High frequency rainfall data disaggregation with a random cascade model

Identifying regional differences in hyetographs in Sweden

Louis Rulewski Stenberg

ABSTRACT

High frequency rainfall data disaggregation with a random cascade model - Identifying regional differences in hyetographs in Sweden

Louis Rulewski Stenberg

The field of urban hydrology is in need of high temporal resolution data series in order to effectively model and analyse existing and future trends in extreme precipitation. When high resolution data sets are, for any number of reasons, not available for a given location, the technique of disaggregation using a random cascade model can be applied. Previous studies have demonstrated the relevance of random cascades in the context of rainfall data disaggregation with temporal resolutions usually down to 1 hour. In this study, an attempt at disaggregation to a resolution of 1 minute was made. Using newly disaggregated rainfall data for different regions in Sweden, the possibility of clustering rain events into separate regional hyetographs was investigated.

The random cascade model was calibrated using existing municipal rainfall data with a temporal resolution of 1 minute, in order to disaggregate continuous 15 minutes data series provided by the Swedish Meteorological and Hydrological Institute (SMHI). The disaggregation process was then performed in multiple stochastic realisations, in order to correct the uncertainties inherent to the random cascade model. The disaggregation results were assessed by comparing them with calibration data: two main rainfall parameters, EV and ED, were analysed by determining their behaviours and distribution. The possibility of transfering calibration parameters from one station to another was also assessed in a similar manner, again by studying EV & ED for different scenarios. Finally, hyetographs were clustered, compared and contrasted, in order to ascertain previously theorized differences between regions.

This research showed the feasibility of applying a random cascade model to very high temporal resolutions in Sweden, while replicating rainfall characteristics from the calibration data quite well. The analysis of the spatial transferability of calibration parameters yielded inconclusive results, as rainfall characteristics were preserved in some cases but failed in others. Lastly, distinct regional differences in hyetographs were noted, but no clear conclusions could be drawn owing to the delimitations of this study.

Keywords: Random cascade model, disaggregation, high frequency, rainfall data, hyetograph, spatial transferability, regional differences, Swedish climate, clustering.

Department of Earth Sciences, Program for Air, Water and Landscape Science, Uppsala University, Villavägen 16, SE-75236 Uppsala, Sweden.

REFERAT

Disaggregering av högupplösta regndata med en slumpmässig kaskadmodell - Identifierande av regionala skillnader i hyetografer i Sverige
Louis Rulewski Stenberg

Inom småskalig hydrologisk modellering finns det idag ett behov av dataserier med hög tidsupplösning för att effektivt kunna modellera och analysera både aktuella och kommande trender hos extrema regnhändelser. När högupplösta dataserier är otillgängliga vid en önskad mätplats kan disaggregering med hjälp av en slumpmässig kaskadmodell tillämpas. Tidigare forskning har visat att kaskadmodeller är användbara för disaggregering av regndata med en tidsupplösning av 1 timme. I denna studie disaggregerades dataserier med syftet att uppnå en tidsupplösning av av 1 minut. För att kunna analysera eventuella skillnader mellan regioner klustrades även hyetografer med de framtagna dataserierna.

Den slumpmässiga kaskadmodellen kalibrerades med befintlig kommunal data med en tidsupplösning på 1 minut, för att sedan kunna disaggregera 15 minuters data från SMHIs databaser. Disaggregeringen genomfördes i ett antal olika stokastiska realisationer för att kunna ta hänsyn till, och korrigera, de inneboende osäkerheterna i den slumpmässiga kaskadmodellen. Disaggregeringsresultaten bedömdes genom en jämförelse med kalibreringsdata: två regnegenskaper, regnvaraktighet (ED) och regnvolym (EV), analyserades för att kunna bestämma deras fördelningar och beteenden. Kalibreringsparametrarnas överförbarhet analyserades också med hjälp av ED & EV för olika scenarier. Slutligen klustrades hyetografer för att fastställa potentiella skillnader mellan regioner.

Studien påvisade möjligheten att använda en slumpmässig kaskadmodell till höga tidsupplösningar i Sverige. Modellen lyckades återskapa regnegenskaper från kalibreringsdata vid disaggregeringen. Möjligheten att överföra kalibreringsparametrar från en station till en annan visade sig dock inte vara helt övertygande: regnegenskaper återskapades endast i vissa fall, men inte i samtliga. Slutligen konstaterades regionala skillnader i hyetografer, men tydliga slutsatser kunde inte dras på grund av underliggande begränsningar med studien.

Nyckelord: Slumpmässig kaskadmodell, disaggregering, högfrekvens, regndata, hyetograf, rumslig överförbarhet, regionala skillnader, Svenskt klimat, klustring.

Institutionen för geovetenskaper, Luft- vatten- och landskapslära, Uppsala Universitet, Villavägen 16, 75236 Uppsala, Sverige.

PREFACE

The present master's thesis represents the culmination of my studies within the Master's Programme in Environmental and Water Engineering at Uppsala Universty and the Swedish University of Agricultural Science (SLU). It has been conducted in collaboration with Tyréns AB, under the supervision of Johan Kjellin. Gabriele Messori, senior lecturer & associate professor at the Department of Earth Sciences, Program for Air, Water and Landscape Sciences; Meteorology, has served as subject reader. Monica Mårtensson, lecturer at the same department, was my examiner.

I would very much like to thank Johan Kjellin for giving me the opportunity of studying a new and exciting discipline, as well as providing inspiration and ideas throughout. A warm thank you to Jonas Olsson at SMHI, whose guidance and ideas helped shape the present study (despite his full schedule), for allowing me to use the random cascade model central to this study. I also wish to thank Gabriele Messori for his feedback and insight during the entire project. I am also grateful to Jimmy Olsson and everyone at the Tyréns office in Uppsala for welcoming me during these trying times.

A big thank you to my family for their help and support. I finally extend my deepest gratitude to Elin Haapaniemi, stuck with me at the home office for the entire semester, for her support and input.

Uppsala, February 2021.
Louis Rulewski Stenberg

IV

POPULÄRVETENSKAPLIG SAMMANFATTNING

Den globala uppvärmningen och dess konsekvenser har varit en central fråga i samhällsdebatten under de senaste åren. Forskning inom området visar en tydligt ökande trend av extrema regn, både i frekvens och intensitet. Förhöjda flöden på grund av skyfall medför slitage av befintlig VA-infrastruktur och kan i värsta fall leda till ökande översvämningsrisker med uppenbara konsekvenser för samhället. Det är därför av största vikt att forskare inom meteorologi och hydrologi, eller såväl som ingenjörer inom VA-branschen, har tillgång till högkvalitativ regndata med en hög tidsupplösning. Detta för att, med hög noggrannhet, kunna modellera nuvarande och framtida flöden för att kunna anpassa samhällbyggnaden och kunna framtidssäkra svenska städer.

En påvisad metod för att skapa högupplöst regndata på platser med bristande serier är tillämpningen av en så kallad slumpmässig kaskadmodell. Kortfattat bygger metoden på en kaskadprocess där en viss kvantitet - som till exempel regn - förgrenas successivt och omfördelas då tidsupplösningen förändras. Omfördelningen innebär att processen börjar från en storskalig startupplösning och forsätter successivt tills en småskalig slutupplösning nås. I denna modell används en förgrening till två grenar, vilket innebär att från en total regnmängd delas mängden vatten ansluten till första och andra halvan av perioden till två nya grenar, osv. Den totala ursprungliga regnmängden bevaras för varje förgrening. Sådana modeller har används för tidsupplösningar upp till 1 timme: denna studie syftar till att visa om det är möjligt att höja upplösningen till 1 minut för regndata i Sverige. Det är också av intresse att veta om kalibreringsparametrar kan överföras från en stad till en annan: detta kommer också att undersökas i studien.

Ytterligare ett syfte med studien är analysen av möjliga regionala skillnader i empiriska regntyper, så kallade "hyetografer". En hyetograf visar hur ett regn fördelas över tiden, och kan till exempel visa när ett regn brukar ha sin toppintensitet. Idag använder sig SMHI av nationella hyetografer för analysen av regn i Sverige, men ny forskning har antytt regionala skillnader i hyetografer. Denna teori ska besvaras i följande studie.

Resultatet visade att kaskademodellen fungerade relativt bra i Sverige, även vid väldigt höga tidsupplösningar. Samma slutsats kunde inte dras gällande överförbarheten av kalibreringsparametrarna: i vissa fall var det möjligt att använda parametrar från en stad i disaggregeringen för en annan, men inte i samtliga fall. Jämförelsen av hyetografer visade stora skillnader mellan regioner. Slutligen är det viktigt att poängtera att studien har vissa begränsningar på grund av dataserierna som användes: serier som användes för kalibrering av modellen är separerade i tid till serierna som skulle disaggregeras, vilket kan leda till osäkerheter i resultaten. Inga säkerställda slutsatser kunde därav dras beträffande regionala skillnader.

CONTENTS

1 INTRODUCTION

The field of urban hydrology is in constant need of time series with high temporal resolution. A complication with such time series lies in finding data sets that are not only of a high enough temporal resolution, but also exhibit sufficient length to be of use in contemporary modelling. This question has been prevalent for decades: Schilling (1991) for instance, calls for data series of 20 years or longer with a 1 min temporal resolution, to be applied for the modelling of urban drainage systems with a spatial resolution of 1 km^2. More recently, Berne et al. (2004) came to the same conclusions for urban hydrological studies in southern France. In their study, they found that for small catchment areas of 10 ha (quite a significant area in an urban context), urban modelling would require temporal resolutions of 1 min, in line with the assessments made in Schilling (1991).

The need for high temporal resolution data series is highly relevant in Sweden, where 87.4% of the country's citizens inhabited urban areas in 2019, urban areas being defined as human settlements of $\geq$ 200 inhabitants (Statistikmyndigheten, 2019). The scientific community is united in its assessment of the consequences of climate change in the future, and projections show increases in both the intensity and the regularity of extreme rain events in Europe (Madsen et al. (2014), among others). In Sweden, Olsson et al. (2017) estimated that future extreme precipitation volumes would increase by 10 to 40%, rainfall events that would undoubtedly cause strain and potential damage to existing (and future) infrastructures as well lead to higher risks of pluvial flooding in urban areas, with undeniable economical and societal consequences.

In assessing future extreme rain in Sweden, Olsson et al. (2017) found clear regional differences in precipitations patterns, which lead to the subdivision of Sweden into four meteorological regions. The same study highlighted the need for high resolution rainfall time series and called for an expansion of existing short-term precipitation measurements. Hyetographs were clustered on a national level, distributing rain events based on their temporal distribution rather than their geographical location.

While continuous data series exist in Sweden, various gaps in the time series are not unusual (Hernebring, 2006). Hydrodynamic modelling requires continuous series showing precipitation characteristics at a resolution of 5 minutes or lower to be of relevance and actionable (Licznar et al., 2011).

The purpose of the present study is to evaluate these aforementioned regional differences in hyetographs, while simultaneously attempting to create new, high temporal resolution data series for Sweden. The latter will be achieved through the use of a random cascade model, developed specifically for the Swedish climate by J. Olsson (1998). Applying the cascade model to different meteorological regions in Sweden, existing 15 min rainfall data series will be disaggregated to a higher resolution of 1 min.

The application of random cascade models for the purpose of disaggregating rainfall time series has been tested and proven in multiple studies, see, for instance, Hershenhorn and Woolhiser (1987), Schertzer and Lovejoy (1987), and Over and Gupta (1994), J. Olsson (1998) and J. Olsson (2012), Menabde and Sivapalan (2000) and Güntner et al. (2001), to name a few. The

model used in this study has historically only been applied to time scales ranging between 1 week to 1 hour. Here, the possibility of going even further - to a temporal resolution of 1 minute - will be assessed.

1.1   Aims & objectives

As mentioned above, hydraulic modelling of extreme flows, infiltration rates or even flooding using high resolution time series are necessary in order to future-proof existing infrastructures such as storm drainage systems. Also, while earlier studies have shown regional differences in precipitations in Sweden, the use of regionalized hyetographs is a new approach to this issue. Therefore, following thesis has in essence two overarching objectives:

- To explore the feasibility of disaggregating data down to very high temporal resolutions (1 minute) using an existing random cascade model developed by J. Olsson (1998) and improved upon by Güntner et al. (2001).

- To verify previously theorized regional differences in hyetographs in Sweden.

The latter point, hinted at by Olsson et al. (2017), will be achieved by clustering rain data series obtained through disaggregation into hyetographs by applying the K-means clustering method with five clusters.

In order to achieve these goals, this study will attempt to answer the following questions:

- Can a random cascade model produce reliable and statistically significant rain data at high temporal resolutions?

- Is calibration data from one region, used in the random cascade model, transferable to a different region? Are calibration parameters applicable between different regions?

- Are there any differences in hyetographs in Sweden from one region to another, or is the current method of national hyetographs still relevant?

2 BACKGROUND

2.1 Rainfall data collection

Tipping bucket rain gauges

One of the most commonly used automatic rainfall measuring device in the world is the tipping bucket. As its name suggests, the device is equipped with two container: rain is automatically distributed between the them by rotating, or "tipping" around an axis. When a certain rain quantity is reached in one bucket (usually between 0.1 and 0.2 mm), the bucket tips and rainfall is collected in the other. The resulting number of tippings per time unit is a measure of rainfall intensity (SMHI, 2018).

An important shortfall of these rain gauges shows in colder climates, as precipitation as snow needs to melt before each tipping. This is the main reason SMHI doesn't use this method for rainfall data collection in Sweden, as northern regions experience heavy snowfall during the winter months (SMHI, 2018). However, as SMHI points out, municipalities and other institutions may use tipping buckets, which is the case for the calibration data used in this study. The use of tipping buckets as a rainfall gauge in participating cities is one reason only three could be used for calibration: missing data points for winter months skewed whole time series.

Automatic precipitation measurements

In this study, calibration data provided from municipal measurements were collected with tipping-bucket rain gauges, while SMHI's 15 min data was was gathered by automatic stations.

The devices used in SMHI's network throughout Sweden are weather precipitation rain gauge supplied by Geonor, for real-time rain monitoring. A collecting bucket is suspended on wires kept in pendulum by a magnet. Changes in frequency as precipitation is collected are then translated into weight, and thereby precipitation amounts collected (SMHI, 2018). The rain gauge is equipped with wind shields and, importantly, uses an antifreeze to melt solid precipitations, and therefore doesn't require any additional heating (GEONOR, n.d.).

2.2 Defining rain events & extremes

As mentioned earlier, for the purpose of ease of use and for the sake of continuity, the same data and definitions used in Hernebring (2006) as well as in Olsson et al. (2017) are used in this study. Rain events are quantitatively described by the following parameters:

- *Dry period length*. In other words the duration during which no precipitation occurs so that the next measurements are counted as a new rain event. In this study, 1 hour between each event is used.

- *Intensity boundaries*. The rain quantities required to find the start and end of a given event. In this study, 1 mm/min is used.

- *Lowest intensity*. Lower bound (average over a given event's duration) required in this study: 0.1 mm/h.

Extreme rainfalls are defined by SMHI as "rain events with at least 1 mm/min or 50 mm/h" (SMHI, 2017). This definition is reflected in the rain parameters listed above, and applied in this study.

2.3   Disaggregation using random cascade models

Emerging in the late 1960s to early 1970s for the modeling and analysis of turbulence (Mandelbrot, 1974), and leaping forward during the 1990s, data disaggregation using cascade models has since been shown to be a valid method to study rainfall time series, as seen in a multitude of studies. Applications of rainfall disaggregation are varied and profuse in the fields of hydrology. Jebari et al. (2012), for example, used rainfall disaggregation as a tool for soil loss estimation, by generating highly detailed rainfall data. A lack of high resolution data series at specific locations makes the use of stochastic models to create new series or lengthen existing ones of high interest. An important aspect that needs addressing is the quality of obtained data series: statistical properties need to be preserved as expressed for instance in Menabde et al. (1999) or Molnar and Burlando (2005), among others.

A seminal study in the field of temporal rainfall disaggregation in Sweden is presented in J. Olsson (1998), where a new model was used for disaggregation of rainfall data from 16 hours to 1 hour. Combining empirical characteristics of temporal rainfall with cascade scaling, the model assumed a dependency between the cascade generator (see Section 3.3 below) and rainfall volume & position in the rain progression. The model reproduced both the scaling behaviour of the data as well as the intermittency of rainfall characteristics, as discussed in Güntner et al. (2001).

Scale invariance

The model developed by J. Olsson (1998) for rainfall disaggregation differs from earlier models in its exact conservation of mass (in the case of rainfall analyses, rainfall volume) between each cascade level. Other cascade models have been tested in many different studies: Schertzer and Lovejoy (1987) for example, showed the scaling and intermittency of rain properties in a cascade process. Such cascade processes, where volume is only preserved on average, are called *canonical* cascade models. The random cascade model, conserving mass exactly, is therefore understood to display *micro-canonical* properties (not unlike micro-canonical thermodynamic ensembles in statistical mechanics). An important aspect of micro-canonical cascade is the assumption that the statistical properties of the model, i.e. the probabilities P and the probability distribution $W_{x/x}$, are constant over all time scales (each cascade level): they are scale-invariant (J. Olsson, 1998; Güntner et al., 2001). This property is in fact a major motivation behind the development of micro-canonical random cascade models (Licznar et al., 2011).

However, while finding similar scaling behaviours shown above and confirming the scale invariance of some structures in rainfall data series over different scales, de Lima (1998) found it difficult to determine the scaling range for high-resolution data series. Indeed, breaks in the scaling were found, which they assigned to measuring devices unable to capture extremes in

rainfall intensities, both high and low.

Timescales

J. Olsson (1998) and Güntner et al. (2001) demonstrated the practical applicability of a random cascade model for rainfall disaggregation in southern Sweden for timescales down to 1 hour. Furthermore, rainfall disaggregation with random cascade models has already shown clear potential in representing spatial as well as temporal rainfall variability over different time scales. As Gupta and Waymire (1993) point out, cascade models are especially well suited for the analysis of extremes and intermittency in rainfalls and cloudbursts. This was also demonstrated in J. Olsson (1998) and Güntner et al. (2001). It is, however, the first time the latter's model will be tested and applied to such high resolutions: while several studies apply disaggregation to obtain sub-daily rainfall data series, as demonstrated in Bárdossy and Pegram (2016) and Müller-Thomy (2020) for instance, few go under 1 hour, let alone 1 minute.

2.4 Hyetographs: representing rainfall data

One of the principal objectives for this study is the clustering and subsequent comparison of empirical hyetographs between different regions in Sweden. Hyetographs are a convenient method for studying rainfall, showing the relationship between rain intensity and rain duration. Rainfall intensities are in this case plotted against time to represent a rains behaviour over time. The methodology behind the clustering to hyetographs in this study is detailed more thoroughly in Section 3.4 below. An example of a hyetographs is shown in Figure 2.1, created with the same rainfall data used in Olsson et al. (2017).
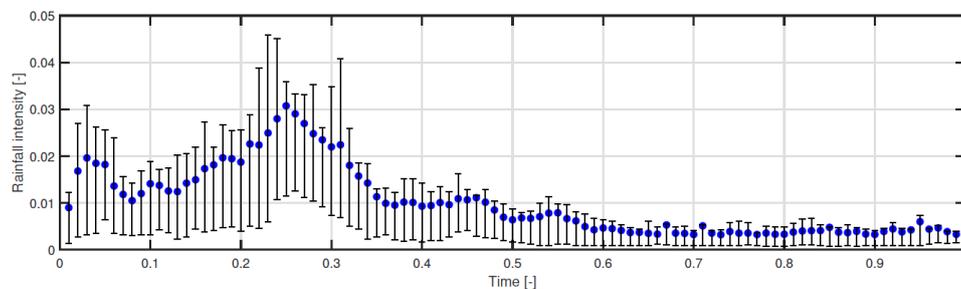
Figure 2.1. Hyetograph created by Olsson et al. (2017) for longer rain events ($\leq$ 90 min). Blue dots show the mean intensities for each time step, whiskers the 25th & 75th percentiles.

2.5 Regional differences in hyetographs

In her Master's thesis, Litsmark (2020), building on Olsson et al. (2017) among others, compared hyetograph distributions with the geographical location of rain gauges and found certain regional differences. These regional tendencies in Sweden were found based on data with 5 min resolution: their findings called for a deeper investigation and analysis of these trends using higher resolution data, which is one purpose of this study. Moreover, definitive conclusions

couldn't be reached on the grounds of too few rain events, hence a need to create more data. The hyetograph distributions found by Litsmark (2020) are shown in Table G.1 in Appendix G below.

2.6 Municipal data series

Extreme rain events are a direct cause to flooding, and especially so in urban areas where water-proof surfaces make infiltration impossible. High resolution rainfall data has therefore been of high demand in the last decades for the proper dimensioning of sewage and stormwater system, in a effort to minimize strains on infrastructures, and ultimately urban flooding.

In an effort to provide and compile such data for urban areas in Sweden, Hernebring (2006) conducted a study on 15 municipalities throughout the country. The regional differences found in Hernebring's study were later corroborated by Olsson et al. (2017) and represent the basis for this study.

2.7 Spatial transferability of calibration parameters

Econopouly et al. (1990) studied and proved the feasibility of supplementing data series for certain regions with calibration parameters obtained from different geographical locations. This procedure was done for different data series within the same climatological region in the U.S. (while displaying dissimilar precipitation patterns). The spatial transferability of calibration parameters for a random cascade model similar to the one used in this study was also studied by Güntner et al. (2001), however, mixed results were obtained. For stations located in semi-arid climates, transferability was confirmed while stations in temperate climates (much like southern and central Sweden) displayed noticeable regional differences.

3 METHODS

3.1 Meteorological regional differentiation in Sweden

In a recent study of extreme rainfalls on the behalf of the Swedish Meteorological and Hydrological Institute (SMHI), Olsson et al. (2017), found clear regional patterns in short-term cloudbursts in Sweden. Olsson's analysis yielded four regions: north (N), middle (M), southeast (*SÖ* in Figure 3.1, SE in this study) & southwest (*SV* in Figure 3.1, SW in this study), as shown in Figure 3.1. The same regional division was used in this study in choosing stations to compare after disaggregation and clustering to hyetographs.

Data disaggregation and the subsequent statistical analyses and hyetograph clusterings were performed on a selection of stations across Sweden, as shown in Table 3.1. The grouping of multiple stations together within each region was performed in an attempt to analyze an approximately equal number of rain events.

The random cascade model used in this study is calibrated against data series with the target temporal resolution (1 min). In this case, 1 minute municipal rain data, as collected in Hernebring (2006), was used for model calibration. Three stations were therefore chosen for model calibration: Helsingborg, Malmö (SW) and Växjö (SE), marked in red in Figure 3.1. These three places were selected on the basis of having existing continuous 1 minute municipal rain data, as well as 15 minutes rain data provided by SMHI's measuring stations.
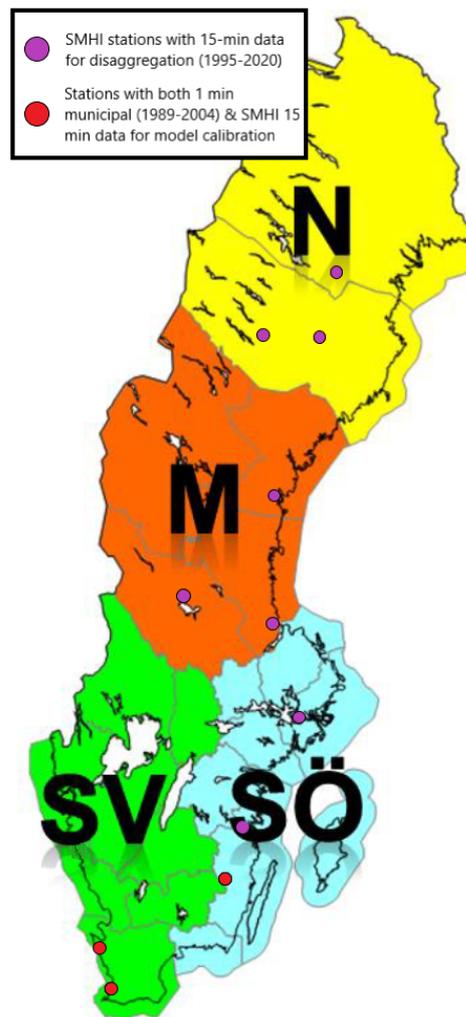

Figure 3.1. Sweden's four meteorological regions as defined by SMHI (Olsson et al., 2017). The different stations used in the following study are juxtaposed: SMHI's recording stations shown in purple, and calibration stations in red.

Table 3.1. Selected stations within each respective region.

Region	Station
SE	Växjö + Horn + Adelsö
SW[i]	Malmö + Helsingborg
M	Gävle + Mora + Sundsvall
N	Vilhelmina + Lycksele + Arvidsjaur

[i] Results from Borås are also presented in Section 4.4 and in Appendix B. However, in an effort to compare similar amounts of rain events between regions, Borås was removed from region SW in some cases.

While other sources of rainfall data exist in Sweden, such as the Swedish Transport Administration's (Trafikverket) or the Swedish University of Agriculture's (SLU) networks of measuring stations, Hernebring's data was chosen in a effort to keep consistency with the earlier studies that inspired the present study, namely Olsson et al. (2017).

3.2 Data pre-processing

3.2.1 Municipal data: tipping bucket rain gauges

Data series created by Hernebring (2006) were specifically chosen for their relative continuity and ease of use (being already reviewed, statistically validated etc.), for the calibration of the random cascade model, displaying the target temporal resolution for disaggregation. However, some pre-processing was needed for the purpose of disaggregation. For instance, data collection changed over time in some municipalites with evolving technologies, leading to different volume resolutions or different measuring methods altogether. It was therefore necessary to select a time interval common to a maximum amount of municipalities with identical volume resolution. Three municipalities matched these criteria while also having SMHI rainfall stations, making them perfect candidates for calibration and validation of the random cascade model and the disaggregation process: Växjö (SE), Malmö (SV) and Helsingborg (SV). Unfortunately, no municipal series matching these criteria were found in regions North and Middle, due to meteorological as well as material constraints (series having irregular rainfall measurements, or there being no SMHI stations, for example).

After the selection detailed above, municipal series were processed following the steps shown in Figure 3.2 below:

- Creation of a continuous time series between 14-Jan-1989 00:00:00 and 17-Aug-2004 23:59:00, with a 1 minute time step with zeroes as rainfall values.

- Concatenation of municipal tipping bucket values and the aforementioned continuous time series.

- Addition of volumes in the case of heavier rainfall recording multiple tippings per minute.

3.2.2 SMHI - meteorological data collection

SMHI collects a range of different meteorological and hydrological data as well as statistical parameters, which are accessible through public databases. For the purpose of disaggregation, 15 minutes rainfall data was extracted from one of those databases (SMHI, 2020). Rainfall measurements are carried out every 15 minutes, where rainfall refers to the accumulated amount of rain for the duration. Data collection from SMHI's automatic station network is ongoing, but time series are limited to the interval 01-Aug-1995 06:00:00 to 01-Jun-2020 06:14:00, when this study began.

The methodology used in the present study for pre-processing data to be disaggregated is straightforward and can be summerized by the following:

- Collection of relevant station data series from SMHI's database.

- Data review: detection and removal of erroneous data and anomalies such as wrong time intervals.

- Addition of zeroes to obtain a continuous series with 15 minutes interval for the duration.

- Division of each value by the measurement device's volume resolution: input data needs to be in multiples of the volume for this specific model in Matlab.

- Division of the whole series into 9 smaller series: eight series of 100 000 points and a ninth of 70 817. This subdivision of the series to be disaggregated is only needed because of a lack of computational power in the present case, and isn't necessary for the actual disaggregation procedure. In a private conversation [1], Jonas Olsson of SMHI did not see any noteworthy drawback with this sectioning before disaggregation.

- Disaggregated sections are then stitched back together into one whole 1-min series over the entire range studied.

A simplified flowchart in Figure 3.2 below shows the pre-processing procedure as a whole in order to create correct input files for the specific random cascade model applied in a Matlab environment for this study.

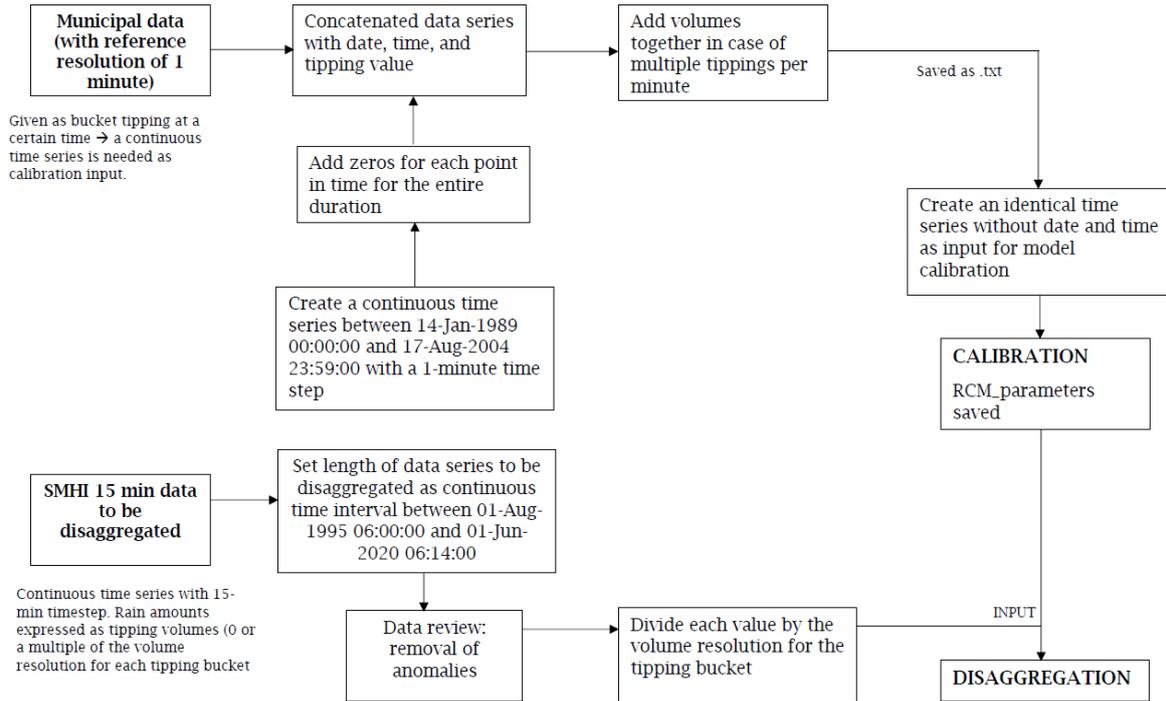[1]Olsson, J. (October 2020). Personal communication.

Figure 3.2. Simplified flow chart showing the various steps taken to prepare for data disaggregation.

3.2.3 Setting data series lengths

As mentioned in Section 1 above, the municipal tipping-bucket series used for model calibration were only available from the 1990s up to 2004, while data collection at SMHI's automatic rainfall measuring grid is ongoing. The possibility of completely different rain events occurring within the different time intervals is acknowledged, and is accounted for in the results. In an effort to correct this, and comparing somewhat similar data sets, disaggregated time series were reduced to the same number of data points as the calibration data when (and only when) validating the disaggregation results against calibration data.

3.3 Micro-canonical temporal rainfall disaggregation with a beta-distributed generator

A practical application of micro-canonical random cascade models within the fields of climatology and hydrology is the disaggregation of continuous rainfall time series. The procedure involves successive branching by which a quantity - rainfall, for instance - is redistributed as the time resolution successively scales down from a starting resolution r_S to a targeted small-scale resolution r_T. An attractive property of micro-canonical cascade models is their ability of exact conservation of mass: in this case, rainfall volumes are conserved at each cascade level.

The model used in this study was developed explicitly for the purpose of temporal rainfall disaggregation by J. Olsson (1998) and subsequently fine-tuned by Güntner et al. (2001). The

model was implemented in a Matlab environment created to attain time scales of ~ 1 week to ~ 1 hour. Further minimal adjustments were made for this study in order to reach and assess the feasibility of even smaller time scales down to 1 minute.
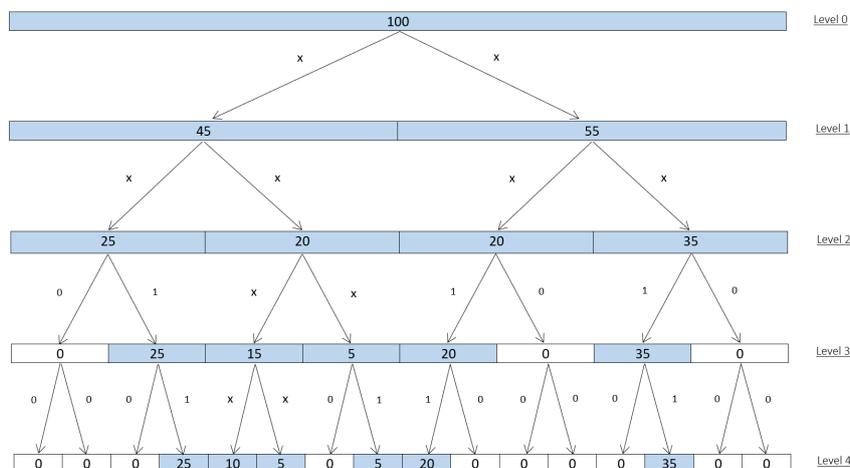

Figure 3.3. Schematic representation of a 1-dimensional cascade model, with a hypothetical starting rainfall of 100 mm, adapted from J. Olsson (1998)

Figure 3.3 above gives a simplified representation of the application of a cascade model used in this study for a similar 4-step disaggregation scheme with an initial rainfall height of 100 mm. The redistribution of rainfall to the first and second half of the period is shown for each level, with wet periods marked in blue and dry periods in white.

A branching number of 2 was used in this study, which means that the total rainfall in a given period with a certain resolution is redistributed between the amounts associated with the first and second halves of the period. Each branching is associated with the multiplicative weights $W_{1,2}$, with $0 \leq W_{1,2} \leq 1$. The weights W_1 & W_2 are therefore used to assess the rainfall volume at the next two finer steps in the branching. As mentioned above, the sum of the weights is always equal to one at each branching level i.e. the rainfall volume is always preserved: reaggregating disaggregated data would yield the original time series (Müller and Haberlandt, 2018).

