Theoretical and Applied Climatology (2025) 156:512
https://doi.org/10.1007/s00704-025-05792-6

RESEARCH

Statistical analysis of monthly precipitation in Sweden using the Tweedie distribution

Jesper Rydén[1]

Received: 9 April 2025 / Accepted: 31 August 2025

Abstract

Statistical models for precipitation is of importance in several applications in climatology and agriculture. In this paper, we study a statistical distribution, the Tweedie distribution, to analyse monthly precipitation. Results are found in the literature for sites found in Australia. Here we investigate the amount of precipitation at several sites in Sweden, where various climate zones are found according to the Köppen–Geiger climate classification. The Tweedie distribution was fitted by maximum-likelihood techniques and the behaviour of the so-called index parameter was assessed on a monthly basis for the various sites. Moreover, a clustering procedure was applied to possibly cluster the sites according to the index parameter. We found similar features as for the Australian studies concerning magnitude of the estimated index parameter and influence of coastlines. The results indicate that the Tweedie distribution is a relevant model also for sites in Northern Europe.

1 Introduction

There is a need for statistical modelling of precipitation. For instance, increases in rainfall may cause floodings and in worst cases landslides. Moreover, climate change implies several challenges related to precipitation, like droughts, heat waves or wildfires. As a consequence of drought conditions, agriculture is affected with e.g. diminished crop yield. Overall, extreme events could result in major economical and social losses. As a few examples could be mentioned flood events where public water systems fail to provide fresh water, or aviation accidents caused by fog and precipitation reducing visibility. For further details and examples, consult e.g. NASA Goddard Space Flight Center (2016); Beniston et al. (2021); González-Pérez et al. (2022).

Precipitation refers to all forms of hydrometeors that originate from the condensation of atmospheric water vapor and subsequently fall to the Earth's surface under the influence of gravity (NASA Goddard Space Flight Center 2016). In the hydrological cycle, precipitation plays a central role in regulating the global and regional climate system, influencing energy and water budgets, circulation patterns, and ecosystem dynamics. Its occurrence and characteristics depend on cloud micro-physical processes and the thermodynamic structure of the atmosphere. The principal forms include drizzle and rain (liquid precipitation of varying droplet sizes), snow (aggregates of ice crystals), sleet (partially frozen or refrozen pellets), graupel (rimed snow particles), and hail (layered ice formed by repeated accretion cycles within convective storms). Precipitation can be studied in terms of the amount that falls, but also its occurrence over time and space. Concerning temporal aspects, rainfall can be studied at several time scales, depending on the application: daily, monthly, seasonal. In this paper, we focus on monthly precipitation.

Modelling precipitation with probability distributions needs careful assessment. In principle, we face a continuous random outcome, but there is also the possibility of observing exact zeros. The continuous part is often right-skewed, and the gamma distribution has in many situations proven itself useful (Martinez-Villalobos and Neelin 2019). A statistical approach which simultaneously models these continuous and discrete features was presented by Dunn (2004), and will be considered in this paper. That methodology builds upon the Tweedie distribution, which is found within the framework of exponential dispersion models (EDM). From a meteorological point of view, daily as well as monthly precipitation were investigated at sites in Australia.

✉ Jesper Rydén
jesper.ryden@slu.se

[1] Swedish University of Agricultural Sciences, Uppsala, Sweden

Published online: 17 September 2025

The approach has an advantage in that regression models can be constructed, introducing covariates of interest. As an example, see Yunus et al. (2017), where daily rainfall was investigated using climatological predictors.

The aim of this paper is to examine precipitation at locations in Sweden by the Tweedie distribution. Actually, there are, to the author's knowledge, few studies in the literature that employ this approach to situations outside Australia. Sweden has a diverse climatology. Moreover, at some locations precipitation falls (partly) in the form of snow. Interest is also in examining influence from geographical factors like latitude or elevation of the recording station. More precisely, this paper studies the variability of one of the parameters in the Tweedie distribution, the so-called index (or power) parameter, from a seasonal as well as a geographical point of view. This is a further development of the framework presented by Dunn (2004). Further, an attempt is made of classifying stations based on their estimated index parameters, month by month. Recently, researchers in agriculture have begun to classify the climate of a region using empirical methods and clustering alogorithms, cf. Abbasi et al. (2022). However, identifying climatic zones have been a concern for scientists already in Ancient Greece (Feddema 2005). Development of different types of climatic classification has been proposed e.g. by Köppen (1900) and De Martonne (1941). A review of the works by de Martonne is provided by Broc and Giusti (2007). Results by de Martonne were used in an early study for Sweden concerning humidity (Tamm 1959).

