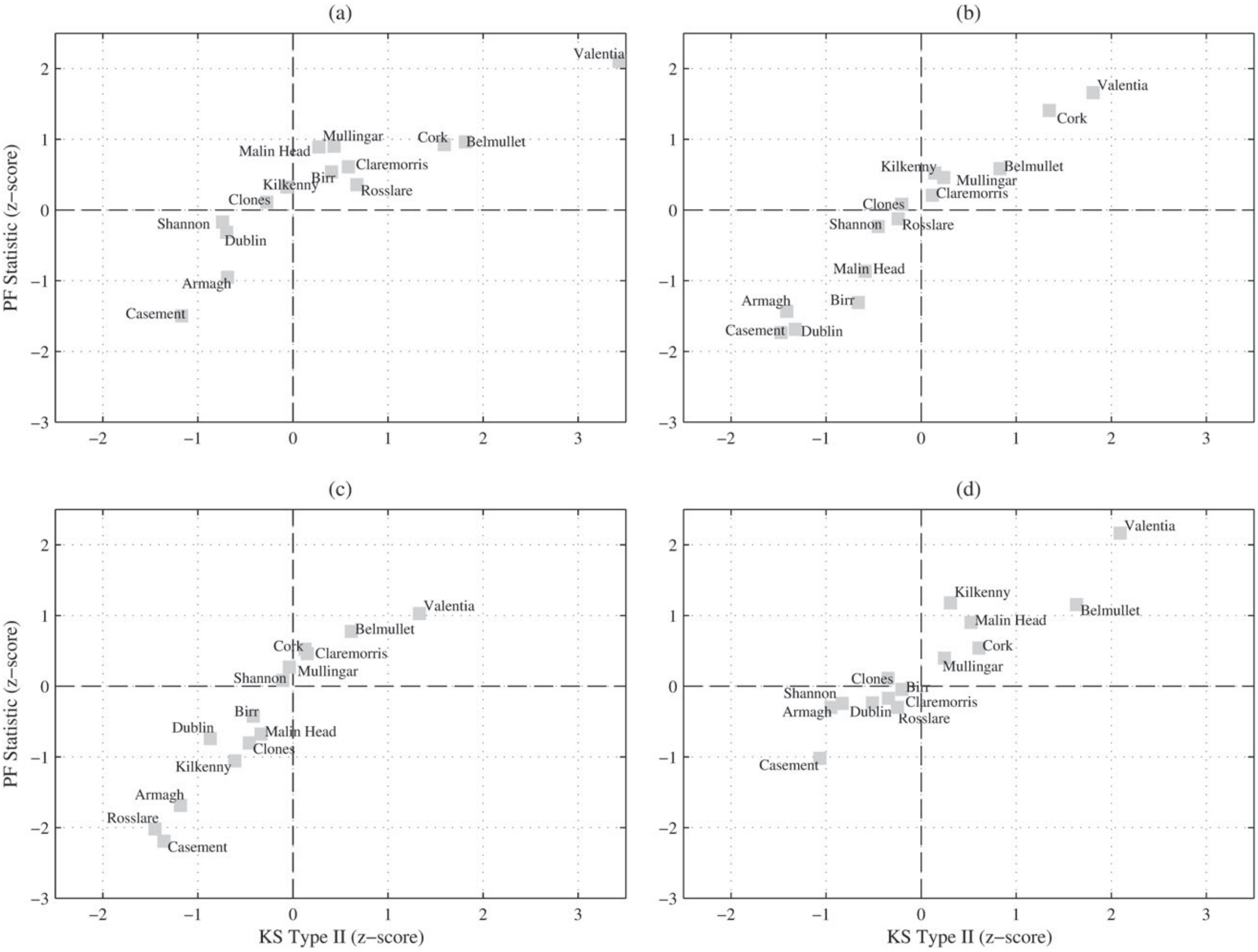

Figure 16. Seasonal scatter plots – DJF (a), MAM (b) JJA (c), SON (d) – of the standardized *z*-scores returned for the KIR (18 CTs) classification when applied to records of daily precipitation from 14 synoptic stations. The values for individual stations are standardized using the results for all stations and each season.

size and location of the domain may have on performance (Beck *et al.*, 2013).