Combinations of W_1 & W_2 are shown in equation 1, the cascade generator specifying each multiplicative weight:

$$W_1, W_2 = \begin{cases} 0 \text{ and } 1 & \text{with the probability } P(0/1) \\ 1 \text{ and } 0 & \text{with the probability } P(1/0) \\ W_{x/x} \text{ and } 1 - W_{x/x} & \text{with the probability } P(x/x) \end{cases} \quad (1)$$

where $0 < W_{x/x} < 1$, in the event of a x/x-distribution (see Figure 3.3) and $P(0/1) + P(1/0) + P(x/x) = 1$ (Güntner et al., 2001). Together, these probabilities are called P in this study, when applicable. The cascade generator divides the series to be analyzed into non-overlapping intervals (called "boxes" in this study) and repeats the procedure until the aimed resolution is

11

attained. For a total volume V in any given box, two volumes are assigned according to:

$$V_1 = W_1 \text{ x } V, \quad \text{assigned to the 1st half}$$
$$V_2 = W_2 \text{ x } V, \quad \text{assigned to the 2nd half} \tag{2}$$

The volume assignation described in equation 2 above was then repeated in the next resolution doubling and so on and so forth until the desired resolution is reached. The results from an analysis of the probabilities P and of the weight distributions are further explored for this specific study in Section 3.5 below.

The model described by J. Olsson (1998) is parameterized by four parameters as defined in equations 3 through 11.

The probability distribution of $W_{x/x}$ generally follows a theoretical distribution and a probability density function for W, the symmetric single-parameter beta distribution shown in equation 3, as proposed by Menabde and Sivapalan (2000):

$$f(W) = \frac{1}{B(a(r))} W^{a(r)-1}(1-W)^{a(r)-1} \tag{3}$$

where B is the beta function, given by equation 4:

$$B(a(r)) = \int_0^1 x^{(a(r)-1)}(1-x)^{(a(r)-1)} \, dx \tag{4}$$

with $a(r)$, the time resolution dependent shape parameter of the distribution, parametrized by the scaling law:

$$a(r) = a_s r^{-H} \tag{5}$$

where a_s is a shape parameter related to the resolution r_s and H describing how fast the shape parameter $a(r)$ decreases with a decreasing resolution. Both a_s and H are estimated from data in J. Olsson (1998), a novelty in the realm of data disaggregation. The same is done in the present study.

Equation 3 above is widely accepted to describe the probability density function for the cascade generator and used in multiple studies on rainfall disaggregation using random cascade

models, as shown for example in J. Olsson (1998), Menabde and Sivapalan (2000), Molnar and Burlando (2005) or Licznar et al. (2011).

Breakdown coefficients: BDCs

The cascade generator given in equation 2 above is based, for micro-canonical models, on statistical analyses of the rainfall time series. Menabde and Sivapalan (2000) used classical probabilities (as opposed to statistical moment-scaling shown in other studies, see for instance Over and Gupta (1994) and Over and Gupta (1996): they noted that the theoretical distribution of the weights W_1 & W_2 should be identical to the breakdown coefficients (BDCs) in local rainfall. In the model used for this study, relevant parameter estimation was then performed with the help of a "reverse cascade": starting from the highest resolution r_s, adjacent rain volumes are aggregated two by two in a coarse-graining procedure down to the lowest resolution r_L. Said procedure allowed for the measurement of BDCs according to the following (Willems et al., 2012):

$$BDC_{i,r} = \frac{R_{i,r}}{R_{i,r} + R_{i+1,r}} \tag{6}$$

where $R_{i,r}$ is the rainfall total accumulated over the period i at a resolution r.

BDCs correspond in essence to the weights W redistributing a total rainfall quantity at each step of the cascade. In the case of the disaggregation of a certain data series with a starting resolution r_S calibrated against an existing data series with a target resolution r_T, the extraction of BDCs can be used to estimate multiple parameters related to weights behaviour, and the validation of the disaggregation process (Willems et al., 2012). More specifically, two main parameters were estimated using breakdown coefficients: the distribution parameter $a(r)$ (in other words, the variability of the cascade generator) and the intermittency probability $P_{0/w}$ (the probability that one disaggregation interval is dry was used between scales *n* & *n+1*). The distribution of *W* was shown to be scale-dependent using this method by many earlier studies, Molnar and Burlando (2005) for instance. Breakdown coefficient histograms obtained for this study are presented in Section 4.1 below.

Position & volume classes

During a study of rainfall patterns in southern Sweden, J. Olsson (1998) elaborated a new way of using cascade models in order to not only preserve the scaling behaviour of precipitations, but also to take into account both clustering of rainfall as well as zero values. This proved to be needed as the probabilities *P* seemed to show a dependency on these box characteristics.

The solution they proposed to improve existing micro-canonical models was to change the structure of the generator in order for the generator parameters to reflect two wet box attributes,

namely rainfall volume and rainfall position (J. Olsson, 1998). When it comes to rainfall volume, a common assumption is that the probability of rain during both the first and second halves of the time interval, $P(x/x)$, is higher for a period with heavier rainfall than for one with lesser amounts of rain. As $P(x/x)$ is higher for wet boxes between other wet boxes , and $P(1/0)$ lower for boxes at the start of a rain event than for boxes at the end (and vice versa for $P(0/1)$, a division into position classes was suggested in J. Olsson (1998).
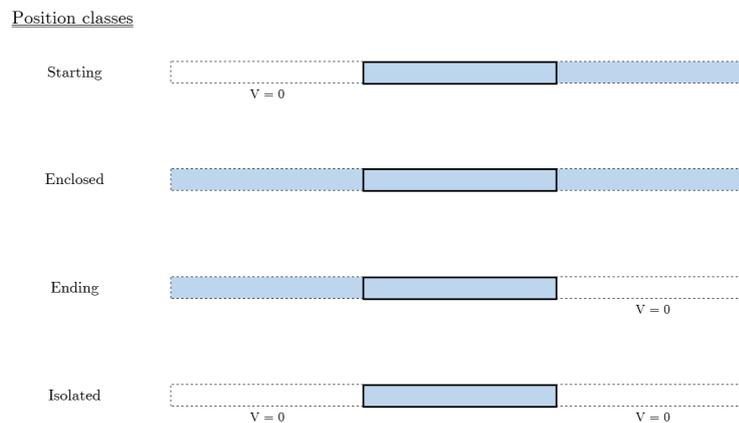

Figure 3.4. Schematic depiction of the four different position classes used in this study. Inspired by Müller-Thomy (2020).

The four position classes proposed in J. Olsson (1998) were also applied to this study, as shown in Figure 3.4 above:

1. **Starting box**: at the start of a sequence, preceded by a dry period and succeeded by a wet one.

2. **Enclosed box**: inside a wet sequence, both preceded and succeeded by wet periods.

3. **Ending box**: at the end of a sequence, preceded by a wet period and followed by a dry one.

4. **Isolated box**: inside a dry sequence, both preceded and succeeded by dry periods.

J. Olsson (1998) proposed using two volume classes, below and above the mean volume, to account for the probabilities' volume dependence. However, using Olsson's own algorithm, Willems et al. (2012) found considerable variations in probabilities with volumes classes, and came to the conclusion that a division into three volume classes gave an even better representation of volume dependence and was therefore justified. The same three classes were used in this study, separated by the 33rd and 67th percentiles.

While variations in probabilities with volume classes are shown in Section 4.1 below, Willems et al. (2012) showed that the probabilities' volume dependence could be grouped and described as a linear function instead of separately, according to:

$$P = \mathrm{int} + \mathrm{slo} * \mathrm{vc} \tag{7}$$

14

where *vc* is the volume class, *int* is the intercept at $vc = 0$ and *slo* is the slope of the linear regression.

Furthermore, they also tested *P*'s dependence on cascade step, showing clear linear relationships between *P* and volume class changes with cascade steps. While *slo* remained constant, *int* varied with a certain regularity. Accordingly, they proposed a modification to their earlier defined variables *slo* and *int*, using a fixed slope slo_m, the averaged slope over all cascade steps using a mean regression line based on equation 7, namely:

$$int_m + slo_m * vc \tag{8}$$

Yet another set of calibration variables to account for describe the dependence of the variable *int* on the cascade step, *cs*. *Int* is then expressed using the linear expression:

$$int = c_1 + c_2 * cs \tag{9}$$

Weighting of *P* values and simplifications

Compared with its first form as layed out by J. Olsson (1998), the current cascade model modifies the weighting of some model parameters. Each *P* value is assigned a weight based on the number of boxes used: *P* becomes more accurately estimated with higher resolutions as more boxes are used in their estimation (Güntner et al., 2001). The distribution of $W_{x/x}$ was also simplified somewhat. Because $W_1 = 1 - W_2$, J. Olsson (1998) produced symmetrical distributions when including both W_1 and W_2. The issue is different depending on position classes. For instance, starting & ending boxes have similar variations for *P*(x/x) and reversed variations for *P*(0/1) & *P*(1/0). In this case, *P*(x/x) can be modelled using equation 7, with int defined by equation 9, as used in this study. *P*(0/1) & *P*(1/0) become independent from the cascade step, leading to their modelling using equation 8, with a fixed slope (J. Olsson, 1998). These probabilities are then estimated as:

$$\begin{aligned} P(0/1) &= 1 - (P(x/x) + P(1/0)) \\ P(1/0) &= 1 - (P(x/x) + P(0/1)) \end{aligned} \tag{10}$$

For isolated and enclosed boxes, J. Olsson (1998) and J. Olsson (2012) considers them equal, and estimates both *P*(0/1) and *P*(1/0) as:

$$P(0/1) = P(1/0) = \frac{(1 - P(x/x))}{2} \tag{11}$$

Güntner et al. (2001) corrected this issue of neglecting asymmetries in the empirical distribution of *W*, in order to avoid the creation of random structures within events, by calculating W_2 as: $W_2 = 1 - W_1$.

Stochastic realisations

Disaggregation with a random cascade model being as it name suggests, random, different results should be obtained for different executions of the algorithm. To correct for this uncertainty, multiple realisations for the same studied data series were performed in this study.

Concerning the amount of stochastic realisations, Müller-Thomy (2020) suggests that 30 realisations covers the uncertainty of the model and accounts the random behaviour of disaggregation. Hyetographs were therefore clustered for Växjö and Helsingborg in this study for two different data sets: a disaggregated rainfall series averaged over 30 realisations as well as the same disaggregated series averaged over 100 realisations. These two scenarios were then compared and contrasted with hyetographs clustered from available 1 minute municipal data series. Results for Växjö are presented in Section 4.3, while hyetographs for Helsingborg are shown in Appendix D.

3.4 Clustering to empirical hyetographs

Olsson et al. (2017) studied the temporal distribution of extreme rain events and cloudbursts in Sweden. They clustered rain events into five predefined groups (called clusters hereafter), after a pre-processing of raw rainfall data. The pre-processing in question entailed among other things a division of rain event into three duration classes: 0-60 min, 60-90 min and 90+ min rains. A similar subdivision was made in this study, while limiting the scope of the study to the 90+ min group only, as mentioned earlier.

Events within a certain duration class were then normalized with regards to ED as well as EV, by substracting the mean and dividing the results with the standard deviation, rendering dimensionless hyetographs. ED range therefore between 0 and 1 while EV add up to 1. The resulting event series were then sampled with 100 points along the time dimension. All hyetographs were created using an identical algorithm[2], showing mean values in red and representing the 25th and 75th percentiles with blue boxes.

Finally, rain events were sorted into five clusters, using K-means clustering to partition data sets, see Section 3.4.1 below. Concerning the number of clusters used, Olsson et al. (2017) did an analysis between different subdivisions and came to the conclusion that five clusters were ideal in creating the clear regional boundaries in Sweden used in this study (North, Middle, Southeast, Southwest), as three clusters meshed the southern regions together, and six or more clusters created further subdivisions within regions. Five clusters were therefore used in this study, in order to recreate these same regional divisions.

With respect to nomenclature, the first hyetograph was said to be of "type 1" (representing the very beginning of a rain event), the second, "type 2" (the first quarter) etc. Hyetographs were ordered from type 1 to type 5 for all regions.

3.4.1 K-means clustering

As mentioned above, the K-means method, also called the Lloyd–Forgy algorithm, was applied by Olsson et al. (2017), with the objective to minimize variations within clusters while dividing data set into the most optimal amount of clusters. It assigns n data points to k clusters (defined by so-called *centroids*). Shown earlier, the number of clusters was set to $k = 5$, chosen before the algorithm starts. The algorithm, used with Matlab, then continues with the following steps:

---

[2]Created and provided by Johan Södling of the SMHI for use in Olsson et al. (2017).

1) Initial centroids chosen. Cluster initialization is made with the k-means++ algorithm in Matlab, see below.

2) Point to cluster centroid distances computed for all observations.

3) Observation assignation to the cluster with the closest centroid.

4) Average or the observations computed in each cluster in order to obtain k new centroid locations

5) Steps 2-4 repeated until stable cluster assignments or maximum amount of iterations attained.

In short, data points are assigned to the nearest mean in step 2): the assignment step. In the next step, 3), the algorithm adjusts the means to correspond to the sample means of the data points they represent : the update step (Mackay, 2003).

In order to improve the running time as well as the quality of the results of the Lloyd-Forgy algorithm, the k-means++ algorithm is utilized to find centroid seeds for the clustering. For a number of clusters k, the algorithm operates in the following manner (Mathworks, 2020a):

1) Observations are chosen at random from the data set X: the first centroid, c_1.

2) The distances from the observations to c_1 are computed. The distance between a centroid c_j and an observation is then called m: $d(X_m, c_j)$.

3) The next centroid, c_2, is then selected at random from the data set X with the probability:

$$\frac{d^2(X_m, c_1)}{\sum\limits_{j=1}^{n} d^2(X_j, c_j)} \tag{12}$$

4) The center j is chosen by computing he distances from each observation to each centroid. Each observation is the assigned to its closest centroid.

Step 4 is then repeated until k centroid are chosen.

Some weaknesses with the method should be considered. For one, its inability to represent the size or shape of clusters. As the k-means algorithm only accounts for the distance between data points and the means, weights or breadth of clusters are not taken into account. Furthermore, points are assigned to one cluster exactly, and all points are equal when assigned to a cluster. This means that points located at borders between clusters have no impact on neighbouring points in other clusters with the k-means algorithm, when it is possible that such points should play a role in finding the locations of all clusters (Mackay, 2003).

The validity of the method has been proven in Sweden however, for clustering rain events into separate meteorological regions, as demonstrated by Olsson et al. (2017).

3.5 Model calibration

Calibration of the random cascade model used in this study was carried out based on the assertions made in Section 3.3. In short, a reverse cascade was applied to the municipal data sets with the 1 minute reference resolution and the following parameters were determined and saved before being subsequently used in the actual disaggregation step of the algorithm:

1. **BDCs** were extracted, plotted as histograms and fitted with the beta distribution line for visual representation. The number of histogram intervals was calculated with $k = 1 + 3.3log_{10}n$ (Haan, 1977 in J. Olsson (1998)).

2. **Probabilities** for all four disaggregation steps were calculated and plotted as a function of volume class.

3. **The shape beta parameter a** and its dependence on the cascade steps was evaluated for each position class.

4. **Constants** ($c_{1,2,3,4}$) were calculated and saved to be used specifically for the Matlab scripts.

Calibration parameters for Växjö are shown in Section 4.1 in graphical form. Numerical parameters for Växjö, Helsingborg and Malmö can be found in Appendix A.

3.6 Temporal interpolation

As stated above, SMHI's 15-min data was used as input data to be successively disaggregated. This means that rainfall volumes within 900 seconds were successively divided into intervals of 56.25 seconds. To obtain the target 1 minute resolution, a post-processing was needed. In this case, sixteen 56.25 seconds rainfall values were geometrically interpolated to fifteen 60 seconds values.

3.7 Result validation

In order to evaluate the validity of the results obtained through disaggregation, several methods and indicators were implemented. Two distinct areas were to be analysed: the quality of disaggregated data set and possibility of transferring calibration parameters between regions or cities.

3.7.1 Validation of the disaggregated data

The methodology applied to validate results obtained with a disaggregation algorithm in Matlab was similar for all three calibration municipalities and can be divided into three main categories.

Statistical measurements
A number of statistical measurements were made for Växjö, Helsingborg and Malmö each, and

synthesized in Table 4.1 in Section 4.1 below. The methodology was the same for all three municipalities: only non-zero values were taken from the continuous disaggregated data series (i.e. actual rain quantities). Medians, means, skewness, standard deviations and variances were calculated for municipal data and disaggregated data to be compared and contrasted. For all data series, disaggregation was carried out in 30 stochastic realisations. Results were then averaged before validation.

Quantile plots & histograms

Another way to validate the disaggregation process was to observe if the random cascade model applied with an algorithm in Matlab could accurately reproduce certain characteristics of the 1-min municipal data series used for model calibration. To that effect, two rainfall characteristics were chosen for analysis: EV and ED. Note that while the term "volume" is used for the first characteristic, it actually represents rain quantities in mm. The choice of rainfall characteristics was inspired by Güntner et al. (2001), in a study comparing the effectiveness of rainfall data disaggregation between multiple countries, where similar charateristics were analysed and compared for observed and disaggregated time series, using the same model as in this study.

Two types of visualisations were then produced for both rainfall characteristics at each municipality. The first was quantile-quantile plots (Q-Q plots), used to compare two data distributions, in this case EV and ED for calibration data against disaggregated data. The advantaged of using QQ-plots is threefold, according to Helsel and Hirsch (2002):

- No arbitrary categories are required.

- All data are displayed.

The plotting position (*p*) formula used in this case was the Hazen formula:

$$p = \frac{(i - 0.5)}{n} \tag{13}$$

where the smallest data point is assigned a rank $i = 1$ and the largest $i = n$. This particular formula is most commonly used for comparing two or more data series in Q-Q plots (Helsel and Hirsch, 2002).

In this case, the quantiles-quantiles plots are created using a Matlab function where the quantiles of the data in one series are plotted against the quantiles of another. Here, each point is marked with a cross (+). The red dotted line is a reference line representing the theoretical distribution (Mathworks, 2020b).

The second, histograms, are used in order to give a graphical comparison between municipal and disaggregated data sets. As with the Q-Q plots, ED and EV are compared. Both methods are shown in greater detail for this study in Section 4.1 below.

3.7.2 Spatial transferability

One important area of interest for this study is the possibility of obtaining calibration parameters from one geographical location and use these to disaggregate rainfall series from another

location. Econopouly et al. (1990) and Güntner et al. (2001) suggested that such a transfer is possible in certain climates, but this needed to be tested for Sweden. If proven true for Sweden, areas with fewer or poorer rainfall data could gain high resolution data sets by disaggregating lacking series with calibration data provided by better sources. This could potentially be instrumental for northern regions for instance, where long winters often lead to large gaps in rainfall data series and expensive equipment (such as heated facilities for tipping bucket collection) cannot be reasonably installed at the municipal level. Transferability was analyzed within the same region (SW) but also between regions (SW and SE). To that effect, two types of assessments were made in order to test the spatial transferability of calibration parameters: some statistical measurements and a Wilcoxon rank-sum test, see below. Histograms as well as Q-Q plots showing the probabilities of rain quantities and durations between regions were also created.

Statistical measurements
In a similar fashion to the validation of the disaggregation process, statistical measurements were taken for non-zero values for data series disaggregated with municipal data from the same location, scenario *A*, (15-min SMHI data for Växjö disaggregated with 1-min municipal data from Växjö for example) and for data series disaggregated with municipal data form another region, scenario *B*, (15-min SMHI data for Växjö disaggregated with 1-min municipal data from Helsingborg for example. Again, disaggregation was executed in 30 realisations, and averaged over all stochastic realisations before analysis.

Rank-sum test
Yet another way of assessing eventual similarities between different disaggregation scenarios was the rank-sum test (called Wilcoxon rank-sum test in this study). The Wilcoxon rank-sum test is a non-parametric test used to test when samples are independent from one another in two populations. The test is usually used for the specific purpose of determining whether or not two groups come from the same population (Helsel and Hirsch, 2002). This is accomplished by testing the equality between the medians of two groups. Let *x* and *y* be arrays with disaggregated data from scenario *A* and *B* respectively. The Wilcoxon rank-sum test then tests the null hypothesis (H_0) that *x* and *y* come from independent continuous distributions with the same medians against the possibility that they are not (Mathworks, 2020c). The results from this test are then given as two parameters. The first, the *p*-value, a positive scalar $0 < p < 1$, is the probability that a test statistic is as or more extreme that the observed value as defined by the null hypothesis. The second value, *h* is the result of the test and is returned as:

- $h = 1$, if the null hypothesis is rejected at the 5% significance level.

- $h = 0$, if the null hypothesis fails to be rejected at the 5% significance level.

Q-Q plots, histograms and results of the Wilcoxon rank-sum tests applied to this study for the rainfall characteristics ED and EV for three different tests of spatial transferability for calibration parameters are shown and discussed in further details in Section 4.2 below.

4 RESULTS

The calibration and spatial transferability of the random cascade model when applied to Swedish cities and regions were investigated and validated with different methods. Following this initial analysis, a series of hyetograph were clustered, and distributions calculated for the purpose of regional comparisons.

4.1 Calibration of the random cascade model

4.1.1 Probabilities

The variations of the probabilities were studied as part of the validation of the obtained calibration parameters, and are shown in graphic form in Figure 4.1 for Växjö. Results for Helsingborg and Malmö can be found in Appendix A in Figures A.4a & A.4b respectively. Numerical values for the observed probabilities are also found in the same Appendix, compiled in Tables A.5 through A.7.
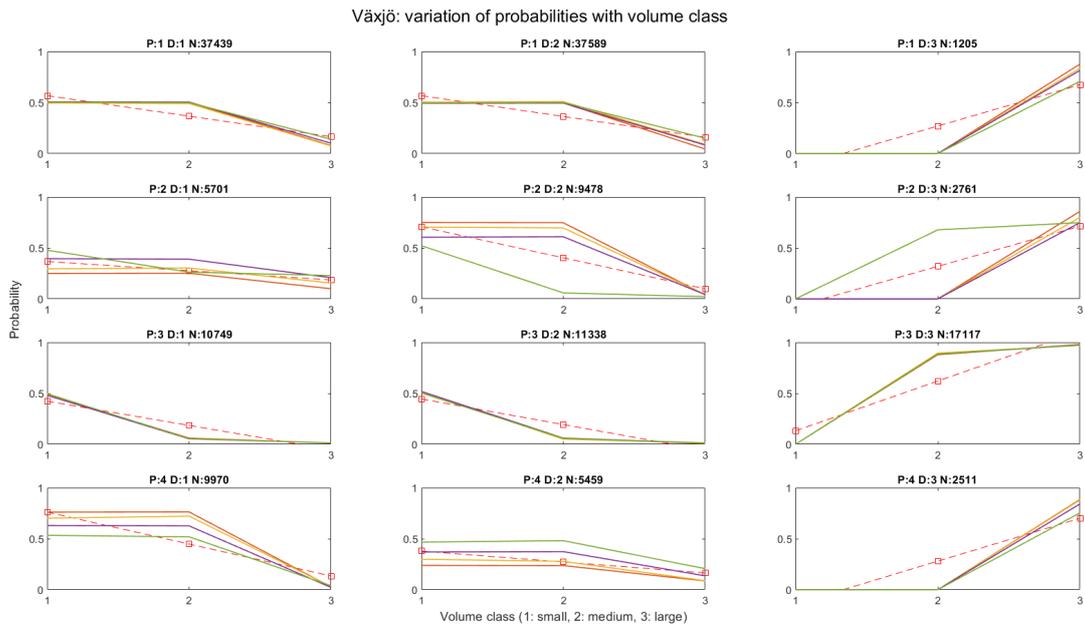

Figure 4.1. Variation of probabilities with volume class for Växjö station. *P* denotes the position type (1: isolated, 2: starting, 3: enclosed, 4: ending), *D* denotes the type of distribution (1: 0/1, 2: 1/0, 3: x/x) and *N* is the total amount of periods for each position and distribution type. The different solid coloured lines represent the different cascade steps during disaggregation. Orange: 1st, Yellow: 2nd, Purple: 3rd, Green: 4th. The dashed red line is the mean of all cascade steps.

As theorized and demonstrated by J. Olsson (1998) and Güntner et al. (2001), the possibility of describing the changes of probabilities *P(0/1), P(1/0), P(x/x)* with the linear function described earlier in Equation 7 is also apparent in this study, as shown in Figure 4.1 above. The dashed line representing the mean variation of probabilities over all cascade steps is linear for all distribution types and volume classes.

Probability variations and changes with volume classes are in this case consistent with earlier experimentation at time scales between 1 week and 1 hour (J. Olsson, 2012). The fact that the same trends in probabilities are obtained in this study suggests the possibility of obtaining relevant disaggregation results at the higher temporal resolution of 1 minute.

While the mean of all cascade steps in Figure 4.1 has a linear relationship with volume class, showing the probabilities' dependence on volume, their dependence on cascade step - in other words the time scale in temporal disaggregation - displays some variations. The intercept *int* varies with some regularity: this behaviour is modeled in the disaggregation step of the algorithm by Equation 9. The fitted intercept *int*'s variations with cascade step are shown in Figure 4.2 for Växjö. Similar trends are shown for Helsingborg and Malmö in Appendix A, Figures A.3a & A.3b respectively. From Figure 4.2, it is apparent that the intercept *int* changes drastically with each cascade step.

Graphs in the third column represent the fraction of (0/1)-divisions of all non x/x-divisions (i.e. $(0/1) + (1/0)$), over all volume classes. *P(0/1)* and *P(1/0)* are then approximately equal, according to J. Olsson (2012) and can then be modeled by Equation 11 shown earlier.
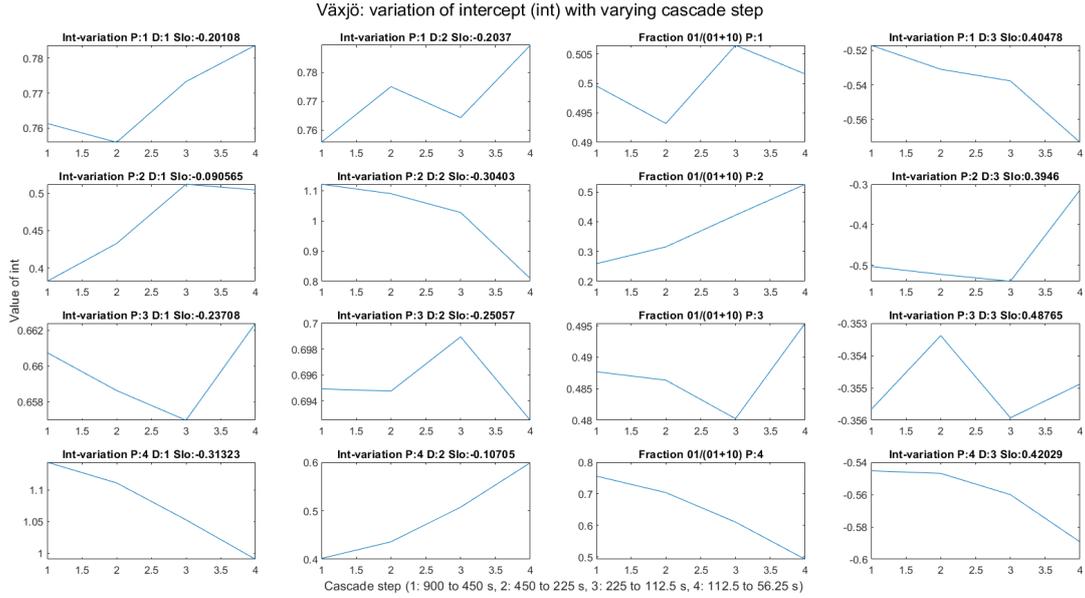
Figure 4.2. Variation of intercept *int* with cascade step for Växjö station. *P* denotes the position type (1: isolated, 2: starting, 3: enclosed, 4: ending), *D* denotes the type of distribution (1: 0/1, 2: 1/0, 3: x/x) and *Slo* the mean slope for each position and distribution.

All in all, variations of probabilities *P* and their dependencies on both volume classes and cascade steps show similar behaviours to J. Olsson (1998) and J. Olsson (2012) as well as Güntner et al. (2001) in all three calibration cities, even at higher temporal resolutions.

4.1.2 $W_{x/x}$ - distribution: histograms

As noted above, the behaviour of BDCs is important in validating and analysing the random cascade model. Figure 4.3 below shows the $W_{x/x}$ histograms for one calibration city, Växjö. Histograms for Helsingborg and Malmö are shown in Appendix A. Contrary to J. Olsson (1998), where $W_{x/x}$ - histograms were symmetrical, due to the inclusion of both weights W_1 & W_2 and $W_1 = 1 - W_2$, here only W_1 was included, by calculating $W_2 = 1 - W_1$ as suggested by Güntner et al. (2001). Values for W_1 are pooled for the entire time range, giving the more accurate histogram more influence. All calibration cities display a similar distribution and are dominated by the 0.5 histograms for all cascade steps and position types.
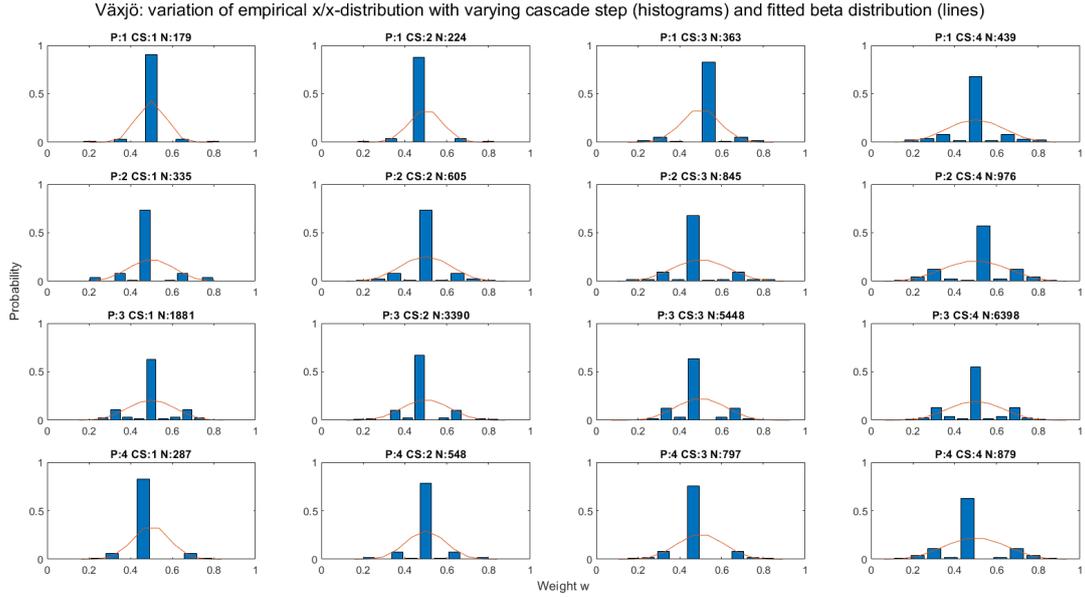
Figure 4.3. Variation of empirical $W_{x/x}$ distributions with cascade step (histograms) and a fitted beta distribution (line) for Växjö station. *P* denotes the position type (1: isolated, 2: starting, 3: enclosed, 4: ending), *CS* the cascade step (1: 900-450 s, 2: 450-225 s, 3: 225-112.5, 4: 112.5-56.25) and *N* the total amount of x/x distributions for each position type and cascade step).

Scaling behaviour of the beta parameter *a* is clear as seen in Figure 4.4. For starting and ending boxes, the variations are quite high, and *a* decreases rather strongly. Similarly for isolated and enclosed boxes, where the decrease is a bit less smooth, but still sloping rather aggressively. This would suggest, according to J. Olsson (2012), that these position types could not be said to display uniform distributions, with a higher probability at the center, as seen in Figure 4.3 for example, with $W = 0.5$ dominating.

It is, however, difficult to compare these results with the ones detailed in J. Olsson (1998) directly, given different data sets, studied time periods and indeed different volume resolutions in the equipment used. Variations shown in Figure 4.4 could very well mean that a constant approximation is sufficient for isolated and enclosed boxes. The behaviour of the beta parameter *a* would need to be explored and analysed further in another study with a similar temporal resolution.
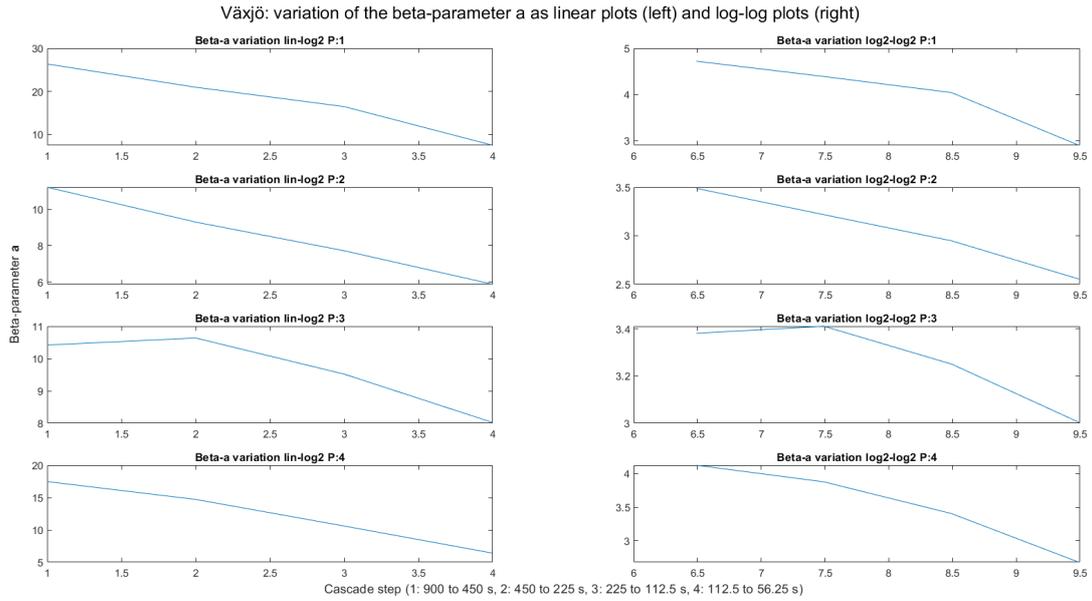
Figure 4.4. Variation of the beta-parameter *a* with cascade step for Växjö station. *P* denotes the position type (1: isolated, 2: starting, 3: enclosed, 4: ending).

4.1.3 Data validation - disaggregation: Municipal data vs disaggregated data

An important aspect of this study was to assess the random cascade model's ability to create disaggregated data series exhibiting similar rainfall characteristics to existing 1-min municipal data. To this end, different statistical measurements were computed and contrasted between each data series, the disaggregated series being set to the same length as the calibration series, as mentioned earlier in Section 3.2.3. The number of non-zero points, median, mean, skewness, standard deviation and variance were calculated for all three calibration cities, as seen in Table 4.1 below.

The different statistical measurement for municipal and disaggregated data yielded somewhat superfluous results, as medians and means would naturally disagree as seen here in Table 4.1, because of the different volume resolutions for each data set. If new municipal rainfall time series were created with the same volume resolution (0.1 mm in the case of SMHI's 15-min data), means and medians would be of interest. However, the number of non-zero values in the data series, N_{nz}, differ in interesting ways between municipal and disaggregated data, and would appear to do so independently from volume resolution. Indeed, a similar conclusion was drawn when using the same volume resolution for two series when extracting N_{nz}. This might be caused by the stochastic properties and characteristics of the random cascade model or indeed the choice of data series. Skewness in the series, standard deviations and variances are all rather similar for all three calibration cities, again giving strength to the hypothesis that disaggregation of rainfall series could prove applicable to such high temporal resolutions.

