John Bjørnar Bremnes
Det norske meteorologiske institutt

Statistical Forecasting of Precipitation Conditional on Numerical Weather Prediction Models
HYDRA - et forskningsprogram om flom

HYDRA er et forskningsprogram om flom initiert av Norges vassdrags- og energiverk (NVE) i 1995. Programmet har en tidsramme på 3 år, med avslutning medio 1999, og en kostnadssramme på ca. 18 mill. kroner. HYDRA er i hovedsak finansiert av Olje- og energidepartementet.

Arbeidshypotesen til HYDRA er at summen av alle menneskelige påvirkninger i form av arealbruk, reguleringer, forbygningsarbeider m.m. kan ha økt risikoen for flom.

Målgruppen for HYDRA er statlige og kommunale myndigheter, forsikringsbransjen, utdannings- og forskningsinstitusjoner og andre institusjoner. Nedenfor gis en oversikt over fagfelt/tema som blir berørt i HYDRA:

- Naturgrunnlag og arealbruk
- Tettsteder
- Flomdemping, flomvern og flomhandtering
- Skaderisikoanalyse
- Miljøvirkninger av flom og flomforebyggende tiltak
- Databaser og GIS
- Modellutvikling

Sentrals aktører i HYDRA er: Det norske meteorologiske institutt (DNMI), Glommens og Laagens Brukseierforening (GLB), Jordforsk, Norges geologiske undersøkelse (NGU), Norges Landbrukshøgskole (NLH), Norges teknisk-naturvitenskapelige universitet (NTNU), Norges vassdrags- og energiverk (NVE), Norsk institutt for jord- og skogkartlegging (NIJOS), Norsk institutt for vannforskning (NIVA), SINTEF, Stiftelsen for Naturforskning og Kulturminneforskning (NINA/NIKU), Norsk Regnesentral (NR), Direktoratet for naturforvaltning (DN), Østlandsforskning (ØF) og universitetsene i Oslo og Bergen.

HYDRA - a research programme on floods

HYDRA is a research programme on floods initiated by the Norwegian Water Resources and Energy Administration (NVE) in 1995. The programme has a time frame of 3 years, terminating in 1999, and with an economic framework of NOK 18 million. HYDRA is largely financed by the Ministry of Petroleum and Energy.

The working hypothes is for HYDRA is that the sum of all human impacts in the form of land use, regulation, flood protection etc., can have increased the risk of floods.

HYDRA is aimed at state and municipal authorities, insurance companies, educational and research institutions, and other organization.

An overview of the scientific content in HYDRA is:

- Natural resources and land use
- Urban areas
- Databases and GIS
- Risk analysis
- Flood reduction, flood protection and flood management
- Environmental consequences of floods
- and flood prevention measures
- Modelling

Central institutions in the HYDRA programme are: The Norwegian Meteorological Institute (DNMI), The Glommens and Laagens Water Management Association (GLB), Centre of Soil and Environmental Research (Jordforsk), The Norwegian Geological Survey (NGU), The Agriculture University of Norway (NLH), The Norwegian University of Science and Technology (NTNU), The Norwegian Water and Energy Administration (NVE), The Norwegian Institute of Land Inventory (NIJOS), The Norwegian Institute for Water Research (NIVA), The Foundation for Scientific and Industrial Research at the Norwegian Institute of Technology (SINTEF), The Norwegian Institute for Nature and Cultural Heritage Research (NINA/NIKU), Norwegian Computing Center (NR), Directorate for Nature Management (DN), Eastern Norway Research Institute (ØF) and the Universities of Oslo and Bergen.
1 INTRODUCTION

Quantifying the uncertainty associated with flood forecasts is important for risk assessments and in making decisions. Due to the large amount of information these forecasts are based on, this is a complicated issue. In Lundquist (1997) the following sources of error are briefly discussed: meteorological forecasts, rainfall-runoff model, initial conditions, transport time, temporary loss of water and discharge rating curves. This report is restricted to the uncertainty of precipitation forecasts.

Today the most accurate meteorological forecasts, at ranges up to 10 days, are produced by numerical weather prediction models (NWPs). These models are based on discretizations of physical laws where the resulting equations are solved by supercomputers. In this deterministic approach, there are mainly two reasons for imperfect forecasts. First, the nature of the equations describing the evolution of the atmosphere is unstable. Along with the uncertainty of the state of the atmosphere at any point in time, this limit the predictability. Second, there are physical processes too small in scale and sometimes too complex to be adequately included in the NWPs. Simplifications (parameterizations) of these processes therefore must be used.