Acknowledgements

We thank the authors of the COST733 catalogue and the developers of the supporting classification software; we also thank Met Éireann and the Armagh observatory for providing the synoptic records used. The authors are grateful to the two anonymous referees for their insightful and helpful comments; a special thanks also to Ole Einar Tveito for his advice during the initial stages of the study.

References

Bardossy A, Duckstein L, Bogardi I. 1995. Fuzzy rule-based classification of atmospheric circulation patterns. *Int. J. Climatol.* **15**: 1087–1097.

Barry RG, Perry AH. 1973. *Synoptic Climatology*. Methuen: London; 535.

Baur F. 1948. *Einführung in die Großwetterkunde* (Introduction to Large Scale Weather). Dieterich Verlag: Wiesbaden, Germany (in German).

Beck C, Philipp A. 2010. Evaluation and comparison of circulation type classifications for the European domain. *Phys. Chem. Earth* **35**: 374–387.

Beck C, Jacobeit J, Jones PD. 2007. Frequency and within-group variations of large scale circulation types and their effects on low-frequency climate variability in Central Europe since 1780. *Int. J. Climatol.* **27**: 473–491.

Beck C, Philipp A, Streicher F. 2013. The effect of domain size on the relationship between circulation type classifications and surface climate. *Int. J. Climatol.*, DOI: 10.1002/joc.3688.

Blair D. 1998. The Kirchhofer technique of synoptic typing revisited. *Int. J. Climatol.* **18**: 1625–1635.

Buishand TA, Brandsma T. 1997. Comparison of circulation classification schemes for predicting temperature and precipitation in the Netherlands. *Int. J. Climatol.* **26**: 461–483.

Calinski T, Harabsz J. 1974. A dendrite method for cluster analysis. *Commun. Stat.* **3**: 1–27.

Casado MJ, Pastor MA, Doblas-Reyes FJ. 2010. Links between circulation types and precipitation over Spain. *Phys. Chem. Earth* **35**: 437–447.

Comrie AC, Yarnal B. 1992. Relationships between synoptic-scale atmospheric circulation and ozone concentrations in metropolitan Pittsburgh, Pennsylvania. *Atmos. Environ.* **26**: 301–312.

Dayan U, Tubi A, Levy I. 2012. On the importance of synoptic classification methods with respect to environmental phenomena. *Int. J. Climatol.* **32**: 681–694.

Demuzere M, Trigo RM, Vila-Guerau de Arellano J, van Lipzig NPM. 2009. The impact of weather and atmospheric circulation on O_3 and PM_{10} levels at a midlatitude site. *Atmos. Chem. Phys.* **9**: 2695–2714.

El-Kadi AKA, Smithson PA. 1992. Atmospheric classification and synoptic climatology. *Prog. Phys. Geogr.* **35**: 96–102.

Enke W, Spekat A. 1997. Downscaling climate model outputs into local and regional weather elements by classification and regression. *Clim. Res.* **8**: 195–207.

Erpicum M, Mabille G, Fettweis X. 2008. Automatic synoptic weather circulation types classification based on the 850 hPa geopotential height. In *Book of Abstracts COST 733 Mid-term Conference "Advances in Weather and Circulation Type Classifications & Applications"*. Krakow, Poland; 22–25 October 2008.

Esteban P, Jones PD, Martín-Vide J, Mases M. 2005. Atmospheric circulation patterns related to heavy snowfall days in Andorra, Pyrenees. *Int. J. Climatol.* **25**: 319–329.

Esteban P, Martin-Více J, Mases M. 2006. Daily atmospheric circulation catalogue for Western Europe using multivariate techniques. *Int. J. Climatol.* **26**: 1501–1515.