The paper is structured as follows. In Section 2, background of the data is given, along with implications for the modelling in the sequel of the paper. A review of the methodology is provided in Section 3; presentation of the Tweedie distribution, related estimation techniques and summary of the classification methodologies employed. Further, in Section 4, results are presented. A summary and concluding dicsussion is given in Section 5.

2 Data

The original quantity of study is daily precipitation (mm). Data were retrieved online from the Swedish Meteorological and Hydrological Institute (SMHI)[1], where 14 stations were selected allover the country. This is a reasonable number of sites to explore; too many may yield too cluttered graphs when summarising reults. See Table 1 for a summary with geographical information, and Fig. 1, left panel, for a map with the stations indicated along with elevations, using a so-called DEM (Digital Elevation Model), as implemented in the R package tmap (Tennekes 2018).

Moreover, attention is paid to climate zones. Sweden is characterized by four Köppen–Geiger climate zones: temperate oceanic (Cfb), temperate continental (Dfb), subarctic (Dfc), and tundra (ET) in the highest mountain areas. For illustration, see Fig. 1, right panel. These zones reflect the pronounced gradients in temperature and precipitation regimes across the country, from the relatively mild and humid conditions in the south and along the western coast, to the cold, continental and snow-dominated regimes in the north. Such climatic heterogeneity has direct implications for hydrological processes, water balance, and the spatio-temporal variability of precipitation. This motivates introduction of the notion of climate zone in this study.

The time period investigated was 1900–2019, except for Station 14 (Kiruna), where the period is 1900-2009. There were no gaps or missing values in the original time series. Note that Station 3, Visby, is located at an island in the Baltic Sea. Furthermore, some of the stations in the northern part (12 and 13), though they belong to the Köppen climate type Dfc, are located along the coastline, as opposed to Station 14 (Kiruna).

In Table 2 is shown summary statistics of precipitation for the selected stations under study, in order to present climatological characteristics at the various sites. The quantity analysed here is total monthly precipitation. Concerning variability in the form of coefficient of variation, we may note a slight increase with increasing latitude.

Based on the original observations of daily rainfall, the analysis in the sequel will focus on monthly statistical

Table 1 Geographical summary for stations investigated

Station	Location	Latitude	Longitude	Elevation (m asl)	Climate zone
1	Lund	55.69 N	13.22 E	26.4	Cfb
2	Karlshamn	56.18 N	14.85 E	50.0	Cfb
3	Visby	57.62 N	18.30 E	46.4	Cfb
4	Borås	57.76 N	12.94 E	135.0	Cfb
5	Jönköping	57.80 N	14.13 E	180.8	Dfb
6	Stockholm	59.34 N	18.05 E	43.1	Dfb
7	Uppsala	59.85 N	17.63 E	23.4	Dfb
8	Falun	60.62 N	15.66 E	80.0	Dfb
9	Särna	61.69 N	13.18 E	425.0	Dfb
10	Sveg	62.01 N	14.30 E	360.0	Dfc
11	Härnösand	62.64 N	17.92 E	90.0	Dfc
12	Piteå	65.26 N	21.47 E	8.3	Dfc
13	Haparanda	65.82 N	24.11 E	13.2	Dfc
14	Kiruna	67.85 N	20.25 E	500.0	Dfc