Table 4.1. Statistical measurements for non-zero calibration and disaggregated data averaged over 30 realisations.

Helsingborg			
Statistical measures	Municipal data	Disaggregated data	Δ
N_{nz}	40223	76690	36467
$Median_{nz}$	0.20	0.10	-0.10
$Mean_{nz}$	0.22	0.12	-0.10
$Skew_{nz}$	9.40	13.7	4.35
Std_{nz}	0.09	0.07	-0.02
Var_{nz}	0.01	0.00	0.00

Växjö			
Statistical measures	Municipal data	Disaggregated data	Δ
N_{nz}	49986	87068	37082
$Median_{nz}$	0.20	0.10	-0.10
$Mean_{nz}$	0.21	0.11	-0.10
$Skew_{nz}$	11.20	13.0	1.83
Std_{nz}	0.09	0.07	-0.02
Var_{nz}	0.01	0.00	0.00

Malmö			
Statistical measures	Municipal data	Disaggregated data	Δ
N_{nz}	29544	66252	36708
$Median_{nz}$	0.20	0.10	-0.10
$Mean_{nz}$	0.22	0.12	-0.10
$Skew_{nz}$	9.28	12.4	3.12
Std_{nz}	0.10	0.10	0.00
Var_{nz}	0.01	0.00	0.00

4.1.4 Data validation - comparing data series in the southeast of Sweden: Växjö. Municipal data vs disaggregated data

As explained in Section 3.7, the efficacy of the random cascade model was tested by comparing and contrasting two rainfall characteristics, namely event volume (EV) and event duration (ED) between municipal calibration data series and disaggregated data series. The results of these comparisons are shown graphically in Figures 4.5 through 4.8 as histograms. As seen here, the disaggregated series tend to overestimate smaller quantities and shorter rainfall durations, while underestimating larger quantities and longer durations to a certain extant. It is however quite evident from all histograms that both EV and ED and their behaviours over whole time series are preserved quite well in disaggregated time series when compared to the calibration data series.

Histograms: EV and ED

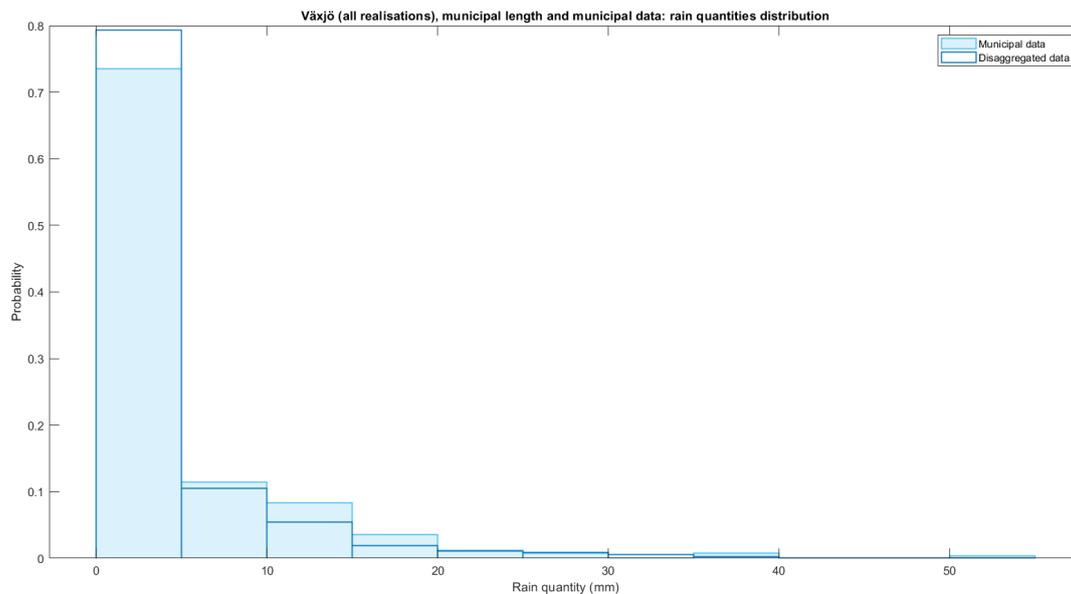Histogram width was set to 5 mm for EV, and 50 min for ED.

Figure 4.5. Disaggregated rainfall series for Växjö, averaged over 30 stochastic realisations (dark outline), and municipal data from Växjö (light blue) for EV.
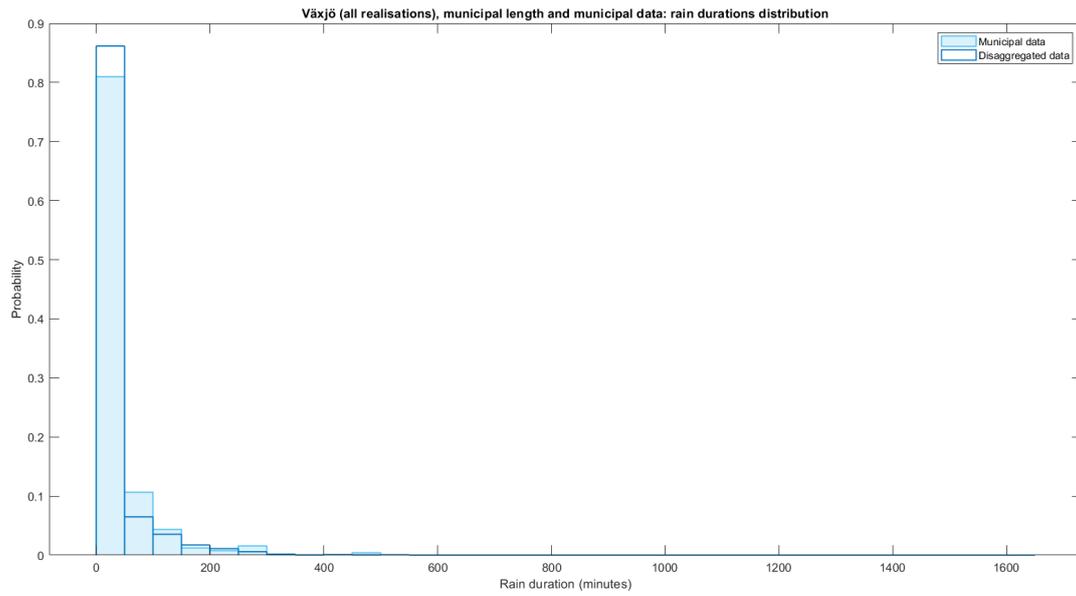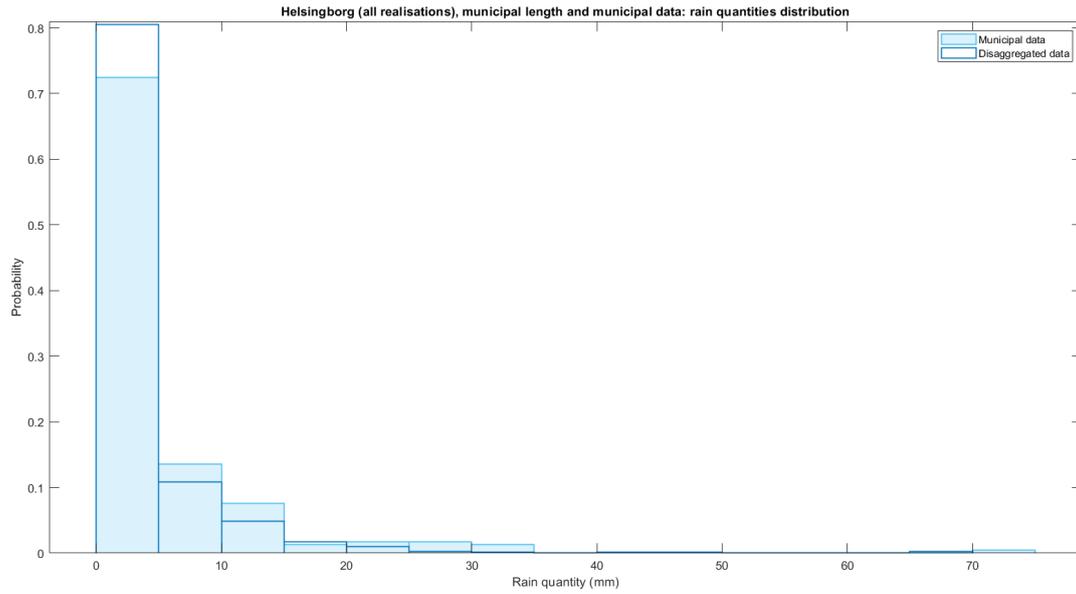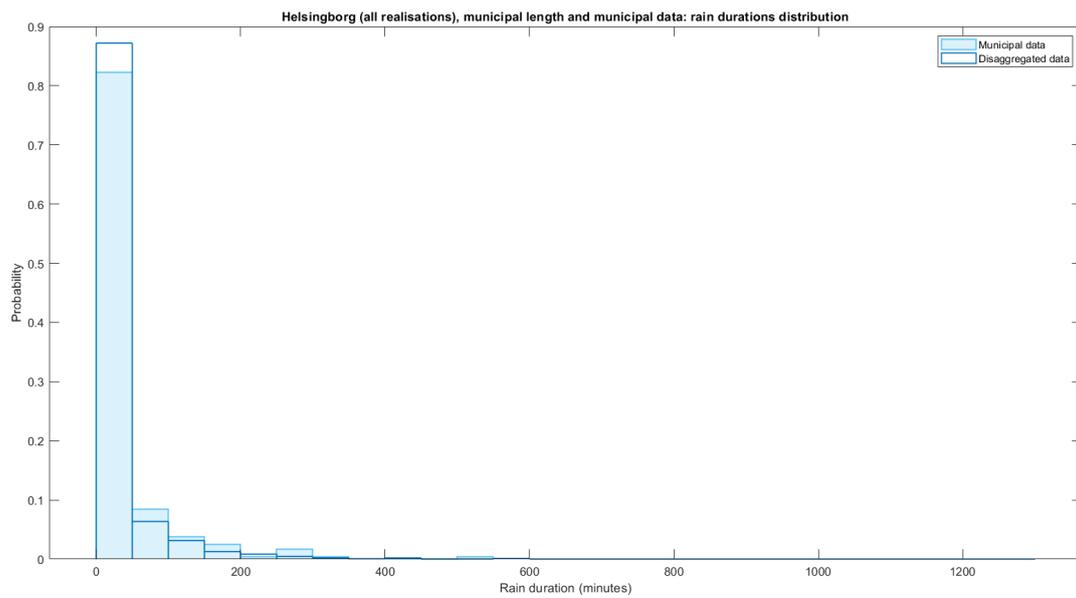
Figure 4.6. Disaggregated rainfall series for Växjö, averaged over 30 stochastic realisations (dark outline), and municipal data from Växjö (light blue) for ED.
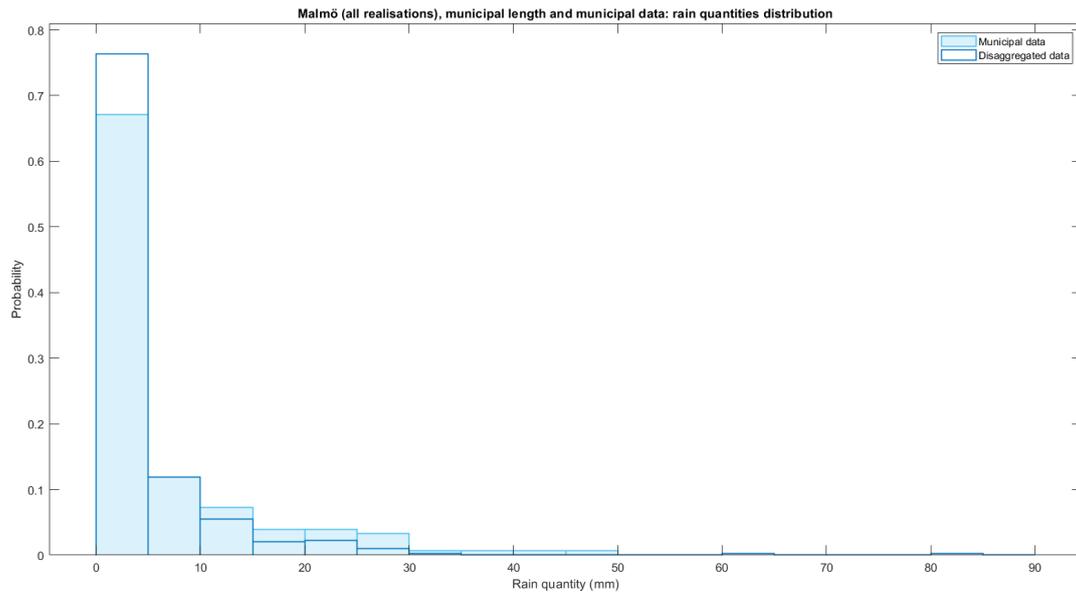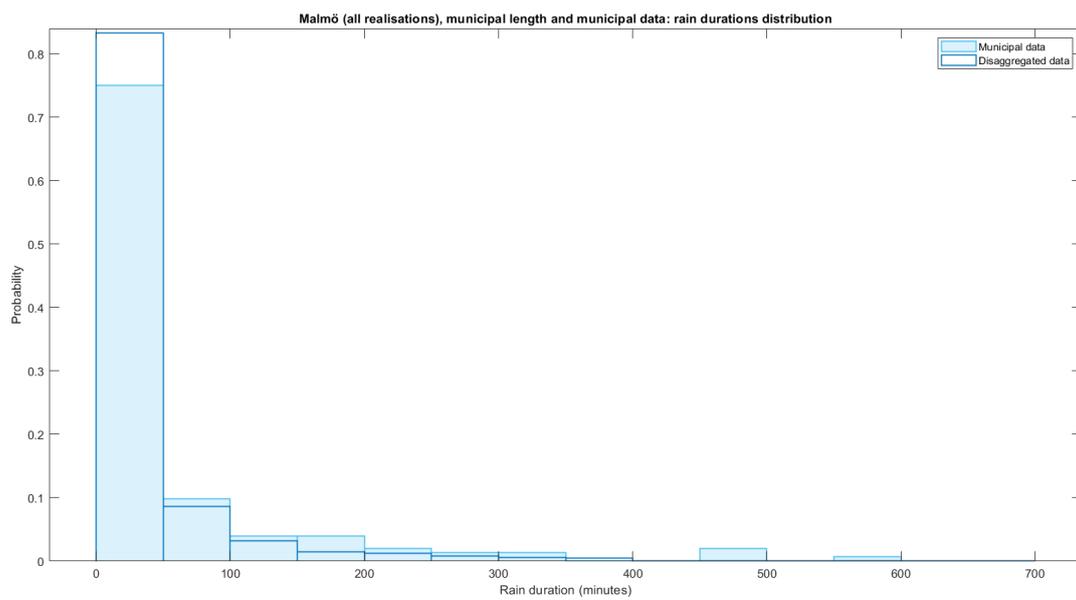
(a) EV

(b) ED

Figure 4.7. Disaggregated rainfall series for Helsingborg, averaged over 30 stochastic realisations (dark outline), and municipal data from Helsingborg (light blue) for EV and ED.

(a) EV

(b) ED

Figure 4.8. Disaggregated rainfall series for Malmö, averaged over 30 stochastic realisations (dark outline), and municipal data from Malmö (light blue) for EV and ED.

Q-Q plots: EV and ED

Behaviours for ED and EV are also shown in the form of quantile-quantile plots. For model validation, disaggregated rainfall series are averaged over 30 realisations and plotted against municipal data. Shown in Figure 4.9 below are the Q-Q plots for Växjö, for parameters EV and ED. Similar plots for Helsingborg and Malmö can be found in Figures F.1 and F.2 Appendix F.
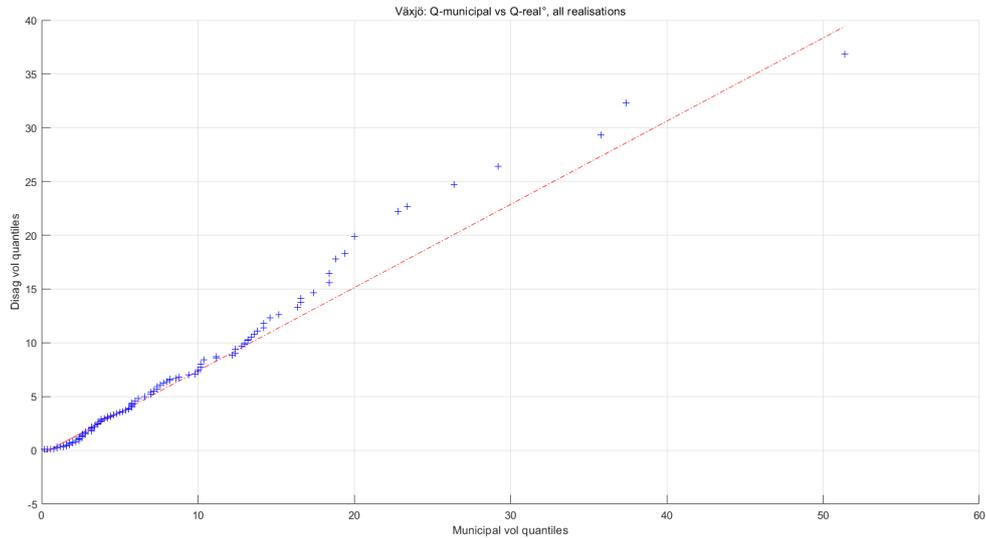
(a) EV
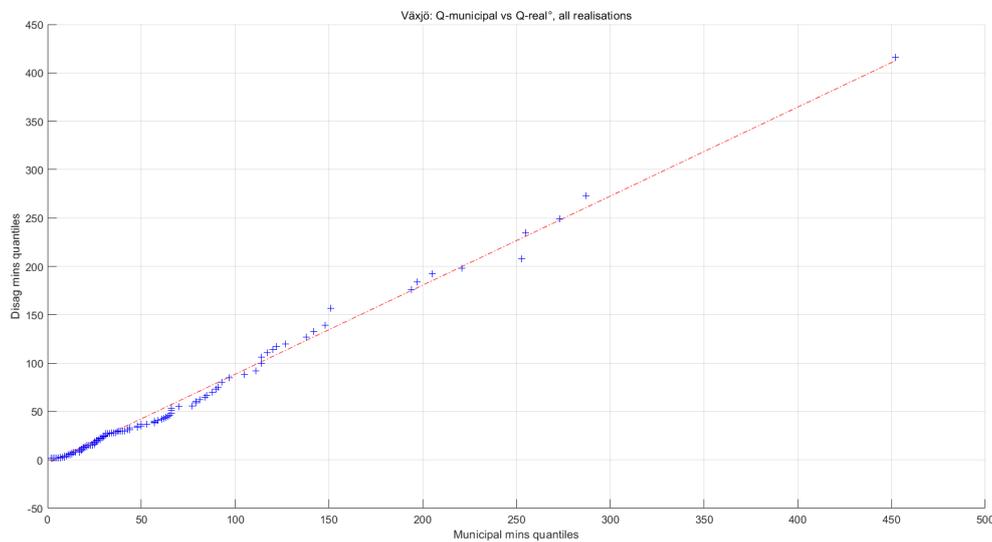
(b) ED

Figure 4.9. Disaggregated rainfall series for Vaxjö, averaged over 30 realisations vs municipal data from Växjö for EV and ED.

31

The Q-Q plots for Växjö found above in Figure 4.9 show the probability distributions for disaggregated series, set to the same length as the calibration series, and the municipal calibration data set for the same city. The parameters EV and ED don't seem to necessarily follow the same distributions for Växjö, an observation that can be made for Helsingborg and Malmö as well, see Figures F.1 & F.2, respectively. It does appear, however, that at lower ranges, the distributions are almost identical, as some degree of linearity can be observed. Some seemingly random bias does occur in both EV and ED, but the bias oscillates between higher and lower values depending on the percentile, without any systematic bias emerging.

At higher quantiles, the results are quite different, with the distribution for disaggregated data being biased towards higher values. Outliers are clearly seen, especially in the Q-Q plots for EV, in Figure 4.9a in the case of Växjö, but are present for all Q-Q plots comparing the distributions of municipal data with disaggregated data.

4.2   Spatial transferability: calibration parameters

An interesting question that arose early in the study of regional differences in Sweden was the prospect of utilizing calibration parameters from on region and apply them to the algorithm for the disaggregation of data series from another region. If proven possible, this would permit a significant increase in the quantity of available rainfall series throughout Sweden. Sweden's northernmost regions, experiencing longer and harsher winters could be provided with high resolution rainfall data disaggregated from existing measuring stations, effectively negating the unpredictable precipitation patterns during winter months.

Disaggregated series for one station were therefore calibrated with municipal data form the same city (called in the case of Växjö "Vxj" or "Vxj/Vxj" in the figures) and calibrated with municipal data from another region ("Vxj/hls" for the case of Växjö (SE) being disaggregated with calibration parameters from Helsingborg (SW), see Figure 4.10). Statistical measurements are shown in Table C.1 in Appendix C below. The same methodology was applied in this case as in the validation of the disaggregation series in Section 4.1.3. The similarities between the disaggregated series Vxj/vxj and Vxj/hls in all measurements in Table C.1 are unmistakable.

The first analysis of the spatial transferability of calibration parameters was performed between the two southern meteorological regions in Sweden: Southwest (encompassing the cities of Helsingborg and Malmö) and Southeast (Växjö). The results seem very promising when seen in the form of histograms, see Figure 4.10 and Figures E.1 through E.3 in Appendix E. Unlike the comparison between disaggregated and municipal series, where disaggregated series were shortened, the comparisons performed between disaggregated series in different regions were carried out with series spanning the same time interval, i.e. 1995-2020.

4.2.1 Spatial transferability between regions: Southwest and Southeast
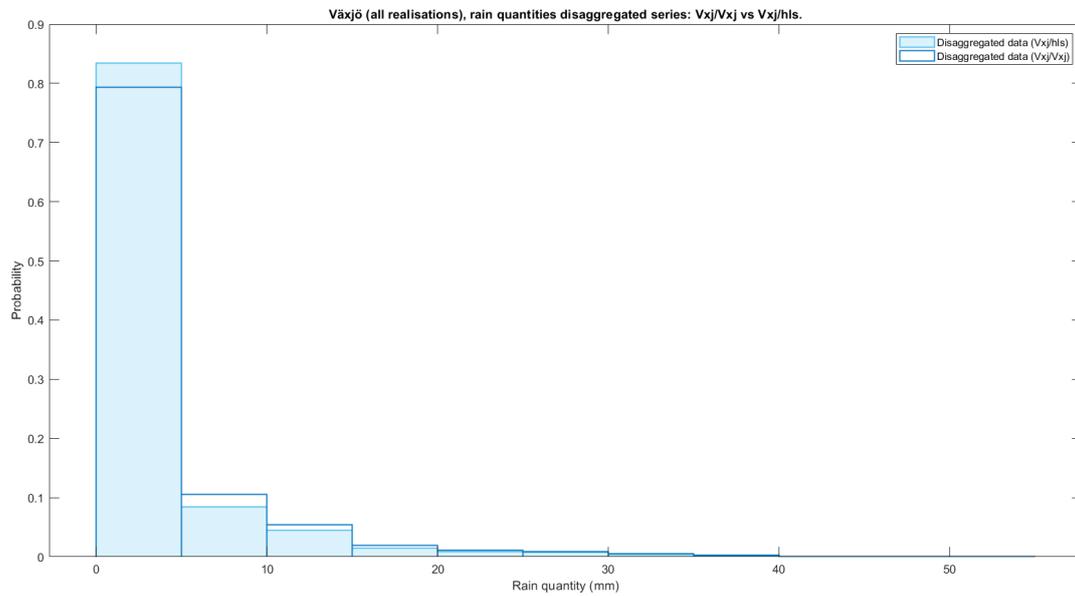
(a) EV
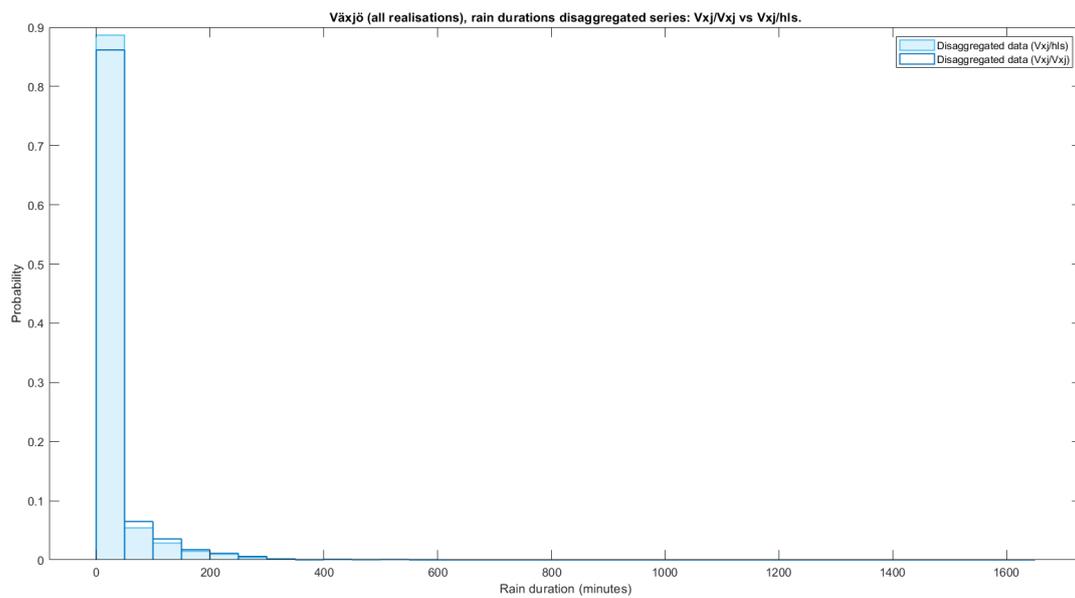
(b) ED

Figure 4.10. Disaggregated rainfall series for Växjö, averaged over 30 stochastic realisations. Calibrated with municipal data from Växjö, SE, (dark outline) and municipal data from Helsingborg, SW (light blue).
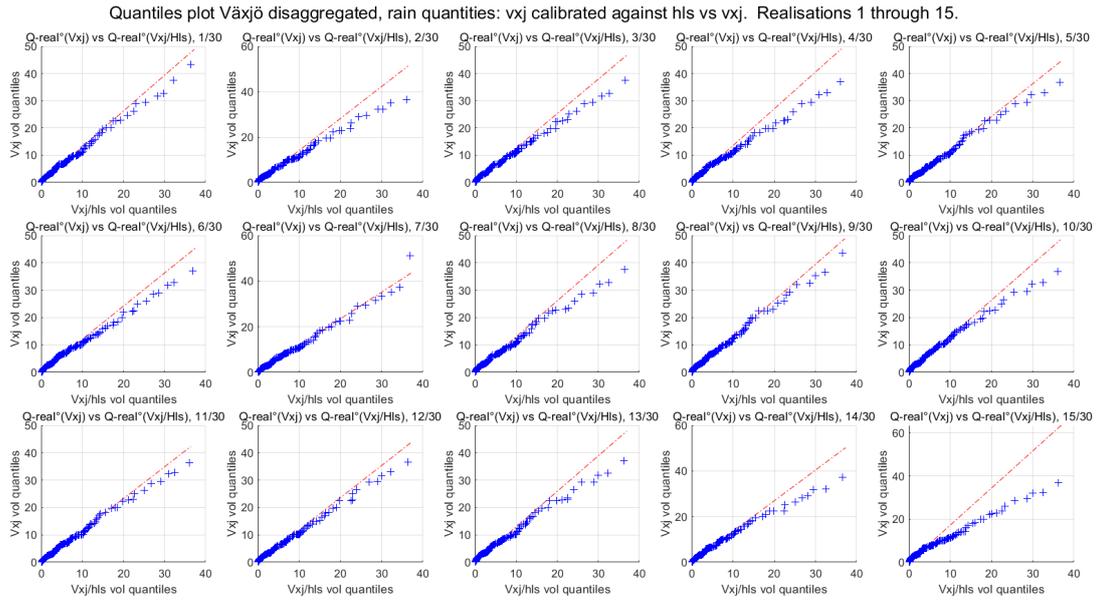
(a) Stochastic realisations 1 through 15
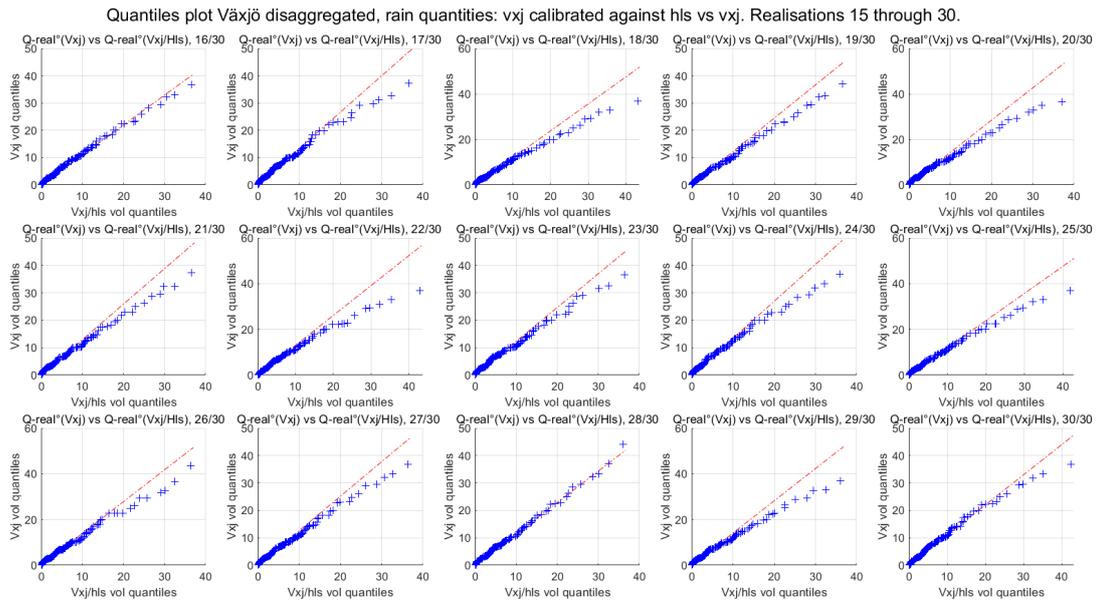
(b) Stochastic realisations 16 through 30

Figure 4.11. Disaggregated rainfall series for Växjö. Quantiles for calibrated with municipal data from Växjö, SE, against quantiles for municipal data from Helsingborg, SW. (EV)

Figures 4.11 above and 4.12 below compare the distributions of EV quantiles and ED quantiles respectively. Q-Q plots for all 30 realisations are presented in an effort to show the disparities present in some cases between realisations.

In the case of the first rainfall characteristic, EV, approximately the same results found in Fig-

ure 4.9 are found here. Indeed, for almost all realisations, a linearity in the points for the lower quantiles is present (only realisation 15/30 seem to diverge in any significant way), indicating relatively similar distributions for SW and SE at lower rain quantities. This would convey the idea that the model behaved "perfectly" in those realisations to give the exact same distributions. Similarly to the municipal vs disaggregated Q-Q plots, the distributions tend to differ at higher ranges. Contrary to the aforementioned municipal vs disaggregated plots, a definite curve towards the right indicates that the Vxj/hls distribution is skewed to the right compared to the Vxj distribution, or has heavier tails.

In terms of ED, the Q-Q plots shown in Figure 4.12, while still showing a linearity at lower ranges, also show a more pronounced right-hand-side tail for the Växjö data disaggregated with calibration parameters from Helsingborg. Outliers are also more prevalent in the Q-Q plots for ED than for EV. For some stochastic realisations (see for instance realisation 7/30), the general trend of the Q-Q plot is substantially flatter than the y=x line, suggesting that the distribution of Vxj/hls has higher variance than the Vxj/vxj.

Similar observations can be made when analysing distributions between different calibration sites within the same region. For region SW, Figures 4.13a & 4.13b show the results of transferability tests performed for Helsingborg and Malmö, in other words within the same region. Like the results of Växjö, a graphical analysis using histogram gives the impressions of very similar distributions for both ED and EV. Figures 4.14 & 4.15 show the distribution of EV and ED respectively for disaggregated data for Helsingborg (SW) calibrated with parameters from Helsingborg plotted against disaggregated data for Helsingborg with calibration parameters from Malmö (SW).

The main difference that can be observed in these Q-Q plots for EV is a general tendency towards somewhat flatter plots, compared to the Vxj/vxj vs Vxj/hls plots. The same linearity at lower volume quantiles are found even in this case. The Q-Q plots for ED, however, show strange behaviours with high volatility between stochastic realisations, and extreme outliers at the higher quantiles. Nevertheless, they do still exhibit some linearity at the lower end.