Esteban P, Ninyerola M, Prohom M. 2009. Spatial modelling of air temperature and precipitation for Andorra (Pyrenees) from daily circulation patterns. *Theor. Appl. Climatol.* **96**: 43–56.

Fleig AK, Tallaksen LM, Hisdal H, Stahl K, Hannah DM. 2010. Inter-comparison of weather and circulation type classifications for hydrological drought development. *Phys. Chem. Earth* **35**: 507–515.

Gerstengarbe FW, Werner PC. 1997. A method to estimate the statistical confidence of cluster separation. *Theor. Appl. Climatol.* **57**: 103–110.

Hulme M, Barrow E. 1997. *Climates of the British Isles present, past and future*. Routledge: London; 454.

Huth R. 1993. An example of using obliquely rotated principal components to detect circulation types over Europe. *Meteorol. Z.* **2**: 285–293.

Huth R. 1996. An intercomparison of computer-assisted circulation classification methods. *Int. J. Climatol.* **16**: 893–922.

Huth R. 1997. Continental-scale circulation in the UKHI GCM. *J. Clim.* **10**: 1545–1561.

Huth R. 2000. A circulation classification scheme applicable in GCM studies. *Theor. Appl. Climatol.* **67**: 1–18.

Huth R. 2010. Synoptic-climatological applicability of circulation classifications from the COST733 collection: first results. *Phys. Chem. Earth* **35**: 388–394.

Huth R, Beck C, Philipp A, Demuzere M, Untstrul Z, Cahynova M, Kysely J, Tveito OE. 2008. Classifications of atmospheric circulation patterns: recent advances and applications. *Ann. N. Y. Acad. Sci.* **1146**: 105–152.

Jacobeit J, Wanner H, Luterbacher J, Beck C, Philipp A, Sturm K. 2003. Atmospheric circulation variability in the North-Atlantic-European Area since the mid-seventeenth century. *Clim. Dyn.* **20**: 341–352.

James PM. 2006. Second generation lamb weather types – a new generic classification method with evenly tempered type frequencies. In: *6th Annual Meeting of the EMS/6th ECAC, EMS2006A00441*. Ljubljana, Slovenia.

Jenkinson AF, Collison BP. 1977. *An initial climatology of gales over the North Sea*. Synoptic Climatology Branch Memorandum No. 62. Meteorological Office, London, 18.

Jones GV, Davis RE. 2000. Using a synoptic climatological approach to understand climate viticulture relationships. *Int. J. Climatol.* **20**: 813–837.

Jones PD, Hulme M, Briffa KR. 1993. A comparison of Lamb circulation types with an objective classification scheme. *Int. J. Climatol.* **13**: 655–663.

Kalkstein LS, Tan G, Skindlov JA. 1987. An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Appl. Meteorol.* **26**: 717–730.

Kassomenos P. 2009. Synoptic circulation control on wildfire occurrence. *Phys. Chem. Earth* **35**: 544–552.

Kassomenos P, Gryparis A, Samoli E, Katsouyanni K, Lykoudis S, Flocas HA. 2001. Atmospheric circulation types and daily mortality in Athens, Greece. *Environ. Health Perspect.* **109**: 591–596.

Kidson JW. 1994a. An automated procedure for the identification of synoptic types applied to the New Zealand region. *Int. J. Climatol.* **14**: 711–721.

Kidson JW. 1994b. Relationship of New Zealand daily and monthly weather patterns to synoptic weather types. *Int. J. Climatol.* **14**: 723–737.

Kirchhofer W. 1973. *Classifications of European 500 mb Patterns. Arbeitsbericht der Schweizerischen Meteorologischen Zentralanstalt*, vol. 43. Schweizerische Meteorologische Zentralanstalt: Zürich, Switzerland.

