A climate classification for corrosion control in electronic system design

Spooner, Max; Ambat, Rajan; Gudla, Helene Conseil; Kulahci, Murat

Published in:
Machine Learning with Applications

Link to article, DOI:
10.1016/j.mlwa.2022.100397

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Machine Learning with Applications 9 (2022) 100397

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

A climate classification for corrosion control in electronic system design

Max Spooner [a,*], Rajan Ambat [b], Hélène Conseil-Gudla [b], Murat Kulahci [a,c]

[a] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark
[b] Department of Mechanical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark
[c] Department of Business Administration, Technology and Social Sciences, Luleå, University of Technology, Sweden

ARTICLE INFO

Keywords:
Climate classification
Corrosion
Electronics
K-means
Clustering
Printed circuit board enclosure
Reliability

ABSTRACT

Climate factors such as humidity and temperature have a significant impact on the corrosion reliability of electronic products. Given the huge geographical variability in climate conditions globally, a climate classification is a useful tool that simplifies the problem of considering climate when designing electronics packaging. Most current guidelines for electronic product design rely on the Köppen–Geiger classification first developed by Köppen over a century ago. Köppen devised a set of heuristics to separate climates to match different vegetation types. These climate classes are unlikely to be the optimal for electronic product design. This paper presents a new climate classification using parameters important for corrosion reliability of electronics. The classification is based on real climate data measured every 3 h during a 5-year period at over 9000 locations globally. A key step is defining relevant features of climate affecting corrosion in electronics. Features related to temperature are defined, but also the amount of time that the difference between Temperature and Dew Point is less than 1, 2 or 3 °C. These features relate to the risk of condensation in electronic products. The features are defined such that diurnal, seasonal and yearly variation is taken into account. The locations are then clustered using *K*-means clustering to obtain the relevant climate classes. This data-driven classification, based on key features for corrosion reliability of electronics, will be a useful aid for product design, reliability testing and lifetime estimation.

1. Introduction

Humidity can influence the performance of electronics by encouraging corrosion to occur on the surface of the Printed Circuit Board Assembly (PCBA) or on other components (Ambat et al., 2018; Wang et al., 2013). Variation in the local climate at the place of use, such as diurnal and seasonal changes in humidity and temperature, can allow condensation to form on sensitive electronics components. Electronic packaging regulates the entry of humidity into the device. However, the level of moisture entry depends on the design (Conseil-Gudla et al., 2018; Jacobsen et al., 2014; Tencer & Moss, 2002). During the diurnal cycling of humidity and temperature, lagging of temperature equilibration inside the device can result in intermittent condensation, which also depends on the hygroscopic nature of the surface (Piotrowska & Ambat, 2020; Schimpf et al., 2009). Therefore, for better design for climate compatibility, knowledge on relevant local climates and those aspects of the climates that are critical to the electronics are important. As the Earth contains a huge variety of different climates, clustering climates into classes based on the above factors will be very useful for electronic design decision making. In this paper, consideration is towards the corrosion reliability of electronic devices, where temperature and humidity levels and variations, and the potential risk of condensation, are key variables (Ambat et al., 2018; Conseil-Gudla et al., 2018). However, the classification proposed in this paper is also useful in any setting where corrosion must be considered.

1.1. Corrosion in electronics

Electronics devices are prone to degradation, and loss of performance resulting in the need for repair or replacement. Recently, the effect of external climatic conditions such as humidity and temperature on the performance of these devices is taken into consideration and is becoming a key factor in design considerations due to ever-growing demand for miniaturization of electronic devices and products, as well as energy efficiency and sustainability. The service of electronic equipment in extreme climatic conditions is aggravating the degradation processes such as general corrosion and short circuiting due to electrochemical migration (ECM), conductive anodic filament (CAF) formation, etc, leading to premature failure of electronic devices and components (Ambat et al., 2018). In terms of financial cost, the global cost of corrosion across all industries was estimated at 3.4% of the global GDP (NACE, 2016). The cost due to corrosion in electronics alone is not estimated but is a significant contributor to the overall

* Corresponding author.
E-mail addresses: mpsp@dtu.dk (M. Spooner), ram@mek.dtu.dk (R. Ambat), helco@mek.dtu.dk (H. Conseil-Gudla), muku@dtu.dk (M. Kulahci).

https://doi.org/10.1016/j.mlwa.2022.100397
Received 22 June 2022; Received in revised form 20 July 2022; Accepted 31 July 2022
Available online 3 August 2022

figure. Hence there is a great potential for economic savings if corrosion in electronics can be mitigated.

The electrochemical corrosion of electronics occurs in a locally formed aqueous cell on a PCBA due to the co-existence of three factors namely: (i) potential bias, (ii) metals/alloys and (iii) humid surrounding air. Presence of contamination, from the outdoor environment (such as salt or dust) (D'Angelo et al., 2017; Song et al., 2012, 2013) or from the manufacturing process (due to, for example, solder flux residues or handling) (Conseil et al., 2014; Veselý et al., 2020) are hygroscopic in nature and will absorb water at relatively low levels of humidity through the process of deliquescence, and will increase the conductivity of the formed electrolyte. The first point of failure is a parasitic circuit. The current can leak through the solution and can result in leak current or surface insulation resistance (SIR) reduction. It can further increase the corrosion and ECM formation (Piotrowska et al., 2018b; Verdingovas et al., 2015). Better PCBA design (surface finish, pitch distance, solder mask type, etc.) (Bahrebar & Ambat, 2021; Mantis et al., 2021; Piotrowska et al., 2018a), and better selection of solder flux composition (weak organic acid type) lowering the surface contamination level, may lead to better humidity reliability, as well as a better protection (packaging) of the electronics parts.

Currently, the general approach to withstand the external environmental loads is to use conformal coatings or potting on the PCBAs, and to place the assemblies inside an enclosure. The moisture barrier of the coatings and potting are affected by the material type, thickness, conformity, cleanliness of the board and adhesion (Mantis et al., 2021). Nevertheless, the moisture will eventually permeate through the coating and reach the interface with the PCBA (Hunt et al., 2006; Rathinavelu et al., 2012). The electronic enclosure is also a physical barrier towards the aggressive environmental loads; however, the electronic enclosure is never perfectly sealed and is exchanging humidity with the surroundings under all combinations of humidity transport modes. Humidity can diffuse through intended or unintended openings and also through polymeric casing walls (Conseil et al., 2016; Conseil-Gudla et al., 2018; Roman et al., 2020). The design of the devices, such as the interior arrangement of electronic parts, presence of thermal mass, volume of plastic parts, etc. will influence the build-up of humidity inside the devices. Forced and natural convection are also used in electronics for cooling purposes, giving rise to exchange of humid air between the enclosure and ambient. Natural convection between the electronic enclosure and ambient also occurs due to ambient temperature and pressure changes (Jacobsen et al., 2014; Tencer & Moss, 2002).

Relatively little literature exists on how to design for protection and robustness of electronics devices in relation to environmental loads. Current approaches to climatic testing of electronics, such as, for example, the damp-heat test and thermal cycling test described in IEC 61189-3:2007, tend to test extremely harsh conditions, rarely observed in natural climates. Improved knowledge of observed environmental climates and classification is a vital step towards helping design engineers to minimize or eliminate many corrosive interactions.

