A NEW INVESTIGATION ON THE PERFORMANCE OF RAINFALL IDF MODELS*

B. GHAHRAMAN** AND S. M. HOSSEINI

1 Dept. of Irrigation, College of Agriculture, Ferdowsi University of Mashhad, Mashhad, I. R of Iran E-mail: bijangh@ferdowsi.um.ac.ir
2 Dept. of Civil Engineering, Ferdowsi University of Mashhad, Mashhad, I. R. of Iran

Abstract– Rainfall intensity at different durations and frequencies (IDF curves) are widely needed in nearly all water projects, e.g., culverts, urban drainage, etc. However, IDF curves are reported at discrete durations and frequencies. To avoid subjective decisions, however, IDF models have been developed and are used universally. In this article, the performance of five commonly used IDF models is analyzed for three synoptic stations in Iran. The results show that there is an overall acceptable performance for all models, as far as the prediction of rainfall intensity from duration and frequency is concerned. However, from a probabilistic point of view, the non-negativity of a cumulative distribution function is not satisfied. Therefore, the non-negativity concept was added as a constraint. Consequently for this case, the performance of the models decreased considerably. It seems that all IDF models currently in use are not efficient, yet more efficient models must be developed to take into account all characteristics of the phenomena.

Keywords – IDF models, CDF, Iran

1. INTRODUCTION

Rainfall intensity at different durations and frequencies (IDF curves) are widely used in the design of nearly all water projects such as culverts, urban drainage systems and so on. Intensity values at a given duration and frequency may be found in the form of tables, curves, or isohyetal contours [1]. To reduce subjective decisions in interpolation among curves, isohyetal curves, or tables, however, IDF models may be used [2].

An IDF model is an empirical equation representing a relationship among maximum rainfall intensity (as dependent variable) and other parameters of interest such as rainfall duration and frequency (as independent variables). The simplest type of these models is a relationship between $i$ (rainfall intensity) and $t$ (rainfall duration) for a given frequency. Chow [3] categorized the following equations for different locations of USA:

\[ i = \frac{a}{t + b} \]  

\[ i = \frac{a}{t^c} \]  

\[ i = \frac{a}{(t+b)^c} \]  

\[ i = \frac{a}{t^c + b} \]

where $a$, $b$, and $c$ are constant parameters related to meteorological conditions. However, other researchers before Chow have also used these equations. The earliest published documents are due to Besson (1931)

---

*Received by the editors July 27, 2003; final revised form October 26, 2004.

**Corresponding author
B. Ghahraman / S. M. Hosseini

Iranian Journal of Science & Technology, Volume 29, Number B3                                                                                     June 2005

(quoted by Remenieras [4]) for Eq. (1) (also known as Steel or Talbot formula), Bernard (1932) (quoted by Linsley et al. [5]) for Eq. (2), and Sherman (1931) (quoted by Murray and Rao [6]) for Eq. (3). All above equations are only valid for a given frequency. By changing frequency values, constant parameters will change accordingly. It is possible to express these parameters by some simple exponential type equations. These equations are functions of local meteorological parameters and frequency [7]

\[ a = k_1 \cdot T^{k_2} \] (5)

\[ c = k_3 \cdot T^{k_4} \] (6)

where T is return period, and k_1 to k_4 are local constant parameters. They did not offer any equation for b. Although Vierra and Souza [7] related c to T, in general c varies just negligibly and it may be assumed constant over all durations. As a result, general IDF models that are modified forms of Eqs. (3) and (4) are expressed as

\[ i = \frac{a_1 \cdot T^{a_2}}{(t+b)^c} \] (7)

\[ i = \frac{a_1 \cdot T^{a_2}}{t^b} \] (8)

Equations (7) and (8) are extremely popular (Bernard, quoted by Linsley et al. [5]; Chow [8]; Ramman and Bandyopadhyia [9]) Raudkivi [10] has reported a different form of IDF model

\[ f(t, T, \theta) = \frac{g + h \cdot \log(T)}{(t+c)^d} \] (9)

Vaziri [11, 12] probably was the first investigator who developed IDF models for Iran. He fitted Eqs. (1) and (2) to data of about 150 stations in Iran. He used data of Iran Meteorological Organization and those of Ministry of Power stations and found that none of these models could be used individually for all durations and all intensities.

