A review of NVE's flood frequency estimation procedures

Donna Wilson, Anne K. Fleig, Deborah Lawrence,
Hege Hisdal, Lars-Evan Pettersson, Erik Holmqvist

9
2011

REPORT

A review of NVE's flood frequency estimation procedures

Norwegian Water Resources and Energy Directorate
2011

Report no. 9 – 2011

A review of NVE's flood frequency estimation procedures

Published by: Norwegian Water Resources and Energy Directorate

Authors: Donna Wilson, Anne K. Fleig, Deborah Lawrence, Hege Hisdal, Lars-Evan Pettersson, Erik Holmqvist

Print: Norwegian Water Resources and Energy Directorate
Number printed: 50
Cover photo: Ranaelva, October, 2011 (Photo: Trine Lise Sørensen, NVE)
ISSN: 1502-3540
ISBN: 978-82-410-0774-3

Abstract: This report reviews the methods currently recommended by NVE for flood estimation in Norway, including the approaches for statistical flood frequency analysis as well as rainfall-runoff modelling methods. For both approaches, various aspects of the methods are considered, including data quality, the assumptions constraining the application of the methods, and the determination of final flood estimates. The current guidelines and practices are found to be reasonable. However, as newer methods and data have become available in the mean time, several areas were identified in which it may be useful to focus attention for the future development of flood estimation.

Key words: Flood frequency estimation, extreme value analysis, rainfall-runoff modelling, PQRUT, NVE-guidelines.

Norwegian Water Resources and Energy Directorate
Middelthunsgate 29
P.O. Box 5091 Majorstua
N 0301 OSLO
NORWAY

Telephone: +47 22 95 95 95
Fax: +47 22 95 90 00
E-mail: nve@nve.no
Internet: www.nve.no
December 2011

Contents

Preface

Flood estimation is important for design and safety assessments, flood defence schemes and spatial planning. Property, health and lives are at risk if defence schemes fail to perform to the intended standard or if flood risks are ignored. Flood estimation is difficult particularly for long return periods as the data available are usually insufficient to define precisely the probability of such large floods and may also vary over time due to environmental changes (e.g. climate change, land use change). This requires that the hydrologist use his or her practical knowledge about the processes involved and applies efficient and robust statistical and modelling techniques to give the best possible estimates of rare flood events. This report reviews the methods currently recommended by NVE for use in developing flood estimates based on flood frequency analysis and on rainfall-runoff modelling. The methods are found to be reasonable relative to feasible alternatives. Areas and issues which could benefit from an update or revision in light of the availability of newer methods, data and knowledge are highlighted, as are potential topics for further research.

Oslo, December 2011

Morten Johnsrud
Director

Hege Hisdal
Head of Section

Summary

Flood estimation is important for design and safety assessments, flood risk management schemes and spatial planning. In Norway, the 200-year flood is used for flood hazard mapping, and the 500-year, the 1000-year and the probable maximum floods for dam safety analysis, depending on the safety class of the dam. Hence, the magnitudes of flood events with low probabilities need to be estimated, and this is, by necessity, undertaken using comparatively short timeseries of observed flood data. Limited data availability accordingly introduces uncertainty in the flood estimates, and this uncertainty is further increased by temporal variation in the timeseries due to natural variability in climate and to environmental changes (e.g. anthropogenically-induced climate change and land use changes such as river regulation). The use of advanced and reliable methods is therefore an important prerequisite for generating reliable flood estimates. This report reviews the methods currently recommended by NVE for flood estimation in Norway, including the guidelines presented in Midttømme *et al*. (2011).

Flood estimation methods can be classified into two groups: (1) statistical flood frequency analysis, which is based on the analysis of observed historical flood events and estimates the magnitudes of floods with a certain return period, and (2) rainfall-runoff modelling, which converts precipitation, and in some cases stored snow, into a surface runoff using a conceptual simulation model of the catchment response. Both approaches are reviewed here, and various aspects of the methods are considered, including data quality, the assumptions constraining the application of the methods, and the determination of final flood estimates. The most recent update of the guidelines for flood estimation (Midttømme *et al*., 2011) have particularly aimed at clarifying the procedures used to incorporate snowmelt in flood estimates and at providing guidance for taking account of climate change. The suggested procedures for regional flood frequency analysis for Norway are, however, still those developed in 1997 (Sælthun, 1997), and the procedures for rainfall-runoff modelling date from the 1980's. The guidelines and practices are found to be reasonable. However, as newer methods and data have become available in the mean time, several areas were identified in which it may be useful to focus attention for the future development of flood estimation, including:

Flood frequency analysis

- Consistency between the methods and analyses applied in the various software programs for flood frequency analysis available from NVE (i.e. Ekstrem and Dagut/Finut);
- Review of the current regions and methods for regional flood frequency analysis, including alternative grouping approaches, improved guidance on the selection of representative stations and consideration of multiple regression equations for adjusting flood frequency estimates for the site of interest;
- Review of approaches for estimating instantaneous flood peaks;
- Development of new procedures for the analysis of non-stationary series;

Rainfall-runoff modelling

- Review of current initial conditions used in model applications, in particular, the specification of antecedent soil moisture;
- Evaluation of methods for simulating combined snowmelt/rainfall events for the Probable Maximum Flood (PMF) and for events of a given return period (where the event represents a joint probability of simultaneous snowmelt and extreme rainfall);
- Evaluation of the estimates obtained with the PQRUT rainfall-runoff model in comparison with HBV, where feasible, and with newer approaches such as long-term continuous and semi-continuous simulation modelling;

Final flood estimates

- Reconciliation of the results obtained using flood frequency analysis and rainfall-runoff modelling.

List of Abbreviations

AEP	Annual exceedance probability
AMS	Annual maximum series
ARF	Annual reduction factor
BMS	Block maximum series
EV-1	Extreme Value-1 distribution (Gumble)
GEV	Generalized Extreme Value distribution
GP	Generalized Pareto distribution
MLE	Maximum likelihood estimation method
MT	Precipitation event with return period T
MOM	Method of moments
PDS	Partial duration series
PMF	Probable maximum flood
PMP	Probable maximum precipitation
POT	Peak over threshold approach
PWM	Probability weighted moments method
Q_d	Mean daily flood flow
Q_i	Instantaneous flood peak
QM	Index flood (*here:* = mean flood)
QT	Flood with return period T
T	Return period
X_T	Growth curve

1 Introduction

Flood estimation is important for design and safety assessments, flood risk management schemes and spatial planning. In the case of dam design, the safety of individual dams is reviewed every 15 – 20 years. Property, health and lives are at risk if defence schemes fail to perform to the intended standard or if flood risks are not properly accounted for in land use planning. However, flood estimation is difficult particularly for events with a low probability and long return periods because the quantities being estimated (e.g. streamflow of the 200-year, 1000-year, Probable Maximum (PMF) floods) must be inferred and may vary over time due to natural variability in climate and to environmental changes (e.g. anthropogenically-induced climate change and land use changes such as increased urbanisation).

Sælthun and Anderson (1986) detail the development of Norwegian flood estimation procedures from 1976 onwards, when a governmental committee was given the mandate to work out regulations for dam design. In its final recommendations from 1979, the committee brought Norwegian procedures in line with internationally accepted methods available at the time (*e.g*. NERC, 1975; Sokolov *et al*., 1976) suggesting the use of both, statistical flood frequency analysis and rainfall-runoff models for flood estimation and the calculation of the Probable Maximum Flood (PMF). Alongside the work of this committee, Wingård (1977) compared distribution functions for flood frequency analysis, and guidelines for flood frequency analysis were prepared (Wingård *et al*., 1978). In the 1980's greater attention was given to the development of procedures for calculation of the PMF, which led to the publication of the first set of guidelines for flood calculations for dam design (Vassdragsdirektoratet, 1986). Since the publication of these early guidelines, their routine application has led to several updates. The latest update (Midttømme *et al*., 2011) was particularly aimed at clarifying the procedures used to incorporate snowmelt in flood estimates and at providing guidance for taking account of the effect of climate change in flood estimation. The suggested procedures for regional flood frequency analysis for Norway are, however, still those developed in 1997 (Sælthun, 1997) and the basic methods for rainfall-runoff modelling date from the 1980s.

The methods used for flood estimation can be classified into two groups:

- Flood frequency analysis (statistical methods)
- Rainfall-runoff modelling

Flood frequency analysis is based on the analysis of observed historical flood events and estimates the magnitudes of floods with a given return period. Rainfall-runoff modelling, on the other hand, converts a rainfall into a surface runoff using a model of the catchment response based on model parameters which are either calibrated based on observed data or are estimated from catchment characteristics. It can hence be used to derive the PMF resulting from the combination of the probable maximum rainfall and snowmelt estimates, or alternatively, the flood resulting from extreme rainfall of a given duration and return period. Flood estimation using the rainfall-runoff method, as it is practised in Norway, is based on a frequency analysis of extreme rainfall events to establish a precipitation sequence which is then used in an event-based rainfall-runoff model to simulate the corresponding runoff hydrograph. The flood estimation procedures currently used in Norway have several important strengths. These include that the procedures (i) are relatively easy to apply, (ii) require data that is either readily available or can be derived for Norwegian catchments from existing databases, (iii) build on a wealth of experience from previous applications, and (iv), not least, are supported by a good dam safety record with respect to the management of flood risks.

This report reviews NVE's flood estimation procedures. The various components of statistical flood frequency analysis are considered in Chapter 2, whereas rainfall-runoff modelling is described in Chapter 3. Chapter 4 briefly discusses the selection of final flood estimates, before conclusions are drawn and recommendations given in Chapter 5. The procedures used to route flood flows through reservoirs or river reaches are not considered as part of this review, nor are flood frequency analysis procedures for urban areas.

2 Flood frequency analysis

2.1 Introduction

Flood frequency analysis is a statistical approach used to determine the magnitude of a flood event with a certain occurrence probability or return period. In contrast to rainfall-runoff modelling, statistical flood frequency analysis is based on observed flood data only, either at the site of interest (*at-site flood frequency analysis*) or from one or several comparable gauged basins within the same region in the case of limited local data availability (*regional flood frequency analysis*). A short data record may also be extended by model simulation and a frequency analysis can then be performed.

A statistical flood frequency analysis is based on the assumption that all events in the observed flood series represent a process that can be described by one single flood frequency distribution. A mathematical function is used to describe the distribution of events, and this function is then extrapolated to give values corresponding to return periods beyond the length of the observed record. Flood frequency analysis can be straightforward if the return period of interest does not significantly exceed the period of observation and all of the observed events are generated by the same flood generating mechanism. Uncertainty, however, increases with increasing return period and, therefore, extrapolation should be avoided as far as possible. If one has to extrapolate, this should be done only as far as necessary and preferably only up to twice the record length. Additional information to provide independent support to the extrapolated values is valuable, but one should always be aware of the uncertainty. Unfortunately, flood records are frequently of insufficient length, and this introduces significant uncertainty into the flood estimates. In general, a *regional flood frequency analysis* using data from several stations can be performed to reduce the uncertainty and to "limit unreliable extrapolation when available data record lengths are short as compared to the recurrence interval of interest, or for predicting the flooding potential at locations where no observed data are available" (Castellarin *et al.*, 2011). The NVE guidelines for flood estimation (Midttømme *et al.*, 2011) related to dam safety, recommend an at-site analysis for stations with at least 50 years of data. The procedure is also recommended for stations with observations of at least 30 years, although here there are some restrictions. In all cases, a comparison of the results with those from nearby stations is recommended.

In extreme value statistics the term ‘return period’, sometimes called the recurrence interval, is often specified rather than the exceedance probability to describe the rarity of an event. The return period, T, is the inverse of the annual exceedance probability (AEP), i.e. AEP = 1/T. For example, there is a 0.5% probability that a flood event with a return period of 200 years will be exceeded in any one year (Faulkner, 1999). The longer the return period, the rarer the event. However, the term ‘return period’ can be misunderstood, as people not familiar with extreme value statistics may believe that this implies that a particular flood magnitude is only exceeded at regular intervals, or that it refers to a fixed period of time until the next occurrence. It should be emphasized that return periods are probabilities and not long-term predictions. In general, it has to be kept in mind that the likelihood of a flood event may vary due to natural variability, to the length of the available data sample used for the estimation, and also due to environment changes, such as changes in land use or climate change.

In the following sections, some general aspects of flood frequency analysis are first introduced (Sections 2.2-2.3), before outlining the procedure for flood frequency analysis currently recommended by NVE for observed river flow data. The suggested at-site analysis is described in Section 2.4. The approaches for a regional flood frequency analysis in the case of no or limited data availability are presented in Section 2.5. Common to both at-site and regional analysis is the possible need to derive the instantaneous flood peak value from the daily mean, and this is described in Section 2.6. Finally, the performance and uncertainty of a flood frequency analysis is discussed (Section 2.7).

2.2 Data quality

Reliable data is an important prerequisite for a reliable flood frequency analysis. Data quality can vary significantly between stations. Possible sources of error in flood peak data include:

- inaccuracies in direct and indirect flow and water level measurements;
- quality of the rating curve for flood flows: Water levels are translated into flow values using a rating curve, but measurements of water levels and flows are rarely performed in extreme floods;
- the oldest data in NVE’s Hydra II database are based on the daily observation of water levels prior to the installation of recording devices. These older readings are assumed to represent daily average values, but may differ to a greater or lesser extent from actual daily average values;
- errors in transferring observations onto the Hydra II database.

All data in NVE’s database are now quality controlled by the hydrometrist before the data are stored. As stated in the guidelines for flood estimation (Midttømme *et al.*, 2011) all data used for a flood frequency analysis should additionally be quality checked. This includes, most importantly, an evaluation of the rating curve quality for high flows and a check of the values of extreme flood water levels for possible registration errors. Information about the rating curve quality can be found in the Dagut and Finut software. If the quality check is very time consuming or otherwise difficult to carry out, it is suggested to focus on the data of the most important stations. Data can be corrupted and missing values are common. Although a quality control of the data is undertaken by the practitioner prior to flood frequency analysis, guidance on the thoroughness of the review required and on the requirements for documentation is needed to increase consistency between the analyses undertaken. There is also a lack of comments about data quality in the Hydra II database. For example, it is known by many that there are problems with some of the early manual readings recorded at Kløvtveitvatn (Station 68.1), but these data are still available for download without warning from Hydra II and could inadvertently be used in flood frequency analysis. In addition, it would be useful to know more about individual data points. For example, which are real observations, which are estimates based on neighbouring stations, and which are modelled data. If the largest floods on record are not observations, it would be useful if these were identified. At present, this can only be checked manually before retrieving the data.

2.3 Data requirements

Current methods for flood frequency analysis assume that the data sample consists of independent, identically distributed (iid) events. The criterion of independence implies that there are no autocorrelations, trends or shifts in the sample (Section 2.3.3). For example, the magnitude of one event should not depend on the magnitude of the previous event, and there should be no systematic or abrupt changes over time due to, for example, climate change or anthropogenic influences in the catchment. The requirement of independent events needs to be considered in the general selection methodology for a series of extreme events, which is commonly either an Annual Maximum Series (AMS) or a Partial Duration Series (PDS; Section 2.3.1).