The quality of numerical forecasts will with increasing forecast range gradually depend relatively more on the accuracy of the initial conditions than on limitations in the NWPs. To investigate the consequences of uncertain initial conditions, a set of small perturbations of the most likely initial condition is computed. For each perturbation a NWP is used, at a lower resolution than normal, to generate a forecast. The resulting ensemble of forecasts then constitute an empirical probability distribution which can be used to make probability forecasts.

In this paper we adopt a statistical approach and construct models which explain the observations by means of output from NWPs. In contrast to deterministic forecasts where only one value is available for each forecast, statistical forecasts consist of probability distributions. From these distributions we may, for example, compute prediction/forecasting intervals, probabilities of interesting events or the precipitation amount with highest probability.
From a physical point of view the uncertainty of the forecasts is clearly dependent on the weather situations and will therefore vary from day to day. Detection of such situations in order to quantify the degree of uncertainty requires detailed output from the NWPs and preferably knowledge of the weaknesses of the models. For simplicity, we will use three NWPs and let the degree of difference in these forecasts represent the uncertainty. The statistical models will, however, be quite general, so inclusion of weather regime information also will be possible.

The fact that daily precipitation data have a point mass at zero makes statistical modelling of precipitation far more complicated than for other meteorological parameters. For this reason the most common approach is to split the problem in two stages. First the occurrence of precipitation is modeled and then the amounts on wet days. The drawback with this approach is that all direct statements about the precipitation amounts are conditional on occurrence of precipitation. It is, at least to the author, unknown how to combine the two models to make an unconditional statistical prediction/forecast.

In this paper we will also try another approach which directly provides e.g. prediction intervals. This model avoids the discrete–continuous mix by assuming that the amounts on dry days are not observed, that is we only know that they are less than a lower threshold for precipitation. We then have continuous data and a quite simple model may be used.

In lack of areal observations and somewhat incomplete output from NWPs, the statistical models are applied to a data set at a gauge station (Flisa). Although we primarily are interested in areal precipitation, this example demonstrates the models ability to make probabilistic forecasts.

The outline of the paper is as follows. Section 2 describes the data in more detail. In section 3 the two statistical methods are presented in a general way. We recommend readers with insufficient statistical background to concentrate on the models and skip technicalities like estimation and prediction. Section 4 contains the results when the models are applied to the data at Flisa. In sections 5 and 6 the methods are discussed and conclusions drawn.

2 DATA AND EXPLORATORY ANALYSIS

The data used in this study are daily observed precipitation from a gauge at Flisa (in the southeast of Norway) and corresponding +30 hours forecasts from three numerical weather prediction models; HIRLAM 0.1°, HIRLAM 0.5° and ECMWF. The period covered is November 7th 1995 to June 14th 1998 excluding days with missing values. In all there are 836 cases.

In figure 1 the observations are plotted against the three NWPs, and it is possible to see that the errors increase with the precipitation amounts. By fixing the forecasts, we can also get an idea of the underlying distribution of the observations given the forecasts.

Although there is a clear dependency between errors and amounts, we might expect that not all forecasts with large precipitation amounts are equally uncertain. Figure 2 shows the estimated joint probability distribution of the errors and the difference between the maximum and the minimum of the three NWPs. (Note that the cases with no observed precipitation are left out for illustrative purposes.) We see that the distribution of the errors given the differences be-
comes wider/flatter with increasing differences. In other words, the forecasts are most uncertain when the differences between the NWPs are large.

3 STATISTICAL MODELS

In this section we describe the two approaches which later will be referred to as alternative I and alternative II. Alternative I comprises the two models in section 3.1, while alternative II is the model in section 3.2. The generic notation \( \pi(\cdot | \cdot) \) will be used throughout to denote conditional probability distributions.

3.1 MODELLING OCCURRENCES AND AMOUNTS SEPARATELY

This approach is well described in Stern and Coe (1984) and more recently in Chandler and Wheater (1998a,b), but neither make use of output from NWPs. In this report we let the limit distinguishing precipitation from no precipitation be 0.1 mm (0.1 mm is treated as no precipitation).