[1] https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer/

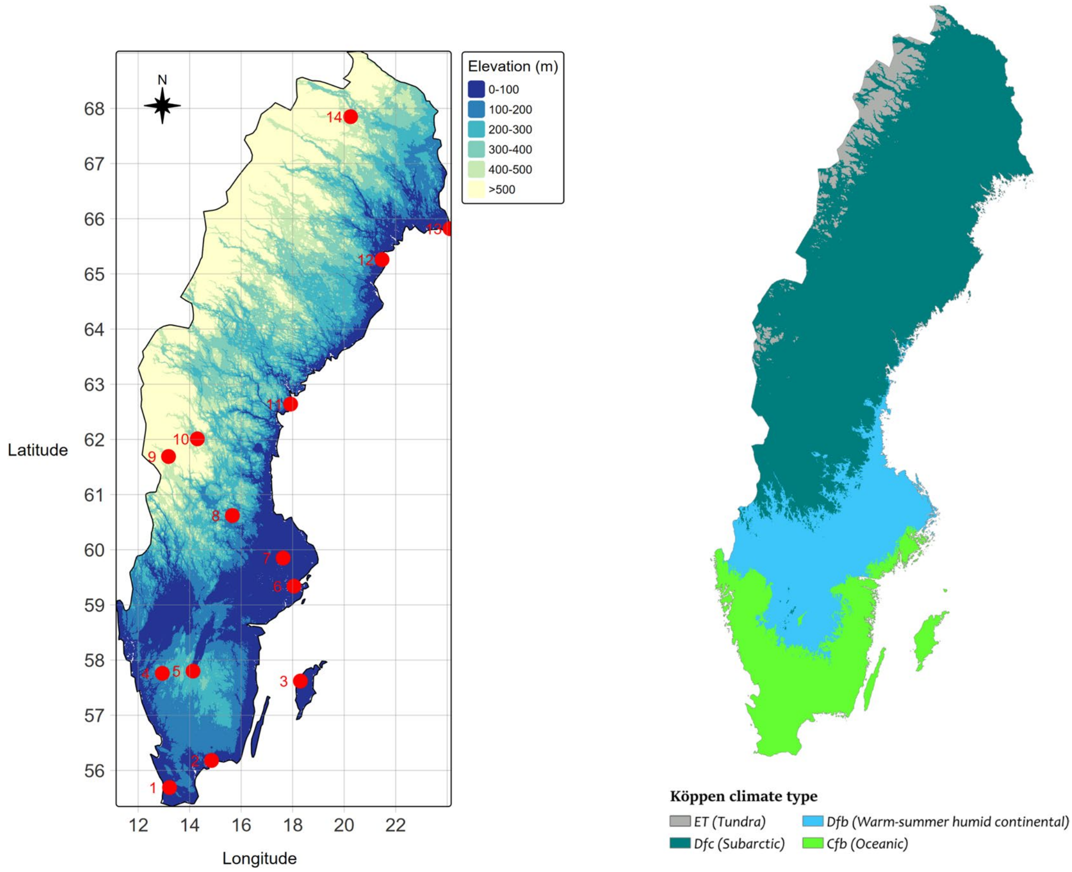

Fig. 1 Left panel: Locations of sites of measurements along with elevations. Right panel: Climate zones of Sweden, following the Köppen–Geiger climate classification. Source: Adam Peterson, Wikimedia Commons, licensed under CC-BY-SA-4.0

Table 2 Summary statistics of monthly rainfall at the stations investigated. SD: Standard deviation. CV: Coefficient of variation

Station	Location	Mean (mm)	Median (mm)	IQR (mm)	SD (mm)	CV (%)
1	Lund	53.0	47.7	37.3	30.9	58.3
2	Karlshamn	48.4	43.7	36.8	28.1	58.1
3	Visby	43.7	38.2	35.0	27.2	62.3
4	Borås	80.5	72.7	60.0	47.6	59.1
5	Jönköping	46.0	41.3	36.5	28.8	62.6
6	Stockholm	46.0	40.0	34.4	29.6	64.2
7	Uppsala	46.0	40.9	34.3	27.9	60.6
8	Falun	48.0	41.9	37.8	31.1	64.9
9	Särna	50.9	43.1	43.5	34.3	67.3
10	Sveg	47.0	38.9	41.4	33.2	70.6
11	Härnösand	57.6	48.8	48.3	39.3	68.3
12	Piteå	44.3	38.9	37.3	30.4	68.6
13	Haparanda	46.8	41.7	36.7	28.0	59.9
14	Kiruna	41.4	30.9	38.7	34.2	82.7

properties of the fitted Tweedie distributions. This is an extension to the approach taken by Dunn (2004), Section 3.5, where statistical properties of daily rainfall in Melbourne was presented (as an illustration) for the month of April. As an important step, for each recording site, the *proportion of zeros* will be computed month by month over the time period. It might be of interest to explore possible relations between this quantity and the fitted index parameter in the Tweedie distributions.

3 Methodology

3.1 The tweedie distribution

The Tweedie distribution is found as a member of the EDM family of distributions (Jørgensen 1987). Within the EDM family of distributions, the probability-density function has the form

$$f(y;\mu,\phi) = a(y,\phi)\exp\left[\frac{1}{\phi}(y\theta - \kappa(\theta))\right]$$

where μ is the mean of the distribution, the dispersion parameter $\phi > 0$ and θ and $\kappa(\theta)$ are known entities. From the theory for EDMs, for a random variable Y we have that for the mean, $\mathrm{E}[Y]$, and variance, $\mathrm{V}[Y]$,

$$\mathrm{E}[Y] = \mu = \frac{\mathrm{d}}{\mathrm{d}\theta}\kappa(\theta), \quad \mathrm{V}[Y] = \phi\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\kappa(\theta).$$

Moreover, a variance function can be defined as $V(\mu) = \mathrm{d}\mu/\mathrm{d}\theta$, and it follows from theory that $\mathrm{V}[Y] = \phi V(\mu)$.