The assumption of identical distribution may be violated if the floods are caused by different generating processes. In Norway, this is particularly the case in catchments where some floods are caused by extreme rainfall only and others are caused by considerable snowmelt. A separate analysis of these two types of floods should therefore be considered (Section 2.3.2).

2.3.1 Annual maximum versus partial duration series

At NVE and in general, flood frequency analysis is typically based upon annual maxima series (AMS). This is a special case of the Block Maximum Series (BMS) with a block size of one year. An alternative is to analyse a partial duration series (PDS; also called peak-over-threshold approach, POT) which includes all floods exceeding a predefined threshold value. This would have the advantages of taking into account other major floods in flood-rich years and of preventing the analysis of small or non-flood events in other years. Care has to be taken, however, to assure the mutual independence of the events included in a PDS. This can, for example, be done by requiring a minimum period between the occurrences of two subsequent events included in a PDS (e.g. Engeland *et al.*, 2004) and may also be necessary for AMS in case the year shift is during the high flow season. Several comparative studies (e.g. Madsen *et al.*, 1997; Martins and Stedinger, 2001) suggest that the PDS approach is more precise than the AMS. Cunnane (1989) found that the analysis of annual maxima performed better than the analysis of PDS where the mean number of peaks per year is small (<1.65). Engeland *et al.* (2004) compared the use of BMS with block sizes of 3, 6 and 12 (i.e. AMS) months and PDS for flood frequency analysis at Haugland, a station in south-west Norway with a long data record. They found that both approaches can be used with comparable results. The performance of the PDS approach depends on the chosen threshold, and for the BMS approach they found the model with a block size of three months and seasonal dependency for the fitted distribution parameters to perform best.

2.3.2 Critical season

Large parts of Norway are affected by two types of floods: (1) snowmelt floods (i.e. floods driven by a large volume of melting snow, often in combination with rain) and (2) rainfall floods. Since the two types of floods typically occur during the spring season and summer through autumn/winter periods, respectively, they are often called (1) spring and (2) autumn floods (Midttømme *et al.*, 2011). Due to the different generating processes, it is important that the two different types of floods are analysed separately. A flood rose or summary table of maximum monthly flood peaks is often used to identify the *critical season*, i.e. the season during which the largest floods occur (Pettersson, 2009a; Pettersson, 2009b). For many catchments across Norway the critical season is autumn, and along the coast it is winter. These floods are assumed to be rainfall floods, but in some cases there is also a significant contribution from snowmelt. For large reservoirs and catchments the most critical floods may be due to spring snowmelt coupled with a period of heavy rainfall, and in small catchments summer events caused by heavy precipitation. At some stations, floods with a major snowmelt contribution may predominantly occur during the summer season, depending on the location, altitude and the percentage glacial cover in the catchment.

When it is difficult to distinguish between the two types of flooding or when too few events of one flood type are observed, the NVE guidelines recommend that annual maxima are analysed with no seasonal division (Midttømme *et al.*, 2011). However, the guidelines only caution with respect to underestimation in situations where the largest spring and autumn floods are of a similar magnitude. Unless the analyst can be confident that the values for one season are always less than the other (Figure 2.1a), specifying the magnitude of return period events based on only one season analysis, could result in an underestimation of flood magnitudes for a range of return periods (Figure 2.1b). Where these two types of floods are combined, the extrapolation of the distribution can be affected and, therefore, may be unrealistic. Waylen and Woo (1982) found that in catchments where the annual flood series is generated by more than one distinctive hydrological process, the Gumbel distribution does not provide a satisfactory fit. When the floods associated with a given hydrological process were distinguished and modelled separately, this distribution was found to be adequate. A prerequisite for a reliable fit of the distribution is, of course, that enough extreme events are observed for each flood type.

For most applications, flood frequency estimation is undertaken using the annual flood, since the most damaging floods are generally caused by a combination of rainfall and snowmelt, regardless of whether the floods occur in the spring or autumn, with the exception of Finnmark. In Finnmark, where the topography is more uniform, a warm period can cause large quantities of snowmelt over large areas simultaneously. However, the seasonal distribution of floods should always be considered prior to analyses to ensure that floods for the correct season or the whole year are analysed. For many sites, the identification of the critical season is often thought to be relatively clear, and can be undertaken through the use of flood roses or summary monthly statistics. Cunderlink *et al.* (2004) found, however, that the subjective identification of flood seasons is potentially unreliable. Pettersson (2000) also found, through a flood frequency analysis for the Gaula catchment in Trøndelag that the critical season can differ for different return periods. There are cases in which most annual flood events have occurred in the spring, but the largest event has occurred in the summer/autumn period. In these situations it is important to analyse both spring and summer/autumn floods separately and derive conclusions based on these separate analyses. The spring flood is typically characterised by a high volume flood peak with a high mean annual flood, but moderate growth curves, whereas autumn floods are typically of shorter duration and higher intensity with steeper growth curves (Sælthun and Andersen, 1986), as illustrated by Figure 2.1b. This general tendency suggests that in order to evaluate flood magnitudes for the higher return periods it may in many cases be appropriate (if supported by review of the data) to consider autumn as the critical season.

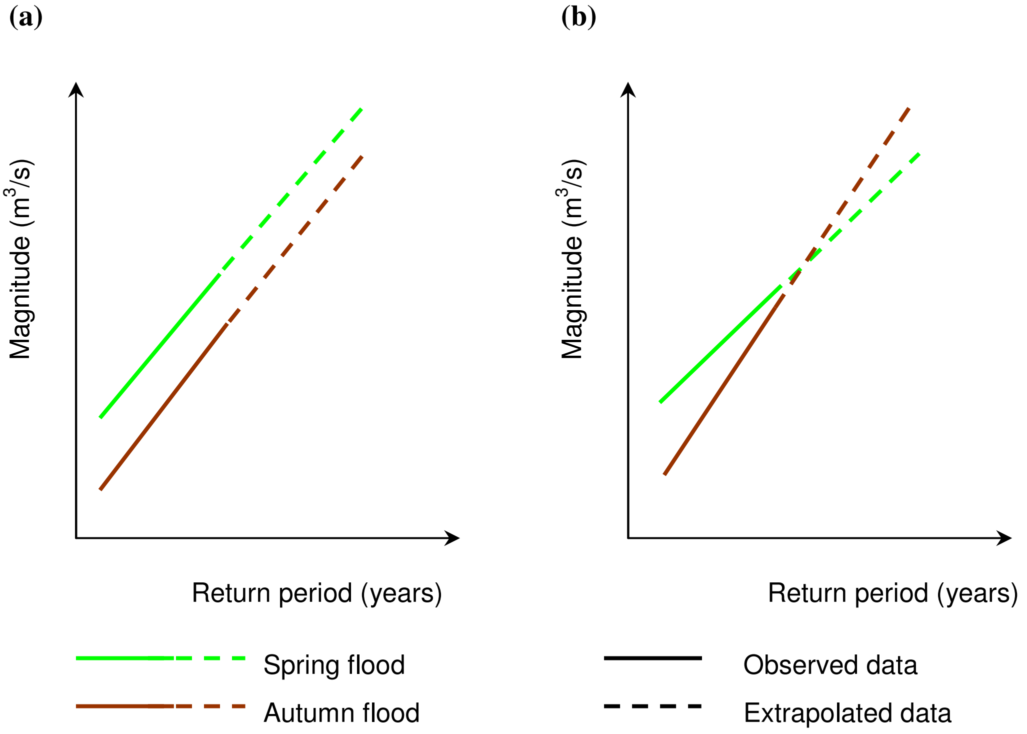


Figure 2.1 Two possible distributions of the spring and autumn floods for a single catchment

2.3.3 Stationarity

Current methods for flood frequency analysis assume that data are stationary (*i.e.* not changing over time). There are currently no systematic analytical procedures for accounting for the effect of environmental change (*e.g.* climate change, land use changes, urbanisation, extensive tree felling) available for use in Norway. In dam safety assessments the effect of environmental change is partly taken into account through the requirement that assessments are repeated every 15 – 20 years. Environmental change can lead to changes in flood frequency. Two main considerations are that:

- historical flood data may not be stationary
- future flood characteristics may not be stationary

If a data series used for flood frequency analysis shows a strong trend, then its flood frequency curve will, at best, represent the average response of the catchment over the period of record, if the estimation of flood frequency is based on annual maxima. It will give a poor representation of current or future flood frequencies (Reed and Robson, 1999). Statistical tests (*e.g.* Mann Kendall, Kendall-Theil) can be used to help determine whether a record displays a significant trend that might indicate non-stationarity. Time series are not, however, systematically investigated for trends prior to undertaking frequency analysis, but shifts are investigated and adjustments made in relation to watercourse regulation. A regulated series is naturalised by maintaining continuity of the water balance, through either adding or subtracting changes in the reservoir water level and transfers in and out of the catchment.

For selected stations in Norway trends in the timing and magnitude of both the spring and autumn floods have been analysed (Wilson *et al.*, 2010). Results suggest that the timing of the spring flood has become earlier at many stations within Norway, but that there are no consistent trends in either the timing of the autumn flood or the magnitude of either the spring or autumn floods.

Lawrence (2010) and Lawrence and Hisdal (2011) investigated projected changes in the magnitude of the 200 year flood using hydrological projections based on input from 13 climate projections derived from various combinations of SRES emission scenarios, global and regional climate models. The input data were used in hydrological models calibrated for each of 115 catchments within Norway, and likely changes in flooding between a 1961-90 reference period and 2021-2050 and 2071-2100 future periods were estimated based on a flood frequency analysis of the simulated runoff time series. Results indicate that the magnitude of the 200-year flood is expected to increase in western Norway and along much of the coast due to increases in rainfall, particularly in the autumn and winter months. More frequent and intense rainfall at a local scale will increase the probability of rapid flooding in small streams and urban areas throughout the country. In more inland and northern areas currently dominated by snowmelt flooding, the magnitude of the 200-year flood is expected to decrease, due to a projected decrease in winter snow storage and an earlier spring snowmelt. In some catchments the critical season may also change from spring to autumn. These analyses suggest that flood magnitudes can be expected to change in the future making non-stationary analyses an important consideration. At present, however, it has not been possible to detect a clear climate change signal in the observed magnitude of annual flood events (Wilson *et al.*, 2010).

Given the expected non-stationarity of flood magnitudes in the future, new approaches are needed for the analyses of non-stationary series. This need applies not only to Norway, but also in general to other countries, and for various types of environmental change. In Norway, climate change projections are now being used to develop guidance for incorporating the effects of climate change into flood estimates. This involves specifying recommended increases in flood estimates on a catchment or sub-catchment basis, where regional climate projections provide an expected increase in the floods of more than 20% by the end of the century (Lawrence and Hisdal, 2011). As a first effort in this regard, the sensitivity of flood inundation maps to an increase in the 200-year flood is being examined for areas where a large increase is expected under a future climate.

Prudhomme *et al.* (2010) recently proposed a scenario-neutral approach to climate change impact assessment on flood risk. This approach is designed to allow evaluation of the fraction of climate model projections that would not be accommodated by specified safety margins, and offers another approach that could be used to help adjust final flows/water levels. The use of projected changes in flooding in flood risk management in general is being investigated within the EU Interreg IVB SAWA project (2008 – 2011), and a comparison of the current status of work by NVE in Norway, by SMHI in Sweden and by regional water boards in the Netherlands can be found in Lawrence and Graham (2010) and Lawrence *et al*., 2011. NVE are also involved in the COST Action ES0901 (European procedures for flood frequency estimation) which is aiming to develop a framework for assessing flood frequency in a changing environment.

2.4 At site analysis

Where sufficiently long data records are available for the site of interest, flood frequency analyses can be relatively straightforward and involves fitting a theoretical statistical distribution to the observed flood data. The Norwegian guidelines for flood estimation (Midttømme *et al*., 2011) recommend performing an at-site analysis where long records (> 30 years) are available. However, for records of 30 – 50 years, only 2-parameter distributions are recommended for this procedure. For stations with more than 50 years of data, 3-parameter distributions can also be applied, if suitable. For all stations, it is recommended that flood frequency analysis for other stations in the region is also performed, in order to compare the results. The Ekstrem, Dagut and Finut statistical software available on Hydra II have been developed by NVE to aid in flood frequency analyses.

2.4.1 Selection of the statistical distribution

Many different statistical distributions are available and commonly applied for flood frequency analysis. For many of them there is no underlying justification for their use other than their flexibility in mimicking the shape of an observed statistical distribution (Coles, 2001). The theoretically correct limit distributions are the Generalized Extreme Value (GEV) distribution in case of AMS and the Generalized Pareto (GP) distribution for PDS. In practise, a number of different distributions are commonly compared, and the flood frequency distribution selected is often that which provides the best fit as described below. The statistical distributions available for use in the Ekstrem, Dagut and Finut software on Hydra II are detailed in Table 2.1. The distributions found to provide the best fit to catchments in Norway are often either the Gumbel (EV1) 2-parameter distribution, or the Generalized Extreme Value (GEV) 3-parameter distribution (Midttømme *et al*., 2011).

However, a frequency distribution should not be selected simply because it provides the best fit to the data, but also the number of parameters of a distribution and knowledge about the properties of the catchment should be considered. Practical application has shown that 3-parameter distributions are very sensitive to outlying events (Cunnane, 1985, Sælthun and Anderson, 1986). The higher the number of parameters, the more flexible the distribution, but also the more easily the distribution can follow particular peculiarities of the dataset. With 2-parameter distributions estimates of the tail quantiles can be severely biased if the shape of the tail of the true frequency distribution is not well represented by the fitted distribution (Hosking and Wallis, 1997). The use of a distribution with more parameters, when these can be accurately estimated, yields less biased estimates of quantiles in the tails of the distribution. One of the advantages of regional frequency analysis, where the pooling of station data creates longer datasets, is that distributions with three or more parameters can be estimated more reliably than would be possible using only data from a single site (Hosking and Wallis, 1997). NVE guidelines therefore specify a minimum period of record for the use of 3-parameter distributions (i.e. 50 years) and recommend the comparison of several distributions, with careful consideration of the influence of outliers (Midttømme *et al*., 2011).

Table 2.1 The statistical distributions available for use in Ekstrem, Dagut and Finut.