(a) Stochastic realisations 1 through 15
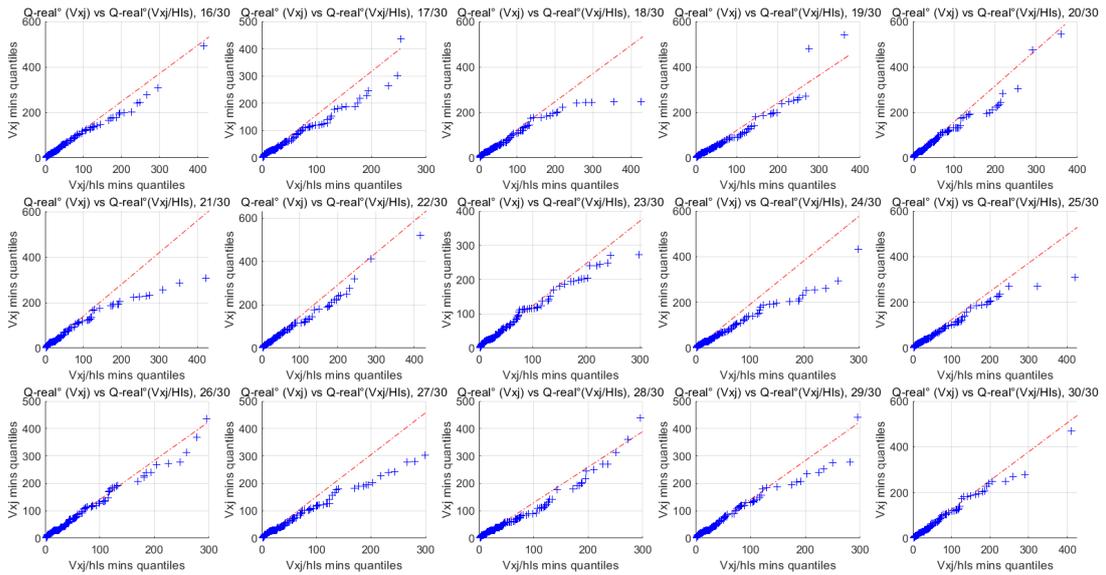
(b) Stochastic realisations 16 through 30
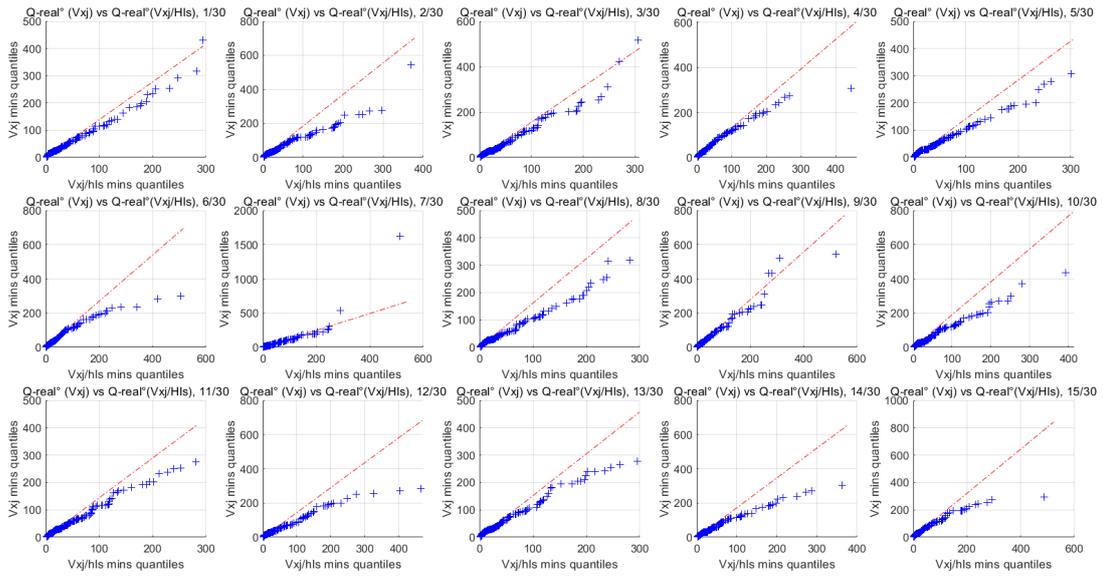
Figure 4.12. Disaggregated rainfall series for Växjö. Quantiles for calibrated with municipal data from Växjö, SE, against quantiles for municipal data from Helsingborg, SW. (ED)

4.2.2   Spatial transferability within the same region: Southwest
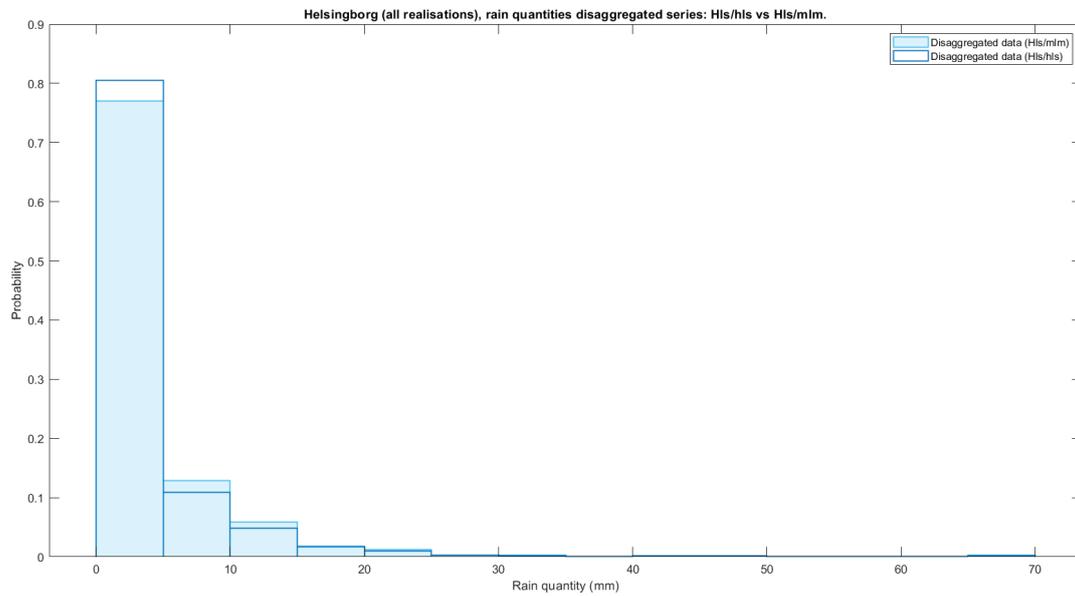
(a) EV
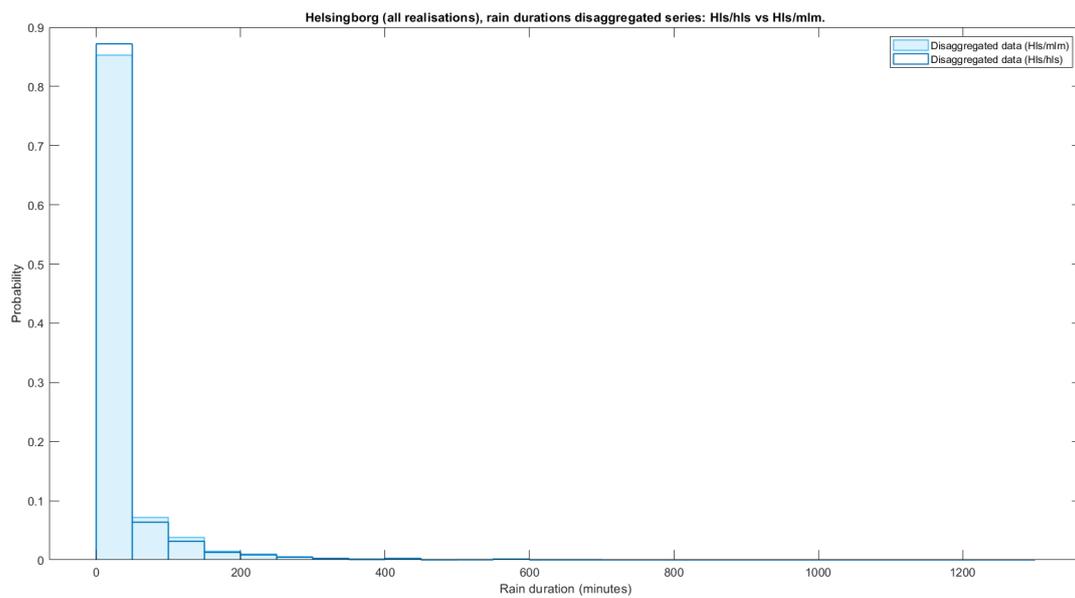
(b) ED

Figure 4.13.   Disaggregated rainfall series for Helsingborg, averaged over 30 stochastic realisations. Calibrated with municipal data from Helsingborg, SW, (dark outline) and municipal data from Malmö, SW (light blue).
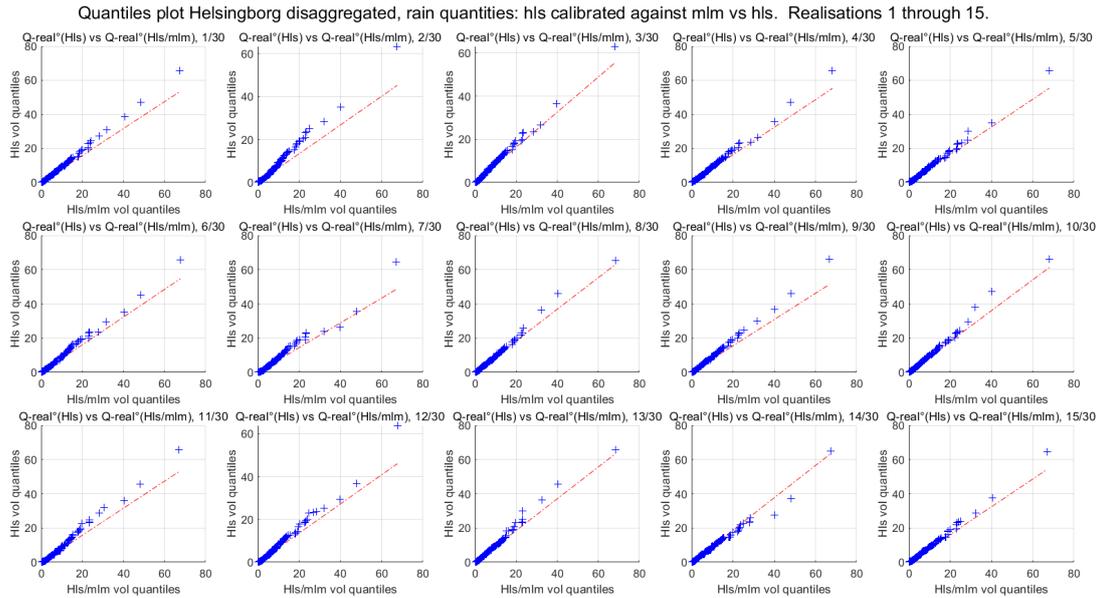
(a) Stochastic realisations 1 through 15
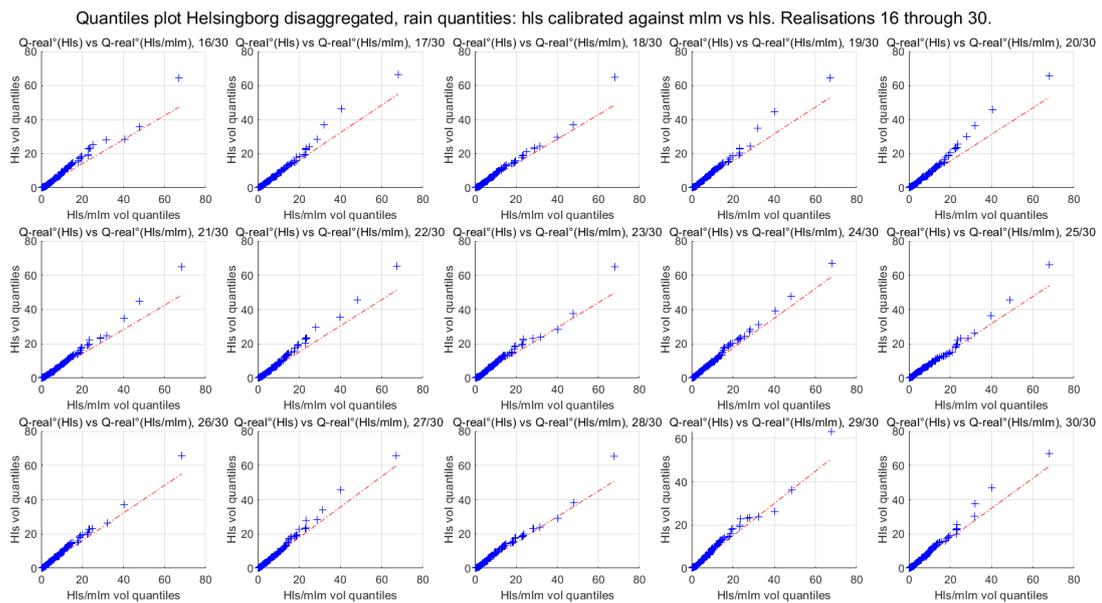
(b) Stochastic realisations 16 through 30

Figure 4.14. Disaggregated rainfall series for Helsingborg. Quantiles for calibrated with municipal data from Helsingborg, SW, against quantiles for municipal data from Malmö, SW. (EV)
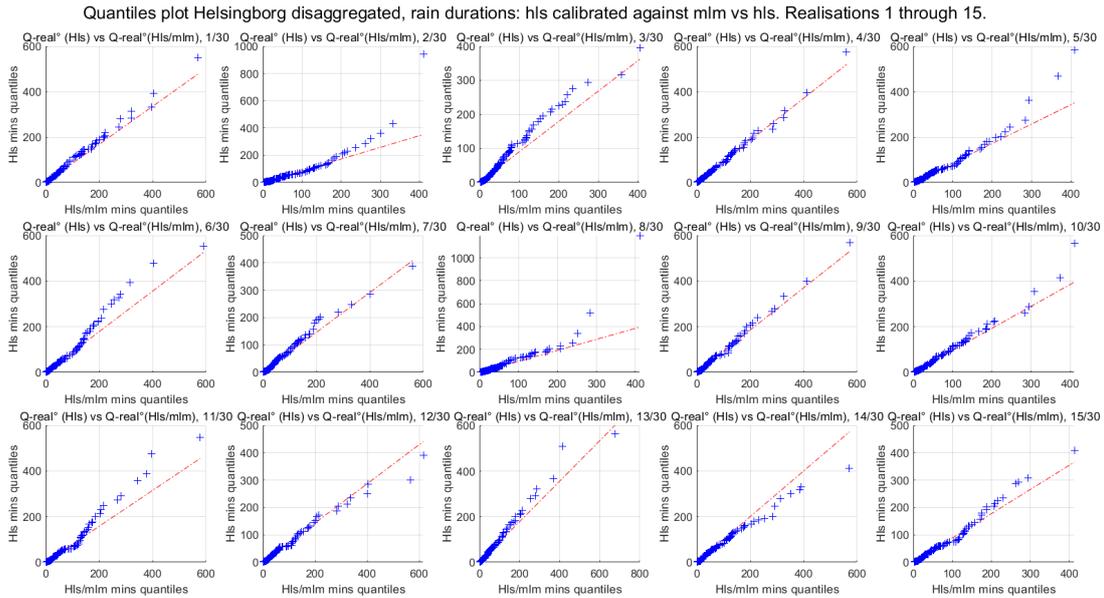
(a) Stochastic realisations 1 through 15
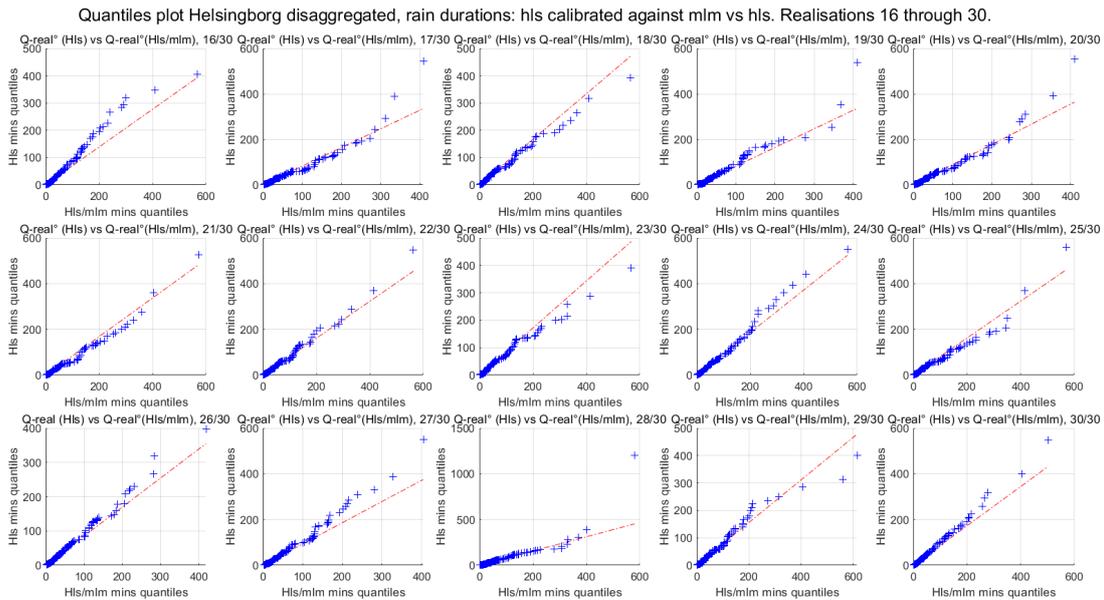
(b) Stochastic realisations 16 through 30

Figure 4.15. Disaggregated rainfall series for Helsingborg. Quantiles for calibrated with municipal data from Helsingborg, SW, against quantiles for municipal data from Malmö, SW. (ED)

Wilcoxon rank sum test results

As stated above, a series of Wilcoxon rank sum tests were made in an effort to assess the spatial transferability of calibration parameters as thoroughly as possible. In this case, the null hypothesis H_0 that x and y come from independent continuous distributions with the same medians against the possibility that they are not (Mathworks, 2020c). h is the result of the test and is returned as:

- $h = 1$, if the null hypothesis is rejected at the 5% significance level.

- $h = 0$, if the null hypothesis fails to be rejected at the 5% significance level.

Table 4.2. Wilcoxon rank sum test: H_0 rejection rates for EV and ED

	Wilcoxon rank sum		
	Vxj/vxj vs Vxj/hls	Hls/hls vs Hls/mlm	Mlm/mlm vs Mlm/vxj
Realisations not rejected (%) for EV	17	43	100
Realisations not rejected (%) for ED	0	33	100

The results from the Wilcoxon rank sum test are shown in the form of percentage of realisations not rejected over 30 realisations in Table 4.2. Results showing the *p*-values and rejection/acceptance of the null hypothesis can be found in Tables C.2 & C.3 in Appendix C for EV and ED respectively.

Table 4.2 shows surprising results for Växjö and Malmö. In the latter, the null hypothesis wasn't rejected for any realisation in the experiment comparing Malmö (SW) disaggregated with parameters from Malmö and the same with paremeters from Växjö (SE) for neither EV nor ED. Contrary to these findings, the H_0 was rejected for all realisation in the case of Växjö (SE) calibrated against Helsingborg (SW), while only a relatively small amount (17%) wasn't rejected. These findings are reflected in the histograms and Q-Q plots for Växjö and Helsingborg shown above and likewise for Malmö in Figures E.4 & E.8 in Appendix E. The rank sum test confirms the results of the Q-Q plots, as discussed in Section 4.2.1 above.

4.3 Hyetographs

The following figures show empirical hyetographs for Växjö for a serie of three different experiments: hyetographs clustered from 15-min SMHI data, disaggregated with 30 stochastic realisations down to 1 minute, the same data disaggregated with 100 stochastic realisations (Figures 4.16 & 4.17) as well as the 1-min municipal data used for the RCM model calibration shown in Figure 4.18. Each hyetograph cluster has been ordered by peak arrival, from earliest to latest. Variations in intensity are noticeable in all hyetographs for Växjö, and are even clearer at and around the peaks. All hyetographs obtained in this study consistently show the same variation patterns, and can be found in Appendix D. Based on the results shown in Section 4.2, the decision to use calibration parameters from Växjö for the disaggregation of SMHI data in cities lacking viable municipal series - i.e., most localities in region MID and all stations in region N - was made.

As seen in Figures 4.16, 4.17 and 4.18, some differences between hyetographs for Växjö are apparent. The first two depict hyetographs clustered with disaggregated data for Växjö, where 30 and 100 realisations were averaged respectively. While the overall shapes of the hyetographs differ slightly, and especially so for the clusters in the middle, peak arrival times as well as intensities are quite similar. It would therefore seem that disaggregation of existing data series does in fact recreate rainfall characteristics and distributions relatively well. Similar hyetographs for Helsingborg can be found in Appendix D, Figures D.12 through D.14. In the case of Helsingborg, differences between hyetographs are a little more pronounced. However, as mentioned earlier, the limitations of this study are quite clear owing to the different time series used, which could create these differences in hyetographs.
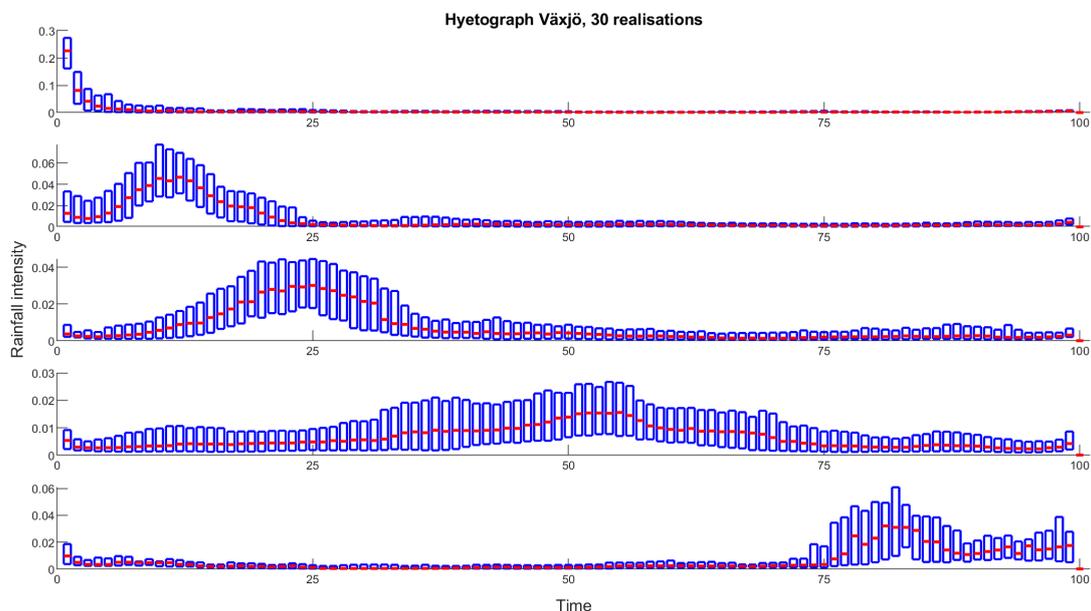
Figure 4.16. Hyetographs in five clusters for Växjö (SE), 30 realisations averaged
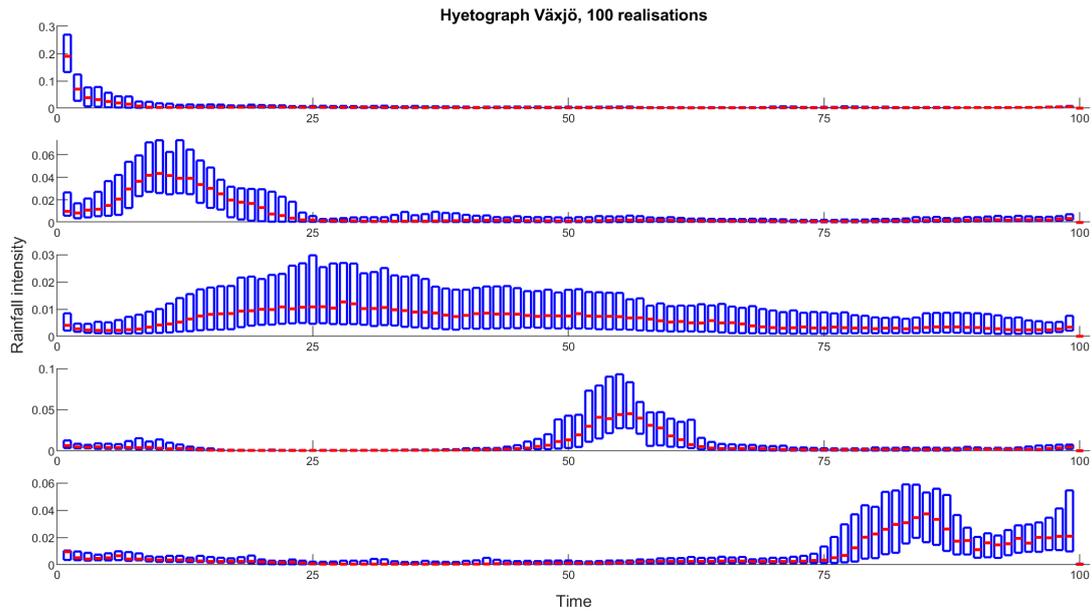
Figure 4.17. Hyetographs in five clusters for Växjö (SE), 100 realisations averaged
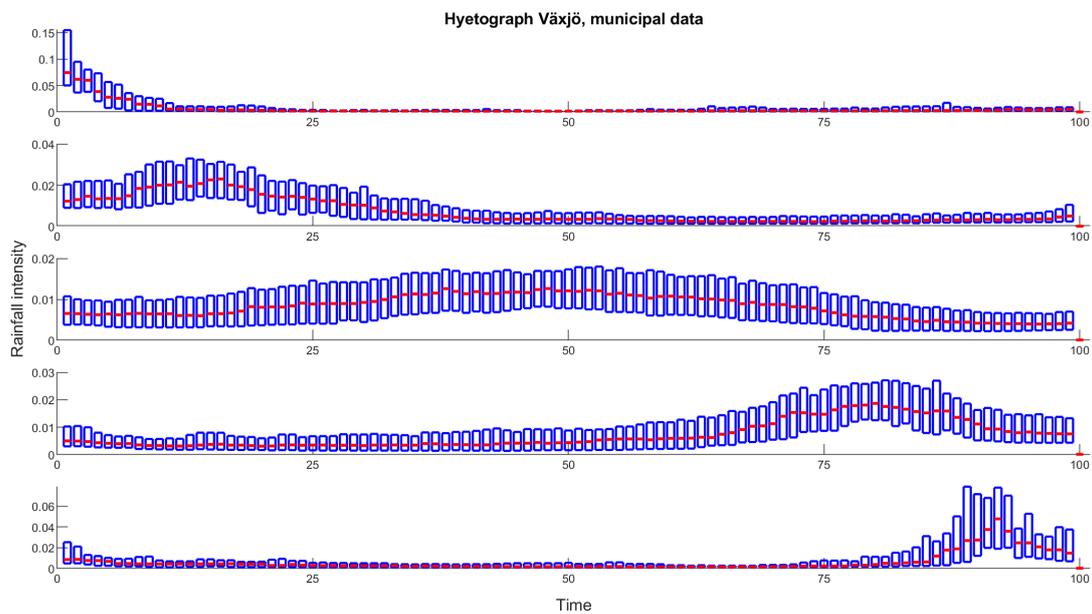
Figure 4.18. Hyetographs in five clusters for Växjö (SE), municipal data.

In some cases, a lack of rain events is apparent in the different clusters, seemingly independently of geographic location. In Figures D.5 & D.4, the hyetographs for Mora (Mid) and Gävle (Mid) are missing values for cluster 1. The hyetograph for Helsingborg (SW), when using the average over 30 realisations is missing values in cluster 3. The issue appears to be

42

solved when using 100 realisations instead. Similarly, hyetographs for some northern stations, while not showing empty clusters, have clusters with only one (or very few, depending on the total amount of events) event, which can be observed in Figure D.2, for example, where cluster 2 is created from only one rain event. These clusters are distinctive in the absence of a percentile spread, there being only one event clustered.

4.4 Regional differences

A key aspect of this study was to analyse hyetograph distributions in the four previously defined meteorological regions in Sweden. The regional differences are also shown in the present study. Note that only longer rain events, defined here as lasting 90 minutes or longer, are used in the analysis. In Table 4.3 below, regional hyetograph distributions for each meteorological region are shown, averaged over 30 realisations.

Table 4.3. Regional hyetograph distributions, averaged over 30 stochastic realisations. The last row displays the results found in Olsson et al. (2017).

Region	Distribution of hyetograph clusters 90+ min (%)					# of events
	Cluster number					
	1	2	3	4	5	
SW	26%	18%	16%	13%	27%	110[ii]
SE	23%	19%	16%	21%	21%	85
MID	33%	24%	7%	22%	14%	62
N	25%	13%	25%	20%	17%	48
Olsson et al. (2017)	18%	27%	40%	6%	8%	133

[ii] Borås included.

The results seem mixed: for one, regional differences seem to be reflected in hyetograph distributions to a certain extent for 90+ min rain events, even with so few data points. Definitive conclusions are, however, harder to reach. For instance, in the case of cluster number 1, regions SW, SE and N all exhibit quite similar distributions, around 23-26%, while being geographically distant (region N being in a subartic climate zone).

When compared to the results in Olsson et al. (2017) shown in the last row in Table 4.3, only one observation can be made: while they found 133 rain events within the 90+ min class for the whole of Sweden, the present study does provide significantly more events to cluster to empirical hyetographs. In terms of cluster distribution, no correlations or similarities between the distributions shown in Olsson et al. (2017) and this study's distributions are evident for any of the clusters.

Another comparison between hyetograph distributions was made, using all 30 realisations together when clustering rainfall data to hyetographs. This created, in effect, longer series with

even more data points. The results are shown in Table 4.4.

Table 4.4. Regional hyetograph distributions, 30 stochastic realisations clustered together. The last row displays the results found in Olsson et al. (2017).

Region	Distribution of hyetograph clusters 90+ min (%)					
	Cluster number					# of events
	1	2	3	4	5	
SW	28%	53%	11%	4%	4%	3289
SW[iii]	14%	28%	32%	6%	20%	1805
SE	9%	27%	4%	15%	44%	2531
MID	41%	35%	19%	0%	4%	1851
N	42%	30%	15%	5%	7%	1446
Olsson et al. (2017)	18%	27%	40%	6%	8%	133

[iii] Borås excluded.

The result presented in Table 4.4 above diverge completely from the earlier ones based on 30 realisations averaged (Table 4.3). They do, nevertheless, create a more nuanced picture over the distribution of hyetographs. Indeed, in this case, distributions in regions MID and N as well as in regions SW and SE seem to resemble each other respectively.

For the three calibration cities, Växjö, Helsingborg and Malmö, the analysis of trends and behaviours with 100 realisations were also made, in addition to the 30 realisations created for all series. This analysis was performed in order to test the hypothesis of additional realisations changing the results, as well as "creating" longer data series. The result of the clustering to hyetographs and their distributions are shown in Table 4.5. Comparing the hyetograph distributions for each city with the distributions found for their respective regions (averaged over 30 realisations, Table 4.3) shows the disparity in the results apparently caused by the number of recorded rain events. Indeed, neither Malmö and Helsingborg (SE) nor Växjö (SE) reflect the same hyetograph distributions found for their respective regions. For example while region SE displays 9% and 27% for clusters 1 and 2 respectively, Växjö has almost the opposite of 19% and 6%. This again invites to further analyses with a much higher number of stochastic realisations. Interestingly, hyetograph distributions for Helsingborg shown in Table 4.5 match the distributions found by Olsson et al. (2017).

Differences in hyetograph distributions are still exhibited, even for the individual calibration stations, lending strength to the findings of Olsson et al. (2017) and Litsmark (2020).

Table 4.5. Hyetograph distributions for Växjö, Helsingborg and Malmö. Here, raw data is disaggregated in 100 stochastic realisations.

Individual stations	Distribution of hyetograph clusters 90+ min (%)					
All (1 through 100)	Cluster number					# of events
	1	2	3	4	5	
Malmö	7%	40%	40%	4%	9%	841
Helsingborg	19%	21%	44%	10%	6%	981
Växjö	19%	6%	8%	59%	8%	924

Hyetograph distributions per realisation for each region can be found in Tables B.1 through B.5 in Appendix B. Distributions for each of the 12 stations studied in this project are shown in Table B.6. A degree of randomness between each realisation is obvious for all distributions and regions, owing to the stochastic nature of rainfall disaggregation.

5 DISCUSSION

5.1 Random cascade model - calibration of the model

A major component of this study was the calibration of the random cascade model and it subsequent application to disaggregate SMHI's 15 min rainfall data series. The feasibility of applying temporal disaggregation to such a small time scale was the subject of uncertainties from the start. The model calibration alleviated some concerns. Indeed, while using different data series than in J. Olsson (1998), J. Olsson (2012), and Güntner et al. (2001), model calibration still yielded comparable results. The behaviours of all calibration parameters were expected, and would seem to confirm the possibility of applying this specific model to higher temporal resolutions.

The results of the BDCs - the weight distrubutions $W_{x/x}$ shown as histograms in Section 4.1.2 - were expected, based on previous studies. Güntner et al. (2001) asserted that such a dominance towards the middle suggests a larger likelihood of rainfall volumes being more equally distributed at higher temporal resolutions. Licznar et al. (2011) however, while also finding a majority of histograms being dominated by BDCs at exactly 0.5, reasoned that this resulted from limitations of the rain gauges used. J. Olsson (1998), also found histograms dominated by 0.5 for the smallest time scales used in that study (8 minutes to 1 hour). However, their results were expected given the nature of his $W_{x/x}$ - histograms.

The analysis of changes and modulations at time scales below the mean duration or rainfall events reflect the internal structure of rainfall events: in the temperate climate of southern and central Sweden, this entails relatively evenly distributed precipitation. It is therefore reasonable to expect the weight distribution $W_{x/x}$ to be grouped around 0.5, as for smaller time scales, rainfalls during the first and last halves of the rainy period are similar (J. Olsson, 1998; Güntner et al., 2001).

Internal asymmetries do show in most of the histograms, lending strength the idea of modifying the expression of the weight distribution $W_{x/x}$ as mentioned in Section 3.3 above. Differences between position types, found in J. Olsson (1998) and Güntner et al. (2001) for instance, are not as clearly present in this study: empirical distributions seem more randomly distributed in the present study. This may be due to the higher temporal resolution, or even to the volume resolution of the tipping-bucket municipal data: J. Olsson (1998) used tipping-bucket gauges with a volume resolution of 0.033 mm, while volume resolutions of 0.2 mm are used in this study, ~ 6 times larger.

One conclusion reached by J. Olsson (2012) was the possibility of modelling the beta parameter a as a log-log linear function such as $log(a) = c_3 + c_4 * log(s)$ where s is the time scale and c_3 & c_4 are constants. While this appears true for the present study, the second conclusion presented by J. Olsson (2012), that a can be approximated as a constant is less obvious in this study, which invites further investigations in the matter.

Disaggregation results were also quite satisfactory as a whole, seemingly preserving the distributions of the defined rainfall parameters ED and EV when compared between municipal and

disaggregated data series.

The results presented in Section 4.1.4 above could very well show that the stochastic nature of the disaggregation process fails to reflect the extremes of Swedish precipitation patterns. However, it is important to keep in mind the limits of this study, the most important one being the different time intervals. While there is overlap between disaggregated (1995-2020) and municipal (1989-2004) series, a combined 22 years are not overlapping, as the calibration data starts 6 years before the SMHI data, and the latter continues past 2004 another 16 years. It is quite probable that some rain events captured during the time interval chosen for disaggregation were not present in the municipal data sets, and vice versa, skweing the results in one way or the other. All in all, the distributions shown with Q-Q plots seem to strengthen the results found earlier and graphically presented as histograms: EV and ED seem to be preserved to a certain extent in the disaggregation process, and quite well at lower ranges.

5.1.1 Training and testing of the RCM

One concern that should be raised involves the data sets used in this study. When constructing models to mirror or predict real behaviours, one should use two different data sets for model calibration ("training" the model), for verifying the model ("testing" the model), with a third data set sometimes used for the so-called validation. In this study, overlapping periods were used for calibrating and testing the model. Here, an ideal scenario would have been to train the model on municipal data from, for example, 1989-1999, continue by disaggregating SMHI 15 min data over 2000-2020 and finally verify the disaggregation results against municipal data from 2000-2020. Such a method would avoid information leakage as well as prevent overestimation of the competence of the model. This ideal scenario could not be applied owing to the rainfall data available in Sweden, but is relevant when discussing the model's ability and skill.