1.2. Climate classifications in engineering

The IEC 60721-1:1990 (1990) standard provides some guidelines for designing electronic enclosures for different climates, based on the climate classification of Kottek et al. (2006). The Köppen–Geiger classification (KGC) of climates is the most widely used and known climate classification, and defines five main climate classes (A - Tropical, B - Dry, C - Temperate, D - Continental and E - Polar). The classification was first presented by the German scientist Wladimir Köppen in 1900, and was later updated by Rudolf Geiger in 1954. Köppen devised the five climate classes to correspond to different vegetation types, and formulated a set of rules to classify a climate based on its mean monthly temperature and precipitation values. The Köppen–Geiger classification further divides its 5 main climate types into 14 sub-types based on the type of seasonal precipitation, and a final subdivision using additional temperature rules results in 31 climate types. The rules used to define the climate types are heuristics arising from the expert knowledge of Köppen and Geiger in an attempt to separate the different vegetation types, at a time when climate data was far more limited than today. Despite this, it remains the most widely used climate classification today, perhaps in part due to "historical inertia" (Cannon, 2012), but also the fact that for most general purposes, it provides an adequate and logical classification of climate. A drawback of the KGC is that the climate classes are not separated based on any inherent structure in the climate data, but rather by imposing external, somewhat arbitrary, rules. For example, the main Tropical class is defined as climates where mean monthly temperatures for all months exceed 18 °C, but it is unlikely that this precise cut-off is supported by statistical distribution of minimum mean monthly temperatures of climates. A number of authors have proposed alternative, more data-driven climate classifications. Zscheischler et al. (2012) apply K-means clustering to monthly average values of climate variables to obtain an alternative to the KGC. Netzel and Stepinski (2016) use the partitioning around medoids algorithm on temperature and precipitation data to obtain climate classes, and Cannon (2012) use multivariate regression trees. Köppen's motivation of separating vegetation types has also been re-examined, and new climate variables that are more directly linked to plant physiology were defined, on which to apply Ward's method of hierarchical clustering (Gardner et al., 2020).

In addition to the methodology behind the KGC, another limitation is that a classification based on temperature and precipitation may not be the most relevant for corrosion related applications. A number of application-specific climate classifications have been proposed. For example, to obtain a climate classification that is more relevant for photovoltaic systems, an extension of the KGC to define classes also based on solar irradiance has been proposed (Ascencio-Vásquez et al., 2019). In the domain of building design, a climate classification is used that focuses on building energy usage by taking into account thermal criteria represented by the number of so-called heating and cooling degree-days (Briggs et al., 2003). In the field of pesticide pollution monitoring, a climate classification was developed for Europe based on climate variables found particularly important in pesticide fate models (Blenkinsop et al., 2008).

For evaluating corrosion of metals and alloys in general, the ISO 9223:2012 (2012) standard defines six corrosivity categories. The classes are determined by the observed corrosion loss of standard metals. In the absence of corrosion data, classes are proposed based on atmospheric conditions, as quantified by the "Time of wetness". This is the number of hours per year that Temperature exceeds 0 °C and relative humidity exceeds 80%. This very coarse classification does not take into account daily and seasonal variation in conditions.

1.3. Structure of this paper

In this paper, a climate classification is developed based on climate features that are critical to corrosion reliability of electronics, although the outcome is also useful for understanding how climate conditions at different locations may affect corrosion more generally. The Methods section outlines how these features are derived from temperature and humidity, to emphasize conditions that represent a condensation risk. The features take into account both seasonal, and diurnal variation as formation of condensation is often linked to diurnal cycles of temperature that lead to heating and cooling of an object (Schindelholz & Kelly, 2012). The new climate classification is contrasted with the currently used KGC, and the properties of the proposed climate classes are discussed in the Results and Discussion section, before concluding the paper with a summary of the main findings and future perspectives. Additionally, the paper outlines a methodology that can easily be adapted to develop climate classifications in other applications.

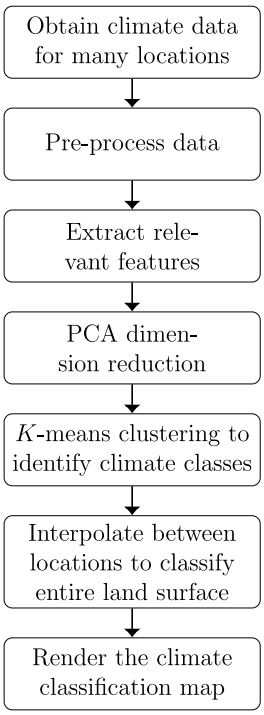

Fig. 1. Overview of the methods used.

2. Methods

An overview of the methods used to develop a new climate classification for electronics reliability is illustrated in Fig. 1. The methodology is highly generalizable to obtain climate classifications for any application. The choice of features in the feature extraction step is key to ensuring that the climate classification is useful for the application of corrosion reliable design of electronic systems. The following sections present details of each step in the methodology.

2.1. Data

To perform the desired climate classification, a dataset that fulfilled the following specifications was required:

1. Climate parameters: For this application, the vital climate parameters were identified as Air Temperature and Humidity (quantified by any of Dew Point, Relative Humidity or Absolute Humidity—if one of these humidity parameter and Temperature is known, then the other humidity parameters can be calculated).
2. Sampling frequency of observations: Must be greater than daily in order to take into account diurnal cycles.
3. Spatial resolution: Data for several thousand locations covering Earth's land surface was required to allow a reasonable extrapolation from the classified locations to all points on the Earth.
4. Time-periods covered: For maximum relevance, recent data is required covering at least one year (to capture seasonal variation) and preferably several years (to capture year-to-year variation).

In general, available climate datasets appear to be subject to a trade-off between sampling frequency of observations and spatial resolution of locations, where datasets are high in one of these attributes but not both. For example, the WorldClim dataset (Fick & Hijmans, 2017), which has previously been used for climate classification (for example by Cannon (2012)) has a very high spatial resolution, but a low sampling frequency of monthly observations, such that diurnal variation is absent. A dataset that met the requirements was obtained from the US National Oceanic and Atmospheric Administration (NOAA) that provides free access to its archive of global historical climate data. The NOAA Integrated Surface Database (ISD) (Smith et al., 2011) contains up to hourly (depending on location) measurements at thousands of stations located around the world, with earliest records over a century old. The stations are generally located at airports. The data is available in a consistent format and includes dew point and temperature. Data from all available locations for the five-year period from 2014 to 2018 were extracted from the ISD-Lite database (a cleaner, easier to use version of the full ISD (Smith et al., 2011)).

2.1.1. Data preprocessing

Data pre-processing consisted primarily of three steps: standardization of sampling frequency for all locations, identifying and removing locations with insufficient data, and outlier detection and removal.

Approximately half of the locations had a sampling frequency of hourly, whilst the sampling frequency was less for the other locations where observations were typically spaced every 3, 6, or 12 h. Therefore, requiring hourly observations would mean having to discard half of the locations. So that more locations could be used, whilst retaining sufficient sub-daily detail, sampling frequency was standardized to 3-hourly for all locations by discarding locations where observations were less frequent than 3-hourly, and down-sampling observations at locations where they were more frequent than 3-hourly.

There were also periods of missing values in the data. These were not filled in using data imputation methods, but instead locations with insufficient data were discarded according to the rule that if a location did not have at least 85% complete observations (both air temperature and dew point) for a year then that location was discarded for that year. This step, along with removal of stations with less than 3-hourly sampling frequency, resulted in removing approximately 40% of the locations originally present in the raw data from the ISD. Next, a second data completeness requirement was applied as not all locations were represented for all years (especially after applying requirement of 85% data coverage per year outlined above). So that features capturing year-to-year variation could be calculated, locations that were only represented for a single year in the dataset were deleted.