Software	**Distribution**
Ekstrem, Dagut/Finut	Log-normal (2 parameter)
Ekstrem	Log-normal (3 parameter)
Ekstrem, Dagut/Finut	Gumbel (EV1) (2 parameter)
Ekstrem, Dagut/Finut	GEV (3 parameter)
Ekstrem, Dagut/Finut	Gamma (2 parameter)
Ekstrem	Gamma (3 parameter)
Ekstrem	Log-Pearson (3 parameter)
Dagut/Finut	Gaussian Normal
Dagut/Finut	Pareto (2 parameter)

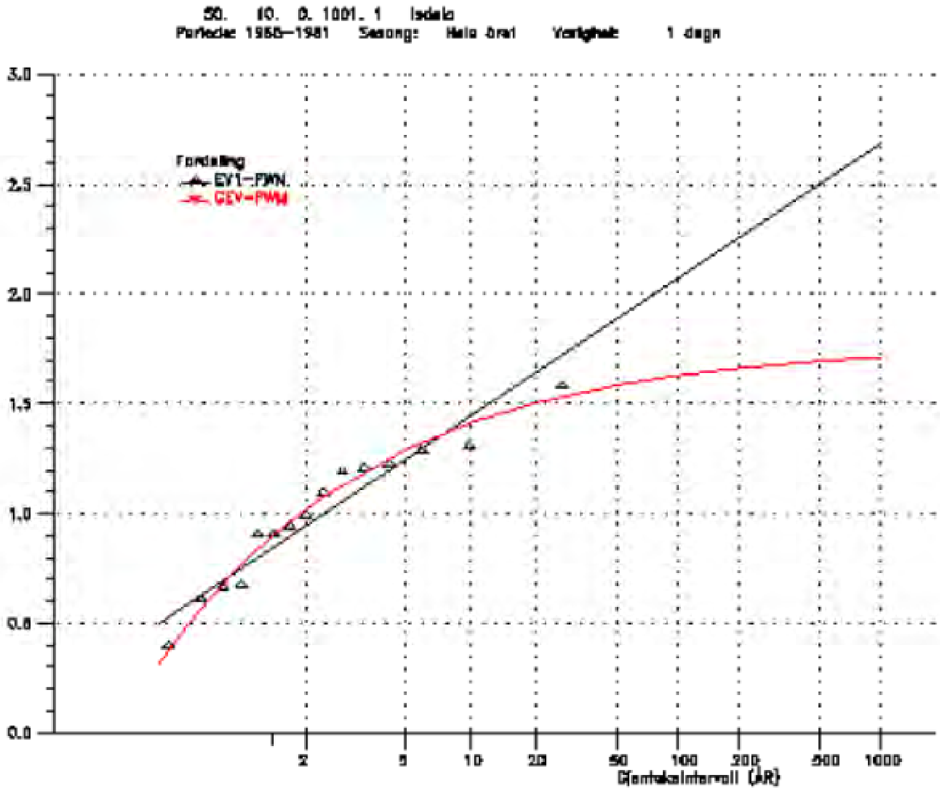


Figure 2.2 Flood frequency curves for Isdøla gauging station using data for the period 1968-1981.

Knowledge about the local characteristics of a catchment may sometimes be needed to judge the reliability of a fitted distribution. For example, the selected statistical distribution can sometimes indicate that there is an relatively low upper bound to flood peaks expected in a catchment (e.g. Figure 2.2 – GEV distribution, although it should be noted that this dataset has fewer values than is recommended (Table 2.1) for use of this distribution). This is nearly always physically unrealistic (Reed and Robson, 1999). However, there are circumstances where this characteristic reflects a real feature, such as attenuation due to floodplain storage.

The selection of the best distribution is often based on visual examination of a plot and is judged by eye, paying particular attention to the fit for the largest floods and the occurrence of outliers. Goodness of fit statistics, such as L-moments as recommended by Hosking and Wallis (1997), can be used to identify the best fitting distribution and to test for acceptability, but these are seldom used in practice. The selection of the flood frequency curve is a subjective decision, and the curve selected is likely to differ if different analysts undertake the same analysis. Figure 2.3 illustrates several flood frequency distributions (available in Ekstrem) fitted to flood event data from Krinsvatn in central Norway. This illustrates the range in flood frequency estimates obtained using different distributions. The variations are:

- 200 yr flood: 260 – 340 m^3s^{-1} (301 m^3s^{-1} ± 14%)
- 1000 yr flood: 234 – 502 m^3s^{-1} (368 m^3s^{-1} ± 36%)

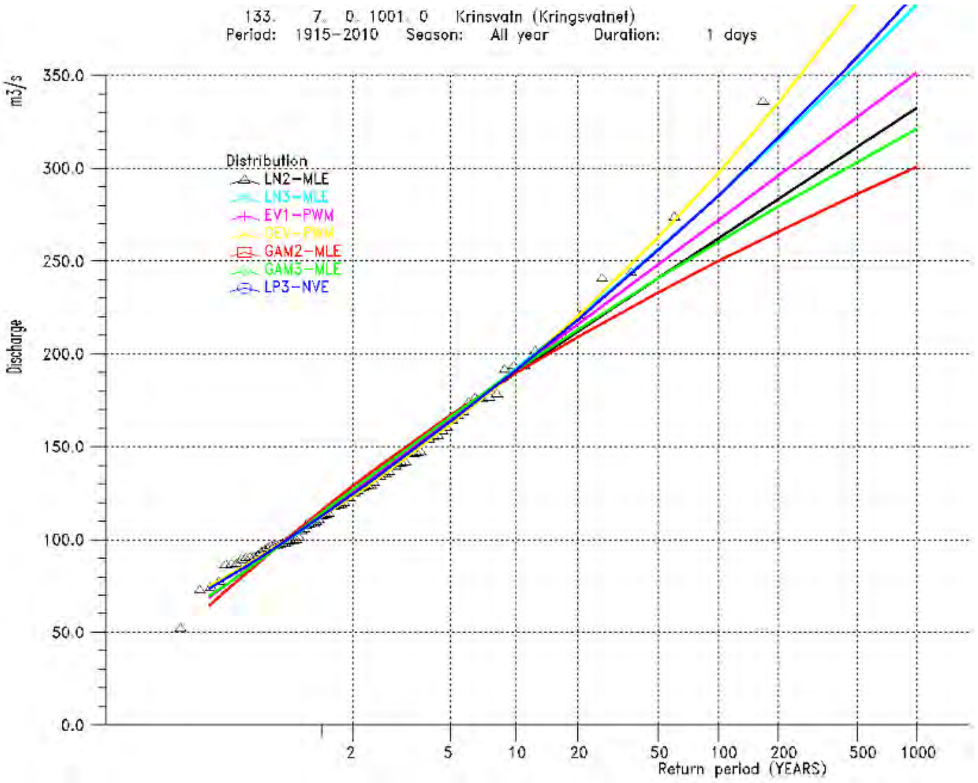


Figure 2.3 Various flood frequency distributions fitted to flood data for Krinsvatn (Central Norway).

The use of goodness of fit statistics can ensure consistency among analysts, but judging the goodness of fit by eye has a key benefit that the fit of the largest floods can be given greater attention, rather than giving equal weight to the fit of all records. The NVE guidelines for flood frequency estimations recommend the comparison of several distributions. However, no procedure to specify the fit of the distributions to the data is suggested. The method commonly applied at NVE is visual comparison of the fitted distributions to the plotted data. For this, first the parameters of the statistical distributions need to be estimated (Section 2.4.2) and then a method for plotting estimated frequencies of the observed data needs to be applied (Section 2.4.3). Both the parameter estimation method and the plotting position formula chosen will influence the fit, and these are discussed in the following two sections.

2.4.2 Parameter estimation

There are several methods available for parameter estimation, including the method of moments (MOM), the probability weighted moments method (PWM, equivalent to L-moments) and the maximum likelihood estimation method (MLE).

Both the MOM and PWM determine the parameters by equating the moments of the data sample with the moments of the statistical distribution. These moments convey information about the location, variance and skewness of the data sample. The PWM often give comparable parameter estimates to the MOM, and in some cases the calculations are simpler. The MOM is a simpler estimation method and is more robust than PWM for small samples with respect to the root means square error of quantile estimates (Engeland *et al.*, 2004). When measured in terms of the bias of the quantile and parameter estimate, however, PWM performs better than MOM (Engeland *et al.*, 2004).

The MLE seeks to determine the distribution parameters that maximise the likelihood of a given observed sample to be the one randomly drawn from the chosen distribution with the estimated parameter values. The MLE is generally considered the most efficient since it provides the smallest sampling variance of the estimated parameters, but iterative calculations to locate the optimum parameters are required and numerical problems can arise during the iteration process and prevent a solution from being found (Reed and Robson, 1999; Rao and Hamed, 2000). However, for small samples the MLE has been found to be less efficient than PWM (Hosking and Wallis, 1997), and it is also found to be less robust in terms of bias and root mean square error of estimated quantiles (Engeland *et al.*, 2004). In other words, for small samples PWM is recommended. The MLE may also perform poorly when the distribution of the observations deviates significantly from the distribution being fitted (Stedinger *et al.*, 1993).

Rao and Hamed (2000) provide a comparison of observed and estimated flows and their 95% confidence intervals for a range of distributions estimated using the MOM, PWM and MLE for parameter estimation. All methods were found to perform well, and none was found to perform consistently better for all distribution types. However, for each distribution only one test dataset was considered, thus precluding general conclusions regarding the performance of each parameter estimation method for each distribution.

At NVE all of the three methods described above are used, and their availability in the Ekstrem, Dagut and Finut software on Hydra II is specified in Table 2.2.

Table 2.2 Parameter estimation methods available in the NVE's Ekstrem, Dagut and Finut software

Software	Distribution	Parameter estimation method
Ekstrem	LogNormal (2 parameter)	MLE
Dagut/Finut		Not specified
Ekstrem	LogNormal (3 parameter)	MLE
Ekstrem	Gumbel (EV1) (2 parameter)	PWM
Dagut/Finut		PWM or MLE
Ekstrem	GEV (3 parameter)	PWM
Dagut/Finut		PWM or MLE
Ekstrem	Gamma (2 parameter)	MLE
Dagut/Finut		MOM or MLE
Ekstrem	Gamma (3 parameter)	MLE
Ekstrem	Log Pearson (3 parameter)	NVE procedure
Dagut/Finut	Gaussian Normal	Not specified
Dagut/Finut	Pareto	PWM or MLE

2.4.3 Plotting positions

Plots are often used to visualise a sample distribution and to identify a good fit between various flood frequency distributions and observed flood magnitudes. As the real frequency distribution of the observed data is unknown, so called "plotting positions" for the data; i.e. estimates for the likely annual exceedance probability/return period of the observed flood magnitudes, need to be found. A frequently used approach is to rank the flood events from largest to smallest, and to assume that each flood magnitude corresponds to the quantile related to its position in the list, i.e. i/n, where n is the number of events and i is the rank of an event. Hence, the largest observation is assigned plotting position $1/n$ and the smallest $n/n=1$ for its annual exceedance probability, *AEP*. The return period, T, of an event is then the inverse of the *AEP*. In practice, there is a range of plotting position formulas available. Most involve the addition of constants to the numerator and denominator, $(i + a)/(n + b)$, in an effort to produce improved estimates in the tails of specific distributions (FEMA, 2007). Some of the plotting positions are optimized for a specific distribution, while others aim to produce either unbiased estimates of exceedance probabilities or quantiles (Stedinger *et al.*, 1993). Examples include the frequently applied Weibull plotting position (Eq. 2.1), which provides unbiased exceedance probabilities for all distributions, and the plotting position by Cunnane (Eq. 2.2), which is approximately quantile-unbiased. One of the first available plotting positions, which is still frequently used, is the Hazen formula. The differences relative to the Weibull and Cunnane formulas are typically modest for i of 3 or more. They can, however, be large for $i = 1$ and $i = n$, i.e. for the smallest and largest events (Stedinger *et al.*, 1993).

Weibull formula:

$$AEP_i = \frac{i}{n+1} \tag{2.1}$$

Cunnane formula:

$$AEP_i = \frac{i - 0.4}{n + 0.2} \qquad (2.2)$$

Hazen formula:

$$AEP_i = \frac{i - 0.5}{n} \qquad (2.3)$$

As different plotting positions plot observed data differently, the choice of plotting position affects the visual judgment of the fit to theoretical distributions. Hence, the use of different plotting positions may result in the choice of a different theoretical distribution. A further point for consideration is that when the goodness of fit between observed data and a flood frequency curve is considered, the error is taken to be the difference between the two values. However, the plotting position for each data point has only been assigned based on the rank of observed values. The error could therefore be in the plotting position assigned, rather than in the observed value (FEMA, 2007).

At NVE, the Gringorten plotting position is most frequently used, as this is available as part of the Ekstrem software:

$$AEP_i = \frac{i - 0.44}{n + 0.12} \qquad (2.4)$$

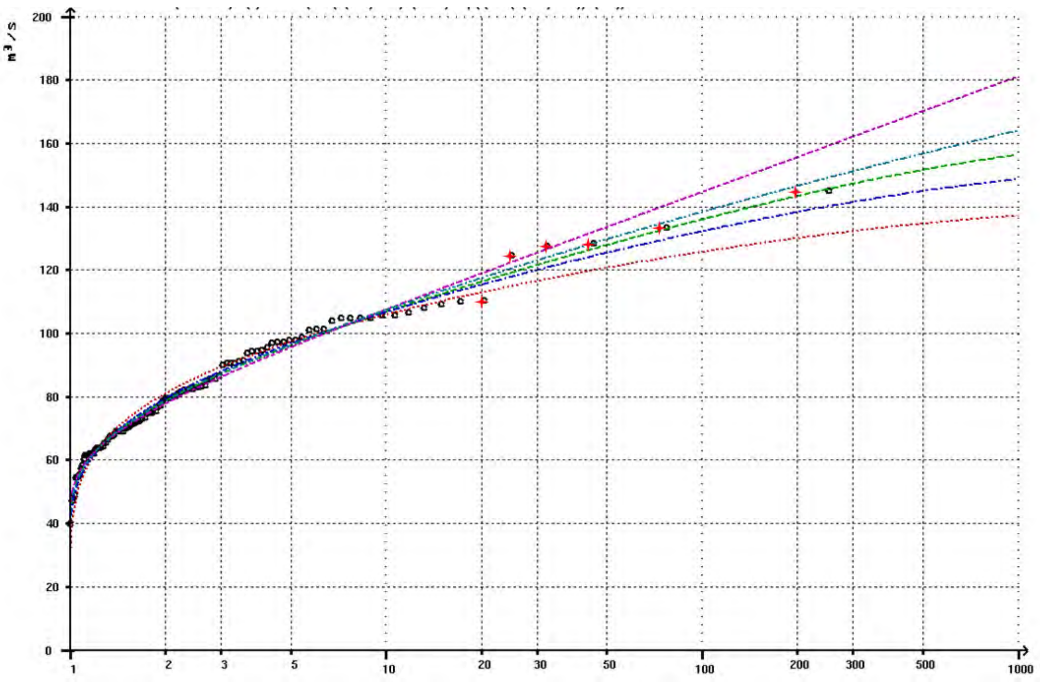


Figure 2.4 The largest annual flood events at Lovatn (1900 – 2000) plotted with plotting positions used in Dagut (black circles) and Extreme (red crosses) together with various flood frequency distributions fitted to the flood data.

The Gringorten formula is optimized to plot the largest observations from a Gumbel distribution. A different plotting position formula is used in the Finut software, which can make another probability distribution appear to fit the plotted data better as compared to the Ekstrem software. An example of how the six largest annual flood events at Lovatn

(1900 – 2000) are plotted with the two different plotting positions from Dagut and Extreme is shown in Figure 2.4, together with various fitted flood frequency distributions. Even though the difference in the resulting plotting positions for lower magnitude events is small, it can be considerable for the largest event.

2.5 Regional analysis

Regional flood frequency analysis is applied when data at a site are insufficient for the reliable estimation of flood quantiles (Cunderlink and Burn, 2001). In this case, flood data from one or several alternative stations with observation records in a "region" are used to improve flood frequency estimation at the site of interest. One could say that space replaces time to increase the sample size and reduce sample uncertainty. A region can be defined geographically or it may comprise stations with similar flow or catchment characteristics. Although a data set constructed with data from different sites may be more heterogeneous, research has shown that more accurate flood frequency estimates are obtained using regional analysis, compared to at-site analysis (Lettenmaier *et al.*, 1987). One has, however, to be careful not to overestimate the reduction in uncertainty, as correlation between the stations reduces the size of the sample comprising independent data.