Now, consider variance functions on the form $V(\mu) = \mu^p$, with p any real number except $0 < p < 1$. An EDM with such a variance function is called a Tweedie distribution, and p is usually denoted as an index parameter. Some standard distributions can be found as special cases. For instance, $p = 0$ corresponds to the normal distribution. As a further example we find that as $p \to 2$ from below, the distribution tends to a gamma distribution. This is of interest in the present paper.

Statistical inference could be carried out by maximum likelihood techniques (Dunn and Smyth 2018), in particular profile likelihood. Such frameworks are implemented in a software package in R, called `tweedie`, see Dunn (2022).

3.2 Hierarchical clustering

With hierarchical clustering is meant an algorithm that groups similar objects into groups called clusters; for an overview in the time-series context, see Liao (2005). In this paper, interest is in whether the estimated values of index parameter p tend to follow geography, in terms of latitude or Köppen–Geiger zones, and hence how groups would cluster.

Crucial in clustering is how to evaluate in a general sense similarity between objects; in a more mathematical setting, the distances between objects. Several distances have been suggested in the literature and are often implemented in software. In this study, the R package `TSclust` was employed (Montero and Vilar 2014). Note that our objects in a statistical sense are time series; after the estimation procedures, we face monthly observations of the estimated index parameter for a number of stations.

We employed two distance measures, as implemented in `TSclust`: (1) Fréchet distance, (2) A distance based on autocorrelation. The former takes into account the ordering of the observations, and does not just treat the series as two point sets (like e.g. the occasionally encountered Minkowski distance) The latter is based on estimated autocorrelation vectors. For details, see Montero and Vilar (2014); Galeano and Peña (2000). Both these measures can be said to be model-free approaches, as opposed to others based on e.g. ARIMA processes.

4 Results

4.1 Assessment of numerical work

Profile likelihood is carried out in the fitting routine from the package `tweedie`. Of special concern is the index parameter p. as mentioned earlier. In Fig. 2, the seasonal pattern for p is shown. Monthly mean values of p are displayed, averaged over all 14 locations considered, with an overall average of $p = 1.56$. We note a peak over time in early autumn.

For assessment of the obtained distributions, quantile residuals can be employed (Dunn and Smyth 2018). To visualise these, a QQ plot is a useful tool (Venables and Ripley 2002). Here, residuals are plotted with sample quantiles on the ordinate versus theoretical quantiles (of a standard normal distribution) on the abscissa. In the ideal situation, the pattern of points should follow a straight line (the identity line). As an illustration, the quantile residuals for Station 3 (Visby) are shown in Fig. 3 for September months. The