3.1.1 MODEL FOR OCCURRENCES

In this case the observations \( Y_1, \ldots, Y_T \) are either zero (dry) or one (wet), i.e. Bernoulli distributed. Further, let \( p_t \) denote the probability of precipitation at day \( t \) and assume that it depends on a set of explanatory variables \( x_{t1}, \ldots, x_{tp} \); in our case these will be simple functions of output from NWPs. Assuming
Figure 2: Estimated joint probability density functions of the errors and the difference between the maximum and the minimum of the three forecasts. Dark gray scales indicate high probability.

independent observations the model may then be written

\[ Y_t \sim \text{Bernoulli}(p_t) \]

\[ \log\left( \frac{p_t}{1-p_t} \right) = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_{ij} \]

for \( t=1, \ldots, T \). The unknown parameters \( \alpha_0, \ldots, \alpha_p \) could be estimated by maximizing the likelihood function. Probabilities of future observations may then be predicted by inserting a new set of explanatory variables into the fitted model.

Model (1) is an example of generalized linear models and general algorithms for estimation and prediction are provided by the statistical software package S-Plus (Becker et al., 1988; Venables and Ripley, 1997). A thorough introduction to generalized linear models is given in McCullagh and Nelder (1989).

3.1.2 Model for amounts on wet days

Precipitation amounts are usually modeled by the gamma or the log-normal distribution. Here, we apply a normal distribution restricted to the interval \((0,1,\infty)\), because the conditioning on the NWPs makes it appropriate. This distribution is characterized by two parameters \( \mu \) and \( \sigma^2 \), say, which physically can be interpreted as the precipitation amount with highest probability (the mode \( \mu \)) and the uncertainty (the dispersion \( \sigma^2 \)). We assume that the mode
and the dispersion depend on some explanatory variables \( x_1, \ldots, x_p \) and \( v_1, \ldots, v_q \), respectively.

A Bayesian framework is used in order to simplify both estimation and prediction/forecasting of future observations. In contrast to the classical framework where all parameters are fixed (but unknown), the parameters in the Bayesian world are random and have their own distributions. These distributions are called prior distributions and reflect the modeller's belief about the unknown parameters. Along with the distribution for the data given all parameters, they define a Bayesian model. By using Bayes' theorem the distribution of the parameters conditional on the data, called posterior distribution, is derived. All inference (e.g. estimation) are based on this distribution. For more information about Bayesian modelling see Carlin and Louis (1996).

Let \( Y_{\text{obs}} = (Y_1, \ldots, Y_T) \) denote the observations on wet days and, assuming conditional independence throughout, the generic model is

\[
\pi(Y_t | \mu_t, \sigma_t^2) \sim \mathcal{N}(\mu_t, \sigma_t^2) \tag{2}
\]

\[
\mu_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_{tj} + \sum_{j=1}^{q} \beta_j v_{tj}
\]

for \( t = 1, \ldots, T \) and where \( \alpha_0, \ldots, \alpha_p \) and \( \beta_0, \ldots, \beta_q \) have mutually independent (almost) flat priors, i.e. distributions with very large variances.

The posterior distribution, \( \pi(\theta | Y_{\text{obs}}) \) where \( \theta = (\alpha_0, \ldots, \alpha_p, \beta_0, \ldots, \beta_q) \) for convenience, may now in principle be calculated, but its complexity makes inference analytically impossible. The Markov chain Monte Carlo (MCMC) methodology has greatly reduced this problem by providing algorithms to sample from this distribution. There are, however, some minor problems because in practice the algorithms need some iterations until samples from the correct distribution are produced. As there is no general result about how many iterations are needed, we have to look at the samples and maybe make diagnostic plots to assess convergence. For inferential purposes these iterations are discarded.

In this paper we use the Gibbs sampler, which is the most popular MCMC method, for parameter estimation. The main reason is that a quite general software package called BUGS (Spiegelhalter et al., 1996) is available and by using it some technicalities are avoided. An overview of MCMC methods and applications can be found in Gilks et al. (1996).

Here, we are not primarily interested in the posterior distribution but more in what we will observe in the future or, more precisely, the posterior predictive density \( \pi(Y_{\text{new}} | Y_{\text{obs}}) \) where \( Y_{\text{new}} \) denotes the future observation. The posterior predictive density is

\[
\pi(Y_{\text{new}} | Y_{\text{obs}}) = \int \pi(Y_{\text{new}}, \theta | Y_{\text{obs}}) d\theta
\]

\[
= \int \pi(Y_{\text{new}} | \theta) \pi(\theta | Y_{\text{obs}}) d\theta
\]