Next, outlier values of temperature and dew point that were clearly due to error had to be removed. To identify outliers, locations were split into 10 groups based on 10 equal bands of latitude, and the distributions of temperature and dew point were examined for each group, one calendar month at a time. Combining locations within bands of latitude and taking one month at a time ensured greater power to detect outliers. For each latitude-band and month, if there was a gap in the observations of at least 1 °C within the lower or upper 5% percentiles, then observations below or above this gap respectively, were labeled outliers and removed from the dataset. The procedure resulted in removing approximately 0.005% of observations. A final error-check was to identify observations where the dew point exceeded the temperature (implying greater than 100% relative humidity). Where this occurred, if the difference was less than 1 °C, then the value of the dew point was replaced with a value equal to that of the temperature. If the difference was greater than 1 °C, then a more serious error was assumed and the observation was deleted. From the temperature and dew point, additional humidity variables of relative humidity and absolute humidity were calculated (Sensirion, 2008; Sonntag et al., 2021).

2.1.2. Data summary

The processed data contained 9024 locations, with Temperature/Humidity observations potentially every 3 h during the five-year period 2014 to 2018. The data contained 113,408,506 observations corresponding to 8.6 GB of memory. The locations are shown in Fig. 2. USA and Europe have the best spatial coverage, but the whole world is reasonably represented. Data for all five years was

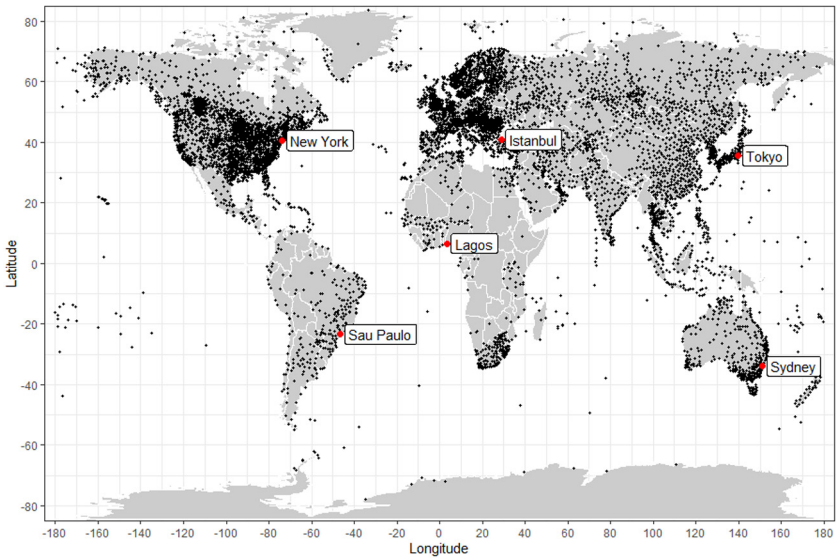

Fig. 2. The 9024 locations used for the classification.

Table 1
Number of locations having each amount of data coverage in number of years in the final dataset.

Number of years	Number of locations
2	590
3	942
4	1248
5	6244

Table 2
Meteorlogical Definition of the seasons.

	N. Hemisphere	S. Hemisphere
1. March – 31. May	Spring	Fall
1. June – 31. Aug.	Summer	Winter
1. Sept. – 30. Nov.	Fall	Spring
1. Dec. – 28. Feb	Winter	Summer

present for the majority of stations, although many stations also had years with no data coverage, as shown in Table 1. If all five years were fully represented by all 9024 locations, then there would have been (4 × 365+366 days)×(8 observations per day)×(9024 locations)= 131,822,592 observations. However, due to periods of missing data, the actual number of observations was 113,408,506 corresponding to a data coverage of 86% of the full potential.

Fig. 3 shows the data for the year 2018 for six locations. Many differences between the six climates can be observed including varying degrees of seasonality (Lagos, being closest to the equator having the least difference between seasons), and differences in temperature and humidity level and variability throughout the year.

2.2. Feature extraction

For each location, the raw data was converted into a number of features that are known to be relevant for corrosion reliability of electronics, and corrosion due to condensation in general. These consisted of features related to temperature and features related to humidity. In both cases, the features were designed to capture day-level patterns as well as seasonal variation and yearly variation. Before further details on the feature extraction are given, a note on how the seasons were defined is necessary. Features were extracted by season, in order to take into account variation in climate at a location, throughout the year. The widely used approach of dividing the year into four three-month seasons was adopted (Trenberth, 1983), as shown in Table 2. Note that the definition depends on whether the location is in the northern or southern hemisphere, i.e., the mean Summer temperature for Sydney would be calculated from the observations in December/January/February, whilst in Istanbul it would be calculated from June/July/August (see Fig. 3) and the resulting summary statistic would be comparable.

The raw data consisted of eight observations per day of each variable, i.e., every 3 h. To ensure that the extracted features captured day-level, season-level, and year-level variation, the chosen variables were first summarized per day. Then each day-level feature was summarized per season. Finally, the season-level features were summarized across the five years to obtain the final set of features. The temperature and humidity features are defined in detail below, and the feature extraction procedure is illustrated in Fig. 4.

2.2.1. Temperature features

Temperature is an important factor influencing the water layer formation on electronic surfaces and corrosion effects. Temperature influences the corrosion reaction rate directly (Cai et al., 2018). In addition, the moisture content in the external atmosphere as well as the temperature drop required for condensation to form are both dependent on the temperature. Condensation processes on an electronic component such as PCBA also depend on the hygroscopic nature of the ionic residues on the surface, and thermodynamically the so-called Deliquescent Relative Humidity (DRH) level and Efflorescence Relative Humidity (ERH) level, both decrease with increase in temperature (Piotrowska et al., 2019). Considering the effect at system level, humidity entry into the device interior due to diffusion through holes and polymer materials packaging also depends on temperature (Conseil-Gudla et al., 2018).

It was important that the temperature features quantified both the average temperature level, as well as the variation within the diurnal cycles. This was achieved by calculating the minimum, mean and maximum temperature per day at each location (each yielding 365 features per year per location). These features were then further summarized by season, by taking the minimum, mean and maximum of the day-level features, within each of the four seasons (resulting in the "summer-mean of the day-minimum", "the winter-maximum of the day-mean" etc.). There was still up to five instances (for each of the five years in the dataset) of each of the season level features at a given location, so in the final stage, the season-level features were summarized by year by taking the minimum, median and maximum of the season level features over the five years. The feature extraction procedure is illustrated in Fig. 4. The procedure resulted in 108 temperature features in total.

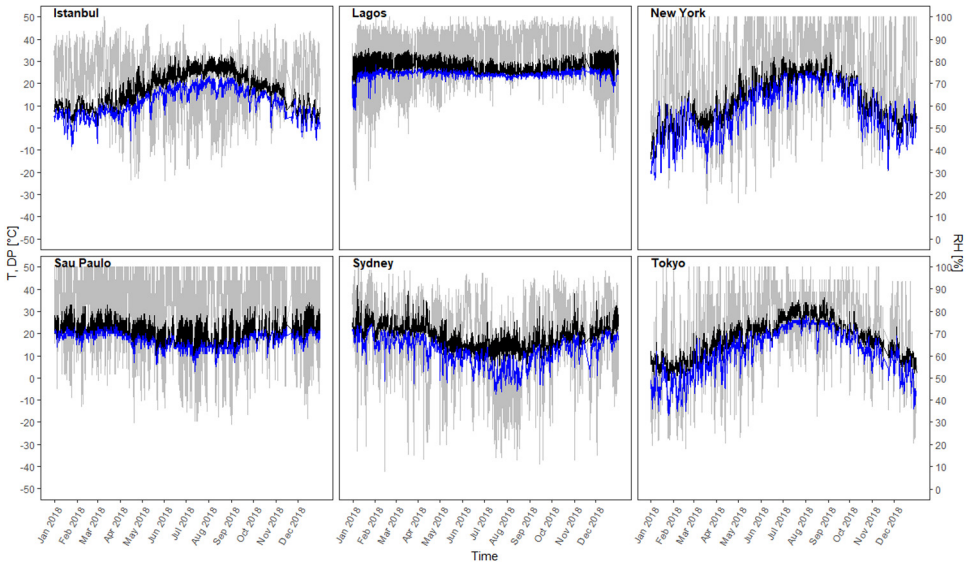

Fig. 3. Temperature (black), dew point (blue) and relative humidity (gray) for six of the locations for the year 2018. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