Alizadeh [13] used Eq. (3) to model rainfall data of Mashhad (data on College of Agriculture Station) corresponding to different durations and at every specified frequency. Following Alizadeh [13], Ghahraman (unpublished surveys) showed that parameters of this model (Eq. (3)), by considering some corrections, exhibit specific trends with frequency. By increasing T, the values of a, and b also increased, while c decreased. Eq. (3) was adopted to model Mashhad rainfall data (airport station) by Ghahraman et al. [14]. These authors correlated a and c parameters of Eq. (3) to frequency, but found that in contrast to the power form of Eqs. (5) and (6), the logarithmic form better fitted the data.

Considering different IDF models, this study aims to evaluate the performance of different IDF models based on rainfall data of Khorasan Province of Iran.

2. THEORY AND METHOD OF RESEARCH

An IDF model may be expressed by the following relationship:

\[ y_{t,T} = f(t, T, \theta) \] (10)

where \( y_{t,T} \) is rainfall intensity at duration t and frequency T, \( f(t, T, \theta) \) defines an IDF model, and \( \theta \) is the parameter set of this model. All IDF models are empirical and the following conclusions can be made from their mathematical structure:

- average rainfall intensity decreases as duration increases, and
- average rainfall intensity increases as return period increases.

Therefore, it is hypothesized explicitly that an IDF model consists of two separate terms

\[ f(t, T, \theta) = f_1(t, \theta_1) \cdot f_2(T, \theta_2) \] (11)
Such multiplication is similar to managing two pdf of two independent functions. In general, an IDF model has a probabilistic nature, as far as frequency is concerned. One can safely assume that observed maximum rainfall intensities corresponding to duration \( t \) are iid (independent and have identical distribution). Therefore, frequency \( T \) corresponding to rainfall intensity at duration \( t \) may be expressed as follows (for example, Tung and Lu [15]):

\[
T = f^{-1}(y_{t,T}; \theta)
\]

As a result, cumulative distribution function (CDF) of an IDF model can be performed as

\[
F(y_t) = P(y_t \leq y) = 1 - \frac{1}{f^{-1}(y_t; \theta)}
\]

It should be easy to verify that the upper limit of any CDF, \( F(y_t=\infty) \), is equal to unity. On the other hand, it is not possible that a CDF be negative. Therefore, all rainfall intensities must be greater than a lower limit. This lower limit theoretically corresponds to a lower value of \( y \) at each duration. This theoretical lower rainfall intensity, in general, depends on \( t \) and \( \theta \) and can be found after substituting \( T=1 \) in an IDF model.

Four common IDF models, restated from before, above CDF and theoretical lower limits are given in Eqs. (14-16). In each equation, we put together the terms for all 4 models, respectively, to save space.

\[
y_{t,T} = \frac{g + h \cdot \log(T)}{(t+c)^d}, \quad y_{t,T} = \frac{a \cdot T^b}{(t+c)^d}, \quad y_{t,T} = \frac{a \cdot T^b}{t^d + c} \quad (14)
\]

\[
F(y_t) = 1 - \frac{1}{y_{t,T} - y}, \quad F(y_{t,T}) = 1 - \left[ \frac{(t+c)^d}{a} \cdot y_t \right]^{-1/b}, \quad F(y_{t,T}) = 1 - \left[ \frac{t^d + c}{a} \cdot y_t \right]^{-1/b}
\]

\[
F(y_t) = 1 - \left[ \frac{t^d}{a} \cdot y_t \right]^{-1/b}
\]

\[
y_{t,\text{min}} = \frac{g}{(t+c)^d}, \quad y_{t,\text{min}} = \frac{a}{(t+c)^d}, \quad y_{t,\text{min}} = \frac{a}{t^d + c}, \quad y_{t,\text{min}} = \frac{a}{t^d}
\]

Although model 4 is a simplification of models 2 and 3 (with \( c=0 \)), we included it in our study because its simple form guarantees its rapid parameter estimations. In addition to these four common models, some researchers tend to use another simple model, with the following format:

\[
y_{t,T} = f_1(t; \theta) \quad \text{(for a given } T)\]

The frequency is not explicitly considered in this model. A common form of this model (Model 5) can be of the following format, called model 5 in this study. Therefore

\[
y_{t,T} = \frac{a}{(t+c)^d} \quad \text{for a given } T
\]