Kruizinga S. 1979. Objective classification of daily 500 mbar patterns. In *Preprints Sixth Conference on Probability and Statistics in Atmospheric Sciences, Banff, Alberta*. American Meteorological Society, Boston, MA; 126–129.

Kysely J, Domonkos P. 2006. Recent increase in persistence of atmospheric circulation over Europe: comparison with long-term variations since 1881. *Int. J. Climatol.* **26**: 461–483.

Kysely J, Huth R. 2006. Changes in atmospheric circulation over Europe detected by objective and subjective methods. *Theor. Appl. Climatol.* **85**: 19–36.

Lamb HH. 1950. Types and spells of weather around the year in the British Isles: annual trends, seasonal structure of years, singularities. *Q. J. R. Meteorol. Soc.* **76**: 393–438.

Lauscher F. 1985. *Klimatologische Synoptik Österreichs mittels der ostalpinen Wetterlagenklassifikation (Synoptic Climatology of Austria based on the Eastern-Alpine Weather Type Classification)*. Arbeiten aus der Zentralanstalt für Meteorologie und Geodynamik, Publikation Nr. 302, Heft 64, Austria (in German).

Leśniok M, Łukasz M, Niedźwiedź T. 2010. Classification of circulation types for Southern Poland with an application to air pollution concentration in Upper Silesia. *Phys. Chem. Earth* **35**: 516–522.

Linderson ML. 2001. Objective classification of atmospheric circulation over southern Scandinavia. *Int. J. Climatol.* **21**: 155–169.

Litynski J. 1969. A numerical classification of circulation patterns and weather types in Poland. *Prace Panstwowego Instytutu Hydrologiczno-Meteorologicznego Inst. Hydrol. Meteorol.* **97**: 3–15 (in Polish).

Lund IA. 1963. Map-pattern classification by statistical methods. *J. Appl. Meteorol.* **2**: 56–65.

Lupikasza E. 2010. Relationships between occurrence of high precipitation and atmospheric circulation in Poland using different classifications of circulation types. *Phys. Chem. Earth* **35**: 448–455.

McGregor GR, Bamzelis D. 1995. Synoptic typing and its application to the investigation of weather air pollution relationships, Birmingham, United Kingdom. *Theor. Appl. Climatol.* **51**: 223–236.

Michaelides SC, Liassidou F, Schizas CN. 2007. Synoptic classification and establishment of analogues with artificial neural networks. *Pure Appl. Geophys.* **164**: 1347–1364.

Mihalakakou G, Flocas HA, Santamouris M, Helmis CG. 2002. Application of neural networks to the simulation of the heat island over Athens, Greece, using synoptic types as a predictor. *J. Appl. Meteorol.* **41**: 519–527.

Milligan GW, Cooper M. 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**: 159–179.

Mooney PA, Mulligan FJ, Fealy R. 2011. Comparison of ERA-40, ERA-Interim and NCEP/NCAR reanalysis data with observed surface air temperatures over Ireland. *Int. J. Climatol.* **31**: 545–557.

Pattison I, Lane SN. 2011. The relationship between Lamb weather types and long-term changes in flood frequency, River Eden, UK. *Int. J. Climatol.* **32**: 1971–1989.

Philipp A. 2009. Comparison of principal component and cluster analysis for classifying circulation pattern sequences for the European domain. *Theor. Appl. Climatol.* **96**: 31–41.

Philipp A, Della-Marta PM, Jacobeit J, Fereday DR, Jones PD, Moberg A, Wanner H. 2007. Long term variability of daily North Atlantic-European pressure patterns since 1850 classified by simulated annealing clustering. *J. Clim.* **20**: 4065–4095.

Philipp A, Bartholy J, Beck C, Erpicum M, Esteban P, Huth R, James P, Jourdain S, Krennert T, Lykoudis S, Michaelides S, Pianko K, Post P, Rasilla Álvarez D, Spekat A, Tymvios FS. 2010. COST733CAT – a database of weather and circulation type classifications. *Phys. Chem. Earth* **35**: 360–373.