A common approach for a regional flood frequency analysis is the so-called *index flood method* (Stedinger *et al.*, 1993). This approach assumes the flood magnitudes of all sites in the region follow the same frequency distribution except for a scaling factor, the *index flood*. The mean or median flood is usually used as the index flood, as these can be more accurately estimated from shorter data records as compared to floods with higher return periods. The normalized regional flood distribution is sometimes called the *growth curve*. The flood frequency curve for the site of interest (Q_T) is then constructed as the product of the index flood (Q_M) and the growth curve (X_T):

$$Q_T = Q_M \,.\, X_T \qquad (2.5)$$

A regional flood frequency analysis based on the index flood method hence comprises three steps: (1) identification of regions or similar sites, (2) calculation of the index flood and (3) calculation of the growth curve. These steps are described in more detail in the following sections. NVE's recommendations for a flood frequency analysis (Midttømme *et al.*, 2011) depending on data availability are summarized in Table 2.3. To derive flood frequency estimates for an ungauged site or a site with limited or poor quality data it is recommended that observed data are used, where possible. If no data are available or if there are other reasons for which the observed data are otherwise not appropriate, the index flood is derived by regional regression analysis and regional growth curves are applied. For small catchments, however, the suggested regional approach is not valid and flood estimation based on rainfall-runoff modelling is recommended. Particular considerations may also be necessary for large (> ca. 1000 km²) and diverse catchments, as these may not fit the regional growth curves and similar catchments with observations may not exist (Midttømme *et al.*, 2011). Depending on the application it may be necessary to do separate analyses for sub- catchments.

Table 2.3 Recommended procedures for at-site and regional flood frequency analysis according to data availability. For regional analysis the procedures for calculating the index flood (mean flood or median flood) and the growth curve are further specified.

	Data available	**Procedure for calculation of the index flood**	**Procedure for calculation of growth curve for target return periods between Q200 and Q1000**
Gauged: long series	> 50 years	(not relevant)	Calculated from 2- or 3-parameter distribution, based on the observed at-site series
	30-50 years	(not relevant)	Calculated from 2-parameter distribution, based on the observed at-site series
Gauged: short series	10-30 years	Calculated from observed at-site series	Calculated by analysis of other long series in the area; possibly extension of series by model simulation
	< 10 years	Calculated by correlation with other series and/or regional regression formulas	Calculated by analysis of other long series in the area; possibly extension of series by model simulation
Ungauged		Calculated for nearby sites and scaled or regional regression formulas	Use of regional flood frequency curves

2.5.1 Identification of homogeneous regions

A crucial step in a regional flood frequency analysis is the identification of appropriate regions. A region may be geographically coherent or may encompass sites that are dispersed and not contiguous. A major prerequisite for the regions is that they fulfil the basic assumption of the index flood method; i.e. that the flood magnitudes of all sites within a group follow the same frequency distribution, differing only by a scaling factor (Tallaksen *et al.*, 2004). Usually, it is impossible to satisfy this theoretical homogeneity criterion exactly, and approximate homogeneity may be sufficient to ensure that the regional frequency analysis is more accurate than an at-site analysis with a smaller data sample (Hosking and Wallis, 1997). Many different methods for defining homogenous regions are available and applied in practice (Hosking and Wallis, 1997, and Tallaksen *et al.*, 2004). The basic concepts include (1) the delineation of fixed, geographically coherent regions according to administrative borders or general knowledge of geographical, hydrological and climatic conditions, (2) the identification of homogenous groups of sites based on different kinds of hydrological or catchment characteristics, and (3) the identification of a suitable group of stations specific to an individual site (sometimes also called *pooling* of sites). The latter is the basis of the so-called "Region of Influence" approach (ROI; Burn and Goel, 2000). For the identification of homogeneous regions, a number of statistical methods are available, such as cluster analysis, split-sample regionalization or empirical orthogonal functions (Tallaksen *et al.*, 2004). These methods can be applied to different types of input data. For flood frequency analysis the grouping is typically based on time series or summary statistics of flood data or other hydrological variables when all considered sites are gauged. Otherwise, proximity or location in terms of latitude and longitude are frequently used as well as climatological characteristics or other catchment descriptors.

Within the ongoing COST Action ES0901 "European procedures for flood frequency estimation, FloodFreq" the methods applied within Europe have been summarized (Castellarin *et al.*, 2011). The most frequently used regionalization scheme is the delineation of fixed, geographically coherent regions according to geographical, hydrological and climatic characteristics. As a grouping procedure, cluster analysis is most commonly applied. Other applied methods include, for example, the Region of Influence (ROI, Burn, 1990) approach used in Italy and UK, and top-kriging (Merz *et al.*, 2005; Skøien *et al.*, 2006; Skøien and Blöschl, 2007), a novel geostatistical method that takes into account the river network structure and catchment area. In Austria, this method has been used to interpolate the 30-, 100- and 200-year flood quantiles over the entire Austrian river network length, i.e. 26000 km, representing 10500 sites. The applicability of the method is dependent on a sufficiently dense network and a sufficient number of nested catchments. The procedure used in the UK is described in Box 1.

Box 1 - UK pooling group approach (Kjeldsen, 2011)

Each site of interest is considered to lie at the heart of a group of gauged catchments to which it is hydrologically similar. All stations in a pooling group influence the resultant growth curve to some extent, but greater weight is given to the catchments judged most similar and with the longest records.

The similarity measure used to identify the sites of a pooling group and to assign weights for calculating the growth curve is "based on catchment area, standard annual average rainfall as recorded in the reference period 1961-1990, an index of flood attenuation from upstream lakes and reservoirs, and an index of upstream extent of flood plains (ratio of 100-year flood plain compared to total catchment area).

Following methodological developments reported by Kjeldsen et al. (2008) and Kjeldsen and Jones (2009b), there is no longer a need for the pooling groups to be homogeneous. The differences of L-moment ratios (L-CV and L-SKEW) between catchments have been taken into account in the underlying statistical model" (Kjeldsen, 2011).

It may sometimes be the case that the most appropriate delimitation is indeed geographical due to the effect of climatic, topographic and maritime influences, but geographical proximity is not necessarily an indicator of the closeness of frequency distributions (Hosking and Wallis, 1997). Merz and Blöschl (2005) compared the predictive performance of various flood regionalisation methods, including multiple regression, kriging and a variant of the region of influence approach, for flood frequency estimation in ungauged catchments in Austria. They found that the best performance was achieved using a geostatistical method that combines spatial proximity and catchment attributes.

In Norway, it is recommended that for sites having no or limited data, flood frequencies are estimated with the help of one or several nearby stations having longer observation records whenever possible (Table 2.3). The choice of these stations is largely subjective, and proximity and catchment area are primarily used as the similarity criteria. The quality and length of the flow record are also considered. Thus, if a record is available for a site within the same river basin, it will usually be included. Sites in neighbouring catchments are often used based simply on their proximity, but checks as to their suitability in other respects are infrequently made. The general reliability of the derived growth curve is, however, assessed by comparison with other catchments in the vicinity. Greater attention to comparisons of the similarities between catchments with respect to catchment characteristics would improve the procedures currently in use. Pettersson (2008) found, for example, that the growth curve is influenced by catchment parameters such as size, lake percentage and the mean specific flood.

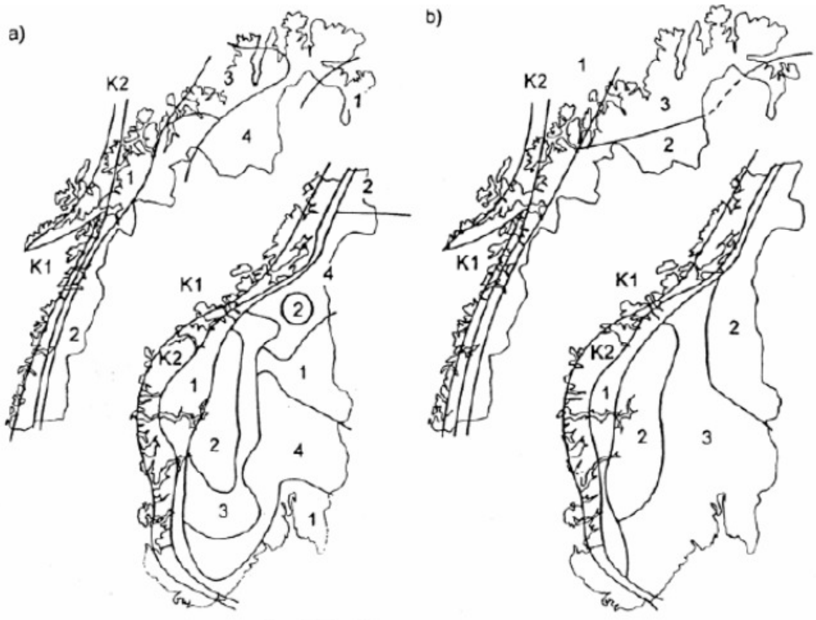


Figure 2.5 Flood regions: annual flood regions (K1 and K2), together with (a) regions for spring floods (V 1-4) and (b) regions for summer and autumn floods (H 1-3; Midttømme *et al.*, 2011).

If no data are available at the site of interest or at a nearby location, regression formulas for the index flood and growth curve available for established regions can be applied. These flood regions have been defined by cluster analysis on the basis of 212 catchments with at least 20 years of observations and no or only minimal influence from regulation (Sælthun, 1997). As it is important to analyse floods generated by different processes separately, the catchments were first separated into four classes according to the season during which the most critical floods (in terms of annual flood peak magnitude) occur: 1) spring floods during the snow-melt season, 2) summer/autumn floods usually generated by heavy rain, 3) annual, i.e. catchments where the occurrence of critical floods is not limited to a particular season but may occur during several seasons of the year, and 4) catchments with a glacier percentage ≥ 5%. Catchments along the west coast of Norway typically belong to the annual flood class, whereas both spring and summer/autumn catchments are present in all other parts of Norway. Separate geographical regions were delineated for the three classes based on a hierarchical cluster analysis with six climatic parameters (mean annual precipitation, the relationship between mean annual precipitation and precipitation with a 5-year return period (%), mean total number of days with snow cover, mean annual snow depth, mean temperature in January and July). The homogeneity within the identified regions was verified with respect to Wiltshire's homogeneity test. This resulted in two annual regions, four spring flood regions and three summer/autumn flood regions (Figure 2.5) as well as a separate glacier region.

Further research would be required to establish if it is possible to improve flood frequency estimation by grouping station data in Norway following other approaches e.g. those used in the UK. This is a general question, in terms of the transferability of flood frequency estimation methods, that the COST Action 0901 (European procedures for flood frequency estimation) aims to address. However, even if other grouping approaches are not suitable for use, the increase in available flood data since 1997 should ideally be used to increase the robustness of flood frequency estimates based on regional analyses.

2.5.2 Index flood

Internationally, the mean or median flood is usually used as the index flood. For an ungauged site or a site with limited data it can either be estimated using flow data from nearby or similar sites or it can be derived using regional regression formulas. In the first case, the index flood can be scaled to the site of interest based on the catchment area. However, other catchment characteristics can also play a significant role which can lead to either an under- or over-estimation of the flood frequency at the site of interest. Therefore, regional formulas for the calculation of the index flood can be used. This is a very practical approach when no data or only limited data exist. However, in general, flood estimates derived from catchment descriptors are grossly inferior to estimates made from flood peak data, even those estimated from short records (Reed and Robson, 1999). Brath *et al.* (2001) compared different methods for estimating the index flood at ungauged sites in Northern Italy and found a regression model linking the index flood to a set of catchment descriptors to be the most efficient approach. Kjeldsen and Jones (2010) recently revised the FEH procedure for the derivation of the index flood at an ungauged site in the UK using catchment descriptors. They found that local factors are probably not sufficiently represented in the FEH regression models (a single model is used for the whole of the UK), and that flood statistics may benefit from the adjustment of estimates using local data from neighbouring catchments. Their results showed geographical proximity to be the most important factor when identifying a good potential donor site, with little benefit gained by identifying donor sites based on hydrological similarity.

In Norway the mean flood is most often used as the index flood, but the median flood is also used. The calculation procedures are summarized in Table 2.3. If no data are available at the site of interest, but data are available for one or several nearby sites, these data are used to calculate the index flood by scaling based on catchment area. If data from nearby sites are not available, a regional regression formula can be used. Such regional formulas based on catchment descriptors, are available for each of the flood regions described in Section 2.5.1 (Table 2.4). However, they are only valid for catchments larger than 20-50 km^2 and should be used with particular caution for catchments smaller than 100 km^2. An upper limit for use of the formulas is not specified.

NVE are currently reviewing the regional equations detailed in Table 2.4 and are considering including new catchment and climate characteristics (e.g. catchment area and 5-year rainfall). However, even with use of the formulas in the recommended range, the smallest catchments will often have a shorter reaction time. A short-duration, intense rain event is more critical for a small catchment than for a large catchment. As a consequence, the scaling of flows (based on catchment area) and the use of regional flood frequency curves to estimate flow in small ungauged catchments may result in an underestimation of flood magnitudes for each return period. Flood formulas also perform less well for large catchments, but there are often several gauged sites within these catchments which can be used to better estimate flood frequencies. It is hoped that by reviewing the equations, they can be made applicable for use in smaller catchments. The procedure of using a single equation for the whole of Norway (rather than separate regional equations) and the adjustment of regional estimates based on nearby station data could also be approaches applicable to Norwegian catchments.

Table 2.4 Regional formulas for derivation of the index flood (Q_M in ls^{-1}km^2)

Spring flood regions	
1	$\ln Q_M = 0.2722 \bullet \ln S_T - 0.1406 \bullet \ln A_{SE} + 0.1006 \bullet \ln A_{SF} + 0.6172 \bullet \ln Q_N + 2.11$
2	$\ln Q_M = 0.0930 \bullet \ln S_T - 0.0816 \bullet \ln A_{SE} + 0.0281 \bullet \ln A_{SF} + 0.5076 \bullet \ln Q_N + 3.59$
3	$\ln Q_M = 0.3066 \bullet \ln S_T - 0.0220 \bullet \ln A_{SE} + 0.0939 \bullet \ln A_{SF} + 0.3252 \bullet \ln Q_N + 3.09$
4	$\ln Q_M = 0.1848 \bullet \ln S_T - 0.0137 \bullet \ln A_{SE} + 0.0873 \bullet \ln A_{SF} + 0.5143 \bullet \ln Q_N + 2.77$
Autumn flood regions	
1	$\ln Q_M = 1.2805 \bullet \ln Q_N - 0.2267 \bullet \ln(A/L_F) + 0.0664 \bullet A_{SE} + 0.0053 \bullet S_T + 1.00$
2	$\ln Q_M = 1.2910 \bullet \ln Q_N - 0.1602 \bullet \ln(A/L_F) + 0.0508 \bullet A_{SE} + 0.0065 \bullet S_T + 0.65$
3	$\ln Q_M = 1.2014 \bullet \ln Q_N - 0.0819 \bullet \ln(A/L_F) + 0.0268 \bullet A_{SE} + 0.0013 \bullet S_T + 1.07$
Glacier and annual flood regions	
BRE	$\ln Q_M = 0.0119 \bullet Q_N - 0.0848 \bullet A_{SE} + 0.0165 \bullet L_F + 5.81$
K1	$\ln Q_M = 1.5212 \bullet \ln Q_N - 1.1516 \bullet \ln P_N - 0.0569 \bullet A_{SE} - 0.0093 \bullet L_F + 8.80$
K2	$\ln Q_M = 1.1524 \bullet \ln Q_N - 0.0463 \bullet A_{SE} + 1.57$

Where: A = catchment area (km^2), Q_N = mean specific annual runoff (ls^{-1}km^2), P_N = mean annual precipitation (mm), A_{SE} = effective lake (%), A_{SF} = exposed bedrock (%), L_F = catchment length (km), S_T = gradient of the main river (m/km).