Although frequency is not explicitly included in the model, it may be assumed that at least one element of the set \( \theta \) is related to frequency. Therefore, a general IDF model including its upper and lower limits, can be studied accordingly. In this case \( \theta \) and the IDF model, should read, respectively, as

\[
\theta = f_2(T; \lambda)
\]

\[
y_{t,T} = f_1(t; f_2(T; \lambda)) = f(t, T; \theta, \lambda)
\]

As a result, \( T \) and CDF can be found afterwards. However, Eq. (19) generally has a form such that explicit computations of \( T \) and CDF are hindered. Therefore, we called this model an indirect model.
Unknown parameters in any IDF model must be found in such a way that the model outputs best fit to a given IDF table. Therefore, an objective function may focus on the minimization of deviations as follows:

$$Min: D(\theta) = \sum_{T} \sum_{t} \left( y_{t,T} - f(t,T;\theta) \right)^2$$  \hspace{1cm} (21)

where $y_{t,T}$ is the observed intensity value in the IDF Table, and $f(t,T;\theta)$ is the estimated intensity. This objective function is valid for the first four IDF models. On the other hand, the following objective function must be considered for model 5 (Eq. (18))

$$Min: D(\theta) = \sum_{t} \left[ y_{t} \left( \frac{a}{(t+c)^d} \right)^2 \right]$$  \hspace{1cm} (22)

No constraint has been applied to the minimization procedure described by Eqs. (21) and (22). This procedure, called “without constraint approach” in this study, follows a conventional regression analysis procedure and does not consider the probabilistic nature of IDF models. However, if the probabilistic nature of the IDF models is considered, due to non-negativity of cumulative distribution function (CDF), the following constraint must be satisfied as well:

$$y_{t,\min}(\theta) \leq y_{i,t} \leq 1 \quad \text{for all durations} \quad (23)$$

where $n$ is the number of observations, $y_{t,\min}(\theta)$ is the theoretical lower limit of rainfall intensity at duration $t$, and $y_{i,t}$ is the $i$th observed rainfall intensity at duration $t$. This second approach is called with constraint case in this study.

Absolute relative error, both maximum and mean absolute relative error, can be used as criteria for evaluating the performance of any IDF model. Absolute relative error is defined by the following relationship:

$$Absolute\ Relative\ Error = \frac{\text{Observed value} - \text{Estimated value}}{\text{Observed value}}$$  \hspace{1cm} (24)

The optimization procedure was applied to data of three rain gage stations (introduced later), in order to find a better understanding of different IDF models. Lingo software (Industrial LINGO/PC, Release 3.0 [25 July 95], Copyright [C] 1995, LINGO System, Inc., 1415 North Dayton St., Chicago, IL 60622) was used to conduct required optimization and regression analysis.

3. DATA SET USED FOR THIS STUDY

Torbat-Heydarieh, Sabzevar, and Mashhad are three autographic rain gage stations in the Khorasan Province of Iran that have more than 25 years of data, which is an adequate record length. Other stations in this area have either low record length or incomplete data. The IDF of all these stations are reported by the Applied Center for Meteorological Research [16]. Table 1 shows the main characteristics of these rain gage stations.

<table>
<thead>
<tr>
<th>Station</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Altitude (m)</th>
<th>Record period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Torbat-Heydarieh</td>
<td>59-13</td>
<td>35-16</td>
<td>1450.8</td>
<td>1969-1993</td>
</tr>
<tr>
<td>Sabzevar</td>
<td>57-43</td>
<td>36-12</td>
<td>977.6</td>
<td>1967-1993</td>
</tr>
<tr>
<td>Mashhad</td>
<td>59-38</td>
<td>36-16</td>
<td>990.0</td>
<td>1969-1993</td>
</tr>
</tbody>
</table>

4. RESULTS AND DISCUSSION

The results of applying an optimization procedure, introduced as “with constraint” and “without constraint”, to data of different selected stations are reported in the following subsections. The results for the Torbat-Heydarieh station are discussed in more detail.
Torbat Heydarieh

a. Performances of the models

The general performance criteria of the first four IDF models for the Torbat-Heydarieh station are presented in Table 2. In general, taking into account the non-negativity constraint (Eq. (23)) in the optimization algorithm, the number of iterations increased, as well as the objective function value (Eq. (21)). Mean and maximum absolute relative error values also increased. Model 1 performs slightly better than the other 3 models. A comparison between the IDF table data [16] and the outputs of model 2 (t ≤ 6 hr) is made in Fig. 1. It can be concluded that, by incorporating the constraint, the performance of the model becomes unsatisfactory. The figure supports that as the return period increases (frequency decreases), a better match is found between the IDF model outputs and the IDF table data. It is usually assumed that intensity values at higher return periods are of a higher standard error of estimates [17], however, this point does not hold here.