Note that \( \pi(Y_{\text{new}} | \theta, Y_{\text{obs}}) = \pi(Y_{\text{new}} | \theta) \) since the observations \( Y_{\text{obs}} \) add no further information when the parameters \( \theta \) are known.
Let \( \theta_1^*, \ldots, \theta_m^* \) be samples from \( \pi(\theta \mid Y_{obs}) \) generated by the Gibbs sampler. An estimate of the posterior predictive density is then given by

\[
\hat{\pi}(Y_{new} \mid Y_{obs}) = \frac{1}{m} \sum_{i=1}^{m} \pi(Y_{new} \mid \theta_i^*)
\]  

(Gilks et al., 1996, p. 154). To summarize, the algorithm for prediction/forecasting is

1. Use the Gibbs sampler to obtain samples \( \theta_1^*, \ldots, \theta_m^* \) from \( \pi(\theta \mid Y_{obs}) \)

2. For each new set of explanatory variables
   - For each \( \theta_i^* \)
     - Compute e.g. quantiles of \( \pi(Y_{new} \mid \theta_i^*) \)
   - Use the mean of each quantile as estimate for the corresponding quantile of the predictive distribution

### 3.2 A MODEL FOR BOTH OCCURRENCE AND AMOUNTS

Using two models, separately constructed, complicates the decision process in that we directly only can make conditional statements about the precipitation amounts and equip this with the probability of precipitation. To combine these two into one statement is not necessarily simple and theoretically it is somewhat unclear how this should be done. In potential flood situations, however, the problem can to some extent be reduced since the probability of precipitation often is close to one and the conditioning may therefore be neglected.

In this section we propose a unifying model by assuming there is a nonphysical process slightly different from the real one. The difference is that the dry days are treated as censored data, i.e. we only know that the amounts in these cases are smaller than the threshold (and possibly negative).

The observation \( Y_t \) is related to the nonphysical process \( W_t \) by

\[
Y_t = \begin{cases} 
0 & \text{if } W_t \leq 0.1 \\
W_t & \text{otherwise} 
\end{cases}
\]  

for \( t = 1, \ldots, T \). The Bayesian model for \( W_t \) is

\[
\pi(W_t \mid \mu_t, \sigma_t^2) \sim \mathcal{N}(\mu_t, \sigma_t^2) \quad (5)
\]

\[
\mu_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_{tj}
\]

\[
\log(\sigma_t^2) = \beta_0 + \sum_{j=1}^{q} \beta_j v_{tj}
\]

independently for \( t = 1, \ldots, T \) and where \( \alpha_0, \ldots, \alpha_p \) and \( \beta_0, \ldots, \beta_q \), ideally, are mutually independent flat priors. Often, more than half of the observations are censored so the priors have to be more informative to get reasonable estimates.

The estimation is similar to that in the last subsection, but note that here the normal distribution is not restricted and that all observations are used; that is, the censored observations are unknown and simulated from normal distributions restricted to \((-\infty, 0.1]\) at each iteration of the Gibbs sampler. Approximate quantiles of the predictive distribution for the latent process are easily computed and by using (4) these are transformed to the real world.

6
4 RESULTS

The statistical models described in the previous section were applied to the data at Flisa. To be able to validate the predictions, the data were divided into training and test sets. The data in the period November 7th 1995 to March 31st 1998 (766 cases) were used to estimate the parameters in the models, while the data from April 1st to June 14th 1998 (70 cases) were used to test the models ability to predict/forecast new observations.

4.1 ALTERNATIVE I

The first part deals with the probability of precipitation. Model (1) was fitted with various selections of explanatory variables, all based on the three numerical precipitation forecasts, and evaluated by means of the deviance (McCullagh and Nelder, 1989, p. 118) and the ability to forecast new observations. This resulted in the following \( p = 4 \) explanatory variables:

\[
\begin{align*}
  x_1 &= \text{mean of the three forecasts} \\
  x_2 &= \text{one, if all forecasts are less than 0.1mm, else zero} \\
  x_3 &= \text{one, if one or two of the forecasts are greater than 0.1mm, else zero} \\
  x_4 &= \text{one, if at least two forecasts are less than 1mm, else zero}
\end{align*}
\]

Note that other possibilities exist. The estimates of the parameters and the predicted probabilities of precipitation for the test period are shown in table 1 (left) and in figures 3 and 5, respectively. We also counted the number of errors in predicting precipitation / no precipitation. Alternative I clearly was the best with only 11 errors; HIRLAM 0.1° had 24 errors, HIRLAM 0.5° 23, ECMWF 18, and alternative II 18 errors. The S-Plus code is included in appendix A.1.