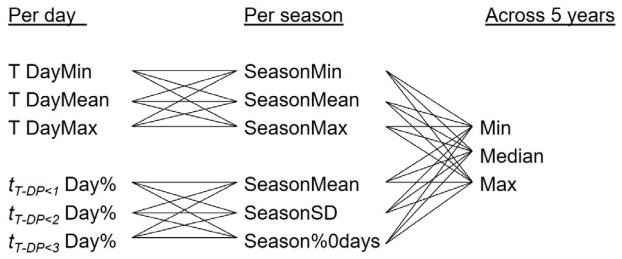

Fig. 4. Illustration of the feature extraction procedure. Each possible path from left to right represents a feature extracted for a particular season. There are $2\times3\times3\times3=54$ paths from left to right in the figure, which when repeated for each of the 4 seasons yields the 216 features used in the analysis.

2.2.2. Humidity features

Humidity level is also a key factor for corrosion reliability of electronics (see Section 1.2). There are many ways to quantify humidity, such as with relative humidity, absolute humidity, and dew point. The property of humidity that has most relevance for the corrosion reliability of electronics or in general for corrosion is its potential for causing condensation on a surface. A small difference between the atmospheric dew point and the air temperature means that conditions on the electronics surface itself are more likely to become conducive to water layer formation, due to differential temperature or hygroscopic residues (Piotrowska & Ambat, 2020). Even a brief time period of close proximity between temperature and dew-point may result in condensation on the electronics surface (Conseil-Gudla et al., 2022). As a direct measure of risk of condensation, three new binary humidity variables were defined to indicate the time (t) when the temperature–dew point difference (T–DP) is less than 1, 2 or 3 ℃, or $t_{T-DP<1}$, $t_{T-DP<2}$, or $t_{T-DP<3}$ respectively:

$$t_{T-DP<i} = \begin{cases} 1, & \text{if } T-DP < i \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The $t_{T-DP<i}$ features, taken together with the temperature features, also indirectly quantify levels of humidity in terms of the dew point, absolute humidity, and relative humidity. For example, a rule of thumb states that each 1 ℃ increase in the difference between the dew point and the temperature corresponds to approximately a 5% decrease in the relative humidity, starting with 100% relative humidity when the dew point and temperature are equal (Lawrence, 2005). This implies that $t_{T-DP<1}=1$, $t_{T-DP<2}=1$, and $t_{T-DP<3=1}$ conditions correspond approximately to values of relative humidity exceeding 95, 90 and 85% respectively, although the exact value of relative humidity will depend on the temperature. These humidity levels correspond to critical relative humidity levels on PCBA surfaces caused by the presence of flux residues of various DRH levels depending on the type of flux chemistry employed for the manufacturing process (Surface DRH reported to vary between approximately 80%–98% RH) (Piotrowska et al., 2018b).

Features were then extracted from the $t_{T-DP<i}$ variables in a similar manner to the temperature features, by summarizing at day level, then at season level and finally across the five years. However, as the $t_{T-DP<i}$ variables were binary, taking the maximum and minimum per day would not be very informative. Instead the percentage of observations per day where $t_{T-DP<i=1}$, for $i=1,2,3$ was calculated. Then, within each season, the mean of the day-percentages, the standard deviation of the day-percentages and the percentage of days when the day-percentage was zero, were calculated. Finally, the features were summarized across the five years by taking the minimum, median and maximum. This yielded 108 humidity features in total and so temperature and humidity were equally represented in terms of number of features.

2.3. Principal component analysis of features

Feature extraction resulted in 216 features for each of the 9024 locations. The aim was to cluster the locations into clusters of similar climate using the K-means algorithm. This method is often sensitive to "the curse of dimensionality" where the difference between the nearest and the farthest point to a given point becomes increasingly meaningless in higher dimensions (Hastie et al., 2009). Certainly, the 216 dimensional feature space (despite representing a considerable reduction in the starting dimensions of eight observations per day for five years corresponding to 14,608 observations for each variable) was likely to be problematic for the K-means algorithm. In addition, many of the features were highly correlated. Therefore, it was natural to apply principal component analysis (PCA) to the features in order to reduce the dimensions further. PCA finds a linear transformation of the data to obtain fewer dimensions that are uncorrelated, whilst retaining as much relevant information as possible (Jolliffe, 2002). Let $\tilde{\mathbf{X}}$ be the $n\times p$ matrix of $p=216$ features at the $n=9024$ locations. Then, let $\mathbf{X}$ be the scaled version of Let $\tilde{\mathbf{X}}$ obtained by taking each column of Let

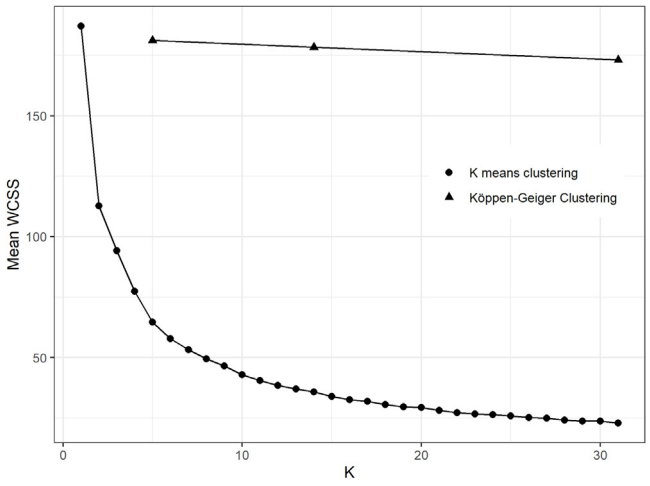

Fig. 5. Mean WCSS against the number of clusters, K, for K-means (circles) and for Köppen–Geiger (triangles).

$\tilde{\mathbf{X}}$, subtracting the mean, and dividing by the standard deviation. Then PCA decomposes $\mathbf{X}$ as

$$\mathbf{X}_{n\times p} = \mathbf{T}_{n\times L}\mathbf{P}'_{L\times P} + \mathbf{E}_{n\times p} \tag{2}$$

where $L \leq p$ is the number of retained components, $\mathbf{T}$ contains the component scores, $\mathbf{P}$ contains the loadings and $\mathbf{E}$ contains the error. The aim of PCA is to retain as few components, L, as possible without losing too much information. Many methods for selecting L have been proposed Hastie et al. (2009). In this work, L was selected as the minimum number of components required to explain 85% of the variation. This resulted in a value of $L = 5$ retained components, which explained 86.6% of the variance in $\mathbf{X}$.