<table>
<thead>
<tr>
<th>Torbat-Heydarieh</th>
<th>Sabzevar</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Duration class, hr</strong></td>
<td><strong>Duration class, hr</strong></td>
</tr>
<tr>
<td><strong>t ≤ 6</strong></td>
<td><strong>t ≤ 2</strong></td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* N=Number of iterations  OF=Objective function
  MEARE =Mean absolute relative error  MAARE = Maximum absolute relative error
  a: without constraint, and b: with constraint

This is because we did not fit a distribution function to original data, and therefore, 2 or 100 year return periods are the same in this regard.

In spite of a nearly good fit between IDF data and the model in the without constraint case, Fig. 2 shows some malfunctions. This figure shows that ignoring the non-negativity constraint leads to negative CDF for a number of data points, which does not make sense from a probabilistic point of view. On the other hand, including the constraint supports the non-negativity axiom, though its overall performance is not as good as that of the other case.
Fig. 1. Fitting IDF model 2 at different return periods (t<=6 hr) for Torbat-Heydarieh station. + sign indicates observed data of IDF table, while solid and dashed lines refer to without constraint and with constraint cases, respectively.

Fig. 2. Cumulative distribution function (CDF) of IDF model 2 for Torbat-Heydarieh rainfall intensity at some selected time durations (t<=6 hr). + sign indicates observed data of IDF table, while solid and dashed lines refer to without constraint and with constraint cases, respectively.
Not all durations and frequencies lead to identical absolute relative errors. Figure 3 shows that there is a systematic trend for mean absolute relative error as a function of either duration or frequency. Maximum absolute relative error shows a similar trend (data not shown).

![Graphs showing absolute relative errors](image)

**Fig. 3.** Mean absolute relative error when IDF model 1 is applied to different time period classes of Torbat-Heydarieh station. Solid and dashed lines refer to without and with constraint cases, respectively.

**b. Splitting time duration**

Splitting the data into two separate classes, in terms of duration $t > 2$ and $t > 2$ hours, produced better performance for the first 4-models (Table 2). Splitting is supported by physical reasoning, because storms with greater than 2 and those with less than 2 hour duration have different mechanisms [18]. Better performance in fit, however, could not resolve the CDF negativity problem for the without constraint case. The IDF model 4 was the worst one amongst the others, especially when compared with IDF models 2 and 3. This is acceptable because model 4 has one parameter less than other models, and therefore, less flexibility for fitting to data points.

**c. Model 5**

The performance criteria for IDF model 5 are shown in Table 3. Mean and maximum absolute relative error and objective function are quite low, indicating the good performance of the model. Despite this good performance, it is not possible to make any clear statement on the behavior of CDF. The parameters of this model (a, c, and d in Eq. (18)) showed a systematic trend with the return period (data not shown). Therefore, their functional relationships were constructed. Eqs. (25) to (27) are the relationships for three duration classes of $t \leq 6$, $t \leq 2$, and $t > 2$ hours, respectively.
The constant parameters $a_i$ and $b_i$ were determined by regression analysis (not reported here). This model is more complex than the commonly used first four models, which is an extended form of a simple one to cover simultaneously all return period classes in only one set. Such a complex form hinders the direct computation of $T$ and CDF. Table 4 shows the overall performance criteria for this generalized IDF model. Comparing this table with Table 2 indicates that this model performs worse, when compared with model 1, and does not show any advantage over the other 3 models 2, 3, and 4. Therefore, it is not recommended for use in practice.