Next, the precipitation amounts on wet days were modeled using (2). In all there were 281 days with precipitation and of these 251 days were in the training set. The search for appropriate explanatory variables resulted in two explanatory variables; one for the mode \( \mu \) and one for the dispersion \( \sigma^2 \). These were

\[
\begin{align*}
  x_1 &= \text{mean of the three forecasts} \\
  v_1 &= (\text{maximum forecast} - \text{minimum forecast}) - c
\end{align*}
\]

where \( c \) is a constant which centers the data. For the mode \( \alpha_0 \) was fixed to zero and \( \alpha_1 \) to one. The estimates of \( \beta_0 \) and \( \beta_1 \) are shown in table 1 (right). Figures 3 and 5 show 97.5% percentiles, modes, and 2.5% percentiles of the predictive distributions for amounts given precipitation. The BUGS program for parameter estimation and the S-Plus code for prediction/forecasting is included in appendix A.1.

4.2 ALTERNATIVE II

Model (4) for unconditional precipitation amounts is similar to model (2), and consequently almost the same explanatory variables could be used. For the mode, however, some additional variables had to be included. The \( p+q = 4+1 \) explanatory variables were
### Table 1: Estimates of the parameters in Alternative I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>α₀</td>
<td>-2.62</td>
<td>0.32</td>
</tr>
<tr>
<td>α₁</td>
<td>0.314</td>
<td>0.065</td>
</tr>
<tr>
<td>α₂</td>
<td>-1.52</td>
<td>0.30</td>
</tr>
<tr>
<td>α₃</td>
<td>-0.369</td>
<td>0.14</td>
</tr>
<tr>
<td>α₄</td>
<td>-0.355</td>
<td>0.14</td>
</tr>
</tbody>
</table>

### Table 2: Estimates of the parameters in Alternative II.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>α₂</td>
<td>-3.65</td>
<td>0.77</td>
</tr>
<tr>
<td>α₃</td>
<td>-2.11</td>
<td>0.60</td>
</tr>
<tr>
<td>α₄</td>
<td>-4.98</td>
<td>0.77</td>
</tr>
<tr>
<td>β₀</td>
<td>2.72</td>
<td>0.11</td>
</tr>
<tr>
<td>β₁</td>
<td>0.157</td>
<td>0.029</td>
</tr>
</tbody>
</table>

The disadvantage of using two models to fit daily precipitation data (alternative I) is, as already mentioned, that all direct statements about precipitation amounts are conditional on occurrence of precipitation. By using alternative II the condition is avoided. From an applied point of view, the choice of method is mainly a question of whether conditioning on precipitation causes inconveniences. We will skip this discussion and go into more details of each of the two approaches.

### 5 Discussion

The disadvantage of using two models to fit daily precipitation data (alternative I) is, as already mentioned, that all direct statements about precipitation amounts are conditional on occurrence of precipitation. By using alternative II the condition is avoided. From an applied point of view, the choice of method is mainly a question of whether conditioning on precipitation causes inconveniences. We will skip this discussion and go into more details of each of the two approaches.

#### 5.1 Alternative I

There are no great statistical difficulties in fitting models to binary precipitation data. The challenges are mainly associated with the selection and construction...
Figure 3: Alternative I: Predicted probabilities of precipitation (dashed line), predicted 97.5% percentiles, modes, and 2.5% percentiles for amounts given precipitation (all solid lines), and observations (points).

Figure 4: Alternative II: Predicted probabilities of precipitation (dashed line), predicted 97.5% percentiles, modes, and 2.5% percentiles for amounts (all solid lines), and observations (points).
Figure 5: Alternative I: Predicted probabilities of precipitation (dashed line), predicted 97.5% percentiles, modes, and 2.5% percentiles for amounts given precipitation (all solid lines), and observations (points).

Figure 6: Alternative II: Predicted probabilities of precipitation (dashed line), predicted 97.5% percentiles, modes, and 2.5% percentiles for amounts (all solid lines), and observations (points).
of explanatory variables. Of particularly interest, is whether yesterdays observations and/or NWP result should be included. We have chosen not to, because there was no clear pattern in the data and, besides, it would complicate the forecasting procedure since not all observations are available immediately.

Although the results at Flisa are promising, it is reasonable to expect even better results if more output from the NWPs become available. For instance, weather situations where NWP forecasts are bad, could then possibly be detected and taken better into account.

The first thing to consider when modelling precipitation amounts on wet days, is the choice of probability distribution. Usually, the gamma distribution is preferred, but here we have tried out a restricted normal distribution. The main advantage of this distribution is that the parameters are easily related to explanatory variables. By using the gamma distribution, this is more difficult, but, on the other hand, the shape of the gamma distribution is more flexible. This is especially important when dealing with outliers (cases when, in practice, all NWP forecasts fail). On the basis of this study it is difficult to tell which distribution to prefer. In any case, the fit to the restricted normal distribution should be thoroughly examined. More work is needed to find adequate methods for model checking.