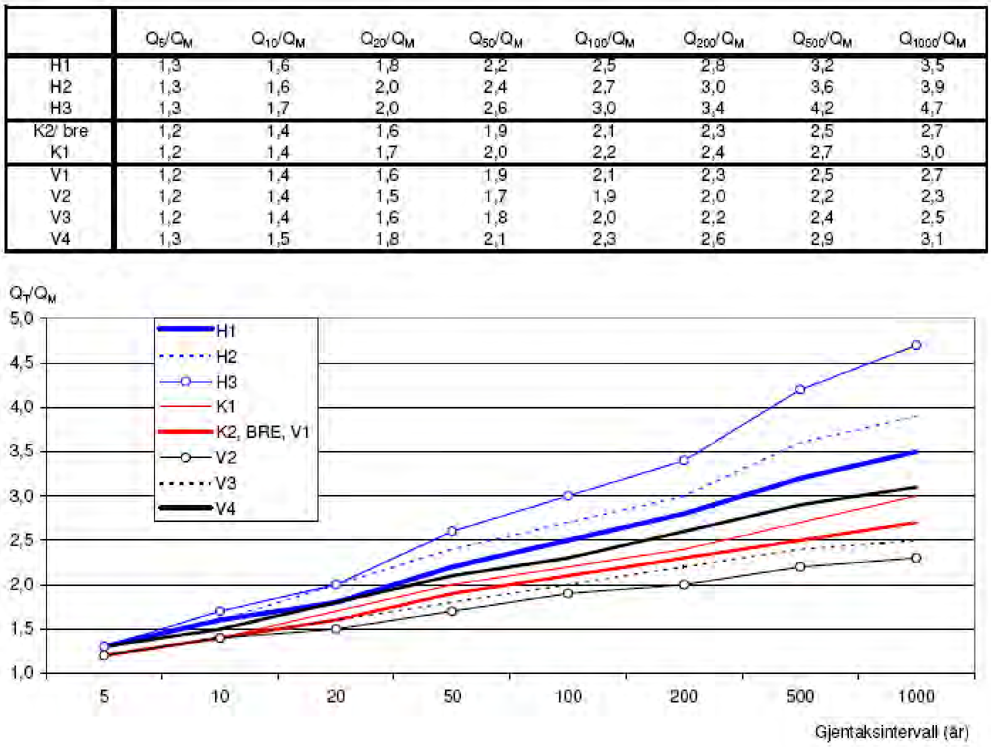
Estimates of the index flood are frequently transferred from a gauged site to the site of interest by scaling based on catchment area. Although the index flood is heavily dependent upon catchment area, the relationship is not linear. Pettersson (2008) found that the specific mean flood decreases with increasing catchment size. The specific mean flood within Norwegian catchments typically lies within the range 200-600 ls^{-1}km^2. In larger catchments, i.e. >500 km^2, the mean specific flood tends to range to a maximum of

400 ls^{-1}km^{2}. In very small catchments, the mean specific flood can range from 100 to 2000 ls^{-1}km^{2}. This suggests that, where possible, a similar sized catchment to the site of interest should be used for estimation of the index flood, and this is the approach generally adopted.

NVE (Midttømme *et al.*, 2011) advises that a minimum of 10 years are used to calculate the index flood, but Pettersson (2008) found that robust values are only obtained with a minimum of 30 years. For records less than 30 years, the index flood value was found to vary widely depending on the period used. A standard minimum period of 30 years is usually used for the calculation of climatological and hydrological averages, and Pettersson (2008) recommended that this is also applied to calculate the index flood.

2.5.3 Growth curve

The growth curve can either be derived using nearby station data or, in the absence of long series from stations nearby, fixed regional growth curves. Regional growth curves (Figure 2.6) have been defined for all Norwegian flood regions shown in Figure 2.5, and their definition has been based on the same data set as the regions (Sælthun, 1997). When data from several nearby sites are available, NVE (Midttømme *et al.*, 2011) recommends that the regional growth curve can be obtained by estimating the distribution for each site separately and combining the at-site estimates (following division by the index flood) to give a regional average. This may increase the robustness of the estimates, but in practise, individual at-site analyses tend to be undertaken, with one selected as the best, rather than combining results to give a regional growth curve.



	Q_5/Q_M	Q_{10}/Q_M	Q_{20}/Q_M	Q_{50}/Q_M	Q_{100}/Q_M	Q_{200}/Q_M	Q_{500}/Q_M	Q_{1000}/Q_M
H1	1,3	1,6	1,8	2,2	2,5	2,8	3,2	3,5
H2	1,3	1,6	2,0	2,4	2,7	3,0	3,6	3,9
H3	1,3	1,7	2,0	2,6	3,0	3,4	4,2	4,7
K2/ bre	1,2	1,4	1,6	1,9	2,1	2,3	2,5	2,7
K1	1,2	1,4	1,7	2,0	2,2	2,4	2,7	3,0
V1	1,2	1,4	1,6	1,9	2,1	2,3	2,5	2,7
V2	1,2	1,4	1,5	1,7	1,9	2,0	2,2	2,3
V3	1,2	1,4	1,6	1,8	2,0	2,2	2,4	2,5
V4	1,3	1,5	1,8	2,1	2,3	2,6	2,9	3,1

Figure 2.6 Regional growth curves (Midttømme *et al.*, 2011).

2.6 Instantaneous flood peak

Mean daily flow data are traditionally used for flood frequency analysis. In some cases the recurrence interval of a daily average may be smaller than for the instantaneous flood peak. Where flood magnitudes have been estimated using maximum mean daily flow values, instantaneous flood peaks must also be estimated (Midttømme *et al.*, 2011).

Where instantaneous flood peak data are available it is recommended that a flood frequency analysis of these data is performed. Such analyses are crucial, as they convey valuable information about the maximum size of the flood peak, rather than the 24-hour average value. However, observations of instantaneous flood peaks are uncertain due to occasional missing peak data values or to the influence of ice. It is a common problem that gauging stations suffer from failure or overtopping in extreme flood conditions, and these are the data that are critical to the reliability of the resultant flood estimates. Short observation periods also increase the uncertainty of instantaneous flood frequency estimates. Some stations have been equipped with continuous stage recorders, which are able to record data at fine time resolutions since the 1960s, while at other stations installation occurred later. Another frequent source of uncertainty is the upper part of the rating curve, as it usually has been necessary to extrapolate the curve due to the lack of spot measurements at the highest observed water levels. Due to these potential problems with the accuracy of instantaneous flood data, it is important that any data used to derive flood frequency estimates are reviewed before use. However, different users frequently review the Finut database and remove different records on data quality grounds. It would be beneficial for all users and for the quality of the resultant assessments if NVE were to undertake a thorough review of the data and demarcate a portion of the database which is suitable for use in instantaneous flood frequency assessments. In an ongoing project at NVE, flood frequency analyses based on instantaneous values are being performed for all small catchments with at least 10 years of high resolution data. Flood events with a return period up to 20 years are currently being estimated and compared to daily mean estimates. As part of the project, the reliability of the observed flood events is compared to corrections made in the corresponding daily series. However, identified problems are currently only documented on paper.

Where instantaneous flood peak data are not available for a site, it is recommended that scaling is performed based on the relationship between the daily flows and the instantaneous peak flows for the largest floods in the catchment or a comparable catchment (Midttømme *et al.*, 2011). Appendix 2 of Midttømme *et al.* (2011) also details the observed ratios for 106 gauging stations, which can be adopted for a site of interest. This assumes that the growth curve for daily flow at the site of interest is the same as, or is at least similar to, the peak flow curve at another site within the catchment or a comparable catchment. Such analyses make the best use of available data, but the degree of uncertainty is likely to be large, particularly given that the ratio between peak flow and the corresponding daily flows at the same site can vary greatly between individual flood events (Sælthun and Anderson, 1986).

If data are not available it is recommended to use the formulas in Table 2.5 which estimate the instantaneous flood peak based on catchment descriptors. It is acknowledged however that the equations detailed in Table 2.5 can produce unrealistic values, especially in large catchments and catchments with a high lake percentage (Midttømme *et al.*, 2011). Careful use of these equations is therefore required. As part of the above mentioned comparative project for flood frequency analysis in small catchments, NVE are currently reviewing these equations with the aim of developing regional formulas for the calculation of instantaneous flood peaks.

Table 2.5 Regression equations for the ratio of the instantaneous flood peak (Q_i) and the maximum mean daily flow (Q_d)

Spring flood:	$Q_i / Q_d = 1.72 - 0.17 \bullet \log A - 0.125 \bullet A_{SE}^{0.5}$
Autumn/summer flood:	$Q_i / Q_d = 2.29 - 0.29 \bullet \log A - 0.270 \bullet A_{SE}^{0.5}$

A = catchment area

A_{SE} = effective lake percentage

2.7 Performance of flood frequency analysis

The performance of flood frequency analysis varies from catchment to catchment depending on the availability of data and the representativeness of flood formulas. Where a long series of reliable flood event data are available for the site of interest, flood frequency analysis is the best method of estimating flood quantiles. In case of a regional analysis the similarity of the chosen alternative sites to the site of interest plays a major role, and when using the regional formulas to derive the index flood, catchment size is one of the key factors affecting the performance of flood frequency analyses as the regional formulas perform best for medium-sized catchments.

In general, it is important to keep in mind that there always will be uncertainty in the flood estimates. The uncertainty is in particular large when events of a large return period need to be estimated based on a small or no data sample. Such limited sample sizes risk being unrepresentative of the true flood frequency distribution. The occurrence of a large flood or the absence of a flood in a year, can greatly affect the results, especially when small samples are used to estimate low probability events (Hosking *et al.*, 1985). Uncertainty increases with increasing return period. Estimates for rarer floods (>200 years), which are often the target return periods of interest, have large uncertainties. These low frequency – high magnitude floods require significant extrapolation beyond the observed data series and rely heavily on the statistical distribution adopted.

Figure 2.7 illustrates the impact of using short records to estimate flood frequency statistics at Øye ndf. in Western Norway. Data are available for this site for the period 1916-2010. In Figure 2.7 flood frequency estimations based on 18 years of observations from five different periods are shown. The estimates based on the different periods deviate considerably. This is particularly the case when comparing the estimate based on 1952 – 1969 with the other periods, as the most extreme flood occurred during this period. But even when comparing only the estimates of the remaining four periods, estimates for the 200 year flood vary between approximately 100 and 160 m^2/s, and for the 1000 year flood between approximately 100 and 200 m^2/s.

The guidelines (Midttømme *et al.*, 2011) recommend carrying out flood frequency analyses for several stations in a region, both to verify that the individual series do not provide extreme distributions and to provide an overview of the regional pattern. However, the uncertainty of a flood estimate is usually not conveyed, even though some procedures to address (parts of) this uncertainty are available within NVE's flood frequency software. Within Dagut/Finut a bootstrap function to estimate and plot confidence intervals on fitted distributions is available. Confidence limits are a function of sample size and distribution parameters. An example for Øye ndf. (1921 – 1950) is shown in Figure 2.8, where a GEV distribution has been fitted and the 5%- and 95%-confidence intervals have been calculated using the bootstrap method with 1000 iterations. Ekstrem does not calculate uncertainty bounds, but it is possible to calculate these separately using, for example, the R Statistical package, or other software.

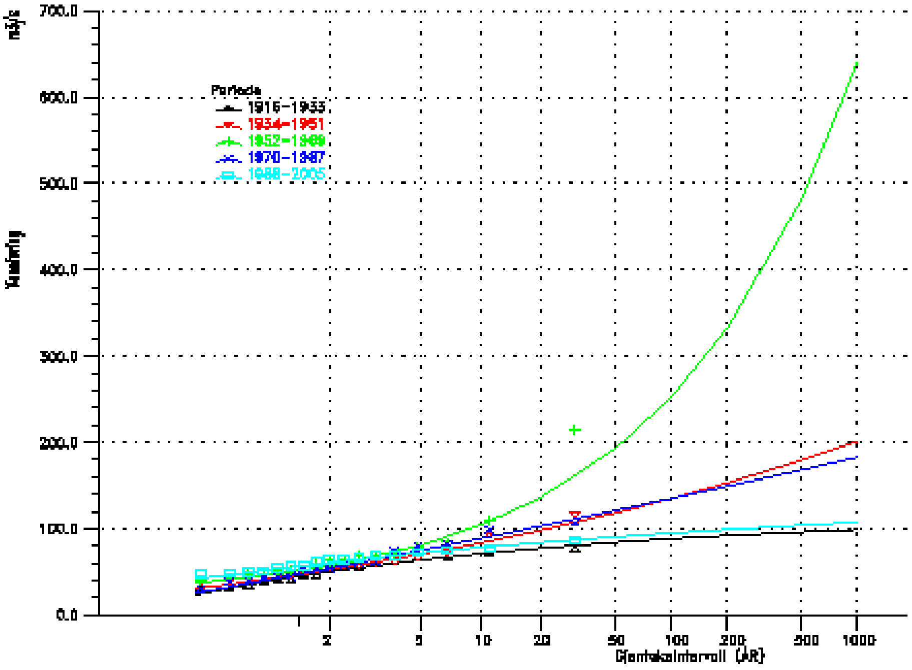


Figure 2.7 Flood frequency analysis for Øye ndf., Western Norway for five different 18-year periods.

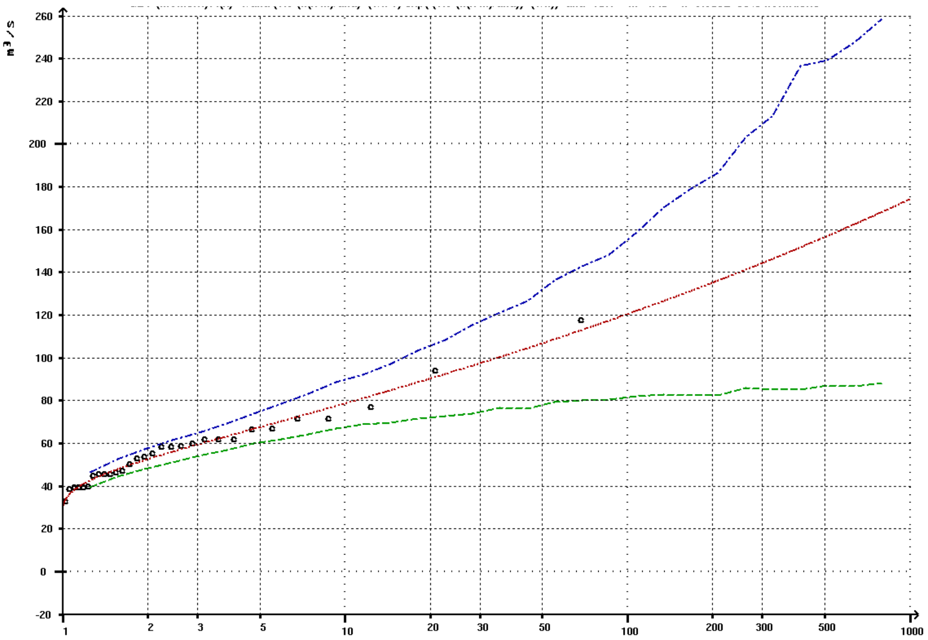


Figure 2.8 Flood frequency analysis for Øye ndf. (1921 – 1950), showing a fitted GEV distribution (red) together with the 5%- (blue) and 95%-confidence intervals (green).