2.4. Clustering of station locations with K-means

Using feature extraction followed by PCA, the five years of climate data at each location were summarized by $L = 5$ latent variables. The next step was to classify the climates into a limited number of climate classes, such that locations assigned to the same class had a more similar climate than locations assigned to different classes. There exist many algorithms for clustering data. An overview of clustering methods is given by Saxena et al. (2017). One of the most popular algorithms for clustering data is the K-means algorithm (Hastie et al., 2009). This algorithm assigns observations to K clusters by aiming to minimize the sum of squared distances of each observation to its cluster center-point, known as the within-cluster sum of squares (WCSS). The algorithm begins by randomly selecting K points to be the K cluster centers. Then the following two steps are repeated until convergence is achieved, i.e., the cluster assignments do not change, or until the number of allowable iterations is reached:

1. Each observations is assigned to the cluster of the closest cluster center
2. The cluster centers are re-calculated by taking the mean of the observations within each cluster

The number of clusters, K, must be specified before the algorithm is applied. The choice of K depends on the purpose of the clustering. In some cases, external factors may dictate the appropriate K to use. For example, if the data must be split into a specific number of segments for a particular problem. More commonly, K must be selected based on the data itself. Then, the usual approach is to apply K-means for all values of K of possible interest, and select the final value based on an evaluation of the total WCSS corresponding to each value of K (Hastie et al., 2009). The Total WCSS will inevitably decrease, as K increases. However, the idea is that Total WCSS will decrease at a faster rate for K smaller than the optimum value, and at a slower rate for K greater than the optimum value, so that the optimum value of K should be revealed as an noticeable bend, or elbow, in the plot of Total WCSS vs. K. The elbow will be more obvious in cases where the data has an inherently clustered structure. Fig. 5 shows the WCSS against K for the K-means clustering applied to the PCA scores. The Köppen–Geiger clustering referred to in the Fig. 5 is discussed in the following section. WCSS decreases rapidly as K increases from 1 to 5, and decreases less rapidly as K increases above 5. This suggests that $K = 5$ is a reasonable number of clusters for this data, although values of K in the range 4 to 10 could be considered equally plausible based on the plot. However, the familiar and widely used (in for example, IEC 60721-1:1990) KGC has 5 main classes suggesting that 5 has proven to be a reasonable and practical number of classes to use.

The K-means algorithm names the K clusters with the numbers 1 to K, but the ordering is arbitrary. Therefore, once the clusters were identified, they were relabeled by calculating the mean absolute latitude of all locations assigned to each cluster, and labeling the clusters with the numbers 1 to K in ascending order of this value.

2.5. Nearest neighbor interpolation of global land surface

Having successfully clustered the 9024 stations into 5 climate classes, the final step was to extend this classification to every point of land on the earth's surface. This was done by constructing a grid of cells covering the earth. The dimensions of each grid cell were 5 arc minutes of latitude and longitude resulting in 12 cells per degree of latitude or longitude. The climate map from Kottek et al. (2006) was used as a template to construct this grid, including the categorization of each grid cell as either land or ocean. Then the nearest station to each cell in the grid was identified based on Euclidean distance in three-dimensional space under the assumption of a spherical earth. The cell was then assigned the climate class of the nearest station. In this way, the entire landmass of the earth could be classified as illustrated in Fig. 6.

3. Results and discussion

The Climate classification using parameters important for corrosion obtained from the methods described in the previous section is presented in Fig. 6(b). This classification is hereafter referred to as the Corrosion Reliability Climate Classification (CRC). No geographical information was used in the clustering, but as locations that are geographically close, also tend to be similar in climate, the climate zones appear as large regions on the map. The geographical separation between the climate classes is well defined, although classes 3 and 4 appear to mix slightly in North America and Europe. This may be due to the larger number of locations present in this part of the world (see Fig. 2), allowing for a much more detailed geographical classification.

Before further discussion of the proposed climate classification, and the properties of each climate class, a comparison of this new classification to the traditional KGC is presented.

3.1. Comparison to Köppen–Geiger classification

Several approaches were applied in order to compare the proposed CRC to the KGC: visual inspection of the two classification maps, cross tabulation of land area coverage by each climate type according to each classification, evaluation of the WCSS corresponding to each classification framework, and finally, examination of the distribution of mean temperature and $t_{T-DP<3}$ of locations assigned to each class in the two systems.

The KGC map from Kottek et al. (2006) is shown in Fig. 6(a). A visual comparison to the CRC in Fig. 6(b) suggests that CRC zones 1

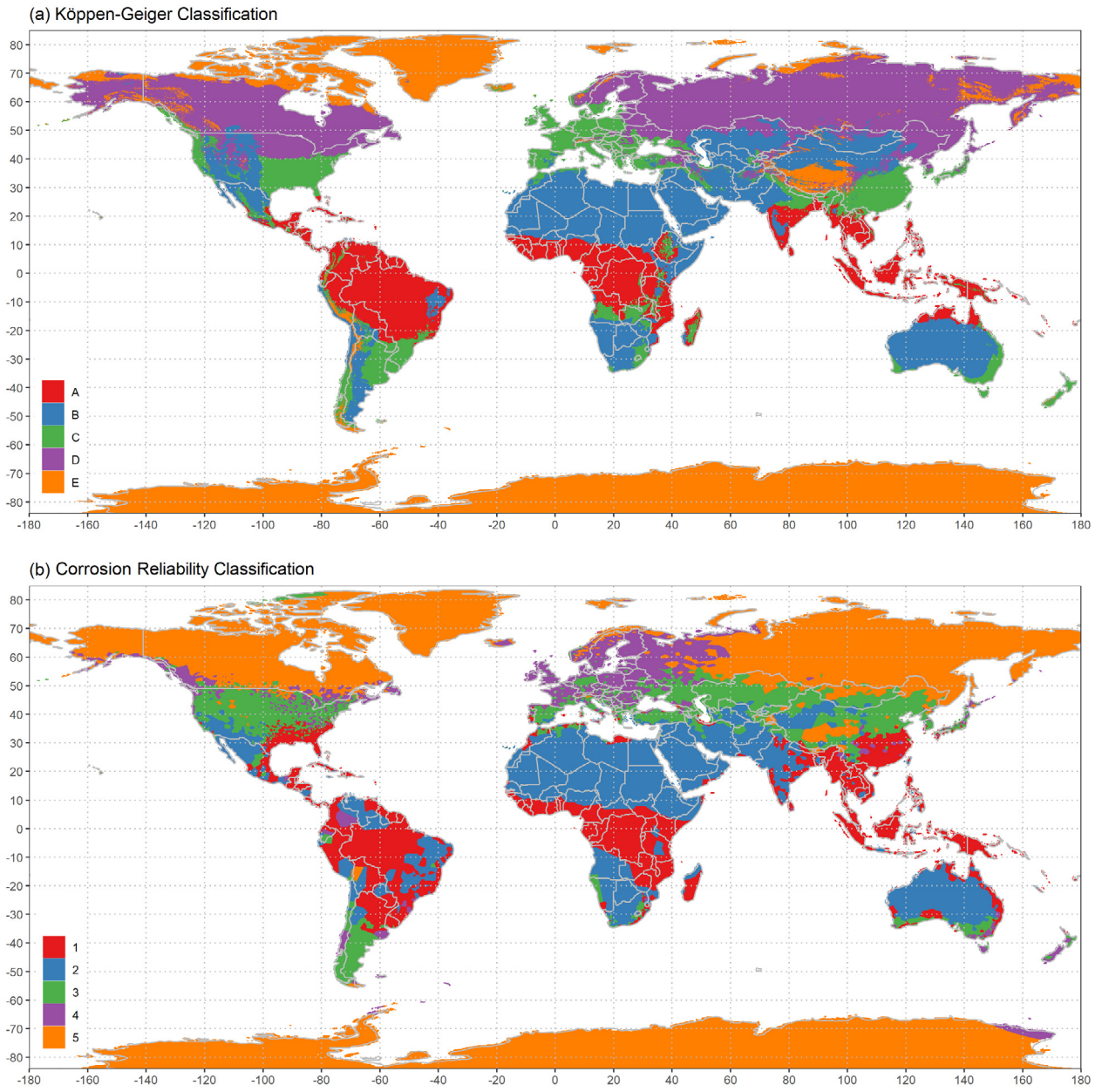

Fig. 6. (a) Köppen–Geiger climate classification and (b) Proposed climate classification for electronics reliability.

Table 3
Cross tabulation of land area covered by each CRC zone (1 - 5) and each KGC zone (A - E) in units of 100,000 km^2.