Several combinations of explanatory variables for the mode (μ) were tried out, but fixing the mode to equal the mean of the forecasts and then add adjustments seemed to be the best general strategy. In this case, however, it resulted in no adjustments; mainly because the mode was not of main interest. The selection of explanatory variables for the dispersion (σ²) was easier since not much potential information was available. We have used the difference between the largest and smallest forecast in a log-linear manner. A possible drawback is that the uncertainty may become too large or too small on some occasions. If so, a transformation of the explanatory variable ought to be considered. It would also be interesting to see how a large ensemble of forecasts could be used to construct new explanatory variables.

5.2  ALTERNATIVE II

The fact that more than half of the observations were treated as censored values caused some minor problems. First, the prior distributions of the βs had to be more informative to make the estimation algorithm work. Second, the output from the Gibbs sampler were strongly autocorrelated for some parameters; significant correlations at lags up to about 30 were not unusual. A few reparameterizations of the model were tried out, but did not seem to reduce the strong dependency. The consequences are not serious, but the assessment of convergence becomes more difficult and more iterations have to be carried out.

The selection of explanatory variables followed the same procedure as in the last subsection, except that situations where the probability of precipitation could be low had to be taken special care of by adding indicator variables to the mode/mean (μ).

5.3 AREAL PRECIPITATION

Hydrologic rainfall–runoff models use the total amount of precipitation in a catchment (or its average) as input; that is, the variation within the catchment
is not of main interest. In our context this means that, ideally, observed areal precipitation should be used to fit the models, but, as we know, such data are not directly available. Instead, spatial interpolation of point observations followed by integration over the catchment may be used to get estimates.

In this report we have for simplicity only used observations at a single station (Flisa). Our results are then only valid for this station or the (small) area it represents. The same statistical models could, however, also be applied to estimated areal observations, but preferably the observational errors should explicitly be taken into account. This is usually obtained by introducing a new level in the hierarchical Bayesian models. We will not pursue this issue further here, but only remark that the spatial aspect somehow has to be considered.

6 Concluding Remarks

In this report we have presented two methods for forecasting precipitation in terms of probability distributions. From an applied point of view, these are distinguished by that in one of them the predictive probability distribution for precipitation amounts is conditional on occurrence of precipitation. In practice this may be a drawback, but as this is an area of current research, we may hope for progress in the near future. In any case, the second method produces unconditional predictive distributions.

Both approaches were applied to data from the gauge station at Flisa. Although the validation of the models is somewhat informal and insufficient, the results are promising. It is reasonable to assume that both models could be used at other locations with small adaptations, and that they could handle areal observations. In the latter case, we ought to mention that complicated spatio-temporal modelling also should be considered.

Acknowledgements

Comments from colleagues at DNMI and Thomas Skaugen at NVE were appreciated and hopefully made the report more readable.

References


Chandler, R.E., and Wheater, H.S. (1998b). Climate change detection using generalized linear models for rainfall – a case study from the west of Ireland.
II. Modelling of rainfall amounts on wet days. Technical report, University College London, Department of Statistical Science.


A COMPUTER CODES

A.1 ALTERNATIVE I

PRECIPITATION / NO PRECIPITATION

```r
# S-Plus code.
# read the Flisa data and put it into a data frame
# columns: date, hirlam0.1, hirlam0.5, ecmwf, observation
flisa <- read.table("Flisa.data", header=T)
# remove rows (dates) containing NAs
flisa <- flisa[!is.na(flisa[,1]),]
# create explanatory variables
y <- flisa[,5]>.1
z1 <- flisa[,2]
z2 <- flisa[,3]
z3 <- flisa[,4]
z4 <- apply( cbind( z1>.1, z2>.1, z3>.1 ), 1, sum )
z5 <- apply( cbind( z1<1, z2<1, z3<1 ), 1, sum )
x1 <- apply( flisa[,2:4], 1, mean )
x2 <- z4==0
x3 <- (z4==1) | (z4==2)
x4 <- (z5==2)
flisa.var <- data.frame( y=y, x1=x1, x2=x2, x3=x3, x4=x4 )
```