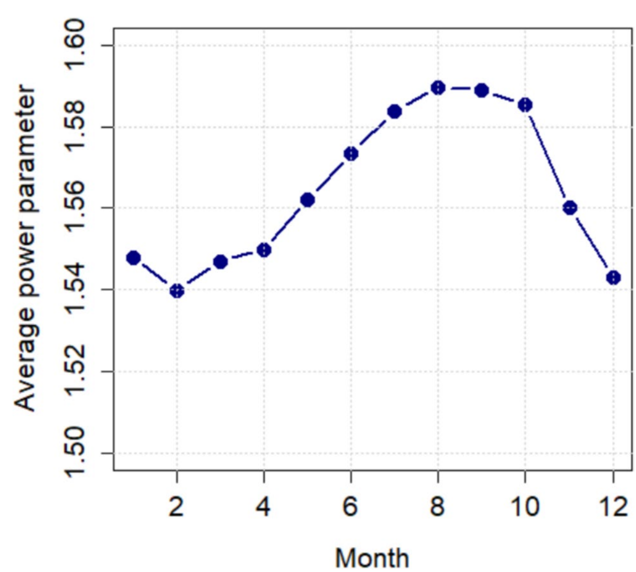

Fig. 2 Monthly average values of the index parameter p over the 14 sites

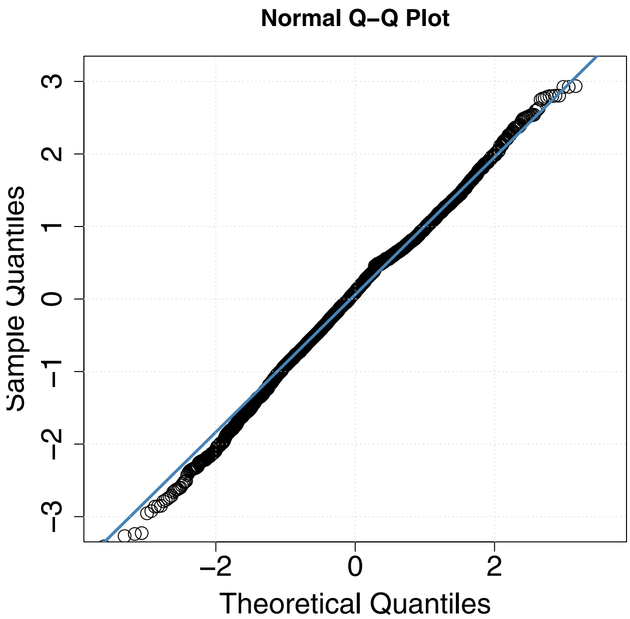

Fig. 3 QQ plot of residuals after fitting of Tweedie distribution (Station 3 (Visby), September months)

model fit is highly satisfactory. Similar features are found for other sites,

4.2 Seasonal and regional behaviour

The estimated monthly values of the index parameter p are displayed in Fig. 4 in the form of a spaghetti plot for the 14 stations. In general, there is a tendency to a concave shape over the year (cf. the averages shown in Fig. 2). A closer examination (not presented in this paper) shows a close connection to the pattern of precipitation; the higher the precipitation, the higher the value of the index parameter. The peak value usually appears in autumn.

Turning again to Fig. 4, the lower curves (also denoted by lower numbers) belong to locations in the southern part of Sweden. The correlation between annual median values and latitude for the investigated stations is positive and statistically significant ($r = 0.62$, $p = 0.02$). A model fitted with multiple regression with the latitude and elevation of the station as covariates resulted again in latitude being significant, but not elevation.

We now focus on the proportion of zeros, starting with simple descriptive statistics; cf. Section 2. A motivation for the use of Tweedie distribution is the (possible) presence of days with zero precipitation. Indeed, such days are present during all months, at all sites. As an example, we demonstrate the monthly behaviour for Station 3 (Visby), see Fig. 5. We note that the proportion is considerably higher than zero regardless of month, and that this proportion is at its highest in the month of May. The same pattern follows for all stations, regardless of location and climate zone.

Furthermore, it might be of interest to examine possible relationships between the proportion of days with zero precipitation and the value of the index parameter. However, estimated correlations of these two quantities differ in magnitude and sign. These vary between -0.87 (Station 4, Borås) and 0.60 (Station 3, Visby). Moreover, further exploration of possible relationship between latitude and the mentioned correlation was carried out, resulting in no significant correlation between these quantities.

4.3 Clustering

We here present the results using hierarchical clustering. Two distance measures were applied: Fréchet dissimilarity index and ACF dissimilarity index (R: `TSclust::diss`), cf. Section 3.2. Moreover, the number of clusters to cluster to were chosen as 2 and 3, respectively, for the sake of easier interpretation.

In Fig. 6, left plots, results for the Fréchet index are displayed. We note that for the two-class scenario (upper panel), stations 1, 2 and 5 form one class. These are located in the south. In the three-class scenario, still stations 1, 2 and 5 form one class.

Results for the ACF dissimilarity index are shown in Fig. 6, right plots. Again for the two-class scenario in the upper panel, stations in southern parts (lower station numbers) seem to cluster. Turning to a three-class scenario, the difference is minor; Station 13 (Haparanda), in the very north of Sweden, forms a group of its own.

5 Summary and discussion

In several situations in e.g. agriculture and water management, there is an interest in statistical modelling of the amount of precipitation. A problem is that precipitation is a continuous random outcome with exact zeros. Use of the Tweedie distributions can tackle this, as shown by Dunn (2004) when applied on data from Australia. In this paper, we extended the research, by studying several locations in Sweden and in addition examining the crucial index parameter in the Tweedie distribution with respect to location, performing a clustering analysis.