Prudhomme C, Genevier M. 2011. Can atmospheric circulation be linked to flooding in Europe? *Hydrol. Process.* **25**: 1180–1190.

Richman MB. 1986. Rotation of principal components. *Int. J. Climatol.* **6**: 293–335.

Romero R, Sumner G, Ramis C, Genovés A. 1999. A classification of the atmospheric circulation patterns producing significant daily rainfall in the Spanish Mediterranean area. *Int. J. Climatol.* **19**: 765–785.

Rousseeuw P. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**: 53–65.

Samaniego L, Bárdossy A. 2007. Relating macroclimatic circulation patterns with characteristics of floods and droughts at the mesoscale. *J. Hydrol.* **335**: 109–123.

Schiemann R, Frei C. 2010. How to quantify the resolution of surface climate by circulation types: an example for Alpine precipitation. *Phys. Chem. Earth* **35**: 403–410.

Schüepp M. 1957. Klassifikationsschema, Beispiele und Probleme der Alpenwetterstatistik (Classification scheme, examples and problems of the Alpine weather statistics). *La Meteorologie* **4**: 291–299 (in German).

Schüepp M. 1979. Witterungsklimatologie – Klimatologie der Schweiz III (Weather Climatology – Climatology of Switzerland III). Beihefte zu den Annalen der Schweizerischen Meteorologischen Anstalt 1978; 93 (in German).

Stahl K, Siegfreid D. 1999. Linking streamflow drought to the occurrence of atmospheric circulation patterns / Etude sur les relations entre les étiages et les types de circulation atmosphérique. *Hydrol. Sci. J.* **44**: 467–482.

Stehlik J, Bardossy A. 2003. Statistical comparison of European circulation patterns and development of a continental scale classification. *Theor. Appl. Climatol.* **76**: 31–46.

Sumner G, Guijarro JA, Ramis C. 1995. The impact of surface circulations on the daily rainfall over Mallorca. *Int. J. Climatol.* **15**: 673–696.

Sweeney J, O'Hare G. 1992. Geographical variations in precipitation yields and circulation types in Britain and Ireland. *Trans. Inst. Br. Geogr.* **17**: 448–463.

Tveito OE. 2010. An assessment of circulation type classifications for precipitation distribution in Norway. *Phys. Chem. Earth* **35**: 395–402.

Unger J. 1996. Heat island intensity with different meteorological conditions in a medium-sized town: Szeged, Hungary. *Theor. Appl. Climatol.* **54**: 147–151.

Uppala SM, Kållberg PW, Simmmons AJ, Andrea U, da Costa BV, Fiorino M et al. 2005. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**: 2961–3012.

Ustrnul Z. 2006. Spatial differentiation of air temperature in Poland using circulation types and GIS. *Int. J. Climatol.* **26**: 1529–1546.

Ustrnul Z, Czekierda D, Wypych A. 2010. Extreme values of air temperature in Poland according to different atmospheric circulation classifications. *Phys. Chem. Earth* **35**: 403–410.

Vicente-Serrano S, López-Moreno J. 2006. The influence of atmospheric circulation at different spatial scales on winter drought variability through a semi-arid climatic gradient in Northeast Spain. *Int. J. Climatol.* **26**: 1427–1453.

Werner PC, Gerstengarbe FW, Fraedrich K, Oesterle H. 2000. Recent climate change in the North Atlantic/ European sector. *Int. J. Climatol.* **20**: 463–471.

Wilby RL. 2003. Past and projected trends in London's urban heat island. *Weather* **58**: 251–260.

Yarnal B. 1993. *Synoptic Climatology in Environmental Analysis*. Belhaven Press: London.

Yarnal B, Comrie AC, Frakes B, Brown DP. 2001. Review: developments and prospects in synoptic climatology. *Int. J. Climatol.* **21**: 1923–1950.