5.1.2 The choice of municipal data for model calibration

Again, an issue with the present study was the use of different time intervals for municipal (1989-2004) and SMHI (1995-2020) rainfall series. In order to certify the results presented in Section 4, a comparison between similar time series would be interesting. The data produced by Hernebring (2006) was used in a effort to be consistent with Olsson et al. (2017) that divided Sweden into the meteorological regions used in the present study, but the use of longer time series could also shed some more light to the applicability of the random cascade model in Sweden. One complication arose from using these data sets: for model calibration, cities with both a SMHI station as well as existing municipal data series had to be chosen. This effectively limited the study to three calibration cities exhibiting these properties.

Other institutions have networks of rain gauges throughout Sweden and databases which could be applied to another study. While one major reason for the use of temporal disaggregation of rainfall series in the first place remains the lack of historical data at certain locations of interest, it is acknowledged that comparing the results obtained in this study with calibration results from other databases would be beneficial.

A potential source of error in this study builds on the data series used. Indeed, as mentioned above, the municipal series used for model calibration stem from tipping-bucket rain gauges with a volume resolution of 0.2 mm. After model calibration, parameters from the municipal series are used for the disaggregation of SMHI's 15 min data - with a volume resolution of 0.1 mm. The effect of such a disparity in volume resolutions was not examined in detail in this study and could potentially lead to erroneous results. This should be investigated in another study, by using other calibration series with the same volume resolution for example.

5.2   Spatial transferability of calibration parameters

The results for the spatial transferability of calibration parameters within region SW were quite surprising, especially when opposed to those between regions SE and SW. That distributions between regions would appeared to be more similar than within a region is perplexing to say the least, and would require further investigation. Given the geographic locations of the three calibration cities, the opposite results would've been expected. Again, the effect of using different time intervals between municipal and disaggregated series should be assessed even in this case. The fact could be that extreme events were observed in Malmö for instance, that never happened in Helsingborg, explaining the skewed distributions. Indeed, this is reflected in earlier studies, as Olsson et al. (2017) also pointed out an 850 year storm over Malmö that occurred for a period of 48 hours in 2014 without affecting Helsingborg.

Another assumption to be tested was the spatial transferability of calibration parameters. Proven for certain climates by Econopouly et al. (1990) and Güntner et al. (2001), this theory had to be tested for Sweden. The results were mostly promising, especially between regions SE and SW. While some stochastic realisations yielded skewed distributions, especially for the parameter ED, most showed reasonable results. A tendency to overestimate smaller rain quantities while underestimating larger ones was present in comparisons between regions. Further testing would need to be performed, especially for regions MID and N. In the present study, it was assumed that calibration parameters from Växjö could be applied to the northern regions, but no quantitative research was made in this case. Further exploration of the possibility of using southern calibration parameters for the disaggregation of northen data series is warranted.

5.3   Hyetographs

The investigation of extreme precipitation in Sweden, conducted by Olsson et al. (2017) is quite unique for Sweden specifically, as similar studies are few and far apart. While in their case, hyetographs were clustered based on data covering Sweden in its entirety (based on the 15 municipal data series compiled by Hernebring (2006)), regional hyetographs were clustered here. The data sets used in this study also differ in the amounts of rain events collected, thanks to the use of disaggregation. The 90+ min hyetographs in Olsson et al. (2017) are clustered together using 133 events in total, while hundreds more are found and used in this study. For instance, a total of 305 events were studied in this study in the case of data averaged over 30 stochastic realisation, and 7633 events were used when clustering 30 realisations together. Despite these differences, their study yielded some conclusions that are applicable to the present study as

well.

First and foremost, Olsson et al. (2017) found that peaks are often located in the first half of the rain events: indeed, the occurrence of types 1 (the very beginning of the rain event) & 2 (the first quarter of the rain event) was almost three times larger than that of types 4 & 5 in their study. This fact is also reflected in the present study, where an overwhelming majority of hyetographs display their peaks in hyetographs of types 1 & 2, with the notable exceptions of Mora and Borås (see Figures D.5 & D.11 respectively), where hyetographs of type 1 are missing entirely, suggesting some anomalies in those cases.

Moreover, hyetographs in Olsson et al. (2017) displayed quite significant variability compared to the mean, for all hyetograph types, with even greater variations at or close to the peaks. The same observations can be made in the present study. The fact that hyetographs for both 30 and 100 realisations show relatively similar uncertainties around the peaks is an indication that more stochastic realisations should be made before clustering.

In short, the fact that intensity variations, peak arrival times and peak intensities seem to be preserved or emulated with a certain degree of fidelity is very promising. Further studies, perhaps with longer time series, closer in time, should be conducted in order to explore the possibility of rainfall data disaggregation in Sweden.

Lastly, while 30 realisation might be able to capture any variations in the stochastic process when in comes to calibration parameters according to Müller-Thomy (2020), as mentioned earlier in Section 3.3, a larger amount of realisations should be made when clustering to hyetograph. In a working meeting[3], Olsson speculated that hundreds, if not thousands of realisations could be made for the purpose of clustering empirical hyetographs in different regions, the purpose of which would be the creation of robust and representative hyetographs. Olsson's assumptions seem to be proven true in this study, as shown for instance in the case of Helsingborg: compare and contrast Figures D.12 & D.13.

5.3.1 Regional differences in hyetographs

The overarching aim of this study was to, through disaggregation of 15 min data series into 1 min series by a random cascade model, cluster hyetographs for different regions in Sweden in order to compare and contrast their distributions as well as confirm the existence of regional differences. This goal was achieved during the course of this study. Differences in hyetograph distributions between regions are undeniable, as shown in Tables 4.3, 4.4 and 4.5 as well as the results shown in Appendix B. However, these results should be questioned and cannot be accepted without any reflection. As is also shown in all tables found in Appendix B, hyetograph distributions exhibit great fluctuations between each stochastic realisation. The cause of the regional differences could very well be the stochastic nature of the disaggregation process, or even the K-means clustering, and not in regional differences rainfall characteristics and precipitation trends. This fact doesn't invalidate the results of this study, it simply reinforces the need for further studies using thousands of realisation for the clustering of empirical hyetographs.

---

[3]Olsson, J. (December 2020). Personal communication.

For the case of 30 realisations clustered together, shown earlier in Table 4.4, a relative resemblance can be seen for all clusters in the case of regions Mid and North, while the southern regions seem only to agree for the first two clusters. The difference between North and South is pronounced, and such a dichotomy would make sense given the meteorological and climate divisions in Sweden. The geography of Sweden, lends to more precipitation in regions SW and to a certain extent Mid: similarities between these two regions would've been expected in this study, but aren't found.

5.4   Delimitations

Choice of time series

A major objective with this study is to obtain high resolution rainfall data for stations where such data isn't available because of, for example, deficient equipment or gaps in data series. The study conducted by Hernebring (2006) collected tipping bucket measurements for 15 participating municipalities throughout Sweden with a temporal resolution of 1 minute. These series, spanning between 1989-2004, are used to calibrate the model used for the disaggregation and analysis of more contemporary data from SMHI's automatic 15-min stations ranging between August 1995 and June 2020. The difference in time series between calibration and disaggregation data may lead to poorer statistics when comparing rain characteristics such as EV or ED. Different kinds of events could happen during different intervals, or certain types of rainfall could happen for one time interval and not the other, skewing the results. However unlikely, the possibility that both calibration and disaggregation time series do give a representative sample of the climate at each station without omitting any possible rain events remains, but it is of importance to note this limitation.

Calibration data: volume resolutions

Another limit to this study is the use of rainfall series collected from different types of devices. While in all cases, data was collected by tipping bucket rain gauges (see Section 2 for more details), municipal data collection devices has a volume resolution of 0.2 mm, while SMHI's data collection devices used for the 15-min series has a volume resolution of 0.1 mm. The impact of such a difference is not studied here, but it is acknowledged that there might be one.

Rain events and seasonality

This study will focus first and foremost on extreme rain events. Hyetographs will therefore be generated for long rain events, defined as events lasting longer than 90 minutes, see Section 3.4 for more detailed information on rain event groups. Furthermore, seasonal variations are not analyzed in this study.

Noteworthy is a weakness in SMHI's 15-min data: taken with fixed 15-min intervals, measurements will tend to underestimate rainfall intensities for rain events with 15 minute duration or shorter. This undervaluation is estimated to be of the order of 15% for 10 mm events (Olsson et al., 2017).

6 CONCLUSIONS

The purpose of this study was twofold:

1) To assess the feasibility of using a random cascade model to disaggregate 15 min data series to a 1 min temporal resolution.

2) To ascertain the existence of regional differences in hyetographs i Sweden, as remarked upon by previous studies.

In order to answer point 1), and evaluate the possibility of disaggregation with a random cascade model to high temporal resolutions, two tests were performed. The first was, in essence, an appraisal of the reliability of the model calibration and the subsequent disaggregation results through statistical means. Based on the available data for model calibration, this study showed that the random cascade model performed quite well in terms of replicating rainfall characteristics as well as rainfall distributions. However, the delimitations of the study outlined in Section 5.4, should be reminded.

The second test was an analysis of the spatial transferability of the calibration parameters from one city to another. The results were in this case inconclusive: While in most cases, rainfall characteristics were preserved, two out of three cities, Växjö and Helsingborg, failed a Wilcoxon rank sum test comparing two data sets with different calibration parameters. Also, spatial transferability between southern and northern Sweden was not investigated.

To answer point 2), hyetographs for all twelve examined cities were clustered based on the average of at least 30 (or 100 in the cases of Växjö and Helsingborg) stochastic realisations of the random cascade model. From these, two main conclusions can be made:

- There are clear differences in hyetographs between region. The regional differentiation detailed in the literature (Olsson et al., 2017) is indeed present in this study.

- The differences found within the scope of this study might not be due to regional meteorological characteristics, but rather to the stochastic nature of the disaggregation process. Therefore, thousands more realisations should be computed and averaged in order to give a more accurate representation of regional differences in hyetographs.

All in all, the prospect of using random cascade models to create high temporal resolution rainfall data is promising and deserves to be studied in deeper detail. Likewise, while differences in hyetograph distributions were displayed, this study proved that additional data are needed to create reliable hyetographs.

6.1 Broader scope and future studies

While 30 realisations might capture any possible statistical variation in calibration parameters when calibrating the cascade model, as shown by Müller-Thomy (2020), it is preferable to use a far greater number of realisations when clustering empirical hyetographs. Using averages over thousands of stochastic realisations would produce a clearer and more significant comparison between regions. The limited number of realisations in this study was due to time limitations and a lack of computational power: a larger study using more powerful hardware could yield better results and would be desirable.

6.1.1 The disdrometer: an alternative way of collecting 1 min rainfall data

Other methods for rainfall collection exist and are being tested in Sweden today. A noteworthy example would be the optical disdrometer used for instance by Uppsala University (CELSIUS, 2021), that has the ability of collecting rainfall data at 1 min temporal resolution. An optical disdrometer has the ability to not only measure rainfall intensities, but also distributions of individual raindrops as well as their velocity, as explained by Van Quyen et al. (2017). Data from such databases are not used in this study however, but could be the basis for model calibration in future studies and could be used for "teaching" the model how to behave and control its results to contemporary 1 min rainfall data series.

6.1.2 Weighted flood risk assessment

Regional hyetographs such as the ones created in this study could be used as a basis for further hydrological simulations, especially in the field of urban hydrology. For example, C. J. Olsson (2019) assessed the influence of storm movement and temporal distribution of rainfall on pluvial flooding. In that study, the very same empirical hyetographs developed by Olsson et al. (2017) mentioned in the present study, were used as rain input. By instead using regional hyetographs for instance, weighted flood risk assessments could be performed. Regional differentiation would give a more nuanced analysis of potential flooding risks, which could lead to more accurate and suitable measures.

References

Bárdossy, A. and Pegram, G.G.S (2016). "Space-time conditional disaggregation of precipitation at high resolution via simulation". *Water Resources Research,* vol. 52, pp. 920–937. DOI: `https://doi.org/10.1002/2015WR018037`.

Berne, A., Delrieu, G., Creutin, J.D., and Obled, C. (2004). *Journal of Hydrology,* vol. 299, pp. 166–179. DOI: `https://doi.org/10.1016/j.jhydrol.2004.08.002`.

CELSIUS Institutionen för geovetenskaper, Uppsala Universitet. (2021). *Uppsala disdrometer.* Available from: `http://celsius.met.uu.se/?pageid=17&meny=7`. Accessed: 2021-01-28.

Econopouly, T.W., Davis, D.R., and Woolhiser, D.A. (1990). "Parameter transferability for a daily rainfall disaggregation model". *Journal of Hydrology,* vol. 118, pp. 209–228. DOI: `https://doi.org/10.1016/0022-1694(90)90259-Z`.

GEONOR (n.d.). *T-200B Series Precipitation Gauge. Manual for 600-mm, 1000-mm & 1500-mm capacity gauges.* Available from: `http://geonor.com/live/products/weather-instruments/t-200b-weather-precipitation-rain-gauge/`. Accessed: 2020-11-04.

Güntner, A., Olsson, J., Calver, A., and Gannon, B. (2001). "Cascade-based disaggregation of continuous rainfall time series: the influence of climate". *Hydrology and Earth System Sciences,* vol. 5(2), pp. 145–164. ISSN: 1607-7938.

Gupta, V.K. and Waymire, E.C. (1993). "A Statistical Analysis of Mesoscale Rainfall as a Random Cascade". *Journal of Applied Meteorology and Climatology,* vol. 32, pp. 251–267. DOI: `https://doi.org/10.1175/1520-0450(1993)032<0251:ASAOMR>2.0.CO;2`.

Haan, C.T. (1977). *Statistical Methods in Hydrology* Iowa State University Press, Ames, Iowa.

Helsel, D.R. and Hirsch, R.M. (2002). *Statistical Methods in Water Resources* Techniques of Water-Resources Investigations of the United States Geological Survey, USGC, pp. 19–118.

Hernebring, C. (2006). "10års-regnets återkomst - förr och nu". *VA-Forsk rapport,* nr. 2006-04.

Hershenhorn, J. and Woolhiser, D.A. (1987). "Disaggregation of daily rainfall". *Journal of Hydrology,* vol. 95, pp. 299–322.

Jebari, S., Berndtsson, R., Olsson, J., and Bahri, A. (2012). "Soil erosion estimation based on rainfall disaggregation". *Hydrology and Earth System Sciences,* vol. 436-437, pp. 102–110. DOI: `https://doi.org/10.1016/j.jhydrol.2012.03.001`.

Licznar, P., Lomotowski, J., and Rupp, D.E. (2011). "Random cascade driven rainfall disaggregation for urban hydrology: An evaluation of six models and a new generator". *Atmospheric Research,* vol. 99, pp. 563–578. DOI: `https://doi.org/10.1016/j.atmosres.2010.12.014`.

Lima, M.I.P. de (1998). *Multifractals and the temporal structure of rainfall (PhD Dissertation).* Available from: `https://library.wur.nl/WebQuery/wurpubs/40455`.

Litsmark, S. (2020). *Investigating the relationship between circulation patterns and cloudburst character in a changing climate (Master's thesis)*. Available from: `http://uu.diva-portal.org/smash/record.jsf?pid=diva2%3A1415510&dswid=-5682`.

Mackay, D.C.J. (2003). *Information theory, inference & learning algorithms* New York, United States: Cambridge University Press, pp. 284–292.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., and Kjeldsen, T.R. (2014). "Review of trend analysis and climate change projections of extreme precipitations and floods in Europe". *Journal of Hydrology,* vol. 519, pp. 3634–3650. DOI: `https://doi.org/10.1016/j.jhydrol.2014.11.003`.

Mandelbrot, B. (1974). "Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier". *Journal of Fluid Mechanics,* vol. 62, pp. 331–358. DOI: `https://doi.org/10.1017/S0022112074000711`.

Mathworks (2020a). *K-means clustering.* Available from: `https://se.mathworks.com/help/stats/kmeans.html`. Accessed: 2020-12-21.

Mathworks (2020b). *Quantile-quantile plot.* Available from: `https://se.mathworks.com/help/stats/qqplot.html`. Accessed: 2021-01-28.

Mathworks (2020c). *Wilcoxon rank sum.* Available from: `https://se.mathworks.com/help/stats/ranksum.html`. Accessed: 2020-12-20.

Menabde, M., Seed, A., and Pegram, G. (1999). "A simple scaling model for extreme rainfall". *Water Resources Research,* vol. 35, pp. 335–359. DOI: `https://doi.org/10.1029/1998WR900012`.

Menabde, M. and Sivapalan, M. (2000). "Modeling of rainfall time series and extremes using bounded random cascades and Levy-stable distributions". *Water Resources Research,* vol. 36, pp. 3293–3300. DOI: `https://doi.org/10.1029/2000WR900197`.

Molnar, P. and Burlando, P. (Oct. 2005). "Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model". *Atmospheric Research* 77, pp. 137–151. DOI: `10.1016/j.atmosres.2004.10.024`.

Müller-Thomy, H. (2020). "Temporal rainfall disaggregation using a micro-canonical cascade model: possibilities to improve the autocorrelation". *Hydrology and Earth System Sciences,* vol. 24, pp. 169–188. DOI: `https://doi.org/10.5194/hess-24-169-2020`.

Müller, H. and Haberlandt, U. (2018). "Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology". *Journal of Hydrology,* vol. 556, pp. 847–864. DOI: `https://doi.org/10.1016/j.jhydrol.2016.01.031`.

Olsson, C. J. (2019). *The influence of storm movement and temporal variability of rainfall on urban pluvial flooding: 1D-2D modelling with empirical hyetographs and CDS-rain (Master's thesis)*. Available from: `http://uu.diva-portal.org/smash/record.jsf?pid=diva2%3A1331080&dswid=2095`.

Olsson, J. (1998). "Evaluation of a scaling cascade model for temporal rainfall disaggregation". *Hydrology and Earth System Sciences,* vol. 2, pp. 19–30. DOI: `https://doi.org/10.5194/hess-2-19-1998`.

Olsson, J. (2012). *Random Cascade Model: documentation,* tech. rep. SMHI.

Olsson, Berg, P., Eronn, A., Simonsson, L., Södling, J., Wern, L., and Yang, W. (2017). "Extremregn i nuvarande och framtida klimat". *Klimatologi,* vol. 47. ISSN: 1654-2258.

Over, T.M. and Gupta, V.K. (1994). "Statistical Analysis of Mesoscale Rainfall: Dependence of a Random Cascade Generator on Large-Scale Forcing". *Journal of Applied Meteorology,* vol. 33, pp. 1526–1542. DOI: `https://doi.org/10.1175/1520-0450(1994) 033<1526:SAOMRD>2.0.CO;2`.

Over, T.M. and Gupta, V.K. (1996). "A space-time theory of mesoscale rainfall using random cascades". *Journal of Geophysical Research Atmospheres,* vol. 101, pp. 26319–26331. DOI: `https://doi.org/10.1029/96JD02033`.

Schertzer, D. and Lovejoy, S. (1987). "Physical Modeling and Analysis of Rain and Clouds by Anisotropic Scaling Multiplicative Processes". *Journal of Geophysical Research,* vol. 92, pp. 9693–9714. DOI: `https://doi.org/10.1029/JD092iD08p09693`.

Schilling, W. (1991). *Atmospheric Research,* vol. 27, pp. 5–21. DOI: `https://doi.org/10.1016/0169-8095(91)90003-F`.

SMHI (2017). *Skyfall och rotblöta.* Available from: `https://www.smhi.se/kunskapsbanken/rotblota-1.17339`. Accessed: 2021-01-27.

SMHI (2018). *Hur mäts nederbörd?* Available from: `https://www.smhi.se/kunskapsbanken/meteorologi/hur-mats-nederbord-1.637`. Accessed: 2020-09-23.

SMHI (2020). *Nederbördsmängd (15 min): alla stationer.* Available from: `https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer/#param= precipitation15MinutesSum,stations=all`. Accessed: 2020-08-30.

Statistikmyndigheten, SCB (2019). *Folkmängd per tätort, andelar i procent efter tätort och år.* Available from: `https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START_ _MI__MI0810__MI0810A/MI0810Tatort03/table/tableViewLayout1/`. Accessed: 2021-01-27.

Van Quyen, L. T., Thang, N. M., Hong Vu, N., Truyen, N. T., Kiesewetter, D., and Malyugin, V. (2017). "The Optical Disdrometer". *Advances in Wireless and Optical Communications (RTUWO),* DOI: `https://doi.org/10.1109/RTUWO.2017.8228499`.

Willems, P., Olsson, J., Arnbjerg-Nielsen, K., Beecham, S., Pathirana, A., Bülow Gregersen, I., Madsen, H., and Nguyen, V.T.V (2012). *Impacts of Climate Change on Rainfall Extremes and Urban Drainage Systems* IWA Publishing, pp. 14–17.
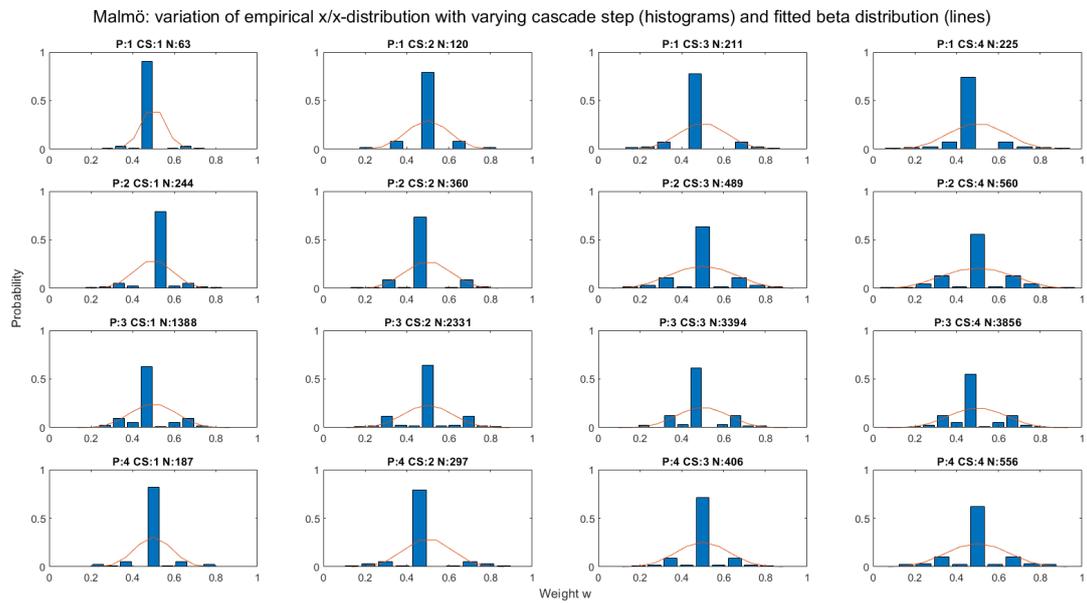
Appendix A Model calibration

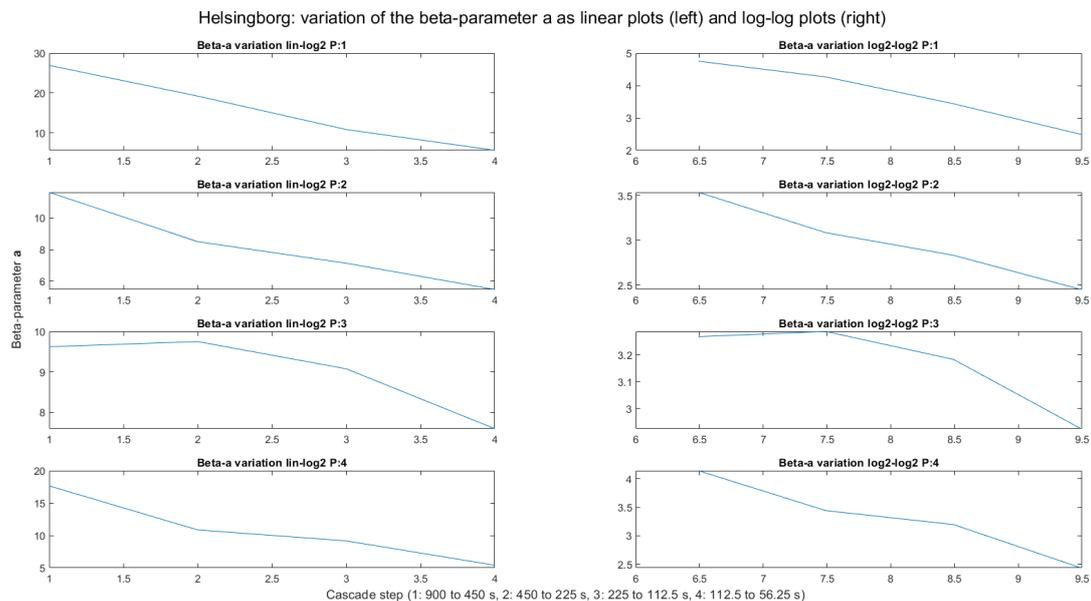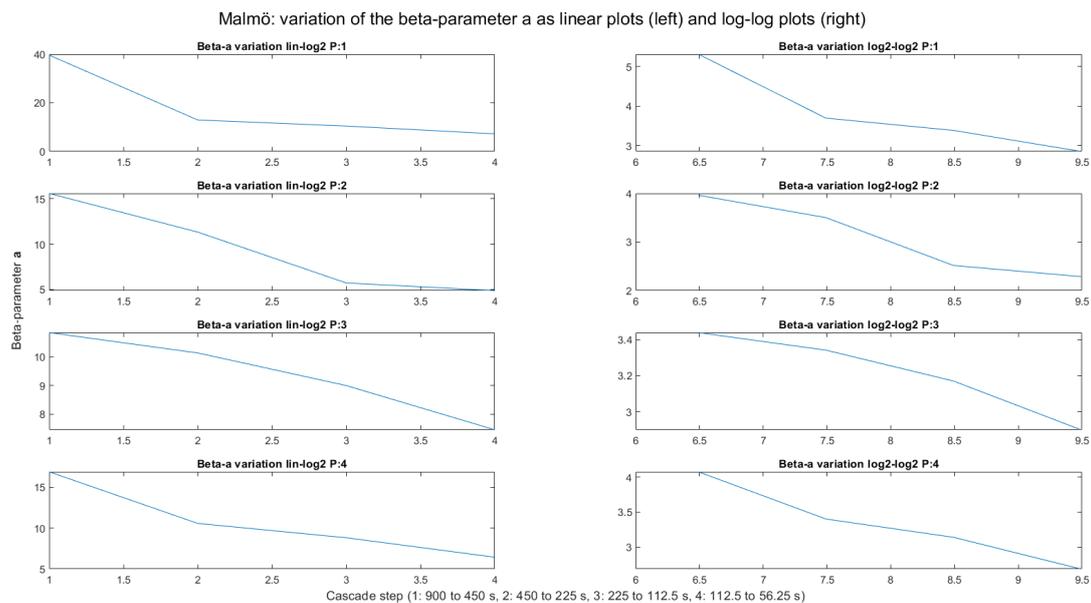*$W_{x/x}$* **- histograms for Helsingborg and Malmö**
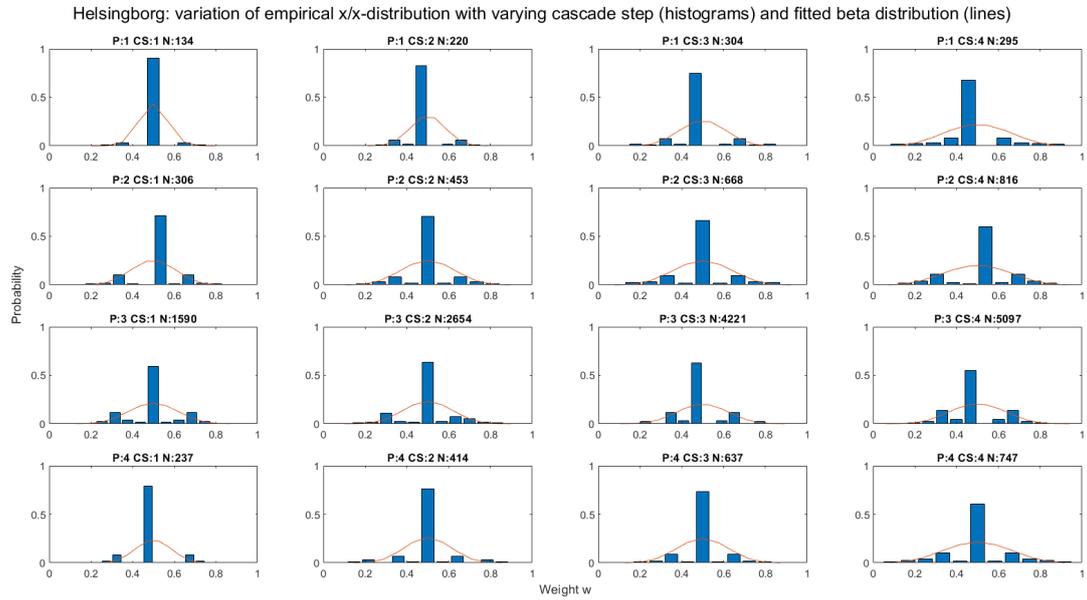
(a) Helsingborg station

(b) Malmö station

Figure A.1.   Variation of empirical x/x-distributions with cascade step (histograms) and a fitted beta distribution (line) for Helsingborg and Malmö stations. *P* denotes the position type (1: isolated, 2: starting, 3: enclosed, 4: ending), *CS* the cascade step (1: 900-450 s, 2: 450-225 s, 3: 225-112.5, 4: 112.5-56.25) and *N* the total amount of x/x distributions for each position type and cascade step.

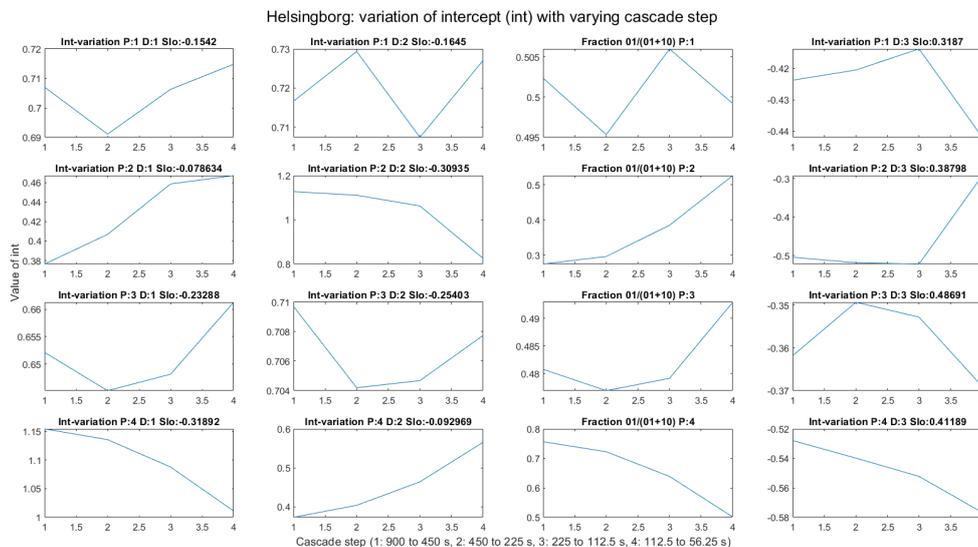# Variations of the beta parameter *a* for Helsingborg and Malmö

(a) Helsingborg station

(b) Malmö station

Figure A.2. Variation of the beta-parameter *a* with cascade step for Helsingborg and Malmö stations. *P* denotes the position type (1: isolated, 2: starting, 3: enclosed, 4: ending).

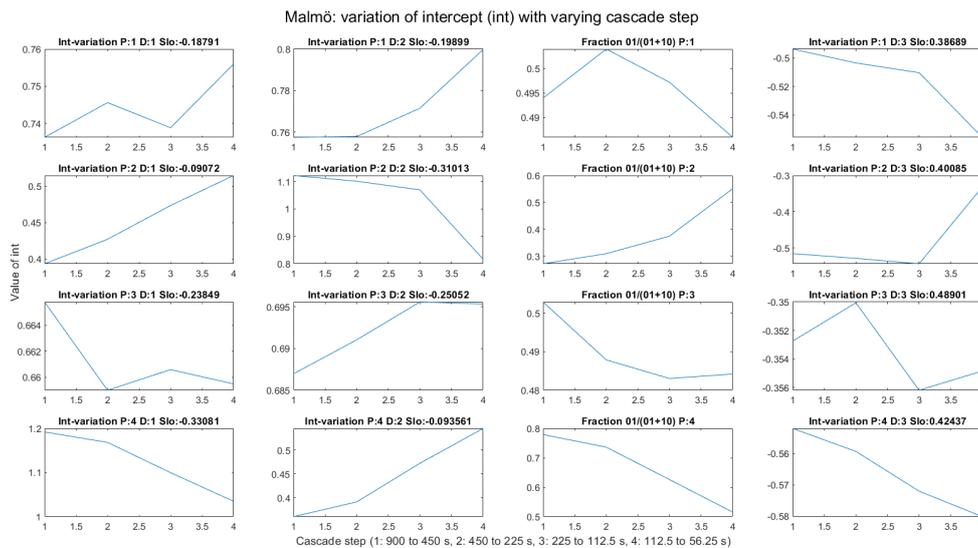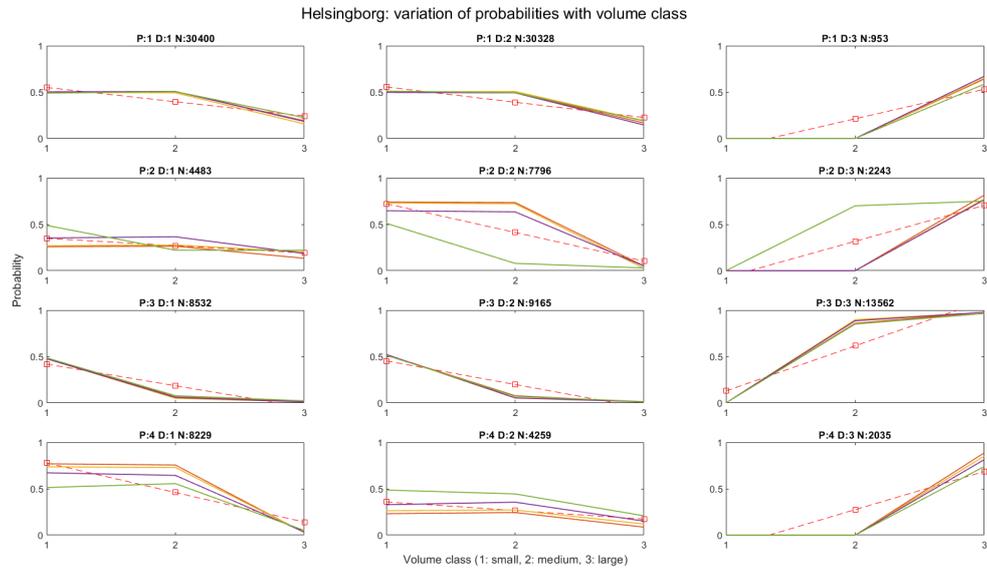**Variations of intercept *int* for Helsingborg and Malmö**

(a) Helsingborg station

(b) Malmö station

Figure A.3. Variation of intercept *int* with cascade step for Helsingborg and Malmö stations. *P* denotes the position type (1: isolated, 2: starting, 3: enclosed, 4: ending), *D* denotes the type of distribution (1: 0/1, 2: 1/0, 3: x/x) and *Slo* the mean slope for each position and distribution.