The findings indicate that the Tweedie distribution provides a good fit to the precipitation data, cf. Figure 3. Moreover, the estimates of the index parameter are interesting to reflect upon. We studied variability over months, as well as geographically. For the 14 stations considered, spread out over Sweden, an overall mean is 1.6 (with two decimals,

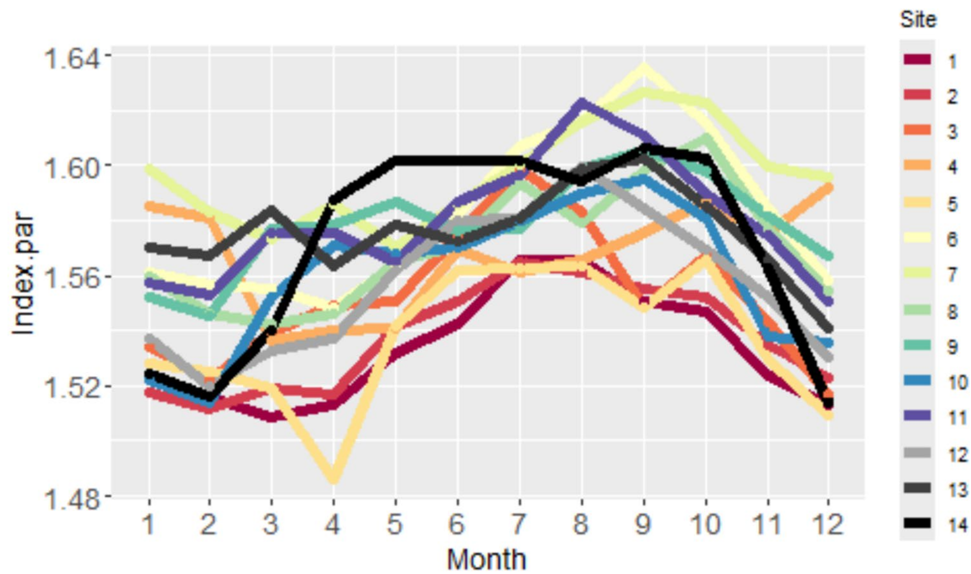

Fig. 4 Monthly values (January to December) of index parameter *p*, based on daily precipitation at 14 locations

1.56). This could be compared to Australian rainfall stations discussed by Hasan and Dunn (2011), where for most stations the value 1.6 was found, with higher values at stations along the coastline. The same features could be viewed here, cf. Figure 4, where two out of the three highest values are attended at coastline (Stations 6 and 11) and similarly two out of the three lowest values are not located along the coast (Stations 1 and 5).

Moreover, through hierarchical clustering, the influence of latitude was investigated. The interpretation is here not clear-cut but depends on chosen dissimilarity index. However, in a scenario with two clusters, the option with Fréchet index results in southern stations clustering to one group.

While this study provides valuable insight on the use of the Tweedie distribution, it is obviously limited by the number of stations considered. However, these were chosen to represent various climatologies of Sweden, not a particular region. Further, additional options for the clustering analysis could have been considered, but since the result was somewhat ambiguous already with two alternatives, the author did not go further. Additional studies could investigate alternative choices.

Another option for modelling continuous positive outcomes that may include exact zeros is given by the Zero Adjusted Gamma (ZAGA) distribution, found in the framework of GAMLSS (Generalized Additive Models for Location, Scale and Shape), see Rigby et al. (2020). This is useful in modelling scenarios where data predominantly follow a gamma distribution but have an excess of zero observations. The ZAGA distribution has been used to fit observed daily streamflow (Rashid and Beecham 2019). However, in the present paper we focused on properties of the Tweedie distribution, in particular the behaviour of its index parameter. In a future publication, the main focus could be a comparison of the Tweedie approach and usage of the ZAGA distribution.

Overall, this study advances the understanding of use of the Tweedie distribution to model monthly precipitation. Most published studies have concerned Australia, and the results presented here gives a first account of the modelling

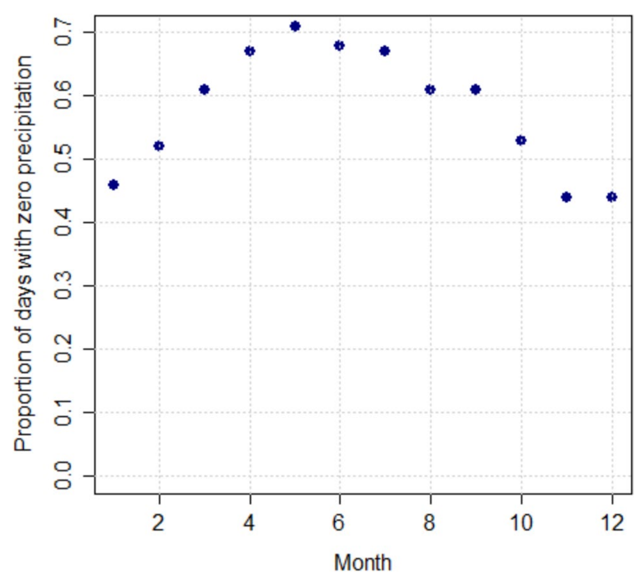

Fig. 5 Monthly values of proportion of days with zero precipitation, Station 3 (Visby)