3 Rainfall-runoff modelling

Rainfall-runoff modelling complements the use of flood frequency analysis for the derivation of flood magnitude estimates. In rainfall-runoff modelling, a rainfall input (which is often also combined with a snowmelt contribution) is converted to a flow output using a model for the catchment response. The main reasons for this approach include (from Killingveit and Sælthun, 1995):

- data series of precipitation are often longer than runoff series
- the climate station network is in some locations more dense than the gauging station network
- precipitation shows stronger regional consistency than runoff

In addition, estimation of the probable maximum flood (PMF), which is used in dam safety analyses in Norway to assess safety against dam break, cannot be undertaken using statistical methods. The application of rainfall-runoff modelling is therefore required.

In Norway, a simple, lumped, event-based precipitation-runoff model (PQRUT) is often applied for dam safety analyses to model low frequency events (*e.g.* 500, 1000-year peak inflow), and the probable maximum flood (PMF). The method and computer program for this model were developed in the 1980's (Andersen *et al.*, 1983) and are still in use, with few modifications. A fairly complete description of the method can be found in Midttømme *et al.* (2011). PQRUT was developed to provide a conceptual model for the rainfall-runoff process using a limited number (3) of adjustable parameters, which in principle can be estimated from readily available catchment characteristics. Conceptually, the model is a simplified version of the HBV model (Bergström, 1976; Sælthun, 1996). However, unlike HBV, the model is event-based and is designed to model storm hydrographs to reservoirs, rather than longer-term seasonal patterns of runoff. The model is often implemented for small catchments using an hourly time-step, reflecting the short concentration times between peak rainfall and peak runoff. The more detailed HBV model is also used in Norway where the data required for calibration of the >15 model parameters are available and the size and concentration time of the catchment are sufficiently large to justify the use of a daily time-step. These requirements, however, severely limit the use of the current version of the HBV model in routine analyses for dam safety which are often concerned with small catchments having rapid response times. The conceptual PQRUT rainfall/snowmelt-runoff model is briefly described in the following section. In PQRUT, the catchment is represented as a simple, lumped, three-parameter inflow-outflow (i.e. 'bucket') model, illustrated in Figure 3.1, where the outflow q (*i.e.* inflow to a reservoir) occurs at either a slower rate ($K_2 * H$) or a faster rate ($K_1 * H$), depending on the value of H (accumulated rainfall and/or snowmelt) relative to a threshold value, T. Discharge is greatest when H is above the threshold value. These three parameters correspond, at least conceptually, to the two recession curve slopes and their boundary from a two-component hydrograph separation, as also illustrated in Figure 3.1.

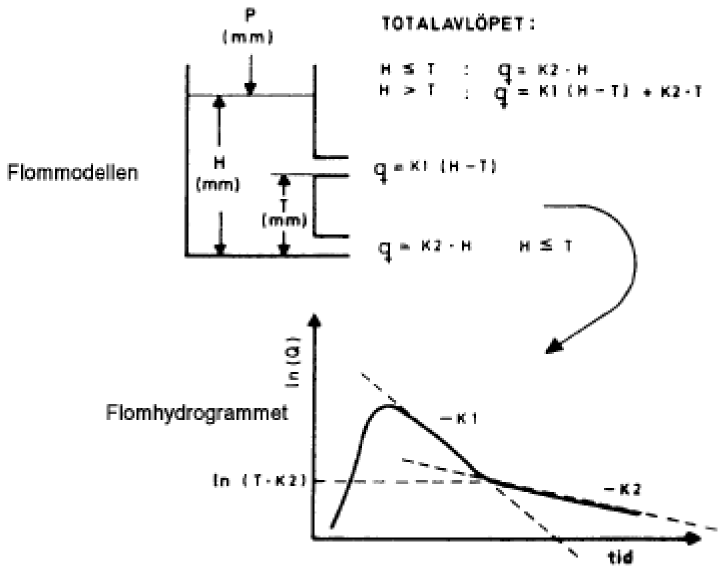


Figure 3.1. The three-parameter precipitation-runoff model used in PQRUT.

3.1 Estimation of model parameters

The three PQRUT model parameters can be calibrated based on observed discharge as a function of observed rainfall (and, where relevant, snowmelt). However, this is rarely done in practise. A reason for this is that such data are generally not available at time steps of less than 1 day. In order to adequately represent peak flows in small catchments with short concentration times, the PQRUT model needs to typically be run using a one- to six-hour time step. A set of three empirical equations was therefore developed to estimate the three model parameters based on the catchment characteristics from 20 catchments (Andersen *et al.*, 1983). These are still in use and are given by:

$$K_1 = 0.0135 + 0.00268{*}HL - 0.01665 * \ln(A_{SE}) \qquad (3.1)$$

$$K_2 = 0.009 + 0.21{*}K_1 - 0.00021{*}H_L \qquad (3.2)$$

$$T = -9.0 + 4.4 * K_1^{\;-0.6} + 0.28{*}Q_N \qquad (3.3)$$

where H_L is a measure of catchment relief , A_{SE} is the effective lake percentage, and Q_N is the normal specific runoff (litre $*s^{-1}$ km^{-2})). The effective lake percentage is defined as $100 * \Sigma (A_i * a_i)/A^2$, where a_i is the surface area of lake i, A_i is the upstream catchment area contributing to lake i, and A is the total catchment area. This formulation is used as it places a larger weighting on lakes which lie further downstream and which therefore have a more significant role in attenuating a flood peak. The empirical estimates for the parameters were developed for relatively small catchments (<500 km^2) and do not perform well for larger catchments or for catchments with a large percentage of lake cover. In general, the parameters are very sensitive to the variable A_{SE}, the effective lake percentage.

The empirical equations used to estimate the three parameters of the PQRUT model have not been re-evaluated since they were developed in 1983. High resolution runoff data is now available for several catchments and could be used to reassess the suitability of this simple three-parameter event-based model at sites where high-resolution rainfall data is also available. Current development work on a gridded version of the HBV model (Beldring *et al*., 2003) and the availability of ARCGIS data for describing catchment characteristics in detail extend the possibility of using an HBV-type model with a sub-daily time-step for modelling arbitrary catchments. This approach is, though, subject to the same constraints as standard HBV and PQRUT modelling in that data are, in principle, required for model calibration. For ungauged areas, robust methods must, alternatively, be developed for extrapolation of the >15 HBV parameters from calibrated catchments.

3.2 Rainfall depth and duration

Event-based simulation of runoff for dam safety analyses with PQRUT uses estimates of extreme rainfall (e.g. 500-yr, 1000-yr, PMP) for a catchment. Estimates of storm depth and duration for each return period by season are provided by the Norwegian Meteorological Institute following the method described in Førland and Kristoffersen (1989) and in Førland (1992). This method applies a similar approach to that developed by NERC (1975) for the UK using a large amount of precipitation data. In that work, growth curves for precipitation depth with long return periods (T = 500, 1000, 10000 and PMP) were developed as a function of the 24-hour precipitation with a 5-year return period (*i.e.* M5). The relationship between M5 and precipitation with other return periods is shown in Figure 3.2.

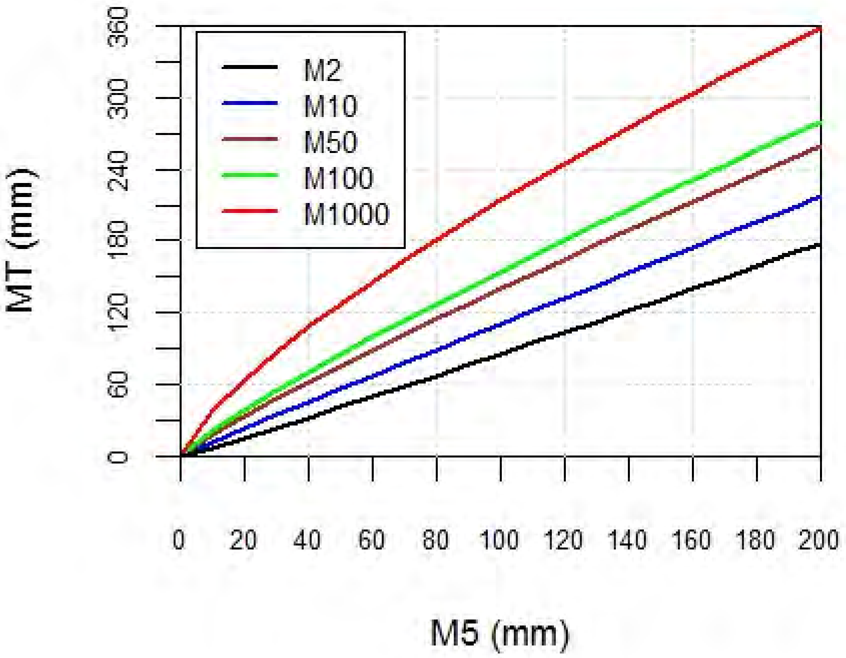


Figure 3.2 The relationship between precipitation events with a 5 year return period, M5, and that with other return periods (based on equations given in Førland, 1992).

For a given return period, a storm can be described by its depth and duration. As the rainfall duration increases, the rainfall depth also increases and *vice versa*, although the relationship is not linear. Establishing the critical storm duration for a catchment is important for deriving estimates of flood magnitude. In catchments where there is no attenuation of the flood response (due to lakes/reservoirs), Reed and Field (1992) found that curves of flood magnitude against storm duration are generally flat. This is because for a given return period changes in storm duration are compensated by changes in storm depth. This means that in non-attenuated catchments, the choice of storm duration is less critical for deriving flood peak magnitude (Reed and Field, 1992), but in general the faster the catchment response, the shorter the critical storm duration. With respect to the catchment response time, NVE (Midttømme *et al.*, 2011) suggests two possible methods for obtaining this: 1) setting this to $1/K_2$, where K_2 is the slower hydrograph recession parameter illustrated in Figure 3.1; or 2) estimating the concentration time based on the catchment length and an assumed surface water velocity of 1 to 2 m/s. However, for the purposes of dam safety analyses, the critical duration for the reservoir is also of importance in setting an appropriate storm duration for the event-based simulation, and this is determined from the hydraulic characteristics of the dam and the estimated peak discharge.

3.3 Storm profile

The design storm depth is distributed over the storm duration, and NVE (Midttømme *et al.*, 2011) recommend that the distribution selected is approximately symmetrical for durations less than 2 days and is skewed for longer durations. The actual choice of the profile is left to the discretion of the practitioner, although the choice must be justified in reporting. Ideally, the storm profile used will be representative of the typical storm profile for extreme precipitation events, and WMO (2009) recommends that the observed storm characteristics of a basin are studied in order to define the correct profile. They also recommend that other probable profiles are considered if these might be more critical than the storm profile which is representative of the catchment. Identifying appropriate storm values takes considerable effort, and in many small catchments with short concentration times, it is not possible due to the lack of appropriate high-resolution data. Sælthun and Andersen (1986), in fact, were unable to identify typical storm profiles for catchments in Norway. This is, though, potentially a topic that might now be fruitfully revisited given the wider availability of high-resolution precipitation and discharge data.

3.4 Areal reduction factors

Point rainfall values are only representative of a very small area, and the average rainfall over a larger area is likely to be much smaller than the point of maximum observed depth (Svensson and Jones, 2010). The spatial and temporal variability of rainfall within a catchment mean that point and areal rainfall of a given duration and with the same probability of occurrence will differ by a factor, with the areal rainfall always being less than a point value. This factor is known as the areal reduction factor (ARF). ARFs are typically displayed as curves which describe the relationship between catchment area and the factor for different storm durations (e.g. Figure 3.3). The factors are largest for short durations and represent statistical averages for fixed areas based on precipitation events with a 2- to 5-year return period. The factors illustrated in Figure 3.3 were originally developed in the UK (NERC, 1975; Keers and Wescott, 1977), but were found to be in good agreement with results from USA (Bell, 1976) and from Norway based on catchments with areas of < 1000 km^2 (Førland, 1987). The main reasons for applying the UK factors directly, rather than developing a similar set for Norway, are that 1) precipitation regimes in Norway are similar to those in the UK; and 2) the UK has a much denser network of precipitation stations with long observation records. In addition, the much more varied topography in Norway makes it difficult to undertake a similar analysis of spatial variations in precipitation regimes (Sælthun and Anderson, 1986).

For large catchments, one of the greatest uncertainties in flood estimation is connected to the application of ARFs, due to temporal variations and to the availability of several different methods for defining the factors which can lead to ARFs with different properties. There are two categories of methods for deriving ARFs: empirical and analytical, which are discussed in detail by Svenssen and Jones (2010). The ARFs used in Norway were derived using an empirical approach which is data intensive and, generally, does not rely on assumptions about the rainfall process. Svensson and Jones conclude that there are advantages in using this approach, including probabilistically correct ARF estimates and applicability over a comprehensive range of spatial and temporal scales. However, there are two empirical approaches which can be used: geographically fixed and storm centred approaches. In the storm centred approach, the region over which the areal rainfall is estimated is not fixed, but changes for each storm. In the UK and for Norway, the geographically fixed approach was used, but Sælthun and Anderson (1986) suggest that storm centre curves are more appropriate for PMF calculations. The curves derived using the storm centred approach tend to have more sharply decreasing ARFs than the geographically fixed curves currently used (Sælthun and Anderson, 1986). It is possible that the ARFs do not decrease sufficiently with catchment size, and there is a need to review the ARFs for Norway based on observational data. It is possible that different curves may be more suitable for different areas of Norway. The general applicability of the ARFs to all return periods should also be established. Witter (1983) found that the influence of seasonal variation and the return period considered contributed to notable differences in the ARFs for a 24-hour storm. Skaugen (1997) found that the difference between the curves of ARFs for frontal and convective rainfall becomes more pronounced for higher return periods. Svensson and Jones (2010) also note that ARFs tend to decrease more sharply for shorter than for longer duration rainfall.