		KGC Zone				
		A	B	C	D	E
CRC Zone	1	224.6	28.7	84.3	0.3	2.1
	2	76.7	316.9	30.4	2.4	5.8
	3	1.1	85.2	63.7	51.0	11.7
	4	4.0	0.5	42.6	53.6	4.9
	5	0.8	7.6	1.9	216.9	205.5

and 2 cover similar regions to KGC zones A and B respectively, whilst there is less of a correspondence between CRC zones 3, 4, and 5 and KGC zones C, D and E. This is also reflected in a cross tabulation of land area covered by the CRC zones against land area covered by the KGC zones (Table 3). Table 3 indicates that CRC zone 3 is a mixture of KGC zones B, C and D. CRC zone 4 is a mixture of KGC zones C and D, and CRC zone 5 is a mixture of KGC zones D and E (Table 3). There is no clear one-to-one correspondence between the KGC and CRC zones. This confirms that clustering based on Temperature and $t_{T-DP<i}$ features results in a very different classification than clustering based on Temperature and Precipitation.

As a measure of goodness of fit of the KGC vs CRC methods, the WCSS (previously used to select K in Section 2.4) was considered. The KGC is a hierarchical classification with three levels, and so the KGC contains within it three different classifications, each with a different number of classes, depending on which level is chosen. At the top level, there are the 5 main climate classes. These main climate types are then subdivided based on seasonal precipitation patterns resulting in the second level of 14 climate classes. These classes are then further subdivided based on temperature features resulting in the third level of 31 climate classes. The WCSS was calculated in the latent space of PCA scores for each of these possibilities in the KGC framework and the values are shown in Fig. 5. This shows that the WCSS is much less in the CRC (K-means) clustering than the KGC clustering, for any K greater than 1. Even the 31-class KGC has a much greater WCSS than the 2-class K-means classification. Of course, the K-means approach directly

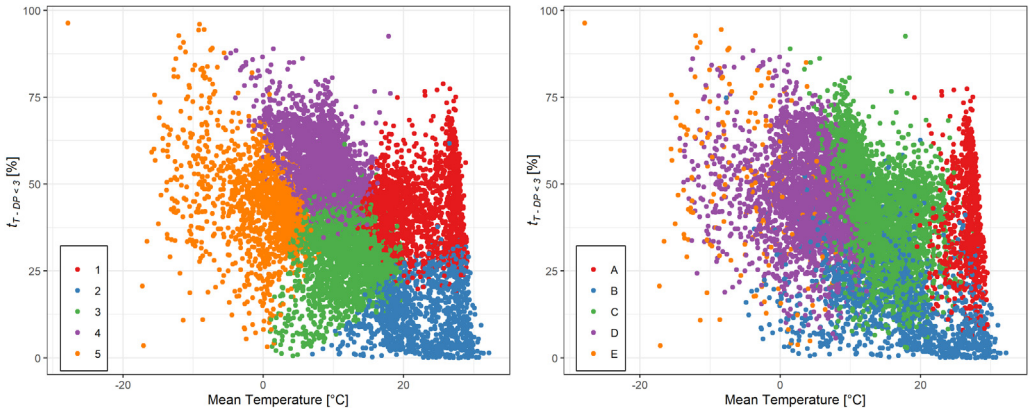

Fig. 7. $t_{T-DP<3}$ vs. mean Temperature per location for the proposed electronics reliability clustering (left) and the Köppen–Geiger clustering (right).

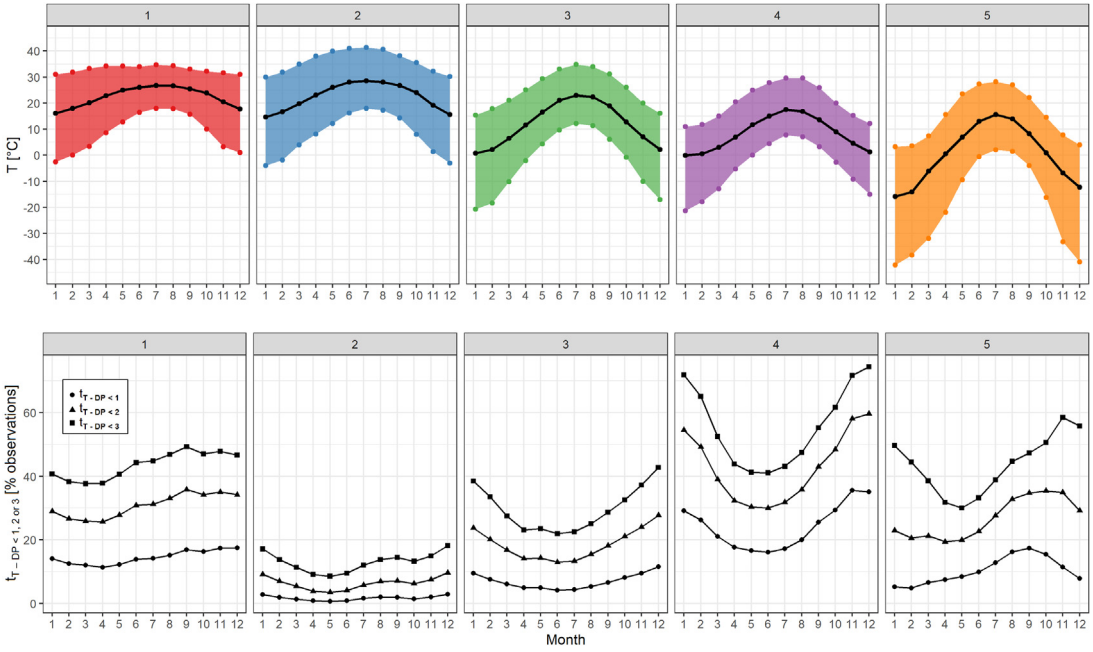

Fig. 8. Summary of average conditions per month in each of the proposed climate zones. Top: Temperature mean (black line) and 95% reference interval (colored band). Bottom: Percentage of $t_{T-DP<1}$ (circles), $t_{T-DP<2}$ (triangles) and $t_{T-DP<3}$ (squares) observations in each month.

minimized WCSS within the latent PCA space, whilst the KGC is based on different features so WCSS is expected to be less for K-means. Even so, it is interesting to note the very large difference in WCSS.

Next, the means of all temperature observations, and of all $t_{T-DP<3}$ observations, were calculated for each location. These two features are a huge simplification of the data compared to the features used in the clustering, which were calculated at day, season and year level. However, they are useful for summarizing in two easy to visualize dimensions. The values are plotted and colored according the 5-class CRC, and the 5-class KGC in Fig. 7. There is a much clearer separation of points in the CRC than the KGC, especially in the $t_{T-DP<3}$ axis. In the KGC, besides climate type B that has generally lower $t_{T-DP<3}$, all the climate types span a large, overlapping range of $t_{T-DP<3}$ values spanning from around 10% to around 80%, and so the KGC is not so useful for distinguishing between different levels of condensation risk. The CRC approach more clearly separates climates into high and low temperature, and high and low condensation risk.