13
# split the data into training and test sets (make indices)
testI <- (flisa[,1] >= julian(4,1,1998)) * (1:nrow(flisa))
testI <- testI[testI>0]
trainI <- (flisa[,1] < julian(4,1,1998)) * (1:nrow(flisa))
trainI <- trainI[trainI>0]

# estimation and prediction (forecasting)
flisa.glm <- glm(y ~ x1+x2+x3+x4, family=binomial, data=flisa.var[trainI,])
summary(flisa.glm)
flisa.pred <- predict.glm(flisa.glm, newdata=flisa.var[testI,], type="response")

PRECIPITATION AMOUNTS ON WET DAYS

# BUGS program to sample from the posterior distribution
model lmm;
const
N = 251;
var
y[N], x1[N], v1[N], mu[N], tau[N], beta0, beta1;
data y, x1, v1 in "lmm.dat";
inits in "lmm.in";
{
  for (i in 1:N)
  {
    mu[i] <- x1[i];
    log(tau[i]) <- beta0 + beta1*v1[i];
    y[i] ~ dnorm(mu[i], tau[i])I(0.1,);
  }
  beta0 ~ dnorm(0,0.0001);
  beta1 ~ dnorm(0,0.0001);
}

# S-Plus code to approximate quantiles of the predictive distribution.
# x = the values of the new explanatory variables (vector)
# theta = the sampled parameters from the Gibbs sampler (matrix)
# pp = percentiles of interest (vector)
# qq = approximate quantiles of the predictive distribution (vector)
u <- x[1]
sigma <- sqrt( exp(- (theta[,1] + theta[,2]*x[2]) ) )
pp <- c(0.025, 0.975)
qqMat <- matrix(0, length(sigma), length(pp))
for (i in 1:length(sigma))
{
  ppAdj <- pp + (1-pp)*pnorm(.1, mu, sigma[i])
  qqMat[i,] <- qnorm(ppAdj, mu, sigma[i])
}
qq <- apply(qqMat, 2, mean)
A.2 ALTERNATIVE II

# BUGS program to sample from the posterior distribution
model lmmCens;
const
N = 766;
var
y[N], y.cens[N], x1[N], x2[N], x3[N], x4[N], v1[N], mu[N], tau[N],
alpha2, alpha3, alpha4, beta0, beta1;
data y, y.cens, x1, x2, x3, x4, v1 in "lmmCens.dat);
inits in "lmmCens.in";
{
for (i in 1:N)
{
mu[i] <- x1[i] + alpha2*x2[i] + alpha3*x3[i] + alpha4*x4[i];
log(tau[i]) <- beta0 + beta1*v1[i];
y[i] ~ dnorm(mu[i], tau[i])I(, y.cens[i]);
alpha2 ~ dnorm(0.0,0.0001);
alpha3 ~ dnorm(0.0,0.0001);
alpha4 ~ dnorm(0.0,0.0001);
beta0 ~ dnorm(-2.0,0.5);
beta1 ~ dnorm(-0.2,2);
}
}