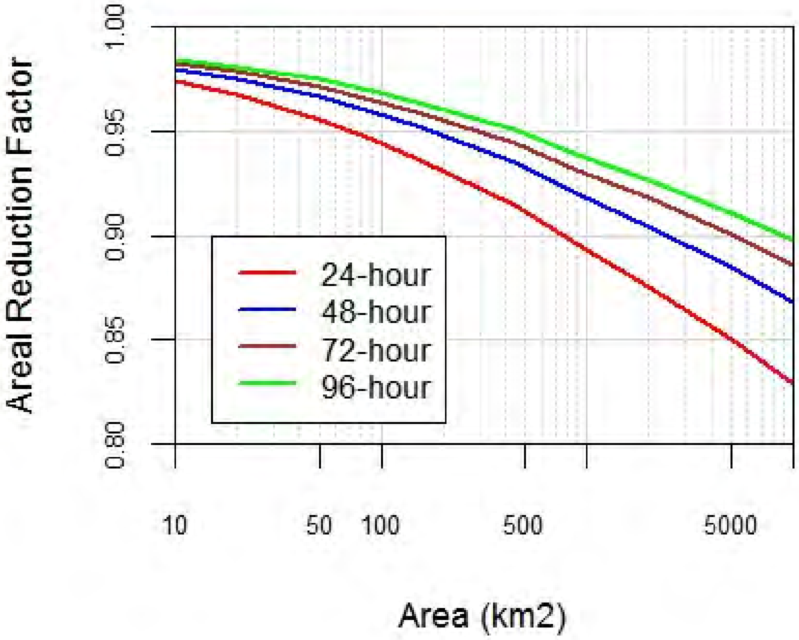


Figure 3.3 Areal reduction factor as a function of catchment size for four storm durations (based on equations given in Keers and Wescott, 1977)

3.5 PMP

The results of a PQRUT simulation of the probable maximum flood (PMF) are dependent on the probable maximum precipitation (PMP) input data and a possible snowmelt contribution. A worst possible scenario is assumed by combining extreme conditions of PMP and snowmelt to give a maximum possible inflow to a reservoir. PMF estimates are required for the assessment of dam safety with respect to dam failure for the highest class dams (Classes 3 and 4) in Norway, and can also be used for Class 1 and 2 dams. In Norway, PMP is estimated by statistical analysis of observed precipitation data (as described in Sections 3.2-3.4), and PMP is calculated by the Meteorological Institute for different seasons and durations. The methods applied by the Meteorological Institute for estimating PMP and its areal reduction factor are to be reassessed and updated in planned work and collaboration with NVE in the near future, as part of a joint PhD research project. This work is relevant both for developing PMF estimates under today's climate and for assessing climate change impacts on dam safety in the future.

3.6 Snowmelt

In many regions within Norway, the critical extreme flood events are generated by a combination of extreme precipitation and simultaneous snowmelt. PQRUT has a simple routine for estimating the snowmelt contribution to runoff, similar to that which has been developed for and implemented in HBV (*e.g.* Sælthun, 1996). As PQRUT is an event-based model (rather than a more long-term seasonal or an interannual model such as HBV), the initial snow cover in PQRUT must be prescribed as an initial condition, given as a snow water equivalent in each of ten equal area elevation zones. The snowmelt is then calculated based on an input air temperature series and a degree day factor, according to:

$$S = Cs * T_L \qquad (3.4)$$

where S is the snowmelt in mm/day, Cs is the degree day factor (see Table 3.1 for recommended values based on the dominant land use) and T_L is the air temperature.

Table 3.1 Recommended degree day factors (Cs) in mm/ °C/ 24 hours (Midttømme *et al.*, 2011).

	Dense forest	Open forest	Bedrock above treeline	Glacier
Without rainfall	1.5	2.0	2.5	3.5
With rainfall	3.0	4.0	5.0	7.0

The values for *Cs* detailed in Table 3.1 were derived in the mid-1980s. Over the past 25 years more data has been collected, offering a good basis for re-evaluation of these values. Degree-day factors have also been found to vary with Julian Day (Dewalle *et al.*, 2002), so it may also be beneficial to investigate whether seasonal variations are important.

To use the snowmelt module in PQRUT, one must also have input temperature values for each time step, and these are generally not available nor would they necessarily be representative of a maximum possible snowmelt event. Therefore, for estimation of PMF it is recommended (Midttømme *et al*., 2011) that a representative temperature (during a rainfall event) is determined from nearby climate stations and adjusted for catchment elevation based on a standard lapse rate. In practice, the maximum temperature observed is often used such that a conservative estimate which can be readily justified on safety grounds is generated. The maximum snow depth and cover must be set for the season of interest based on available sources to determine if snowmelt will continue throughout the storm duration. The snowmelt contribution is often then included in a simulation by adding a fixed amount to the rainfall at each time step (*e.g.* hourly), representing a maximum possible snowmelt contribution (given the assessed conditions) throughout the duration of the storm. This can lead to an accumulated volume contribution from snowmelt (over the course of a 24-hour period, for example) which may be much larger than expected in comparison with observed events.

Snowmelt is an important consideration in deriving flood estimates, but the quantity to add is critically linked to the issue of joint probability. A flood of a particular return period might result from a single flood process or a combination of processes. There is no direct correspondence between a return period event and any particular storm, snowmelt event, ice jam, storm surge or other mechanism. A particular return period event may be produced by a number of mechanisms. For example, when considering a 1000-year rainfall event it is difficult to combine this with an appropriate snowmelt event in order to generate a 1000-year runoff event. Instead of combining an extreme rainfall event with an extreme snowmelt event it may be more appropriate to consider an average snowmelt event or a rainfall event with a lower return period than is required for the final runoff estimate. Nevertheless, this issue can only be addressed with further research into the combined incidences of rainfall and snowmelt events.

The return periods associated with combined snowmelt and rainfall events was the subject of early research by Harr (1981) for two experimental catchments in western Oregon, north-western USA. Climatological and historical records were used to distinguish peak flows resulting from combined rainfall and snowmelt *vs.* those resulting from rainfall alone. For one of the catchments, partial duration series were used to determine the return period for these two types of events. Harr found that for one particular catchment a peak flow of 10 l/s per ha (= 1000 l/(s*km^2)) was 5 times more likely to result from rain-on-snow than rain alone. The peak caused by rain alone was found to have a return period of 15 years, whereas the peaks caused by rain-on-snow were found to have a return period of only 3 years. Such studies could be undertaken for Norwegian catchments using a combination of observed and simulated data. This would help to establish a baseline for guidance regarding appropriate combinations of rainfall and snowmelt events for simulating floods of a given return period. For events with very long return periods, long-term continuous simulations using calibrated hydrological models could also be used to investigate the magnitude of low-frequency events resulting from combined processes. This practice is already applied in methods for design flood analysis used in other countries, and a comparison of those methods with a standard application of PQRUT would provide useful information as to how closely flood magnitude estimates generated with PQRUT approximate, for example, the 1000-year flood estimated from a frequency analysis of a long-term simulated discharge series.

3.7 Soil moisture deficit

The inputs required to simulate a flood event with PQRUT are rainfall and snowmelt values for each time step and appropriate antecedent conditions (i.e. soil moisture and initial discharge). The issue of the joint probabilities of precipitation and snowmelt were discussed in Section 3.6. The initial soil moisture conditions are an added consideration which also affects the resultant probability of a simulated flood event. When simulating a flood event, a catchment is normally assumed to be saturated, thus providing a conservative estimate for these events. NVE (Midttømme *et al.*, 2011) advise that in large catchments especially in eastern Norway, and in Trøndelag and Finnmark, the simulated flood events can be too large if the whole catchment is assumed to be saturated. In some studies the initial soil moisture conditions have been set to 80% to better represent likely initial soil moisture conditions. This approach helps generate more realistic flood estimates, but is rather subjective. It is left to the analyst's discretion and often lacks a justifiable basis for the use of such values. As a result, many analysts take a conservative approach and assume that the catchment is saturated, which may lead to an overestimate of flood magnitudes for a given return period. It is possible to obtain modelled estimates of soil moisture deficit for 115 catchments across Norway, as soil moisture is simulated by NVEs gridded HBV water balance model (Beldring *et al.*, 2003). The model is run with a daily time step, using precipitation and air temperature data as input, and among other components simulates soil moisture, as is also done in the standard HBV model. The principal difference with the gridded version of the model is that the model results are available at the scale of a 1 by 1 km grid for the whole of Norway. However, due to the paucity of observations of soil moisture, these results are generally not as robust as, for example, runoff modelled by HBV, which is calibrated relative to observed discharge. However, HBV results or the results of other soil moisture models such as COUP, could be used in conjunction with observations of soil moisture to establish likely maximum values for antecedent soil moisture on a seasonal basis and are particularly relevant for simulations which consider events occurring during the summer period.

3.8 Model performance

The performance of rainfall-runoff modelling is known to vary from catchment to catchment, with catchment size being a significant factor (*e.g.* Lawrence *et. al.*, 2009). For smaller catchments, rainfall-runoff modelling is generally preferred over flood frequency analysis for cases where observed discharge data are unavailable for flood frequency analysis. However, in larger catchments, rainfall-runoff modelling using PQRUT often produces higher flood magnitudes in comparison with flood frequency analysis. The PQRUT model is recommended for use in catchments from 1 – 800 km^2. In practice both flood frequency analysis and rainfall-runoff methods are often used and compared to derive flood estimates, with results from one method adopted. Event-based rainfall-runoff methods for design flood analyses are also more problematic in catchments where snowmelt is a significant contributor to flooding, due to difficulties with assigning the correct probabilities to the joint incidence of rainfall and snowmelt. In catchments with a minor snowmelt contribution, the rainfall-runoff approach is believed to perform better since there, in principle, is a better correspondence between a particular return period rainfall event and the same return period flood event. As a consequence of spatial variations in the importance of snowmelt there are regional differences in the perceived performance of PQRUT. Generally, for catchments along the west coast of Norway there is greater confidence in the model performance because results tend to be more similar to those derived using flood frequency analysis. For catchments in the south-east of Norway, PQRUT is thought to overestimate flood magnitudes.

Given these perceived variations in the performance of rainfall-runoff modelling for flood estimation, it would be beneficial to further evaluate the performance of PQRUT, particularly in light of other methods that have become available in recent years. Rainfall-runoff modelling may be more appropriate than flood frequency analysis for certain catchments (where, for example, representative data are not available) and for PMF modelling. PQRUT is a simple model which has remained largely unmodified since it was developed in the early 1980s. During this time, a range of alternative models have been developed. In particular, it would be useful to calibrate the model against observed events at high temporal resolution where these data are available. It is technically possible to calibrate PQRUT, but it is very rare that such calibration has been satisfactorily achieved. Continuous and semi-continuous simulation modelling is a further alternative, particularly for larger catchments, whereby long rainfall series are applied to a rainfall-runoff model to generate a long-term runoff series, which may include rarer, larger floods resulting from a combination of processes not observed in the period of record. This more comprehensive approach is still largely confined to research, although it is now about to be adopted in other countries. The continuous simulation approach suffers from several limitations including the suitability of continuous input data derived from weather simulators, the lengthy model runs, and accordingly, the higher costs of undertaking such an analysis. Other, hybrid, semi-continuous methods, based, for example, on rainfall generators linked to regional weather types have also been introduced (e.g. in France; Pacquet *et al.*, 2006) and offer an alternative to both long-term continuous simulation and to event-based modelling. These methods are deserving of further investigation with respect to advantages they may potentially offer for rainfall/snowmelt – runoff modelling of low frequency events. NVE's current participation in COST Action 0901 (European procedures for flood frequency estimation) includes a comparison of PQRUT with methods used in other countries in which combined rainfall and snowmelt are considered.

4 Final flood estimates

Flood frequency analysis and rainfall-runoff modelling are two different methods for deriving flood frequency estimates up to Q_{1000}. Where possible, efforts should be made to reconcile any differences between the flood frequency estimates obtained by the two methods as a test of the suitability of the data and the models used. In practise, the two approaches provide consistent results in some cases, and when this is not the case, possible reasons for the discrepancies are typically discussed. However, this discussion is often based on preconceived ideas as to the performance of the two approaches for either the return period of interest or for a particular part of the country. In particular, the statistical method is believed to produce more reliable results in large catchments (>1000 km^2) and for return periods of up to 1000 years, whereas the rainfall-runoff method is believed to out-perform flood frequency analysis in small catchments and for calculation of the PMF. The PMF is almost always calculated using rainfall-runoff modelling, reflecting the recommended practice. In some circumstances, however, the statistical method has also been used to derive the probable maximum flood (by scaling the 1000-year flood) in large catchments where the rainfall-runoff model used in Norway

(PQRUT) is believed to perform less well. However, a disadvantage of the flood frequency analysis method relative to the rainfall-runoff method is that only peak flows rather than the complete hydrograph are derived. Nevertheless, for applications requiring the full hydrograph, it is feasible to derive a hydrograph shape (Section 3.11) and to scale this to the peak derived using flood frequency analysis. NVE (Midttømme *et al.*, 2011) advises that where there is a large discrepancy in Q_{1000} derived using the two methods, the results derived using rainfall-runoff modelling can be assumed to be of poor quality.

4.1 Uncertainty

It is important that sources of uncertainty in a flood estimate can be identified and that this uncertainty can be quantified and minimised. A reduction in uncertainty can avoid both dangerous under-design and expensive over-design of structures such as dams, embankments, control structures, bridges, culverts and flood protection works (Wiltshire, 1987).

NVE (Midttømme *et al.*, 2011) recommends that sensitivity analyses are undertaken where the underlying data for flood estimation are uncertain. For dam safety assessments the final uncertainty in the flood frequency estimation is addressed by a subjective grading evaluating the quality of the data used and differences between the results based on statistical flood frequency analysis *vs*. rainfall-runoff modelling. The use of long-term rainfall data, which is measured independently of water levels and may span a longer period, could also be analysed alongside flood frequency analyses to give a further indication of result accuracy.

5 Conclusions

A review of the methods used for flood estimation in Norway, including the current NVE guidelines (Midttømme *et al.*, 2011), has been presented in this report. Various aspects of the methods have been considered, including data quality and the fulfilment of underlying assumptions, flood frequency analysis, rainfall-runoff modelling and the determination of final flood estimates. The guidelines and practices are found to be reasonable relative to feasible alternatives. However, as new procedures are developed, a continuous review of the methodologies in relation to new developments presents an opportunity to improve and update the methodologies.

The guidelines provide recommendations with respect to flood estimation related to dam safety (Midttømme *et al.*, 2011), but are not prescriptive. This means that there is a large degree of subjectivity in developing an individual flood estimation assessment, and it is highly likely that different analysts will obtain different results. Efforts to increase the consistency between estimates are desirable, but the use of more prescriptive guidelines would not necessarily increase the reliability of flood frequency assessments. Experienced analysts should be given freedom to make appropriate choices in developing flood estimates. However, there are several areas in which it may be useful to focus attention for the future development of flood estimation, and these are given below according to the sections in which further details can be found in this report.