3.2. Summary of the proposed climate classes

Referring to the mean Temperature and $t_{T-DP<3}$ of all locations in Fig. 7, the five climate classes in the CRC can be characterized as 1: high temperature/high humidity (Hot–Wet), 2: high temperature/low humidity (Hot–Dry), 3: low temperature/low humidity (Cool–Dry), 4: low temperature/high humidity (Cool–Wet), 5: Polar (Cold). Here the descriptors hot/cool/cold/dry/wet are used loosely to describe the zones relative to each other. For a more detailed characterization of the CRC climates, the conditions in each climate class for each month of the year were examined. In the Northern hemisphere, the month numbers start with January as month 1 to December as month 12, whilst in the Southern hemisphere the month numbers correspond to July as month 1 to June as month 12, in order to take into account the opposite seasons of the hemispheres. For each month, and each climate class, all observations from all locations and years were extracted and the mean temperature was calculated. The 2.5 and 97.5 percentiles were calculated to obtain the reference interval containing 95% of all temperature observations. In addition, the percentage of observations satisfying the $t_{T-DP<1}$, $t_{T-DP<2}$ and $t_{T-DP<3}$ conditions was calculated. The results are shown in Fig. 8. The highest $t_{T-DP<i}$ percentages occur in Zone 4 (Cool–Wet zone), during months 11 and 12 with $t_{T-DP<1}$ conditions occurring 35% of the time. The seasonal difference of $t_{T-DP<i}$ is greatest in zone 4. Even so, the zone 4 month with the lowest $t_{T-DP<1}$ percentage (month 6 at 16.1%) almost matches the highest $t_{T-DP<1}$

percentages in zones 1 and 5 (17.5% in month 12 and 17.4% in month 9 respectively), and greater than the highest $t_{T-DP<1}$ percentages in zones 2 and 3 (2.9% and 9.6% respectively).

Although $t_{T-DP<i}$ is very high in zone 4, the mean monthly temperatures are quite low ranging from −0.1 (month 1) to 17.5 ℃ (month 7). The mean monthly temperatures in the Hot–Wet zone (zone 1) range from 16.1 (month 1) to 26.7 ℃ (month 7), and in this zone the $t_{T-DP<1}$ percentage is quite consistent from month to month (ranging from 12.0 to 17.5%). The driest zone is zone 2 (hot–dry), but even here $t_{T-DP<1}$conditions are not completely absent and occur on average 0.7–2.9% of the time (month 5 and month 12 respectively). The so-called cool–dry zone 3 has $t_{T-DP<1}$monthly percentages ranging from 4.2 to 9.6%, which is somewhat higher than the hot–dry zone 2. Finally, the cold zone 5 is much colder than the other zones, with mean monthly temperature of down to −15.9 ℃.

3.3. Use of CRC for designing humidity robust electronic products

The classification shown in this paper is specifically made in connection with corrosion reliability of electronics. The CRC will be more useful for providing general information on aggressiveness of different regions, which will help not only in the design stage for corrosion robustness, but also to predict the relative performance of the device under specific climatic conditions. For example, in electronic systems such as pumps used in cold-water applications, condensation can occur on the pump depending on the combination of dew point of the climate as well as temperature of the cold water. Using the above classification, the possibility of condensation in the region of use can be broadly defined using the dew point variation data as well as cold-water temperature.

4. Conclusion

A new climate classification suitable for aiding design, testing and lifetime estimates of electronic enclosures for providing better humidity protection has been presented. This classification is based on features specifically defined to be relevant for climate reliability of electronics, based on domain knowledge of this application. The classification was shown to more effectively separate climates based on temperature and condensation risk than the currently used KGC. An analysis of each of the new climate classes revealed that $t_{T-DP<1}$ conditions can occur in each climate class at all months of the year, so the risk of condensation must be taken seriously no matter the geographical location.

Although K-means clustering was applied in this work, other clustering methods could have been used instead which would result in alternative climate classifications. However, without an objective response variable (such as product failure data around the world), there is little basis for selecting one clustering method over another. Choice of clustering method for this application remains a topic for future research.

In addition to the specific classification presented in this paper, it is hoped that the methods described can be adapted to a wide range of cases to obtain case-specific climate classifications. Even within the field of electronics reliability, different manufacturers may require different classifications, for example, with differing numbers of climate zones. In this age of cheap computational power, and widely available data, there is no need to be limited to a single climate classification. Climate change provides additional motivation for a flexible classification that can be updated.

The issue of how to test product designs based on the climate classes has not been considered in detail. In future work, it is our hope to expand the classification to define suitable mission profiles of the climate variables for each class. Such mission profiles should more closely reflect the observed climates, and the dynamics of changing conditions over time, in contrast to the extreme conditions often used to test products today. In the present work, the climate classification, and associated class summaries, should prove useful for product design considerations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Current research has been conducted as part of the ELMAC (Electronics Manufactured for Climate) project funded by the Innovation Fund Denmark. The authors would like to acknowledge the commitment and help of the academic and industrial partners.

References

Ambat, R., Conseil-Gudla, H., & Verdingovas, V. (2018). Corrosion in electronics. In *Encyclopedia of interfacial chemistry: surface science and electrochemistry* (pp. 134–144). Elsevier, http://dx.doi.org/10.1016/b978-0-12-409547-2.13437-7.

Ascencio-Vásquez, J., Brecl, K., & Topič, M. (2019). Methodology of Köppen-Geiger-Photovoltaic climate classification and implications to worldwide mapping of PV system performance. *Solar Energy*, *191*, 672–685. http://dx.doi.org/10.1016/j.solener.2019.08.072.

Bahrebar, S., & Ambat, R. (2021). Investigation of critical factors effect to predict leakage current and time to failure due to ECM on PCB under humidity. *Microelectronics Reliability*, *127*, Article 114418. http://dx.doi.org/10.1016/j.microrel.2021.114418.

Blenkinsop, S., Fowler, H., Dubus, I. G., Nolan, B. T., & Hollis, J. (2008). Developing climatic scenarios for pesticide fate modelling in Europe. *Environmental Pollution*, *154*(2), 219–231. http://dx.doi.org/10.1016/j.envpol.2007.10.021.

Briggs, R. S., Lucas, R. G., & Taylor, Z. T. (2003). Climate classification for building energy codes and standards: Part 2-zone definitions, maps, and comparisons. *ASHRAE Transactions*, *109*, 122.

Cai, Y., Zhao, Y., Ma, X., Zhou, K., & Chen, Y. (2018). Influence of environmental factors on atmospheric corrosion in dynamic environment. *Corrosion Science*, *137*, 163–175. http://dx.doi.org/10.1016/j.corsci.2018.03.042.

Cannon, A. (2012). Köppen versus the computer: comparing Köppen-Geiger and multivariate regression tree climate classifications in terms of climate homogeneity. *Hydrology and Earth System Sciences*, *16*(1), 217–229. http://dx.doi.org/10.5194/hess-16-217-2012.

Conseil, H., Gudla, V. C., Jellesen, M. S., & Ambat, R. (2016). Humidity build-up in a typical electronic enclosure exposed to cycling conditions and effect on corrosion reliability. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, *6*(9), 1379–1388. http://dx.doi.org/10.1109/TCPMT.2016.2590779.

Conseil, H., Jellesen, M. S., & Ambat, R. (2014). Contamination profile on typical printed circuit board assemblies vs soldering process. *Soldering & Surface Mount Technology*, *26*(4), 194–202. http://dx.doi.org/10.1108/SSMT-03-2014-0007.

Conseil-Gudla, H., Spooner, M., Kulahci, M., & Ambat, R. (2022). Transient risk of water layer formation on PCBAs in different climates: climate data analysis and experimental study. *Microelectronics Reliability*, *136*, 114655. http://dx.doi.org/10.1016/j.microrel.2022.114655.

Conseil-Gudla, H., Staliulionis, Z., Mohanty, S., Jellesen, M., Hattel, J., & Ambat, R. (2018). Humidity build-up in electronic enclosures exposed to different geographical locations by RC modelling and reliability prediction. *Microelectronics Reliability*, *82*, 136–146. http://dx.doi.org/10.1016/j.microrel.2018.01.013.

D'Angelo, L., Verdingovas, V., Ferrero, L., Bolzacchini, E., & Ambat, R. (2017). On the effects of atmospheric particles contamination and humidity on tin corrosion. *IEEE Transactions on Device and Materials Reliability*, *17*(4), 746–757. http://dx.doi.org/10.1109/TDMR.2017.2771505.