Probabilities for Helsingborg and Malmö

(a) Helsingborg station
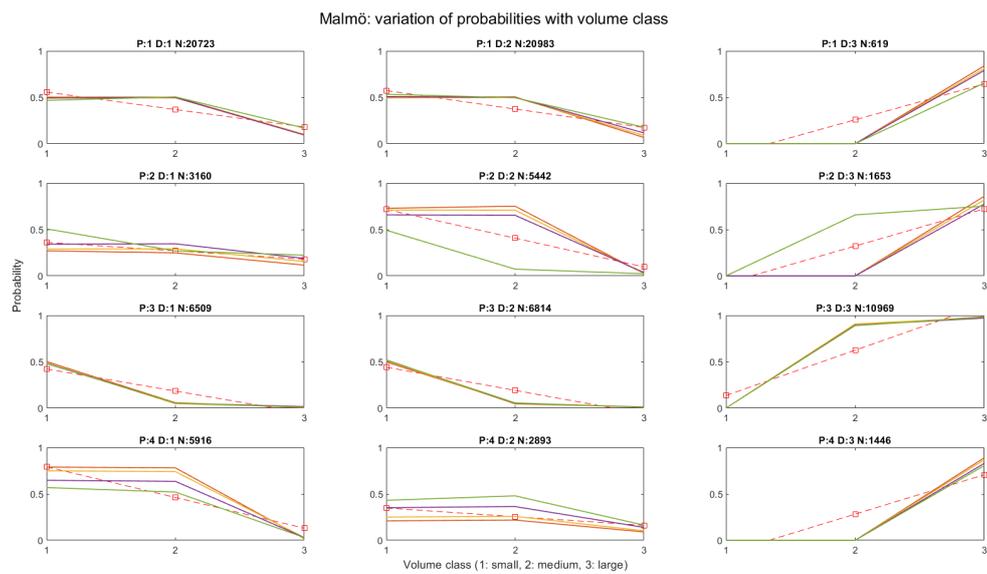
(b) Malmö station

Figure A.4. Variation of probabilities with volume class for Helsingborg and Malmö stations. *P* denotes the position type (1: isolated, 2: starting, 3: enclosed, 4: ending), *D* denotes the type of distribution (1: 0/1, 2: 1/0, 3: x/x) and *N* is the total amount of periods for each position and distribution type. The different solid coloured lines represent the different cascade steps during disaggregation. Orange: 1st, Yellow: 2nd, Purple: 3rd, Green: 4th. The dashed red line is the mean of all cascade steps.

Table A.1. Calibration variables: alim, intm21 & ac.

	Helsingborg			Malmö			Växjö		
alim	7.4919			7.4919			7.4919		
intm21	0.4381			0.4472			0.4719		
ac	3.4309	2.9201	3.1434	3.3095	2.9222	3.1370	3.7745	3.1140	3.2219

Table A.2. Calibration variable: ao.

	Helsingborg				Malmö				Växjö			
ao	25.946	11.535	10.123	17.813	39.641	15.590	10.854	16.859	26.351	11.222	10.420	17.472
	19.138	8.5481	9.8656	11.077	12.937	11.329	10.143	10.561	20.939	9.3010	11.636	14.713
	11.352	7.2560	9.1630	9.1256	10.431	5.7235	9.0018	8.8114	16.440	7.7198	9.5190	10.598
	5.7738	5.5398	7.6323	5.4135	7.2206	4.8832	7.4551	6.4373	4.4456	5.8713	8.0225	6.4060

Table A.3. Calibration variable: slom.

	Helsingborg			Malmö			Växjö		
slom	0	0	0.3177	0	0	0.3869	0	0	0.4048
	-0.0857	0	0.4006	-0.0921	0	0.4126	-0.0988	0	0.4074
	0	0	0.4872	0	0	0.4890	0	0	0.4877

Table A.4. Calibration variable: c1, c2, c3 & c4.

	Helsingborg			Malmö			Växjö		
c1	0	0	-0.3925	0	0	-0.3618	0	0	-0.3999
	0	0	-0.6933	0	0	-0.6981	0	0	-0.6768
	0	0	-0.3290	0	0	-0.3435	0	0	-0.3548
c2	0	0	-0.0039	0	0	-0.0193	0	0	-0.0175
	0	0	0.0236	0	0	0.0219	0	0	0.0202
	0	0	-0.0036	0	0	-0.0012	0	0	-1.7671e-05
c3	7.5435	7.4694	3.5804	15.796	7.7045	4.0748	6.8732	5.4918	3.1883
c4	-0.4385	-0.5589	-0.0371	-1.6155	-0.5677	-0.0978	-0.3317	-0.2594	0.0297

Table A.5. Observed probabilities for Växjö, P(0/1), P(1/0) & P(X/X).

Växjö											
p01o				p10o				pxxo			
val (:,:,1)											
0.4954	0.2507	0.4866	0.7609	0.5046	0.7493	0.5134	0.2391	0	0	0	0
0.4962	0.2953	0.4853	0.7014	0.5038	0.7047	0.5147	0.2986	0	0	0	0
0.5074	0.3947	0.4806	0.6301	0.4926	0.6053	0.5194	0.3699	0	0	0	0
0.5044	0.4778	0.4968	0.5334	0.4956	0.5222	0.5032	0.4666	0	0	0	0
val (:,:,2)											
0.5036	0.2524	0.0611	0.7629	0.4964	0.7476	0.0602	0.2371	0	0	0.8788	0
0.4904	0.3036	0.0546	0.7215	0.5096	0.6964	0.0511	0.2785	0	0	0.8944	0
0.5050	0.3901	0.0538	0.6262	0.4950	0.6099	0.0619	0.3738	0	0	0.8843	0
0.4998	0.2629	0.0552	0.5188	0.5002	0.0596	0.0594	0.4812	0	0.6775	0.8854	0
val (:,:,3)											
0.0784	0.1000	0.0121	0.0278	0.0441	0.0410	0.0077	0.0864	0.8775	0.8590	0.9801	0.8858
0.0746	0.1561	0.0136	0.0305	0.0896	0.0437	0.0150	0.0884	0.8358	0.8003	0.9714	0.8810
0.1011	0.2083	0.0141	0.0222	0.0831	0.0426	0.0121	0.1362	0.8157	0.7491	0.9738	0.8416
0.1405	0.2296	0.0127	0.0395	0.1502	0.0220	0.0114	0.2067	0.7092	0.7484	0.9759	0.7539

Table A.6. Observed probabilities for Helsingborg, P(0/1), P(1/0) & P(X/X).

Helsingborg											
p01o				p10o				pxxo			
val (:,:,1)											
0.4954	0.2582	0.4794	0.7708	0.5006	0.7410	0.5206	0.2292	0	8.197e-04	0	0
0.4895	0.2677	0.4769	0.7387	0.5018	0.7323	0.5231	0.2613	0	0	0	0
0.5027	0.3513	0.4775	0.6736	0.4973	0.6487	0.5225	0.3264	0	0	0	0
0.4895	0.4883	0.4869	0.5145	0.5105	0.5117	0.5131	0.4855	0	0	0	0
val (:,:,2)											
0.5055	0.2651	0.0589	0.7586	0.4964	0.7476	0.0602	0.2414	0	0	0.8701	0
0.4933	0.2760	0.0497	0.7294	0.5096	0.6964	0.0511	0.2706	0	0	0.8976	0
0.5084	0.3639	0.0595	0.6444	0.4950	0.6099	0.0619	0.3556	0	0	0.8857	0
0.5057	0.2208	0.0780	0.5563	0.5002	0.0596	0.0594	0.4437	0	0.6997	0.8510	0
val (:,:,3)											
0.1961	0.1324	0.0133	0.0264	0.1716	0.0514	0.0085	0.0868	0.6324	0.8162	0.9782	0.8868
0.1577	0.1979	0.0114	0.0251	0.1905	0.0243	0.0122	0.1190	0.6518	0.7778	0.9764	0.8559
0.1853	0.1859	0.0107	0.0349	0.1451	0.0508	0.0117	0.1475	0.6696	0.7633	0.9777	0.8176
0.2233	0.2213	0.0212	0.0500	0.1891	0.0296	0.0136	0.2058	0.5875	0.7490	0.9652	0.7443

Table A.7. Observed probabilities for Malmö, P(0/1), P(1/0) & P(X/X).

Malmö											
p01o				p10o				pxxo			
val (:,:,1)											
0.4923	0.2707	0.5022	0.7903	0.5077	0.7293	0.4978	0.2097	0	0	0	0
0.5060	0.2900	0.4868	0.7494	0.4940	0.7100	0.5132	0.2506	0	0	0	0
0.4947	0.3416	0.4805	0.6477	0.5053	0.6584	0.5195	0.3523	0	0	0	0
0.4672	0.5072	0.4829	0.5685	0.5328	0.4928	0.5171	0.4315	0	0	0	0
val (:,:,2)											
0.4956	0.2482	0.0582	0.7815	0.5044	0.7518	0.0509	0.2185	0	0	0.8909	0
0.5019	0.2918	0.0479	0.7414	0.4981	0.7082	0.0444	0.2586	0	0	0.9078	0
0.5006	0.3464	0.0538	0.6360	0.4994	0.6536	0.0518	0.3640	0	0	0.8944	0
0.5046	0.2694	0.0543	0.5204	0.4954	0.0728	0.0561	0.4796	0	0.6578	0.8897	0
val (:,:,3)											
0.0933	0.1162	0.0060	0.0190	0.0667	0.0246	0.0090	0.0905	0.8400	0.8592	0.9849	0.8905
0.1014	0.1549	0.0115	0.0292	0.0878	0.0251	0.0124	0.1023	0.8108	0.8200	0.9761	0.8684
0.0936	0.1886	0.0165	0.0307	0.1161	0.0365	0.0124	0.1391	0.7903	0.7750	0.9712	0.8303
0.1686	0.2240	0.0104	0.0304	0.1773	0.0234	0.0097	0.1638	0.6541	0.7526	0.9799	0.8058

Appendix B Hyetograph distributions

Table B.1. Hyetograph distributions for region SE. In this case, raw data is disaggregated in 30 stochastic realisations.

Region SE	Distribution of hyetograph clusters 90+ min (%)					
Realisation	Cluster number					# of events
	1	2	3	4	5	
All (1 through 30)	9%	27%	4%	15%	44%	2531
1	42%	28%	15%	6%	9%	88
2	21%	6%	16%	2%	55%	95
3	3%	26%	58%	12%	1%	86
4	1%	5%	3%	81%	10%	82
5	49%	3%	15%	24%	9%	79
6	11%	41%	21%	1%	26%	85
7	13%	1%	9%	15%	62%	79
8	30%	6%	19%	44%	1%	80
9	29%	3%	56%	11%	1%	90
10	32%	39%	19%	5%	5%	84
11	30%	7%	2%	44%	17%	89
12	48%	1%	4%	8%	39%	85
13	32%	2%	26%	39%	1%	88
14	4%	52%	22%	2%	20%	81
15	67%	4%	15%	7%	7%	89
16	1%	83%	8%	2%	6%	87
17	41%	10%	6%	3%	40%	78
18	4%	59%	5%	13%	19%	83
19	1%	8%	14%	74%	3%	79
20	9%	3%	18%	35%	35%	91
21	1%	9%	10%	57%	23%	81
22	4%	1%	31%	46%	18%	78
23	6%	81%	6%	6%	1%	85
24	18%	1%	39%	26%	16%	85
25	27%	1%	1%	12%	59%	76
26	84%	3%	5%	1%	7%	86
27	9%	48%	3%	15%	25%	88
28	3%	30%	17%	17%	33%	90
29	1%	6%	23%	4%	66%	79
30	68%	6%	1%	17%	8%	88
Mean (30 realisations)	23%	19%	16%	21%	21%	85

Table B.2. Hyetograph distributions for region SW. In this case, raw data is disaggregated in 30 stochastic realisations.

Region SW	Distribution of hyetograph clusters 90+ min (%)					
Realisation	Cluster number					# of events
	1	2	3	4	5	
All (1 through 30)	28%	53%	11%	4%	4%	3289
1	2%	10%	33%	52%	4%	114
2	28%	1%	62%	3%	6%	116
3	38%	2%	4%	49%	7%	116
4	88%	1%	2%	3%	6%	116
5	43%	3%	3%	8%	44%	110
6	3%	3%	61%	31%	2%	119
7	76%	11%	2%	8%	3%	109
8	2%	1%	5%	37%	55%	111
9	42%	1%	14%	12%	32%	111
10	26%	53%	2%	13%	6%	115
11	1%	43%	12%	1%	42%	97
12	26%	5%	12%	5%	52%	98
13	20%	59%	6%	3%	13%	107
14	1%	21%	7%	23%	48%	118
15	16%	7%	22%	53%	2%	108
16	77%	10%	8%	1%	4%	108
17	27%	39%	8%	25%	1%	103
18	6%	14%	26%	1%	53%	107
19	28%	11%	3%	1%	58%	112
20	28%	46%	13%	8%	5%	100
21	32%	1%	65%	1%	1%	105
22	1%	12%	27%	1%	59%	103
23	50%	3%	7%	1%	39%	112
24	24%	7%	5%	8%	57%	102
25	23%	55%	10%	1%	11%	111
26	7%	51%	32%	9%	1%	117
27	20%	4%	9%	12%	55%	113
28	21%	50%	11%	12%	6%	110
29	6%	6%	18%	1%	69%	115
30	5%	25%	2%	1%	68%	106
Mean (30 realisations)	26%	18%	16%	13%	27%	110

Table B.3. Hyetograph distributions for region SW (Helsingborg and Malmö only). In this case, raw data is disaggregated in 30 stochastic realisations.

Region SW (Hls & Mlm only)	Distribution of hyetograph clusters 90+ min (%)					
Realisation	Cluster number					# of events
	1	2	3	4	5	
All (1 through 30)	14%	28%	32%	6%	20%	1805

Table B.4. Hyetograph distributions for region MID. In this case, raw data is disaggregated in 30 stochastic realisations.

Region MID	Distribution of hyetograph clusters 90+ min (%)					
Realisation	Cluster number					# of events
	1	2	3	4	5	
All (1 through 30)	41%	35%	19%	0%	4%	1851
1	34%	48%	3%	9%	5%	64
2	34%	12%	7%	43%	3%	58
3	60%	5%	2%	23%	11%	62
4	5%	13%	5%	73%	4%	65
5	53%	33%	8%	5%	2%	64
6	43%	28%	2%	6%	22%	65
7	64%	22%	7%	3%	3%	59
8	5%	84%	8%	2%	2%	64
9	23%	6%	29%	23%	19%	62
10	8%	47%	4%	33%	8%	49
11	24%	10%	38%	25%	3%	63
12	26%	2%	3%	46%	23%	61
13	62%	6%	2%	18%	12%	66
14	57%	33%	3%	3%	3%	61
15	54%	12%	4%	13%	16%	68
16	21%	23%	2%	26%	28%	53
17	1%	63%	4%	30%	1%	70
18	40%	22%	2%	7%	30%	60
19	42%	2%	3%	3%	49%	59
20	5%	45%	2%	37%	12%	60
21	45%	22%	2%	20%	12%	60
22	6%	67%	5%	21%	2%	63
23	35%	32%	6%	20%	7%	71
24	7%	31%	10%	35%	18%	72
25	52%	3%	3%	17%	24%	58
26	27%	6%	40%	17%	10%	63
27	2%	12%	2%	33%	52%	66
28	47%	8%	3%	32%	10%	60
29	79%	2%	7%	9%	3%	58
30	18%	12%	7%	47%	16%	59
Mean (30 realisations)	33%	24%	7%	22%	14%	62

Table B.5. Hyetograph distributions for region N. In this case, raw data is disaggregated in 30 stochastic realisations.

Region N	Distribution of hyetograph clusters 90+ min (%)					
Realisation	Cluster number					# of events
	1	2	3	4	5	
All (1 through 30)	42%	30%	15%	5%	7%	1446
1	2%	40%	34%	12%	12%	50
2	13%	6%	58%	19%	4%	48
3	55%	30%	2%	2%	11%	47
4	3%	9%	53%	31%	4%	45
5	7%	7%	2%	43%	41%	46
6	7%	5%	44%	20%	24%	55
7	57%	7%	18%	11%	7%	44
8	55%	4%	10%	27%	4%	52
9	31%	20%	9%	2%	38%	48
10	13%	9%	61%	15%	2%	54
11	2%	9%	58%	27%	2%	45
12	35%	2%	16%	6%	41%	49
13	40%	4%	11%	3%	42%	45
14	40%	11%	40%	2%	7%	45
15	28%	6%	23%	38%	5%	47
16	6%	14%	3%	6%	71%	49
17	5%	5%	14%	69%	7%	43
18	14%	2%	2%	70%	12%	56
19	24%	5%	2%	22%	47%	55
20	45%	6%	9%	38%	2%	47
21	4%	45%	6%	2%	21%	49
22	68%	8%	2%	10%	12%	50
23	38%	21%	8%	20%	13%	53
24	30%	4%	51%	13%	2%	47
25	21%	21%	49%	7%	2%	43
26	9%	4%	81%	2%	4%	55
27	33%	31%	21%	13%	2%	48
28	2%	17%	40%	36%	5%	42
29	41%	21%	15%	21%	2%	47
30	31%	7%	20%	13%	29%	45
Mean (30 realisations)	25%	13%	25%	20%	17%	48

Table B.6. Hyetograph distributions for each station. In this case, raw data is disaggregated in 30 stochastic realisations.

Individual stations	Distribution of hyetograph clusters 90+ min (%)					
All (1 through 30)	Cluster number					# of events
	1	2	3	4	5	
Malmö	23%	25%	8%	38%	7%	826
Helsingborg	43%	3%	32%	22%	0%	979
Borås	52%	0%	11%	36%	0%	1484
Växjö	20%	8%	26%	6%	41%	930
Horn	1%	10%	11%	46%	32%	855
Adelsö	12%	46%	5%	32%	4%	746
Gävle	0%	9%	55%	25%	11%	500
Mora	28%	0%	19%	6%	46%	726
Sundsvall	7%	28%	1%	44%	19%	625
Vilhelmina	3%	21%	53%	4%	18%	403
Lycksele	14%	64%	21%	1%	1%	474
Arvidsjaur	46%	11%	16%	5%	22%	569

Table B.7. Hyetograph distributions for Växjö (SE) and Helsingborg (SW). Shown here are 1-min municipal series used for model calibration.

Municipal stations	Distribution of hyetograph clusters 90+ min (%)					
	Cluster number					# of events
	1	2	3	4	5	
Helsingborg (1991-2004)	43%	12%	30%	11%	6%	478
Växjö (1989-2004)	19%	6%	22%	47%	6%	592
Olsson et al. (2017)	18%	27%	40%	6%	8%	133

Appendix C   Statistical measurements (calibration municipalities)

Table C.1. Statistical measurements for two scenarios for Växjö: calibrated against municipal data from Växjö (Vxj/vxj) and against municipal data from Helsingborg (Vxj/hls). Disaggregated data averaged over 30 realisations.

Statistical measures	Vxj/vxj		Vxj/hls	Δ
	Municipal data	Disaggregated data	Disaggregated data	
N_{nz}	49986	87068	86068	-982
$Median_{nz}$	0.20	0.10	0.10	0.00
$Mean_{nz}$	0.21	0.11	0.12	0.01
$Skew_{nz}$	11.2	13.0	13.06	0.04
Std_{nz}	0.09	0.07	0.07	0.00
Var_{nz}	0.01	0.00	0.00	0.00

Table C.2. Wilcoxon rank sum, EV

Wilcoxon rank sum	Vxj/vxj vs Vxj/hls		Hls/hls vs Hls/mlm		Mlm/mlm vs Mlm/vxj	
EV	H_0	p	H_0	p	H_0	p
1	1	0.0004	0	0.1107	0	0.7372
2	1	0.0001	1	0.0030	0	0.3533
3	1	0.0338	1	0.0230	0	0.2371
4	1	0.0271	0	0.2982	0	0.6215
5	1	0.0052	0	0.1711	0	0.8525
6	1	0.0306	0	0.1063	0	0.3465
7	0	0.0844	1	0.0009	0	0.4803
8	1	0.0025	0	0.2401	0	0.2694
9	1	0.0085	0	0.2404	0	0.6651
10	1	0.0005	0	0.4355	0	0.7839
11	0	0.0958	1	0.0052	0	0.4484
12	1	0.0172	1	0.0001	0	0.9537
13	1	0.0058	0	0.0649	0	0.1998
14	1	0.0044	0	0.0903	0	0.1634
15	1	0.0002	0	0.0575	0	0.6773
16	0	0.0559	1	0.0300	0	0.7299
17	1	0.0002	1	0.0260	0	0.9800
18	1	0.0042	1	0.0043	0	0.5194
19	0	0.0667	1	0.0028	0	0.7238
20	1	0.0007	0	0.1693	0	0.5091
21	1	0.0013	1	0.0005	0	0.9939
22	1	0.0319	1	0.0152	0	0.8699
23	1	0.0075	1	0.0202	0	0.2189
24	1	0.0005	0	0.1078	0	0.5226
25	1	0.0184	1	0.0419	0	0.8058
26	1	0.0002	0	0.1372	0	0.1848
27	1	0.0184	1	0.0407	0	0.2849
28	1	0.0230	1	0.0398	0	0.0906
29	1	0.0004	1	0.0006	0	0.6758
30	0	0.1163	1	0.0302	0	0.4583
Realisations not rejected (%)	17		43		100	

Table C.3. Wilcoxon rank sum, ED

Wilcoxon rank sum	Vxj/vxj vs Vxj/hls		Hls/hls vs Hls/mlm		Mlm/mlm vs Mlm/vxj	
ED	H_0	p	H_0	p	H_0	p
1	1	0.0002	0	0.0630	0	0.7810
2	1	0.0000	1	0.0024	0	0.5014
3	1	0.0031	1	0.0126	0	0.4773
4	1	0.0095	0	0.2298	0	0.6624
5	1	0.0015	0	0.1013	0	0.8654
6	1	0.0089	0	0.0618	0	0.1385
7	1	0.0289	1	0.0002	0	0.4173
8	1	0.0001	0	0.1164	0	0.2635
9	1	0.0032	0	0.0915	0	0.6243
10	1	0.0000	0	0.2400	0	0.7808
11	1	0.0153	1	0.0020	0	0.3670
12	1	0.0048	1	0.0001	0	0.8996
13	1	0.0001	1	0.0337	0	0.0804
14	1	0.0001	0	0.0505	0	0.1718
15	1	0.0000	1	0.0226	0	0.5211
16	1	0.0329	1	0.0058	0	0.3856
17	1	0.0000	1	0.0107	0	0.9748
18	1	0.0011	1	0.0019	0	0.4829
19	1	0.0369	1	0.0010	0	0.6039
20	1	0.0002	0	0.0537	0	0.4330
21	1	0.0006	1	0.0001	0	0.9633
22	1	0.0201	1	0.0054	0	0.6916
23	1	0.0018	1	0.0062	0	0.1876
24	1	0.0000	1	0.0378	0	0.4015
25	1	0.0009	1	0.0143	0	0.7233
26	1	0.0000	0	0.0720	0	0.1394
27	1	0.0024	1	0.0337	0	0.1363
28	1	0.0085	1	0.0045	0	0.1211
29	1	0.0000	1	0.0002	0	0.5603
30	1	0.0484	1	0.0202	0	0.6403
Realisations not rejected (%)	0		33		100	