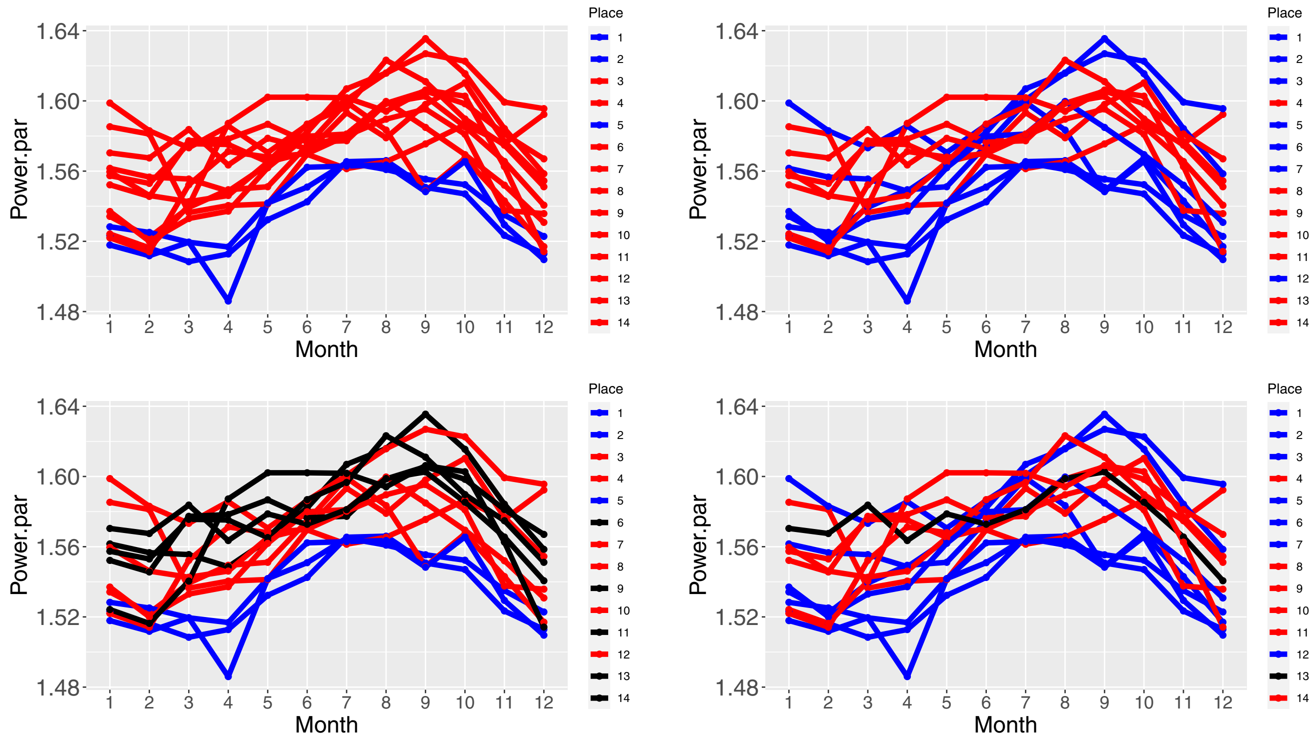

Fig. 6 Left: Clustering based on Fréchet dissimilarity index. Right: Clustering based on an ACF dissimilarity index. Upper plots: Clustering to two groups. Lower plots: Clustering to three groups

possibilites for Northern Europe. The knowledge of rainfall models helps us to develop further modelling and applications in water management, e.g. studies of soil water or forecasts of droughts and floods.

Acknowledgements The author is grateful for comments and remarks from the anonymous reviewers.

Author Contributions One sole author.

Funding Open access funding provided by Swedish University of Agricultural Sciences. Swedish University of Agricultural Sciences; Faculty of Natural Resources and Agricultural Sciences.