Table 3. Performance criteria of IDF model 5

<table>
<thead>
<tr>
<th>$T$</th>
<th>Torbat-Heydarieh</th>
<th>Sabzevar</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>OF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MEARE =0.055, MAARE =0.155</td>
<td>(MEARE=0.057, MAARE =0.161)</td>
</tr>
<tr>
<td>2</td>
<td>209</td>
<td>0.128</td>
</tr>
<tr>
<td>5</td>
<td>301</td>
<td>0.784</td>
</tr>
<tr>
<td>10</td>
<td>402</td>
<td>1.710</td>
</tr>
<tr>
<td>20</td>
<td>470</td>
<td>2.999</td>
</tr>
<tr>
<td>50</td>
<td>547</td>
<td>5.230</td>
</tr>
<tr>
<td>100</td>
<td>597</td>
<td>7.252</td>
</tr>
<tr>
<td></td>
<td>(MEARE =0.010, MAARE =0.060)</td>
<td>(MEARE =0.111, MAARE =0.659)</td>
</tr>
<tr>
<td>2</td>
<td>1595</td>
<td>0.13</td>
</tr>
<tr>
<td>5</td>
<td>1736</td>
<td>0.213</td>
</tr>
<tr>
<td>10</td>
<td>1980</td>
<td>0.582</td>
</tr>
<tr>
<td>20</td>
<td>1897</td>
<td>1.159</td>
</tr>
<tr>
<td>50</td>
<td>2020</td>
<td>2.180</td>
</tr>
<tr>
<td>100</td>
<td>2043</td>
<td>3.156</td>
</tr>
<tr>
<td></td>
<td>(MEARE =0.007, MAARE =0.020)</td>
<td>(MEARE =0.008, MAARE =0.024)</td>
</tr>
<tr>
<td>2</td>
<td>51</td>
<td>0.007</td>
</tr>
<tr>
<td>5</td>
<td>55</td>
<td>0.002</td>
</tr>
<tr>
<td>10</td>
<td>54</td>
<td>0.001</td>
</tr>
<tr>
<td>20</td>
<td>58</td>
<td>0.002</td>
</tr>
<tr>
<td>50</td>
<td>57</td>
<td>0.007</td>
</tr>
<tr>
<td>100</td>
<td>54</td>
<td>0.014</td>
</tr>
</tbody>
</table>

$T$=Return period (yr)  
N=Number of iterations  
OF=Objective function 
MEARE =Mean absolute relative error  
MAARE = Maximum absolute relative error

Sabzevar and Mashhad

Table 2 demonstrates the overall performance of IDF models 1-3 at different time scales for the Sabzevar station. Overall, the findings are similar to the results obtained for the Torbat Heydarieh station. The IDF model 4 and also $t \leq 6$ hr time scale is not considered here because it was shown previously that these cases are not appropriate. The performance criteria of IDF model 5 and its generalized forms are presented in Tables 3 and 4, respectively. These results are a support for previous findings that: (a) inclusion
of the non-negativity constraint, although rational, results in the poor performance of the models, and (b) splitting the time scale $t$ into two distinct classes is suggested.

<table>
<thead>
<tr>
<th>Table 4. Performance criteria of IDF model 5 (extended case)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration class, hr</td>
</tr>
<tr>
<td>$t \leq 6$</td>
</tr>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>Torbat-Heydarieh</td>
</tr>
<tr>
<td>Number of iterations</td>
</tr>
<tr>
<td>Objective function</td>
</tr>
<tr>
<td>Mean absolute relative error</td>
</tr>
<tr>
<td>Maximum absolute relative error</td>
</tr>
<tr>
<td>Sabzevar</td>
</tr>
<tr>
<td>Number of iterations</td>
</tr>
<tr>
<td>Objective function</td>
</tr>
<tr>
<td>Mean absolute relative error</td>
</tr>
<tr>
<td>Maximum absolute relative error</td>
</tr>
</tbody>
</table>

* a: without constraint, and b: with constraint

Similar results were obtained from the analysis of the data of the Mashhad station. To save space, this analysis and numerical values are not reported here.

### 5. CONCLUSIONS

Five common IDF models were fitted to data of three synoptic stations located in Khorasan Province in the eastern part of Iran. As the first approach, the parameters of these models were optimized using a non-linear optimization procedure. It was found that all five models show an acceptable performance, if rainfall intensity is to be computed at the required duration and frequency. However, none of the IDF models satisfied the non-negativity of CDF, which arises from the probabilistic nature of the IDF models. The non-negativity of CDF was added, as a constraint, to the non-linear optimization procedure and the parameters were optimized accordingly. In this new approach, although the non-negativity of CDF was satisfied, the performance criteria of all models decreased drastically. Tung and Lu [15] reported similar results. Therefore, it seems that available IDF models are not efficient and more efficient models must be developed.

**Acknowledgment** - This paper is partly funded by Research Council of Ferdowsi University of Mashhad. By this, the authors acknowledge.

### REFERENCES