Appendix D   Hyetographs
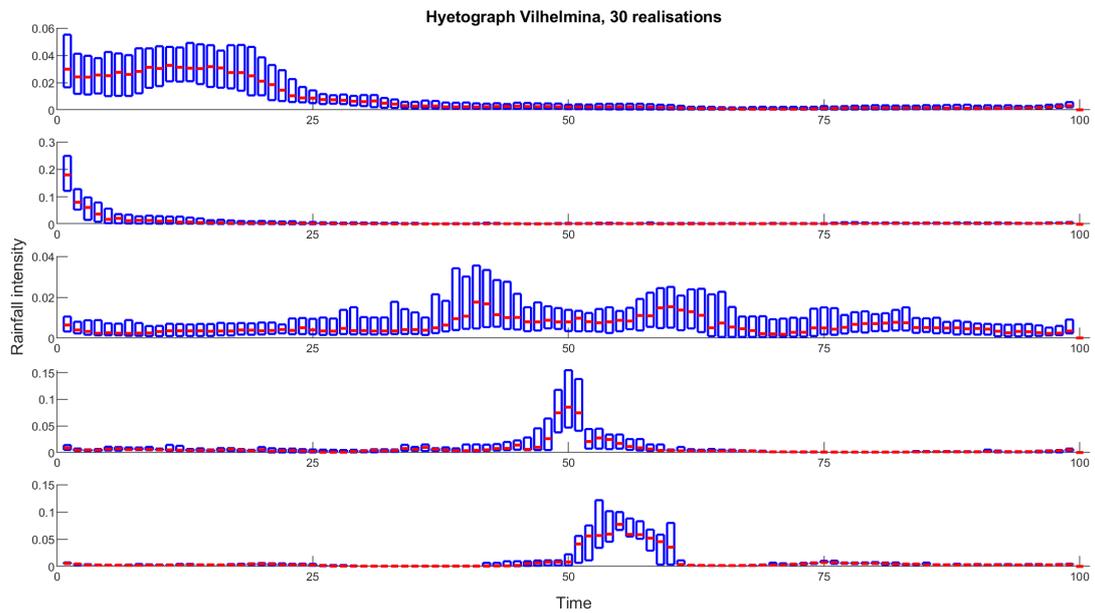
Figure D.1. Hyetographs in five clusters for Vilhemina (N), 30 realisations averaged
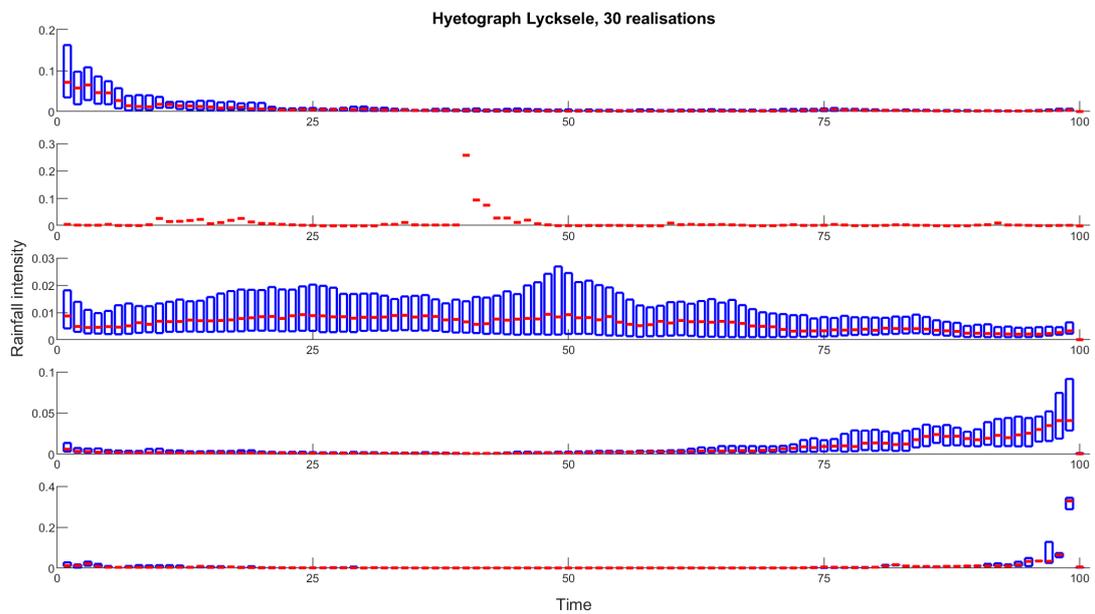
Figure D.2. Hyetographs in five clusters for Lycksele (N), 30 realisations averaged
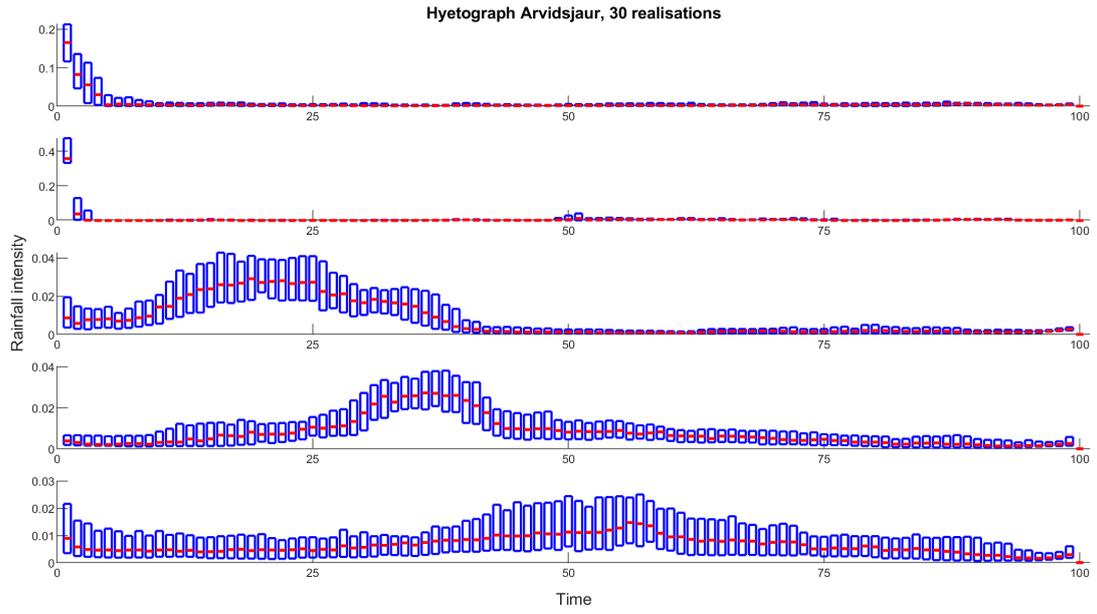
Figure D.3. Hyetographs in five clusters for Arvidsjaur (N), 30 realisations averaged
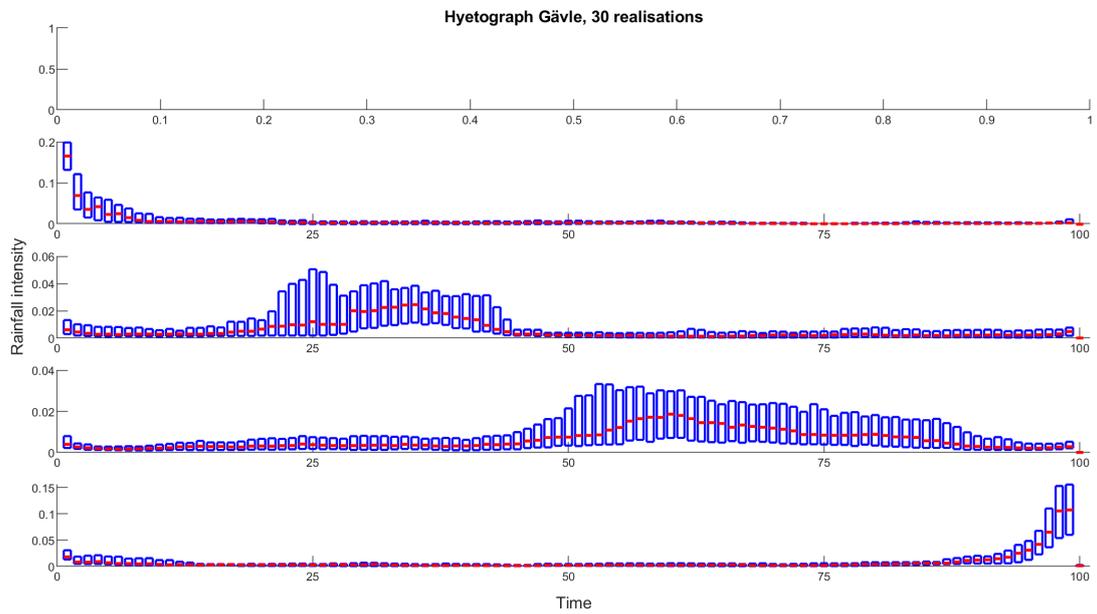
Figure D.4. Hyetographs in five clusters for Gävle (Mid), 30 realisations averaged
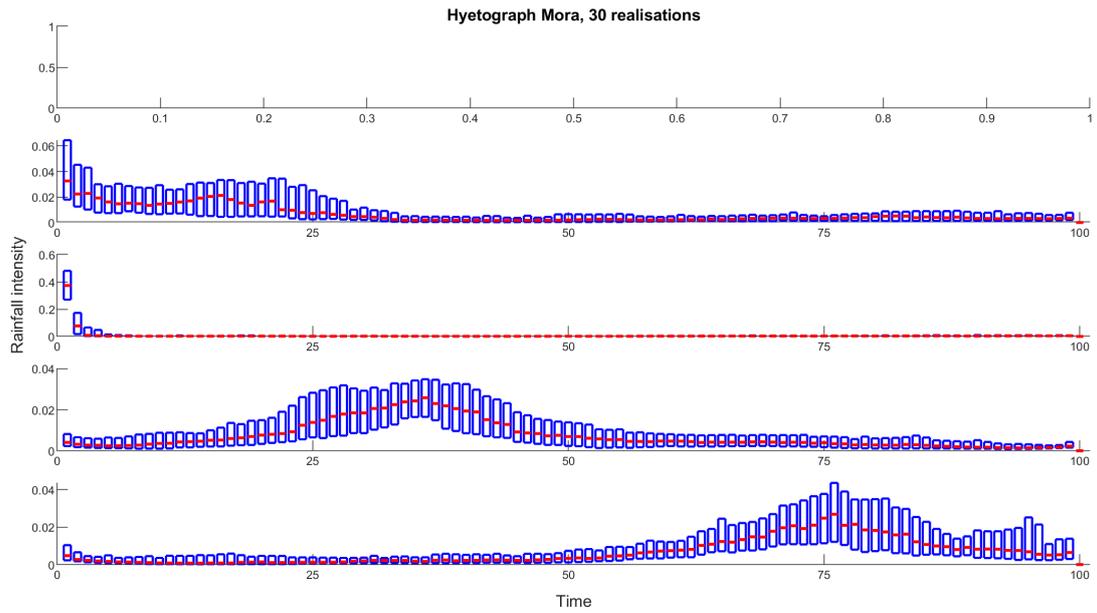
Figure D.5. Hyetographs in five clusters for Mora (Mid), 30 realisations averaged
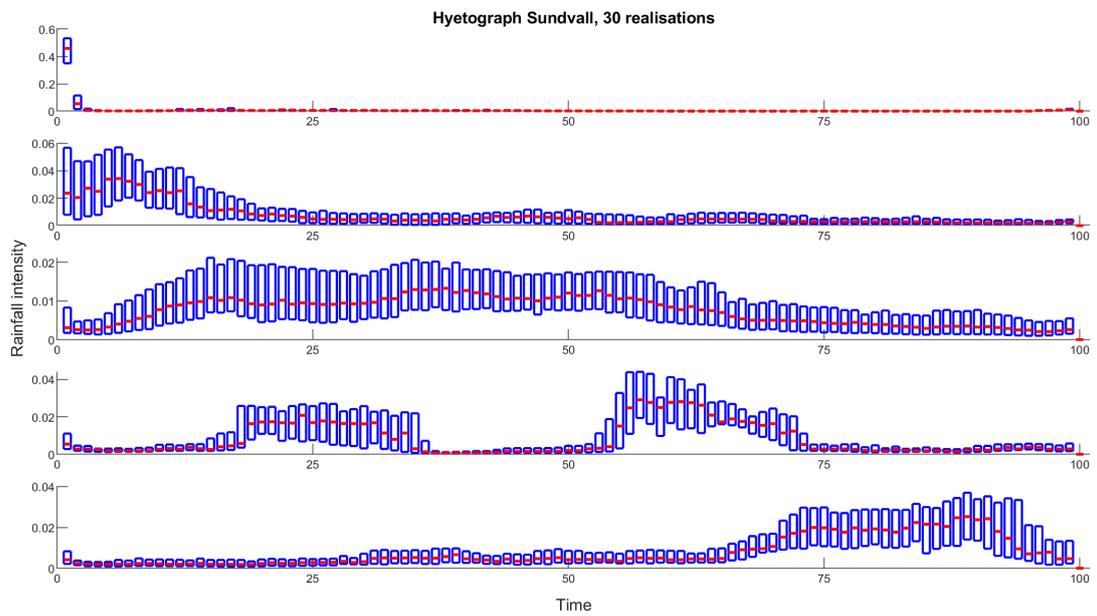
Figure D.6. Hyetographs in five clusters for Sundsvall (Mid), 30 realisations averaged
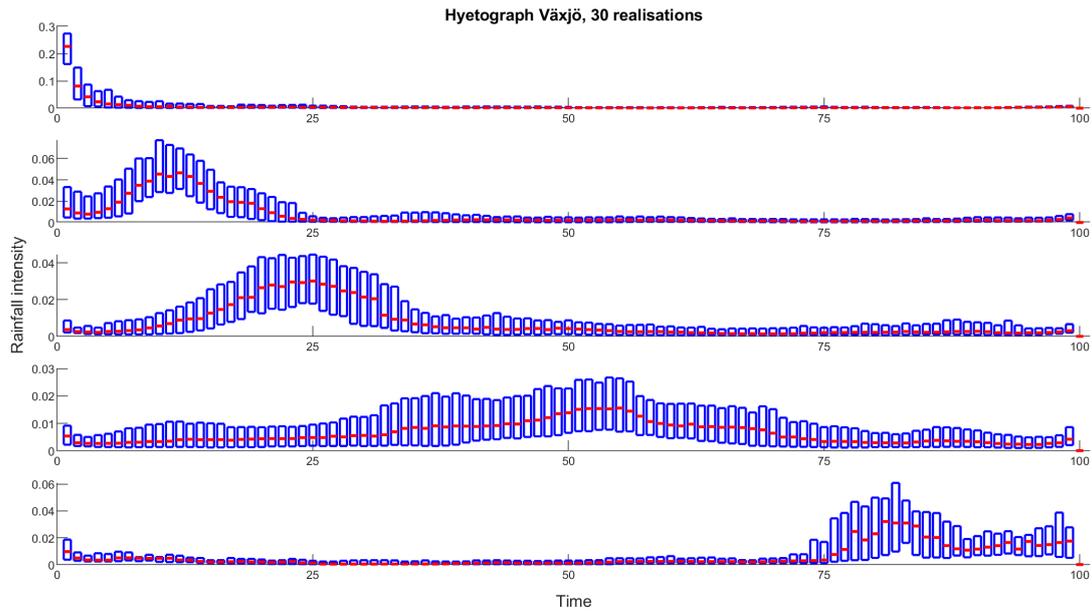
Figure D.7. Hyetographs in five clusters for Växjö (SE), 30 realisations averaged
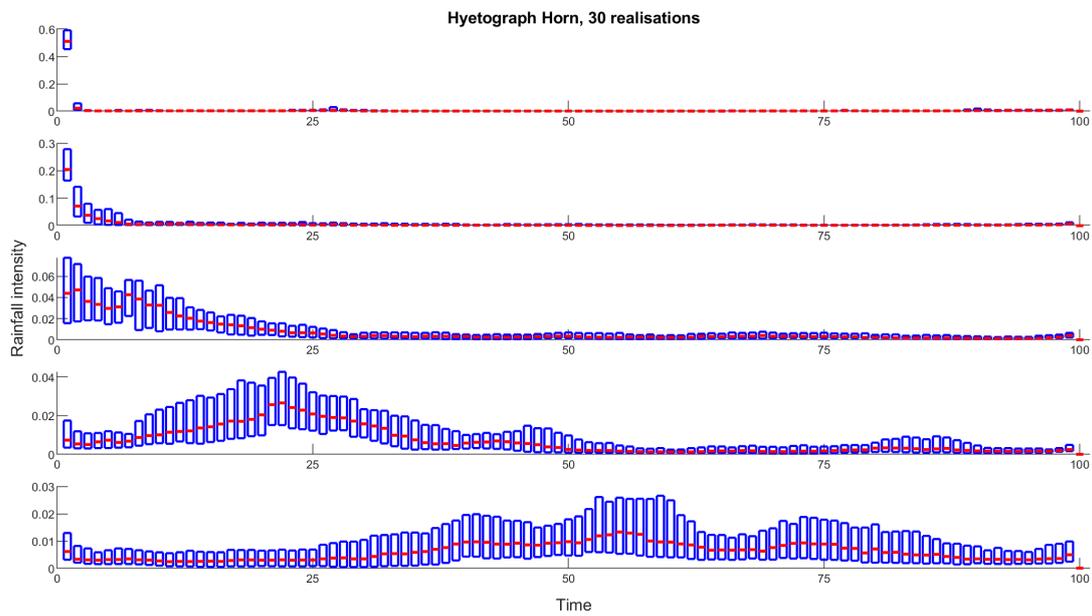
Figure D.8. Hyetographs in five clusters for Horn (SE), 30 realisations averaged
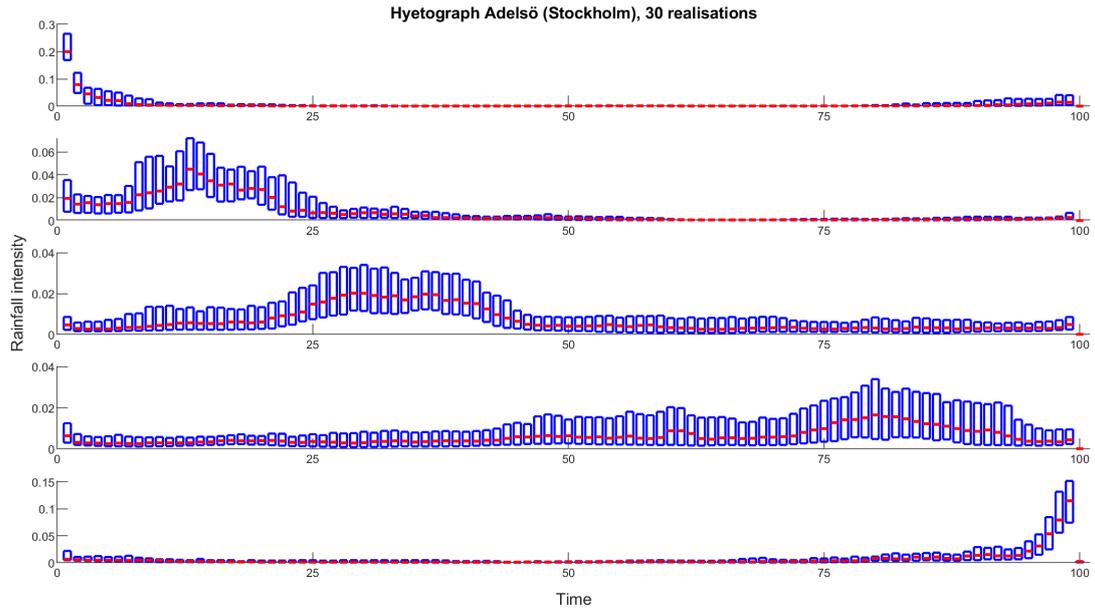
Figure D.9. Hyetographs in five clusters for Adelsö (SE), 30 realisations averaged
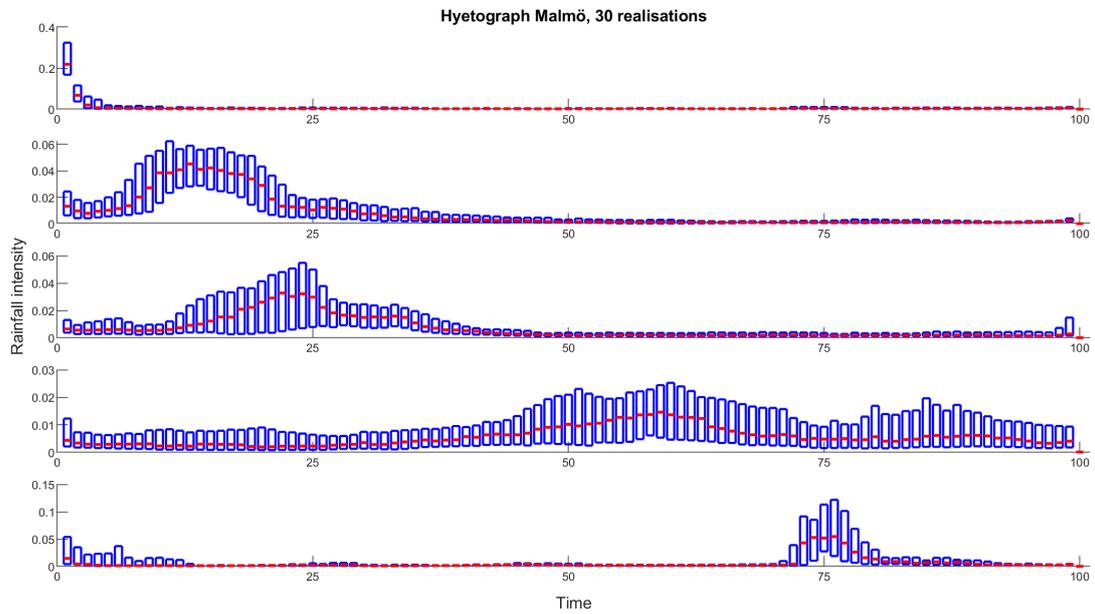
Figure D.10. Hyetographs in five clusters for Malmö (SW), 30 realisations averaged

77
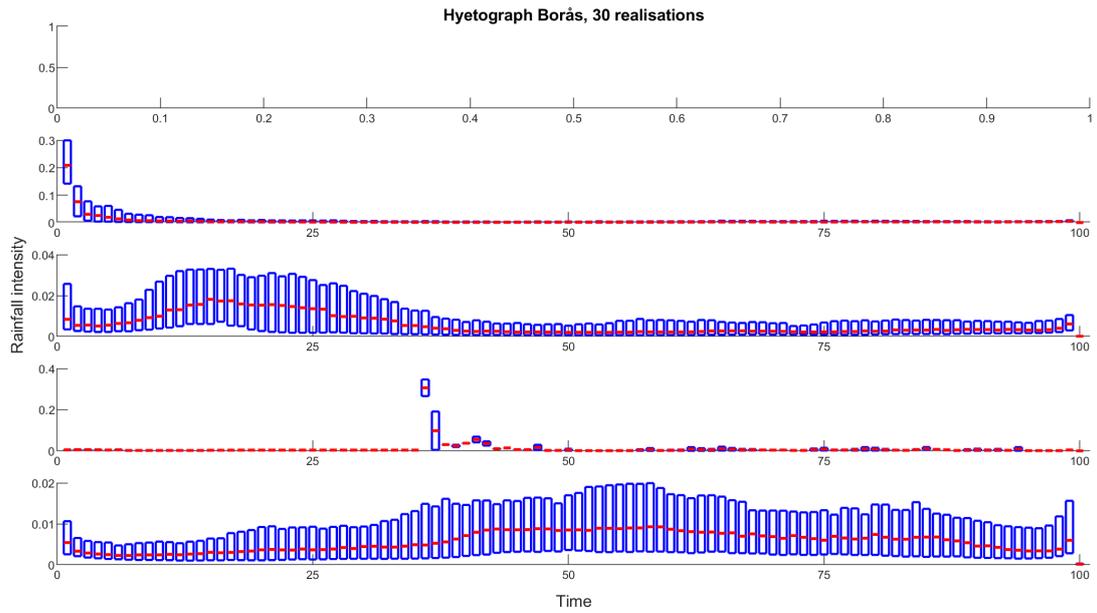
Figure D.11. Hyetographs in five clusters for Borås (SW), 30 realisations averaged
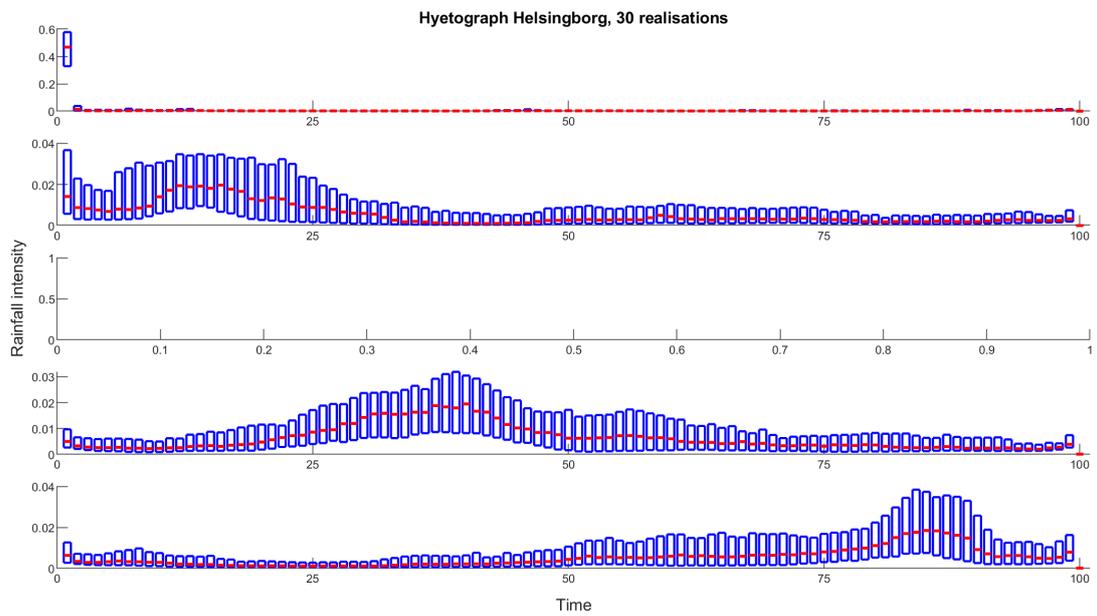
Figure D.12. Hyetographs in five clusters for Helsingborg (SW), 30 realisations averaged
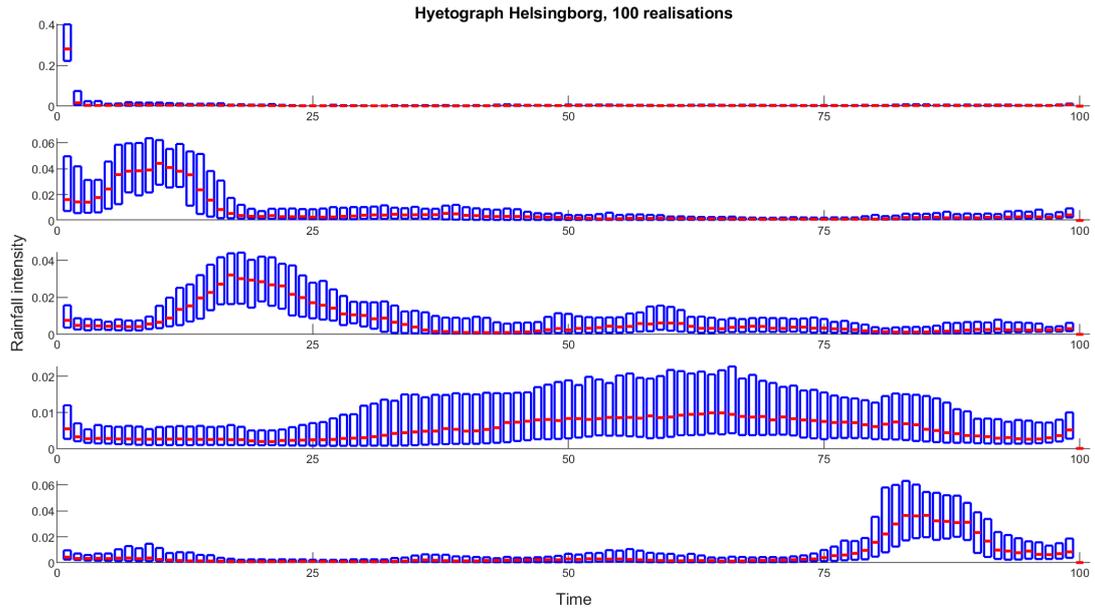
Figure D.13. Hyetographs in five clusters for Helsingborg (SW), 100 realisations averaged
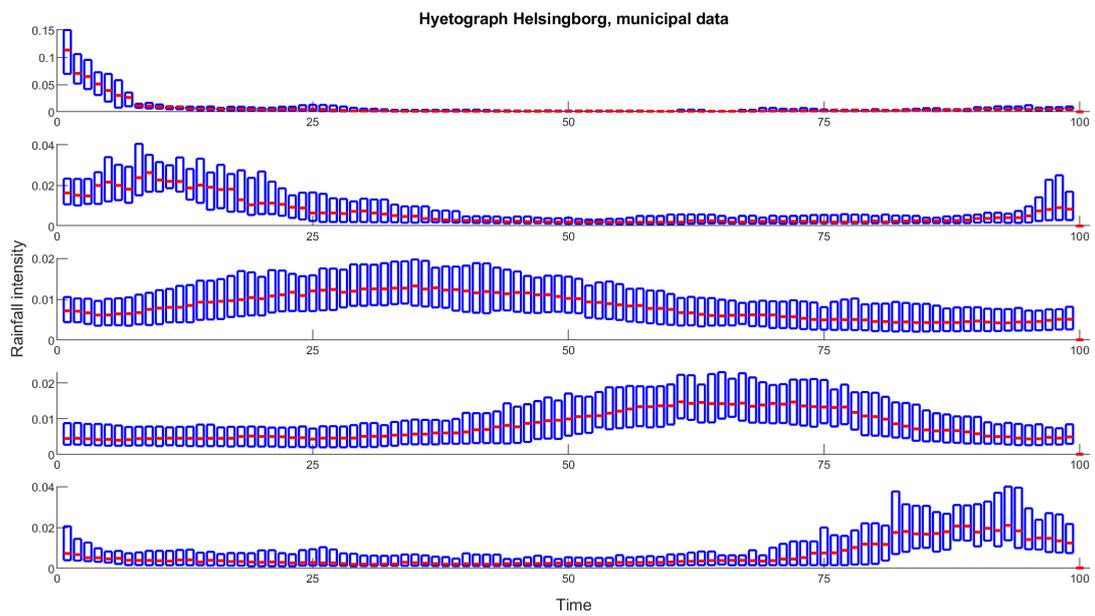
Figure D.14. Hyetographs in five clusters for Helsingborg (SW), municipal 1-minute data (calibration)

79
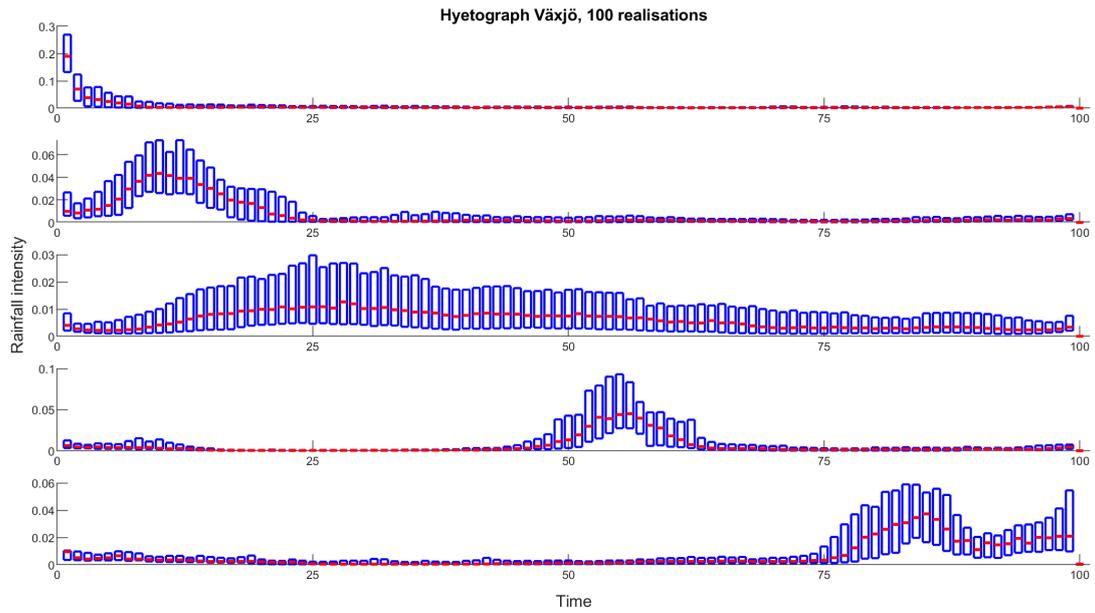
Figure D.15. Hyetographs in five clusters for Växjö (SE), 100 realisations averaged
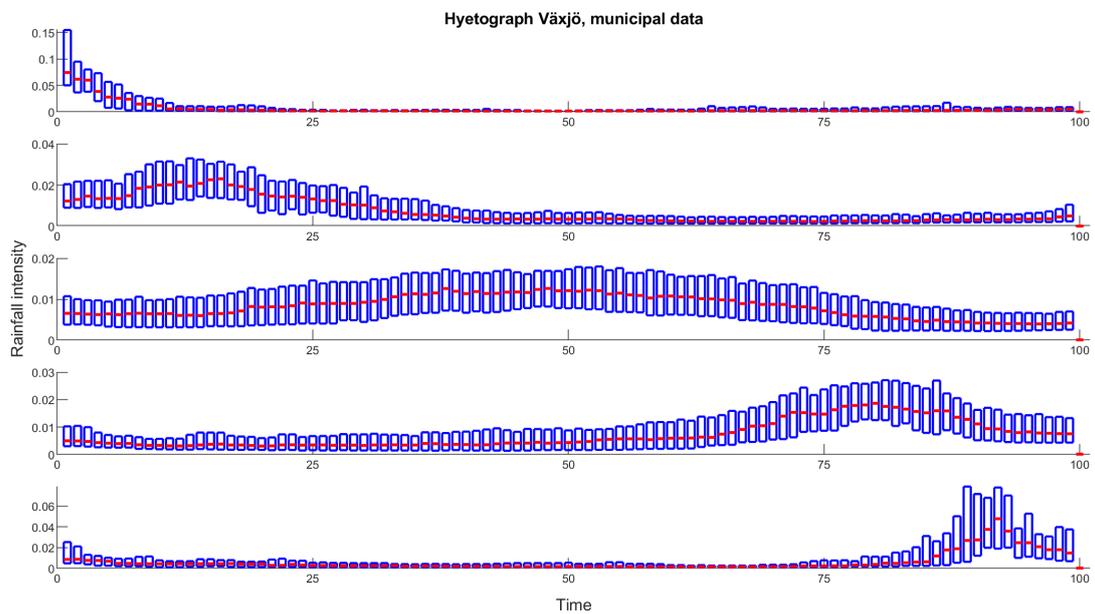
Figure D.16. Hyetographs in five clusters for Växjö (SE), municipal 1-minute data (calibration)

Appendix E   Histograms and Q-Q plots for multiple comparisons.

Comparison between different disaggregation scenarios. EV.
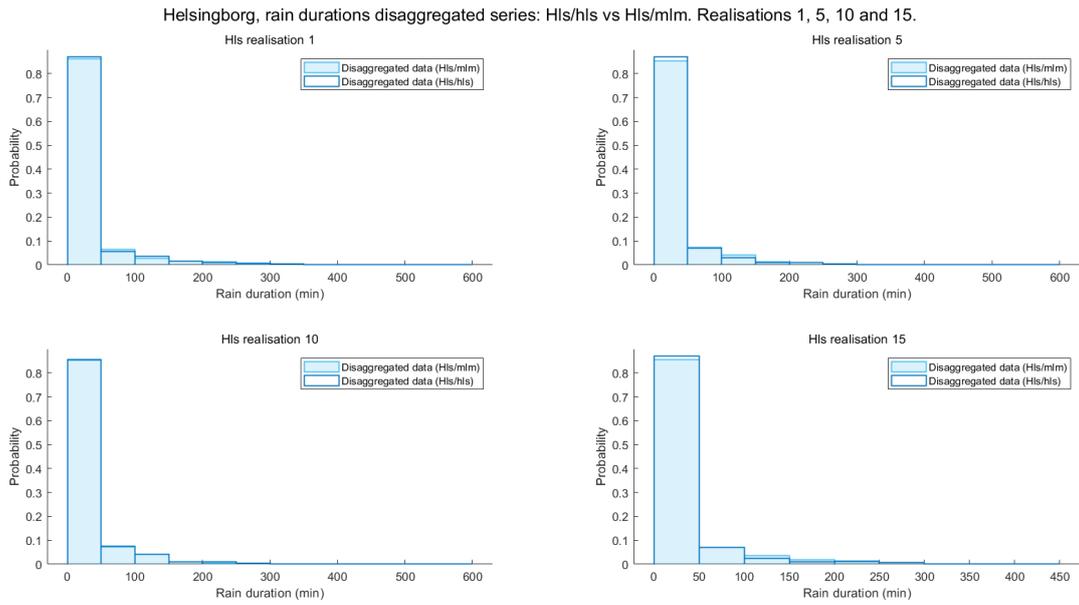

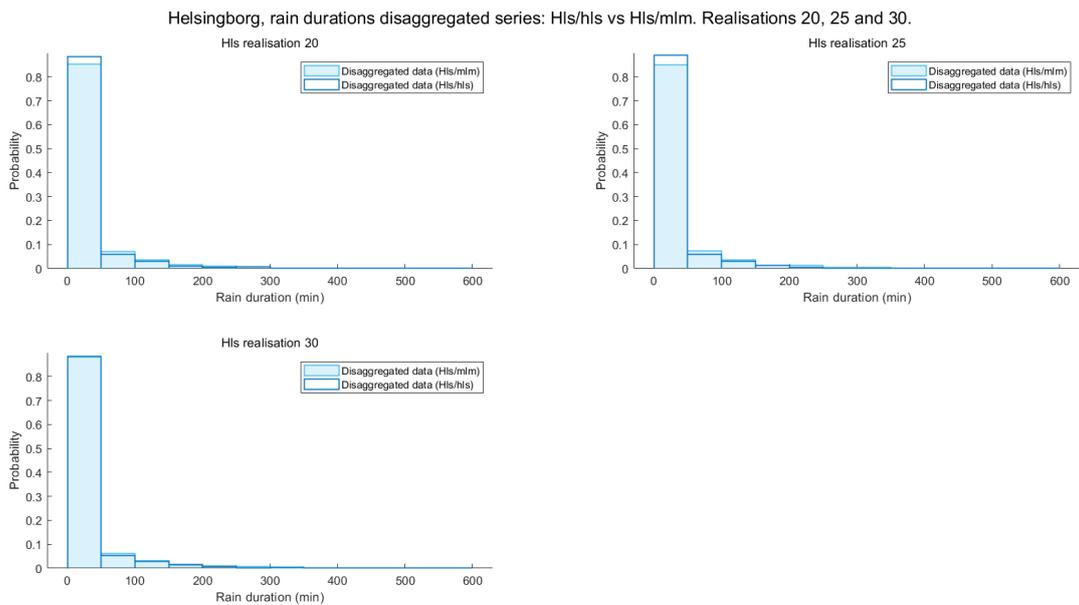(a) Stochastic realisations 1, 5, 10 and 15.