Data Availability Data is available on request from the author.

Declarations

Competing interests The author declares no competing interests.

References

Abbasi F, Bazgeer S, Kalehbasti PR, Oskoue EA, Haghighat M, Kalehbasti PR (2022) New climatic zones in Iran: a comparative study of different empirical methods and clustering technique. Theoretical Appl Climatol 147:47–61. https://doi.org/10.1007/s00704-021-03785-9

Beniston M, van der Wiel K, Bintanja R (2021) Contribution of climatic changes in mean and variability to monthly temperature and precipitation extremes. Commun Earth Environ 2:1–11. https://doi.org/10.1038/s43247-020-00077-4

Broc N, Giusti C (2007) Autour du Traité de Géographie physique d'Emmanuel de Martonne : du vocabulaire géographique aux théories en géomorphologie. Géomorphologie: relief, processus, environment 2/2007. https://doi.org/10.4000/geomorphologie.921

De Martonne E (1941) Traité de Géographie Physique. A Colin., Paris, 6th ed

Dunn PK (2004) Occurrence and quantity of precipitation can be modelled simultaneously. Int J Climatol 23:1231–1239. https://doi.org/10.1002/joc.1063

Dunn PK, White N (2005) Power-variance models for modelling rainfall. In: 20th International workshop on statistical modelling, 10 - 15 July 2005, Sydney, Australia

Dunn PK, Smyth GK (2018) Generalized Linear Models with Examples in R. Springer-Verlag

Dunn PK (2022) Tweedie: Evaluation of Tweedie Exponential Family Models. R package R package version 2(3):5

Feddema JJ (2005) A revised Thornthwaite-type global climate classification. Phys Geogr 26:442–466. https://doi.org/10.2747/0272-3646.26.6.442
Galeano P, Peña D (2000) Multivariate Analysis in Vector Time Series. Resenhas do Instituto de Matemática e Estatística da Universidade de Sao Paulo 4(4):383–403
NASA Goddard Space Flight Center (2016) Understanding Earth. What' s Up with Precipitation?
González-Pérez A, Álvarez-Esteban R, Penas A, del Río S (2022) Analysis of recent rainfall trends and links to teleconnection patterns in California (U.S.). J Hydrol 612:128211. https://doi.org/10.1016/j.jhydrol.2022.128211
Hasan MM, Dunn PK (2011) Two Tweedie distributions that are near-optimal for modelling monthly rainfall in Australia. Int J Climatol 31:1389–1397. https://doi.org/10.1002/joc.2162
Jørgensen B (1987) Exponential dispersion models. J Royal Stat Soc Series B (Methodological) 49:127–145. https://doi.org/10.1111/j.2517-6161.1987.tb01685.x
Köppen W (1900) Versucheiner Klassifikation der Klimate, corzugsweisenach ihren Beziehungen zur Pflanzenwelt. Geogr Z 6:657–679
Liao TW (2005) Clustering of time series data : A survey. Pattern Recognit 38(11):1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025
Martinez-Villalobos C, Neelin JD (2019) Why do precipitation intensities tend to follow gamma distributions? J Atmosph Sci 76:3611–3631. https://doi.org/10.1175/JAS-D-18-0343.1
Montero P, Vilar JA (2014) TSclust: An R package for time series clustering. J Stat Softw 62:1–43
Rashid MM, Beecham S (2019) Simulation of streamflow with statistically downscaled daily rainfall using a hybrid of wavelet and GAMLSS models. Hydrological Sci J 64:1327–1339. https://doi.org/10.1080/02626667.2019.1630742
Rigby RA, Stasinopoulos, DM, Heller, GZ, De Bastiani, F (2020) Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R. CRC Press
Stern RD, Coe R (1984) A model fitting analysis of daily rainfall data. J R Stat Soc Ser A 147:1–34
Tamm OS (1959) Studier över klimatets humiditet i Sverige. Bulletin of the Royal School of Forestry, Stockholm, Sweden, nr 32
Tennekes M (2018) tmap: Thematic Maps in R. J Stat Softw 84:1–39. https://doi.org/10.18637/jss.v084.i06
Venables WN, Ripley BD (2002) Modern Applied Statistics with S (4th ed). Springer-Verlag
Yunus RM, Hasan MM, Razak NA, Zubairi YZ, Dunn PK (2017) Modelling daily rainfall with climatological predictors: Poisson-gamma generalized linear modelling approach. Int J Climatol 37:1391–1399. https://doi.org/10.1002/joc.4784