Fick, S. E., & Hijmans, R. J. (2017). Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. http://dx.doi.org/10.1002/joc.1276.

Gardner, A. S., Maclean, I. M., & Gaston, K. J. (2020). A new system to classify global climate zones based on plant physiology and using high temporal resolution climate data. *Journal of Biogeography*, *47*(10), 2091–2101. http://dx.doi.org/10.1111/jbi.13927.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hunt, C., Mensah, A., Buxton, A., & Holman, R. (2006). Determining conformal coating protection. *Soldering & Surface Mount Technology*, *18*(4), 38–47. http://dx.doi.org/10.1108/09540910610717893.

IEC 60721-1:1990 (1990). Classification of environmental conditions - part 1: Environmental parameters and their severities. *Standard*, Geneva: International Electrotechnical Commission.

IEC 61189-3:2007 (2007). Test methods for electrical materials, printed boards and other interconnection structures and assemblies - part 3: Test methods for interconnection structures (printed boards). *Standard*, Brussels: International Electrotechnical Commission.

ISO 9223:2012 (2012). Corrosion of metals and alloys—Corrosivity of atmospheres—Clasification, determination and estimation, (2nd ed.). *Standard*, Geneva: International Organization for Standardization.

Jacobsen, J. B., Krog, J. P., Rimestad, L., Riis, A., & Holm, A. H. (2014). Climate-protective packaging: Using basic physics to solve climatic challenges for electronics in demanding applications. *IEEE Industrial Electronics Magazine*, *8*(3), 51–59. http://dx.doi.org/10.1109/MIE.2014.2330912.

Jolliffe, I. T. (2002). *Principal component analysis*. Springer-Verlag New York, http://dx.doi.org/10.1007/b98835.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*, 259–263. http://dx.doi.org/10.1127/0941-2948/2006/0130.

Lawrence, M. G. (2005). The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications. *Bulletin of the American Meteorological Society*, *86*(2), 225–234. http://dx.doi.org/10.1175/BAMS-86-2-225.

Mantis, I., Li, F., Jellesen, M. S., & Ambat, R. (2021). Effect of intrinsic PCB parameters on the performance of fluoropolymer coating under condensing humidity conditions. *Microelectronics Reliability*, *122*, Article 114158. http://dx.doi.org/10.1016/j.microrel.2021.114158.

NACE (2016). International measures of prevention, application, and economics of corrosion technologies (IMPACT) report. *Technical Report*, Houston, Texas, USA: NACE International.

Netzel, P., & Stepinski, T. (2016). On using a clustering approach for global climate classification. *Journal of Climate*, *29*(9), 3387–3401. http://dx.doi.org/10.1175/JCLI-D-15-0640.1.

Piotrowska, K., & Ambat, R. (2020). Residue-assisted water layer build-up under transient climatic conditions and failure occurrences in electronics. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, *10*(10), 1617–1635. http://dx.doi.org/10.1109/tcpmt.2020.3005933.

Piotrowska, K., Din, R. U., Jellesen, M. S., & Ambat, R. (2018). Effect of solder mask surface chemistry and morphology on the water layer formation under humid conditions. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, *8*(10), 1756–1768. http://dx.doi.org/10.1109/TCPMT.2018.2792047.

Piotrowska, K., Grzelak, M., & Ambat, R. (2019). No-clean solder flux chemistry and temperature effects on humidity-related reliability of electronics. *Journal of Electronic Materials*, *48*(2), 1207–1222. http://dx.doi.org/10.1007/s11664-018-06862-4.

Piotrowska, K., Ud Din, R., Grumsen, F. B., Jellesen, M. S., & Ambat, R. (2018). Parametric study of solder flux hygroscopicity: Impact of weak organic acids on water layer formation and corrosion of electronics. *Journal of Electronic Materials*, *47*(7), 4190–4207. http://dx.doi.org/10.1007/s11664-018-6311-9.

Rathinavelu, U., Jellesen, M. S., Moller, P., & Ambat, R. (2012). Effect of no-clean flux residues on the performance of acrylic conformal coating in aggressive environments. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, *2*(4), 719–728. http://dx.doi.org/10.1109/TCPMT.2012.2186456.

Roman, A., Wu, B., Han, B., Reinacher, G. M., & Yousef, S. G. (2020). Moisture transport through housing materials enclosing critical automotive electronics. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, *10*(4), 541–550. http://dx.doi.org/10.1109/TCPMT.2020.2977111.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., & Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, *267*, 664–681. http://dx.doi.org/10.1016/j.neucom.2017.06.053.

Schimpf, C., Feldmann, K., Matzner, C., & Steinke, A. (2009). Failure of electronic devices due to condensation. *Microsystem Technologies*, *15*(1), 123–127. http://dx.doi.org/10.1007/s00542-008-0643-y.

Schindelholz, E., & Kelly, R. G. (2012). Wetting phenomena and time of wetness in atmospheric corrosion: a review. *Corrosion Reviews*, *30*(5–6), 135–170. http://dx.doi.org/10.1515/corrrev-2012-0015.

Sensirion (2008). Humidity at a glance, most relevant equations with sample code. https://www.sensirion.com/en/download-center/.

Smith, A., Lott, N., & Vose, R. (2011). The integrated surface database: Recent developments and partnerships. *Bulletin of the American Meteorological Society*, *92*(6), 704–708. http://dx.doi.org/10.2307/26218543, URL http://www.jstor.org/stable/26218543.

Song, B., Azarian, M. H., & Pecht, M. G. (2012). Impact of dust on printed circuit assembly reliability. In *Proc. IPC apex expo conference, Vol. 3* (pp. 1643–1659).

Song, B., Azarian, M. H., & Pecht, M. G. (2013). Effect of temperature and relative humidity on the impedance degradation of dust-contaminated electronics. *Journal of The Electrochemical Society*, *160*(3), C97.

Sonntag, D., Foken, T., Vömel, H., & Hellmuth, O. (2021). Humidity sensors. In T. Foken (Ed.), *Springer handbook of atmospheric measurements* (pp. 209–241). Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-52171-4_8.

Tencer, M., & Moss, J. (2002). Humidity management of outdoor electronic equipment: methods, pitfalls, and recommendations. *IEEE Transactions on Components and Packaging Technologies*, *25*(1), 66–72. http://dx.doi.org/10.1109/6144.991177.

Trenberth, K. E. (1983). What are the seasons? *Bulletin of the American Meteorological Society*, *64*(11), 1276–1282. http://dx.doi.org/10.1175/1520-0477(1983)064<1276:WATS>2.0.CO;2.

Verdingovas, V., Jellesen, M. S., & Ambat, R. (2015). Relative effect of solder flux chemistry on the humidity related failures in electronics. *Soldering & Surface Mount Technology*, *27*(4), 146–156. http://dx.doi.org/10.1108/SSMT-11-2014-0022.

Veselý, P., Bušek, D., Krammer, O., & Dušek, K. (2020). Analysis of no-clean flux spatter during the soldering process. *Journal of Materials Processing Technology*, *275*, Article 116289. http://dx.doi.org/10.1016/j.jmatprotec.2019.116289.

Wang, H., Liserre, M., & Blaabjerg, F. (2013). Toward reliable power electronics: Challenges, design tools, and opportunities. *IEEE Industrial Electronics Magazine*, *7*(2), 17–26. http://dx.doi.org/10.1109/MIE.2013.2252958.

Zscheischler, J., Mahecha, M. D., & Harmeling, S. (2012). Climate classifications: the value of unsupervised clustering. *Procedia Computer Science*, *9*, 897–906. http://dx.doi.org/10.1016/j.procs.2012.04.096.