(b) Stochastic realisations 20, 25 and 30.
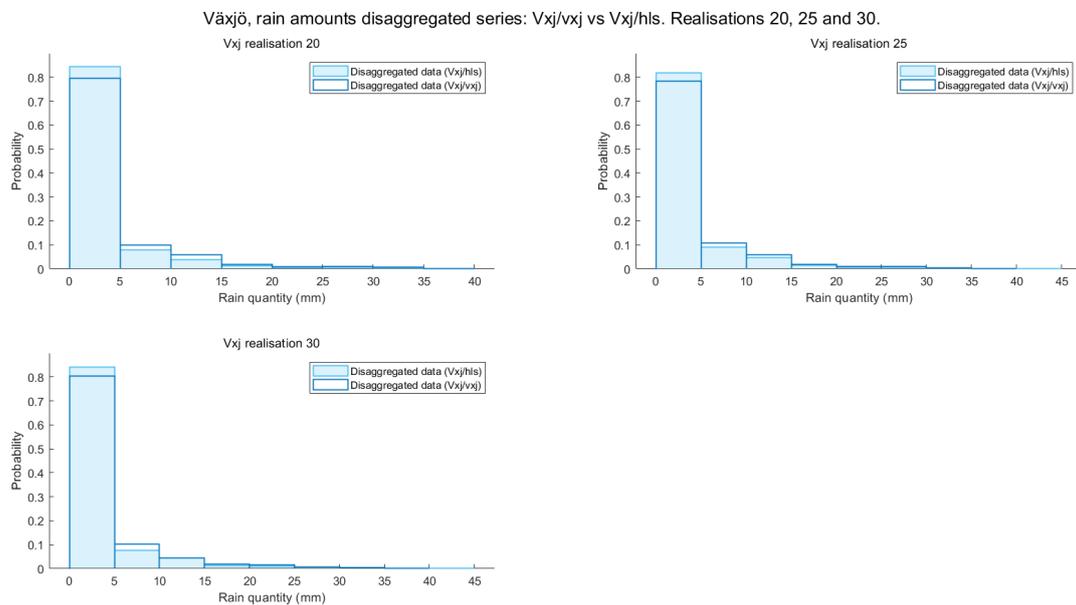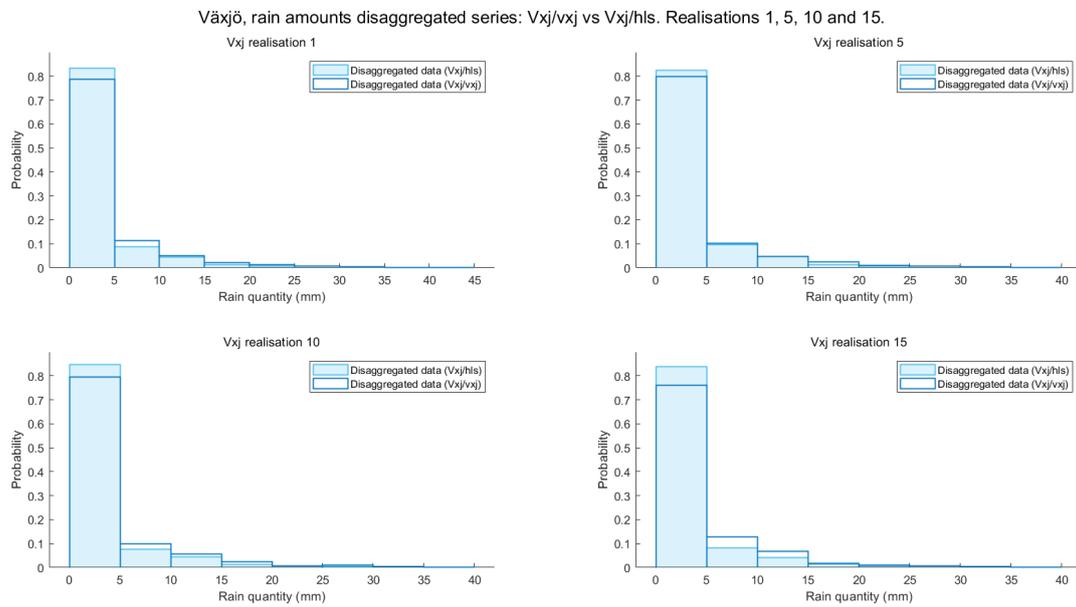
Figure E.1. Disaggregated rainfall series for Växjö. Calibrated against municipal data from Växjö (dark blue outline) and calibrated against municipal data from Helsingborg (light blue).
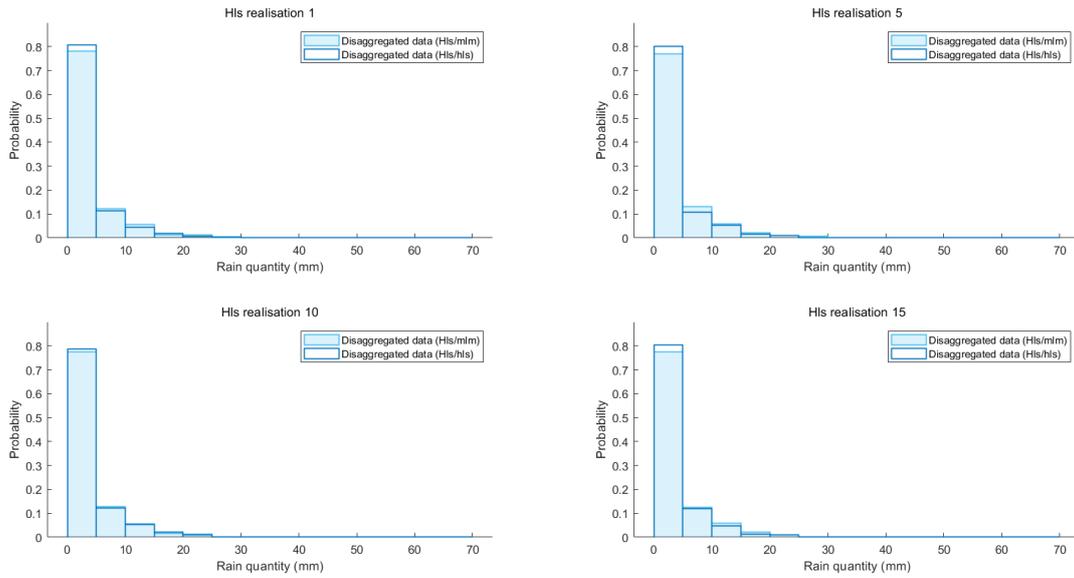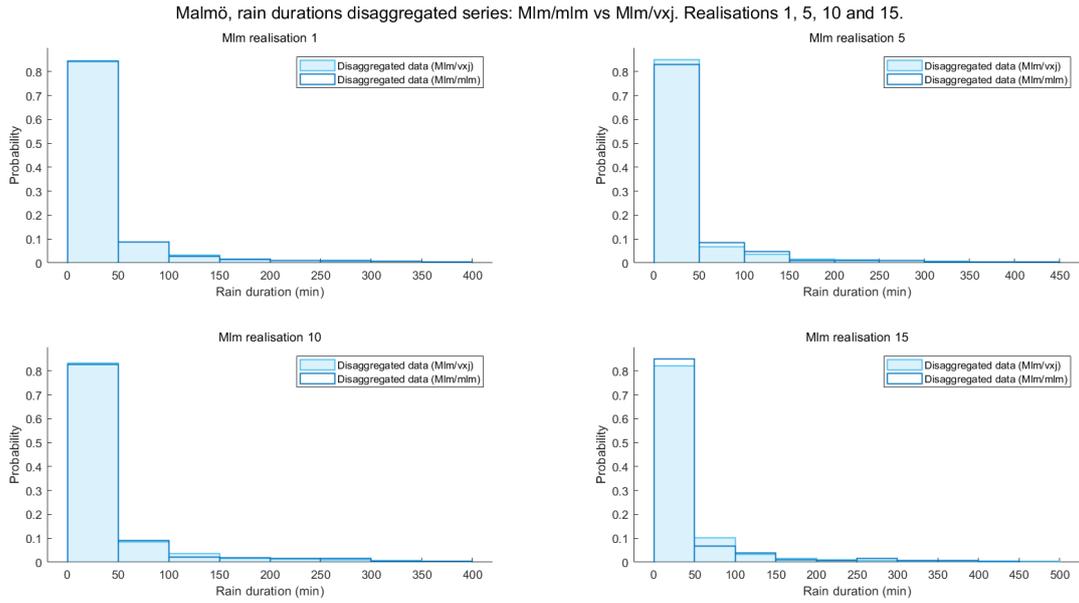
(a) Stochastic realisations 1, 5, 10 and 15.
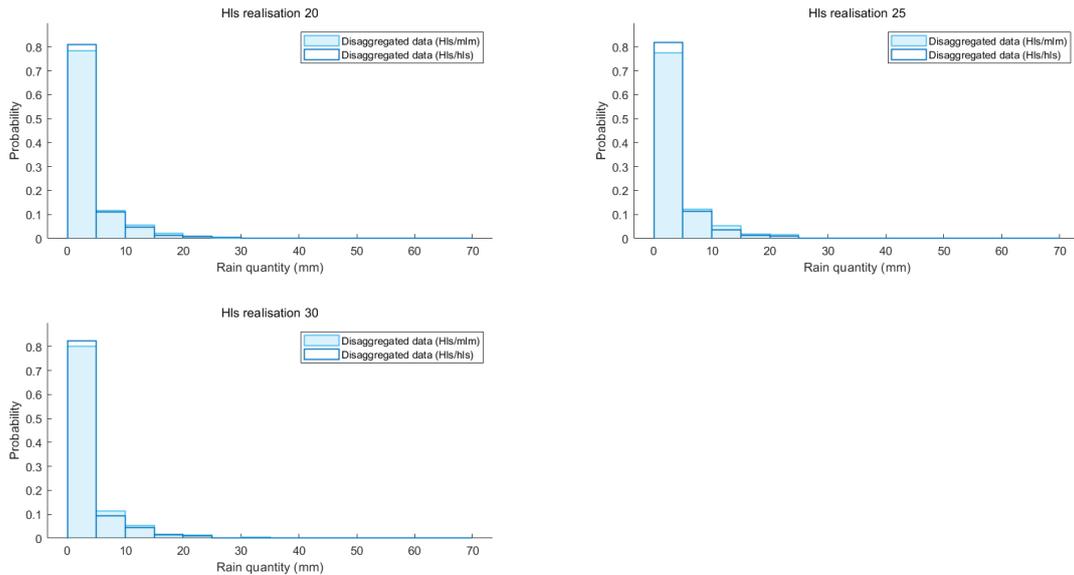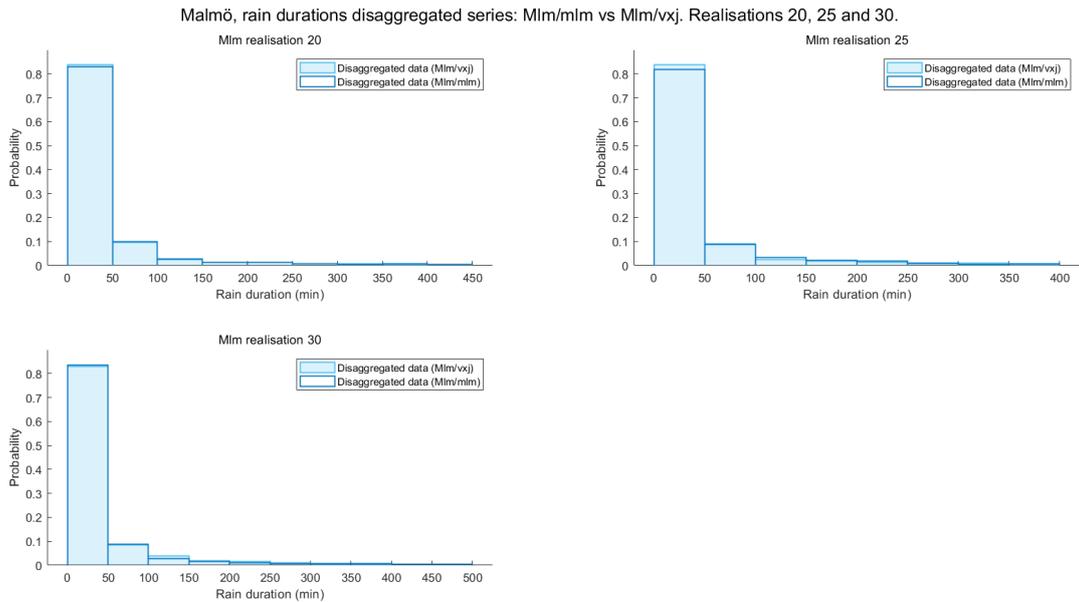

(b) Stochastic realisations 20, 25 and 30.

Figure E.2. Disaggregated rainfall series for Helsingborg. Calibrated against municipal data from Helsingborg (dark blue outline) and calibrated against municipal data from Malmö (light blue).
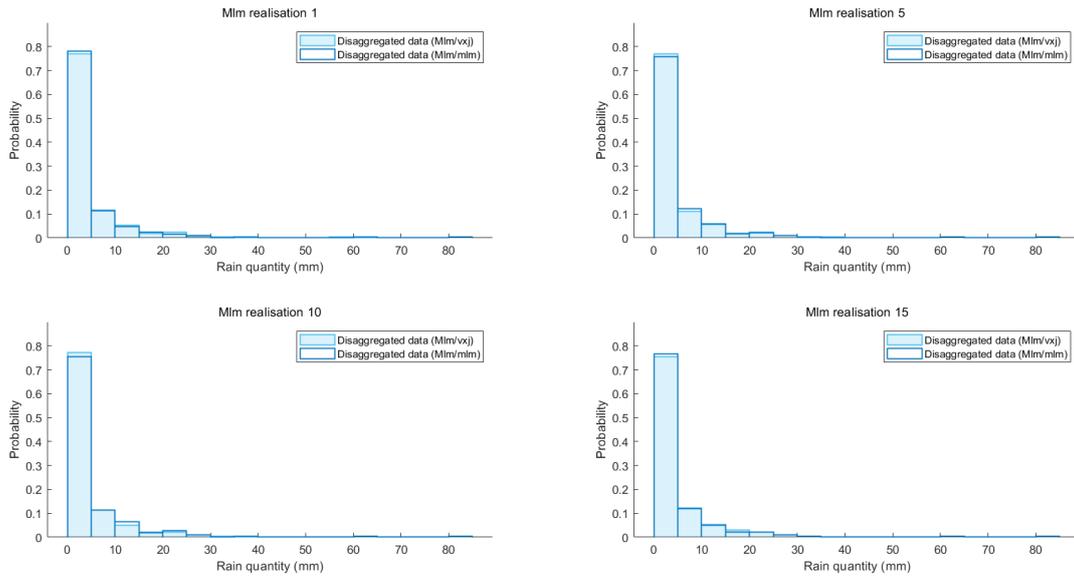
(a) Stochastic realisations 1, 5, 10 and 15.
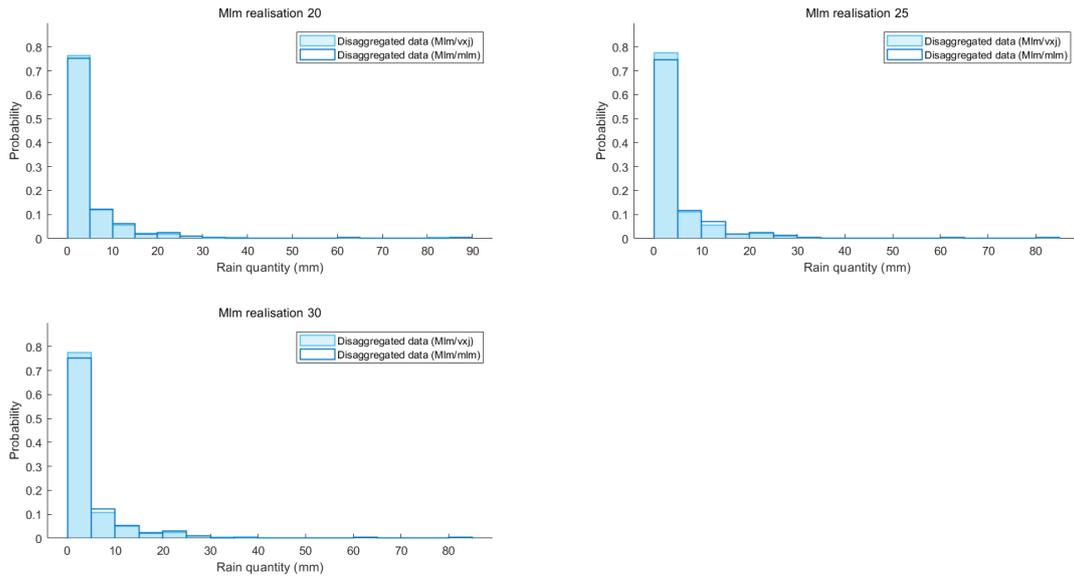
(b) Stochastic realisations 20, 25 and 30.

Figure E.3. Disaggregated rainfall series for Malmö. Calibrated against municipal data from Malmö (dark blue outline) and calibrated against municipal data from Växjö (light blue).
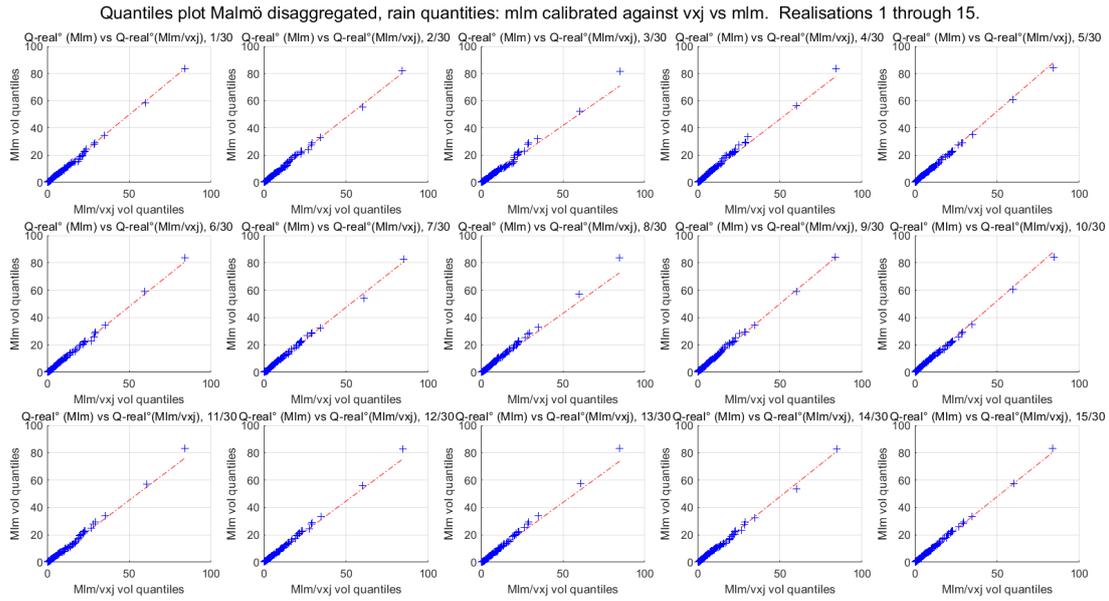
(a) Stochastic realisations 1 through 15
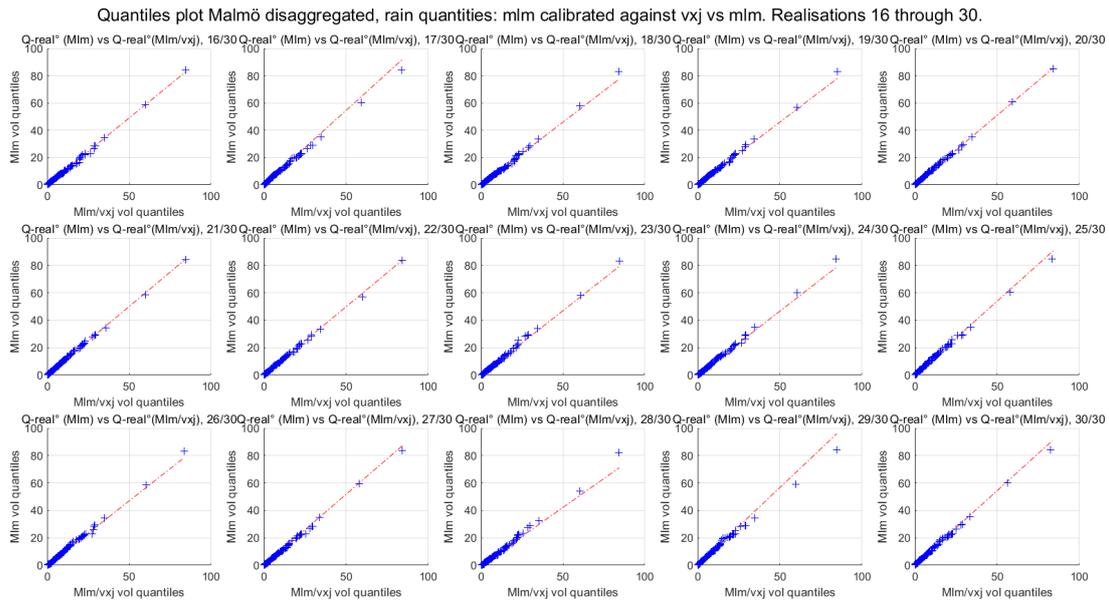
(b) Stochastic realisations 16 through 30

Figure E.4. Disaggregated rainfall series for Malmö. Quantiles for calibrated with municipal data from Malmö, SW, against quantiles for municipal data from Växjö, SW. (EV)

Comparison between different disaggregation scenarios. ED.
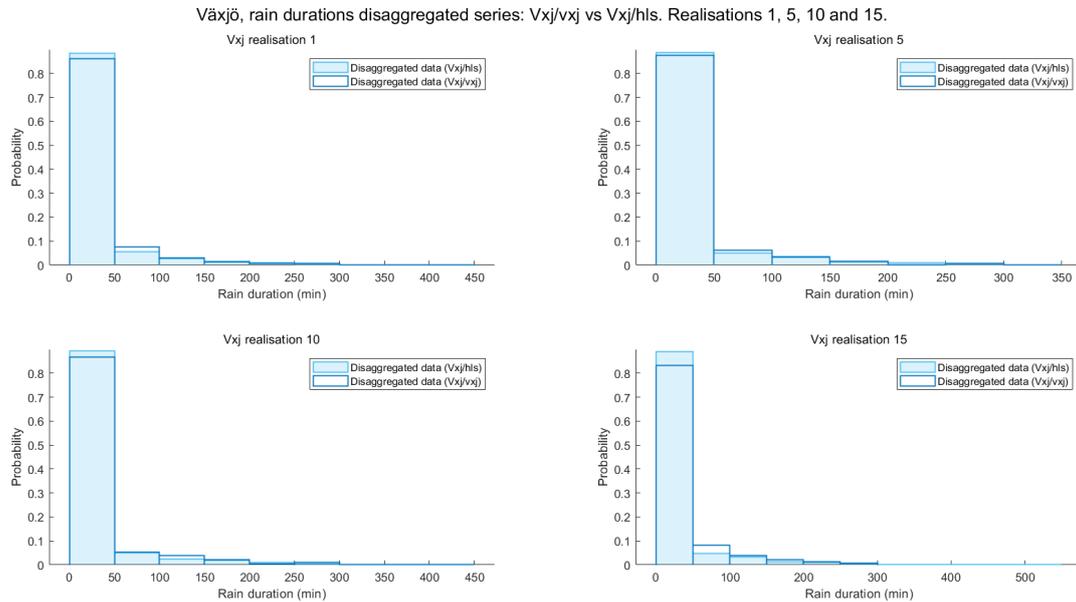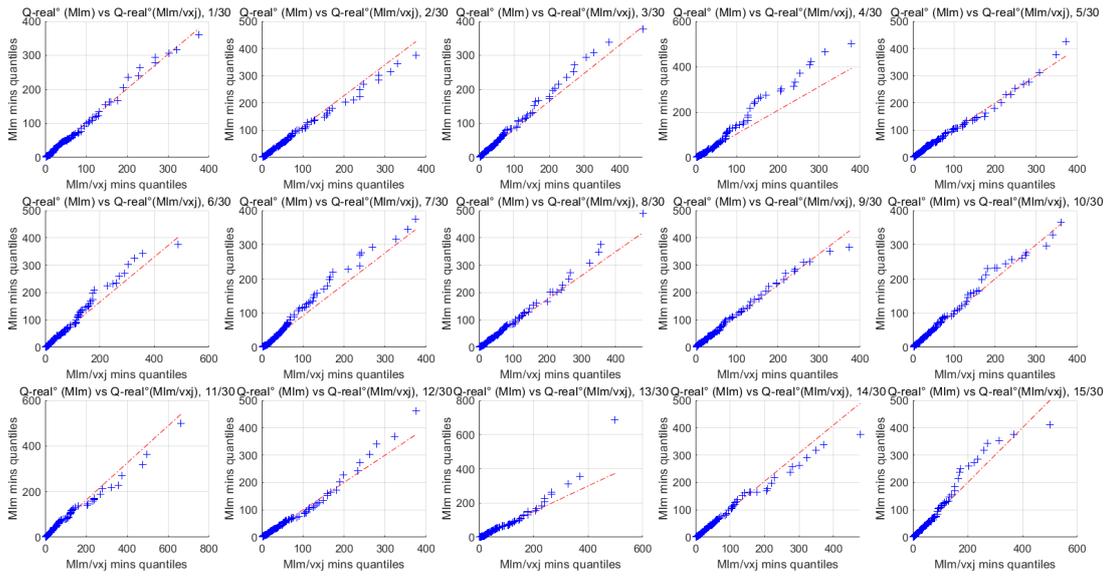

(a) Stochastic realisations 1, 5, 10 and 15.
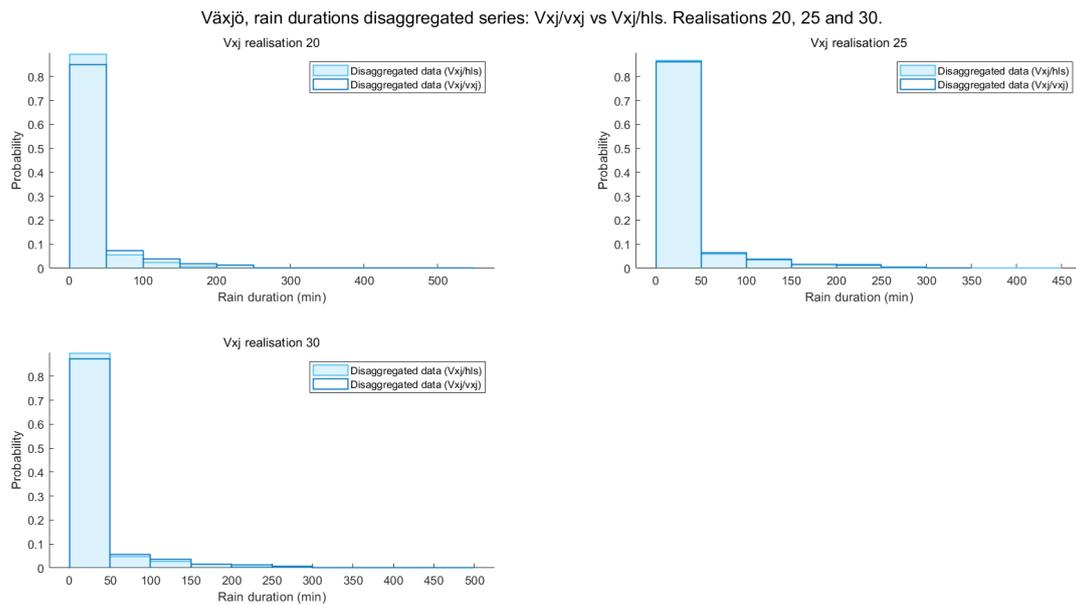
(b) Stochastic realisations 20, 25 and 30.

Figure E.5. Disaggregated rainfall series for Växjö. Calibrated against municipal data from Växjö (dark blue outline) and calibrated against municipal data from Helsingborg (light blue).

(a) Stochastic realisations 1, 5, 10 and 15.

(b) Stochastic realisations 20, 25 and 30.

Figure E.6. Disaggregated rainfall series for Helsingborg. Calibrated against municipal data from Helsingborg (dark blue outline) and calibrated against municipal data from Malmö (light blue).

(a) Stochastic realisations 1, 5, 10 and 15.

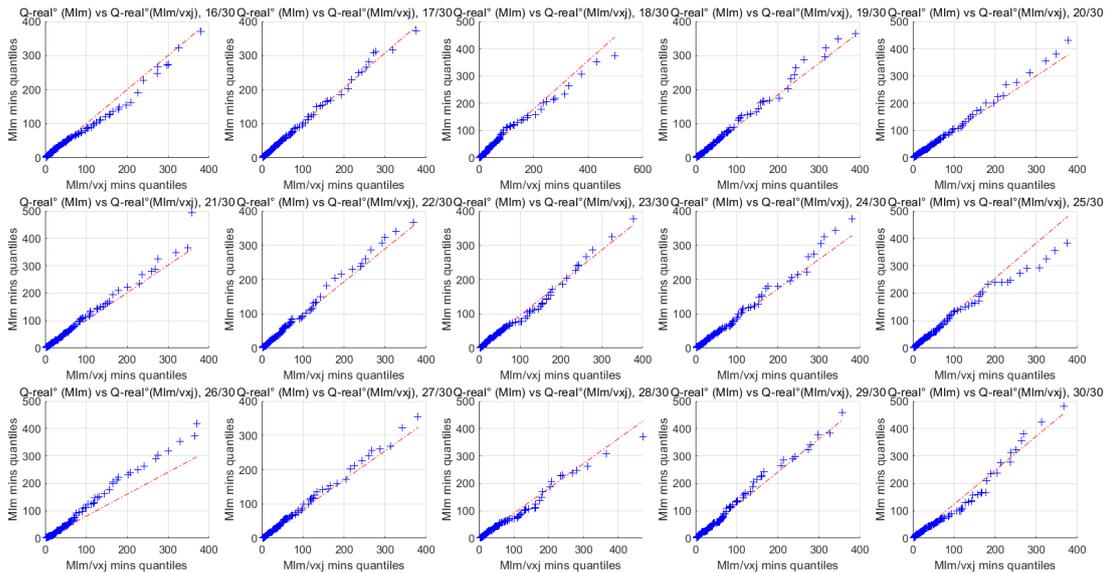(b) Stochastic realisations 20, 25 and 30.

Figure E.7. Disaggregated rainfall series for Malmö. Calibrated against municipal data from Malmö (dark blue outline) and calibrated against municipal data from Växjö (light blue).

(a) Stochastic realisations 1 through 15

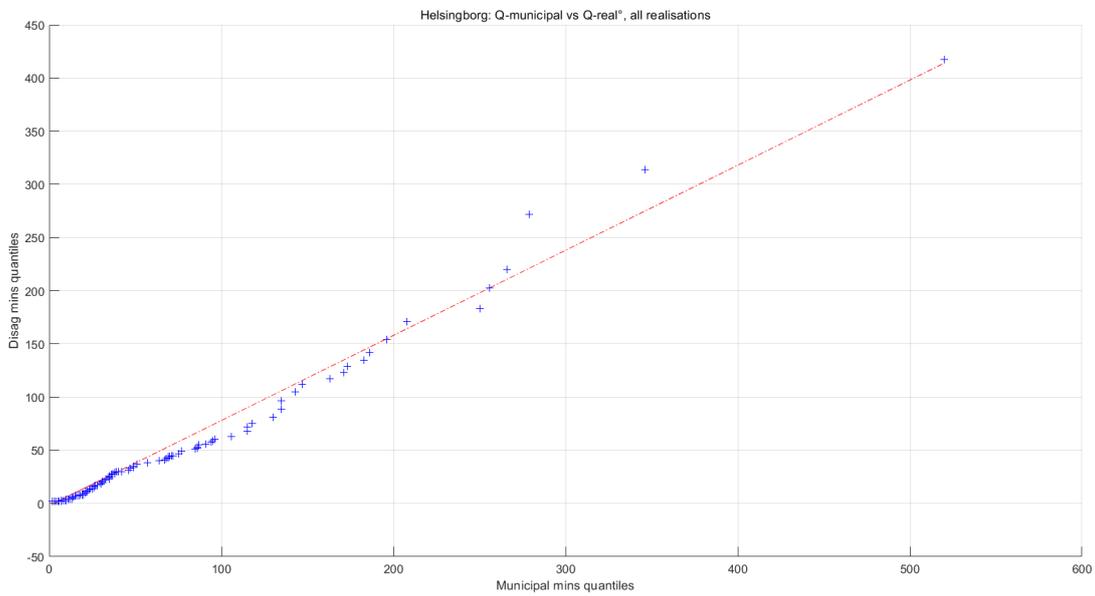(b) Stochastic realisations 16 through 30

Figure E.8. Disaggregated rainfall series for Malmö. Quantiles for calibrated with municipal data from Malmö, SW, against quantiles for municipal data from Växjö, SW. (ED)

Appendix F   Q-Q plots for Helsingborg and Malmö, municipal vs disaggregated

Q-Q plots Helsingborg: municipal vs disaggregated
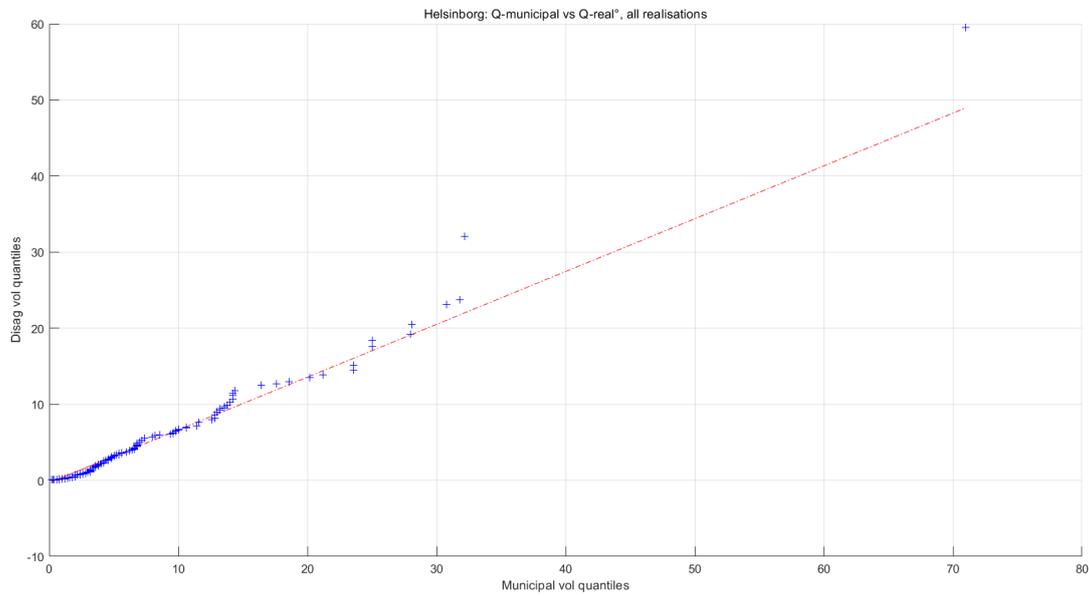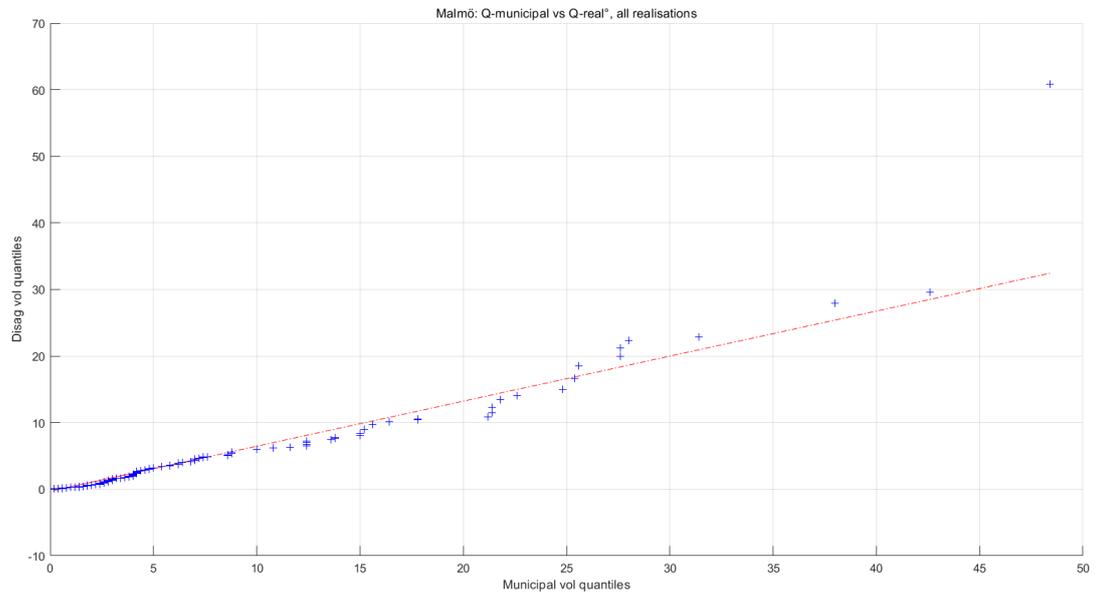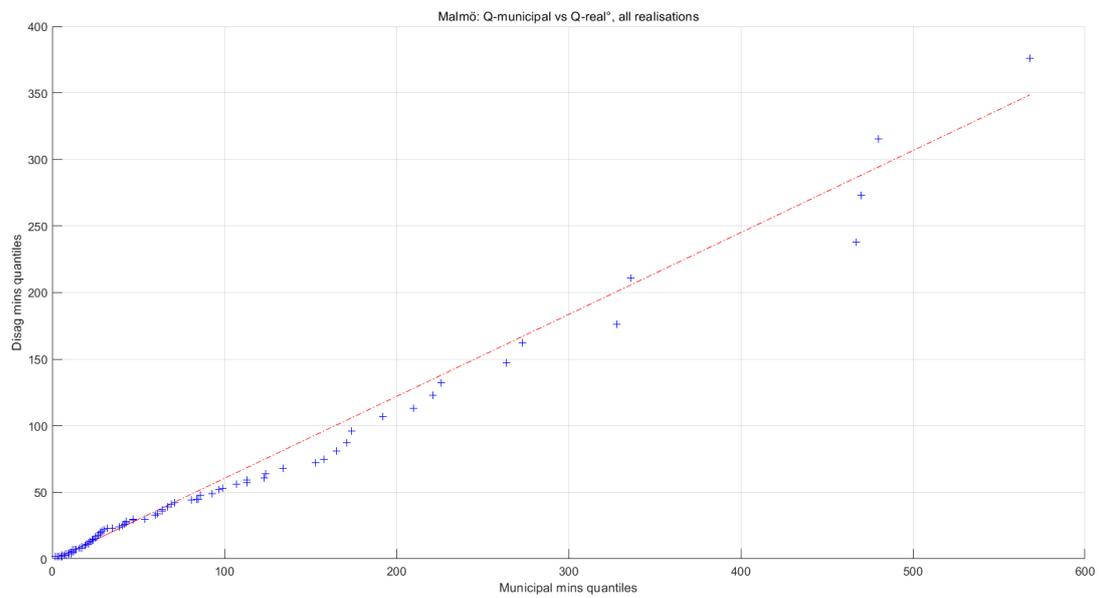

(a) EV

(b) ED

Figure F.1. Disaggregated rainfall series for Helsingborg, averaged over 30 realisations vs municipal data from Helsingborg for EV and ED.

Q-Q plots Malmö: municipal vs disaggregated

(a) EV

(b) ED

Figure F.2. Disaggregated rainfall series for Malmö, averaged over 30 realisations vs municipal data from Malmö for EV and ED.

Appendix G   Previous hyetograph distributions

Table G.1. Hyetograph distributions found by Litsmark (2020) for different cities in Sweden.

City	Cluster number					# of events
	1	2	3	4	5	
Borås	17%	33%	33%	0%	17%	6
Halmstad	42%	25%	25%	0%	8%	12
Helsingborg	0%	33%	44%	11%	11%	9
Jönköping	13%	50%	38%	0%	0%	8
Kalmar	0%	0%	50%	0%	50%	4
Karlstad	0%	20%	40%	20%	20%	5
Malmö	0%	80%	20%	0%	0%	5
Skellefteå	0%	0%	100%	0%	0%	4
Stockholm	50%	0%	0%	50%	0%	4
Sundsvall	25%	50%	25%	0%	0%	4
Uddevalla	22%	33%	33%	11%	0%	9
Uppsala	20%	40%	40%	0%	0%	5
Växjö	44%	22%	22%	0%	11%	9