# S-Plus code to approximate quantiles of the predictive distribution.
# x = the values of the new explanatory variables (vector)
# theta = the sampled parameters from the Gibbs sampler (matrix)
# pp = percentiles of interest (vector)
# qq = approximate quantiles of the predictive distribution (vector)
pp <- c( 0.025, 0.975 )
qqMat <- matrix( 0, length(mu), length(pp) )
for (i in 1:length(mu))
{
qqMat[i,] <- qnorm( pp, mu[i], sigma[i] )
}
qq <- apply( qqMat, 2, mean )
qq <- qq * (qq > 0.1)
<table>
<thead>
<tr>
<th>Nr.</th>
<th>Tittel</th>
<th>Forfatter</th>
<th>Institusjon</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/97</td>
<td>Metoder for kvantifisering av hydrologiske prognosefelts representativitet.</td>
<td>Kai Fjelstad, NLH (diplomarbeid).</td>
<td></td>
</tr>
<tr>
<td>2/97</td>
<td>Effekter av flomsikringstiltak. En gjennomgang av litteraturen.</td>
<td>Magne Wathne, SINTEF.</td>
<td></td>
</tr>
<tr>
<td>4/97</td>
<td>Forslag til kravspesifikasjon av vassdragsmodell.</td>
<td>Lars A. Roald, NVE.</td>
<td></td>
</tr>
<tr>
<td>5/97</td>
<td>A note on floods in high latitutde countries.</td>
<td>Lars A. Roald, NVE.</td>
<td></td>
</tr>
<tr>
<td>6/97</td>
<td>Climate change and floods.</td>
<td>Nils Roar Sælthun, NIVA.</td>
<td></td>
</tr>
<tr>
<td>7/97</td>
<td>Flomdemping i Gudbrandsdalslågen. Programstruktur og systembeskrivelse.</td>
<td>Magne Wathne og Knut Alfredsen, SINTEF.</td>
<td></td>
</tr>
<tr>
<td>8/97</td>
<td>Klima, arealbruk og flommer i perspektiv.</td>
<td>Arnor Njos, Jordforsk.</td>
<td></td>
</tr>
<tr>
<td>9/97</td>
<td>Flood forecasting in practice.</td>
<td>Dan Lundquist, GLB.</td>
<td></td>
</tr>
<tr>
<td>1/98</td>
<td>Bruk av ensembleprognoser til estimering av usikkerhet i lokale nedbørprognoser.</td>
<td>Marit Helene Jensen, DNMI.</td>
<td></td>
</tr>
<tr>
<td>2/98</td>
<td>LANDSKAP OG ESTETIKK –et kulturelt perspektiv.</td>
<td>Oddrun Sæter, Byggforsk.</td>
<td></td>
</tr>
<tr>
<td>3/98</td>
<td>Betydning av vårflommens størrelse for tetthet av laks- og ørretunger i Saltdalselva.</td>
<td>Arne J. Jensen og Bjørn Ove Johnsen, NINA.</td>
<td></td>
</tr>
<tr>
<td>4/98</td>
<td>Pure Model Error of the HBV-model.</td>
<td>Øyvind Langsrud, Arnoldo Frigessi and Gudmund Høst, Norwegian Computing Center (Norsk Regnesentral)</td>
<td></td>
</tr>
<tr>
<td>5/98</td>
<td>Analyse av effekter av urbanisering og avrenningsutjevnende tiltak i Svebestadfeltet – Sandnes kommune.</td>
<td>Jadranka Milina, SINTEF.</td>
<td></td>
</tr>
<tr>
<td>6/98</td>
<td>Virkning av urbanisering på avrenningsforhold i Storånavassdraget.</td>
<td>Jadranka Milina, SINTEF.</td>
<td></td>
</tr>
<tr>
<td>7/98</td>
<td>Metodebeskrivelse for flomsoneanalyse med eksempler fra Flisa og Kirkenær.</td>
<td>Søren Elkjær Kristensen og Astrid Voksø, NVE.</td>
<td></td>
</tr>
</tbody>
</table>
Kontaktpersoner

formann i styringsgruppen: Fylkesmann Erling Norvik
Tlf. 69 24 70 00
E-post: erling.norvik@fm-os.sri.telemax.no

programleder: Arnor Njøs
Jordforsk
Tlf. 64 94 81 70 (Jordforsk)
Tlf. 22 95 90 98 (NVE)
E-post: arnor.njos@jordforsk.nlh.no
E-post: xarn@nve.no.

naturgrunnlag og arealbruk: Arne Grønlund
Jordforsk
Tlf. 64 94 81 09
E-post: arne.gronlund@jordforsk.nlh.no

Noralf Rye
Universitetet i Bergen
Tlf. 55 58 34 98
E-post: noralf.rye@geol.uib.no

tettsteder: Oddvar Lindholm
Norges Landbrukshøgskole
Tlf. 64 94 87 08
E-post: oddvar.lindholm@tfh.nlh.no

flomdemping, flomvern og flomhandtering: Dan Lundquist
Glommens og Laagens Brukseierforening
Tlf. 22 54 96 00
E-post: danlund@sn.no

skaderisikoanalyse: Nils Roar Sælthun
Norsk institutt for vannforskning
Tlf. 22 18 51 21
E-post: nils.saelthun@niva.no

miljøvirkninger av flom og flomforebyggende tiltak: Olianne Eikenæs
Norges vassdrags- og energiverk
Tlf. 22 95 92 24
E-post: oli@nve.no

databaser og GIS: Lars Andreas Roald
Norges vassdrags- og energiverk
Tlf. 22 95 92 40
E-post: lars.roald@nve.no

modellarbeid: Ånund Killingtveit
Norges teknisk-naturvitenskaplige universitet
Tlf. 73 59 47 47
E-post: aanund.killingtveit@bygg.ntnu.no

programadministrasjon: Olianne Eikenæs
Norges vassdrags- og energiverk
Tlf. 22 95 92 24
E-post: oli@nve.no
Hjemmeside: http://www.nve.no

Per Einar Faugli
Norges vassdrags- og energiverk
Tlf. 22 95 90 85
E-post: pef@nve.no