Flood frequency analysis

- Requirements for data inspection prior to use in analyses and the provision of more detailed data quality information on Hydra II (Section 2.2);
- Development of new procedures for the analysis of non-stationary series (Section 2.3.3).
- Consistency between the available distribution functions, parameter estimation methods and the plotting position formulas used in the Ekstrem *vs.* Dagut/Finut software (Section 2.4);
- Possible recommendation of the method of probability weighted moments (PWM) as the preferred parameter estimation method for all distributions (Section 2.4.2);
- Review of the current regions and methods for regional flood frequency analysis, including
  - research of alternative grouping approaches (such as Region of Influence) to establish whether it is possible to increase the robustness of flood frequency estimates in Norway (Section 2.5.1);
  - guidance on the selection of representative stations for estimation of index flood and growth curves with reference to other possible relevant factors (*e.g.* elevation and height distribution, normal runoff, effective lake percentage) in addition to or possibly *in lieu* of station proximity (Sections 2.5);
  - reassessment of the regression formulas used to estimate the index flood, and an assessment of potential benefits from including different/additional catchment characteristics now available, *e.g.* from GIS analyses; Consideration of the use of multiple regression equations also when estimating the index flood based on a similar gauged catchment, rather than scaling based on catchment area (Section 2.5.2);
- Review of the instantaneous flood peak data held in the Finut database and the establishment of a version of the dataset suitable for use in flood frequency assessments (Section 2.6);

Rainfall-runoff modelling

- Review of the choice of representative storm profiles for different regions and possibly seasons in Norway (Section 3.3);
- Re-evaluation of the methods used to simulate combined rainfall/snowmelt events in conjunction with estimation of Q_T and PMF, including
  - assessment of the degree day method and factors currently used to estimate snowmelt; research into suitable combinations of snowmelt and rainfall for simulating discharge of a given return period, including PMF (Section 3.6);
- Specification of typical soil moisture values prior to large flood events for catchments in different regions and for different seasons (Section 3.7);

- Evaluation of the estimates obtained with the PQRUT rainfall-runoff model in comparison with HBV, where feasible, and with newer approaches such as long-term continuous and semi-continuous simulation modelling (Section 3.8);

Final flood estimates

- Reconciliation of the results obtained using flood frequency analysis and rainfall-runoff modelling (Section 4).

NVE are already involved in projects which are aiming to improve various components of flood frequency estimation, including a review of the equations used to derive instantaneous flood peaks using high resolution flow data (Section 2.6), and analyses of possible changes in flood magnitudes under a future climate based on the application of hydrological modelling with flood frequency analyses (Section 2.3.3). In addition the Meteorological Institute have recently initiated a PhD project focused on extreme precipitation and methods for estimating this, and NVE are involved in this work as joint supervisors. NVE are also active participants in the European Science Foundation COST Action ES0901 (Flood Freq), which is comparing and evaluating flood estimation procedures (both flood frequency analyses and rainfall-runoff methods) across Europe.

Acknowledgements

Thomas Væringstad at NVE and Leif Basberg, Arne Carlsen, Dan Lundquist, Einar Markhus and Daniel Fossberg at Norconsult are thanked for their valuable input based on many years of experience in application of the flood estimation guidelines.

References

Andersen, J.H., Sælthun, N.R., Hjukse, T. and Roald, L. (1983) Hydrologisk modell for flomberegninger, NVE Rapport nr. 2/83.

Beldring, S., Engeland, K., Roald, L.A., Sælthun, N.R., Voksø, A (2003) Estimation of parameters in a distributed precipitation-runoff model for Norway, Hydrology and Earth System Sciences, 7, 304-316.

Bell, F.C. (1976) The Areal Reduction Factor in Rainfall Frequency Estimation. Institute of Hydrology, Report 35, Wallingford.

Bergström, S. (1976) Development and application of a conceptual runoff model for Scandinavian catchments. SMHI Report RH07.

Brath, A., Castellarin, A., Franchini, M., Galeati, G. (2001) Estimating the index flood using indirect methods, Hydrological Sciences Journal, 46 (3), 399-418.

Burn, D.H., and Goel, N.K. (2000) The formation of groups for regional flood frequency analysis, Hydrological Sciences Journal, 45 (1), 97-112.

Coles, S. (2001) An introduction to statistical modelling of extreme values, Springer, London.

Castellarin, A., Kohnová, S., Gaál, L., Fleig, A.K., Salinas, J.L., Toumazis, A., Kjeldsen, T.R. and Macdonald, N. (Eds.) (2011) Review of applied statistical methods for flood frequency analysis in Europe. COST Action ES0901 “European procedures for flood frequency estimation - FloodFreq”, Report Working group 2.

Cunderlink, J.M., Ouarda, T.B.M.J., Bobée, B. (2004) Determination of flood seasonality from hydrological records, Hydrological Sciences Journal, 49 (3), 511-526.

Cunderlink, J.M., Burn, D.H. (2001) The use of flood regime information in regional flood frequency analysis, Hydrological Sciences Journal, 47 (1), 77-92.

Cunnane, C. (1985) Factors affecting the choice of distribution for flood series, Hydrological Sciences Journal, 30 (1), 25-36.

Cunnane, C. (1989) Statistical distributions for flood frequency analysis, WMO Operational hydrology, Report No. 33, WMO-No. 718, Geneva, Switzerland.

Dewalle, D.R., Henederson, Z., Rango, A. (2002) Spatial and temporal variations in snowmelt degree-day factors computed from SNOTEL data in the upper Rio Grande basin, Proceedings of the Western Snow Conference, 73-81.

Engeland, K., Hisdal, H., and Frigessi, A. (2004) Practical extreme value modeling of hydrological floods and droughts: a case study, Extremes, 7, 5–30.

Faulkner, D. (1999) Flood Estimation Handbook 2: Rainfall frequency estimation, Institute of Hydrology, Wallingford.

FEMA (2007) Guidelines and specifications for flood hazard mapping partners, FEMA, US.

Førland, E.J. (1987) Beregning av ekstrem nedbør, DNMI rapport nr. 23/87, pg. 72.

Førland, E.J. (1992) Manual for beregning av påregnelige ekstreme nedbørverdier, DNMI rapport nr. 21/92.

Førland, E.J. and Kristoffersen, D. (1989) Estimation of extreme precipitation in Norway, Nordic Hydrology 20, 257-276.

Harr, R.D. (1981) Some characteristics and consequences of snowmelt during rainfall in western Oregon, Journal of Hydrology, 53, 277-304.

Hosking, J.R.M., Wallis, J.R. (1997) Regional frequency analysis: an approach based on L-moments, Cambridge University Press.

Hosking, J.R.M., Wallis, J.R, Wood, E.F. (1985) An appraisal of the regional flood frequency procedure in the UK flood studies report, Hydrological Sciences Journal, 30 (1), 85-109.

Keers, J.F., Wescott, P. (1977). A computer-based model for design rainfall in the United Kingdom. Met Office Scientific Paper No. 36, HMSO, London, UK.

Killingtveit, Å., Sælthun, N.R. (1995). *Hydrology.* (Volume No. 7 in *Hydropower Development*). Norwegian Institute of Technology, Trondheim.

Kjeldsen, T.R. (2010) Modelling the impact of urbanisation on flood frequency relationships in the UK, Hydrology Research, 41 (5), 391-405.

Kjeldsen, T.R. (2011) UK – United Kingdom. In: Castellarin, A., Kohnová, S., Gaál, L., Fleig, A., Salinas, J.L., Toumazis, A., Kjeldsen, T.R. and Macdonald, N. (Eds.) (2011) Review of applied statistical methods for flood frequency analysis in Europe. COST Action ES0901 "European procedures for flood frequency estimation - FloodFreq", Report Working group 2.

Kjeldsen, T.R., Jones, D.A (2010) Predicting the index flood in ungauged UK catchments: On the link between data-transfer and spatial model error structure, Journal of Hydrology, 387, 1-9.

Lawrence, D., Haddeland, I. and Langsholt, E. (2009) Calibration of HBV hydrological models using PEST parameter estimation. NVE Report No. 1-2009.

Lawrence, D (2010) Hydrological projections for changes in flood frequency under a future climate in Norway and their uncertainties. In: Hydrology: From research to water management, NHP Report No. 51, 203-204.

Lawrence, D. and Graham, P. (2010) Climate change projections and flood risk management. *Proceedings of the SAWA Mid-term conference Gothenburg,Sweden, 25-27 May 2010.* Report Nr. 5/2010, Berichte des Landesbetriebes Straßen, Brücken und Gewässer, Hamburg; pp. 18-19.

Lawrence, D., Graham, L.P. and den Besten, J. (2011) Estimating and communicating climate change impacts and uncertainties in flood risk management: Examples from the North Sea Region. Report to the EU Interreg IVB SAWA project. (NVE Report - In press).

Lawrence, D. and Hisdal, H. (2011) Hydrological projections for floods in Norway under a future climate. NVE Report no. 2011/5, 47 pp.

Lettenmaier, D.P., Wallis, J.R., Wood, E.F. (1987) Effect of regional heterogeneity on flood frequency estimation, Water Resources Research, 23, 313-323.

Madsen, H., Pearson, C.P., Rasmussen, P.F. and Rosbjerg, D. (1997) "Comparison of annual maximum series and partial duration series methods for modelling extreme hydrological events 1. At-site modelling," Water Resour. Res. 33(4), 747Y758.

Martins, E.S. and Stedinger, J.R. (2001) "Generalized maximum likelihood Pareto-Poisson flood risk analysis for partial duration series." Water Resour. Res. 37(10), 2551Y2557.

Merz, R., Blöschl, G. (2005) Flood frequency regionalisation—spatial proximity vs. catchment attributes. Journal of Hydrology, 302, 283–306.

Midttømme, G., Pettersson, L.E., Holmqvist, E., Nøtsund, Ø., Hisdal, H. and Sivertsgård, R. (2011) Retningslinjer for flomberegninger (Guidelines for flood calculations). Retningslinjer no. 04/2011. 59 pp. Available at www.nve.no

NERC (1975) Flood Studies Report. Natural Environment Research Council, London.

Norconsult (2009) Nordmarkavassdraget, Dam Åklungen, Frognerbekken. (prepared for Oslo Kommune Friluftsetaten)

Paquet, E, Gailhard, J., Garçon, R. (2006) Evolution of GRADEX method: improvement by atmospheric circulation classification and hydrological modelling. La Houille Blanche 5:80-90.

Pettersson, L.E. (2000) Flomberegning for Gaulavassdraget (122.Z), NVE Rapport nr. 15/2000.

Pettersson, L.E. (2008) Om indeksflom og flomfrekvenskurver, Oppdragsrapport B, nr. 4/2008.

Pettersson, L.E. (2009a) Flomforhold i Sør- og Midt-Norge. NVE Report no. 2009/3, 67 pp.

Pettersson, L.E. (2009b) Flomforhold i Nord-Norge. NVE Report no. 2009/11, 49 pp.

Prudhomme, C., Wilby, R.L., Crooks, S., Kay, A.L., Reynard, N.S. (2010) Scenario-neutral approach to climate change impact studies: Application to flood risk, Journal of Hydrology, 390, 198-209.

Rao, A.R., Hamed, K.H. (2000) Flood frequency analysis, CRC Press.

Reed, D.W., Field, E.K. (1992) Reservoir flood estimation: another look. IH Report No. 114, Institute of Hydrology, Wallingford.

Reed, D.W., Robson, A.J. (1999) Flood Estimation Handbook 3: Statistical procedures for flood frequency estimation, Institute of Hydrology, Wallingford.

Skaugen, T. (1997) Classification of rainfall into small- and large-scale events by statistical pattern recognition, *Journal of Hydrology*, 200 (1-4), 40-57.

Skøien, J.O. and Blöschl, G. (2007): Spatiotemporal topological kriging of runoff time series. *Water Resources Research* 43: W09419.

Skøien, J.O., Merz, R. and Blöschl, G. (2006): Top-kriging - geostatistics on stream networks. *Hydrology and Earth System Sciences* 10(2): 277–287.

Sokolov, A.A., Rants, S.E., and Roche, M. (1976) Floodflow computation, Methods compiled from world experience, The Unesco Press, Paris.

Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, E.: Frequency analysis of extreme events, in: Handbook of Hydrology, edited by: Maidment, D. R., McGraw-Hill, Inc., New York, NY, 18.1–18.66, 1993.

Svensson, C., and Jones, D.A. (2010) Review of methods for deriving areal reduction factors, Journal of Flood Risk Management, 3, 232-245.

Sælthun, N.R. (1996) The Nordic HBV model. NVE Publication No. 07.

Sælthun, N.R. (1997) Regional flomfrekvensanalyse for norske vassdrags. Rapport 14 – 97, NVE

Sælthun, N.R., Andersen, J.H. (1986) New procedures for flood estimation in Norway, Nordic Hydrology, 17, 217-228.

Tallaksen, L. M., Madsen, H., and Hisdal, H. (2004) Frequency Analysis. In: *Hydrological Drought – Processes and Estimation Methods for Streamflow and Groundwater*, edited by: Tallaksen, L. M. and van Lanen, H. A. J., Developments in Water Science, 48, Elsevier Science B.V., Amsterdam, 199–271.

Vassdragsdirektoratet (1986) Retningslinjer for beregning av dimensjonerende flom og påregnelig maksimal flom, V-information 1-86.

Waylen, P., Woo, M-K (1982) Prediction of annual floods generated by mixed processes, Water Resources Research, 18 (4), 1283-1286.

Wilson, D., Hisdal, H., Lawrence, D. (2010) Has streamflow changed in the Nordic countries? – Recent trends and comparisons to hydrological projections. Journal of Hydrology, 394, 334-346.

Wiltshire, S.E. (1987) Statistical techniques for regional flood frequency analysis, Unpublished PhD thesis, Newcastle University.

Wingård, B. (1977) Hvilke(n) flomfrekvensfunksjon er best i Norge?, NVE Rapport nr. 3/77.

Wingård, B. (1978) Regional flomfrekvensanalyse for norske vassdrags, NVE Rapport nr. 2-78.

Witter, J.V. (1983) Statistical areal reduction factors in the Netherlands, IAHS Publication. 147, 25-34.

WMO (2009) Manual on estimation of Probable Maximum Precipitation (PMP), WMO-No. 1045.

This series is published by Norwegian Water Resources and Energy Directorate (NVE)

Published in the Report series 2011

Nr. 1 Representation of catchment hydrology, water balance, runoff and discharge in the JULES and SURFEX land surface models. Tuomo Saloranta (26 s.)

Nr. 2 Mapping of selected markets with Nodal pricing or similar systems Australia, New Zealand and North American power markets. Vivi Mathiesen (Ed.) (44 s.)

Nr. 3 Glaciological investigations in Norway in 2010: Bjarne Kjøllmoen (Ed.) (99 s.)

Nr. 4 Climate change impacts on the flow regimes of rivers in Bhutan and possible consequences for hydropower development. Stein Beldring (Ed.) (153 s.)

Nr. 5 Hydrological projections for floods in Norway under a future climate. Deborah Lawrenc, Hege Hisdal (47 s.)

Nr. 6 Project report for the project Hydrological Flood Forecasting System for Small and Medium Sized Catchments in Serbia, 2009 – 2010. Documentation and technical references. Elin Langsholt (57s.)

Nr. 7 Preliminary Flood Risk Assessment in Norway. An example of a methodology based on a GIS-approach (39 s.)

Nr. 8 GPR-measurements of snow distribution on Hardangervidda mountain plateau in 2008-2011, Galina Ragulina (s.)

Nr. 9 A review of NVE's flood frequency estimation procedures. Donna Wilson, Anne K. Fleig, Deborah Lawrence, Hege Hisdal, Lars-Evan Pettersson, Erik Holmqvist (46 s.)

Norwegian
Water Resources and
Energy Directorate
N V E

Norwegian Water Resources
and Energy Directorate

Middelthunsgate 29
PB. 5091 Majorstuen,
N-0301 Oslo Norway

Telephone: +47 22 95 95 95
Internet: www.nve.no