Journal of Hydrology, 122 (1991) 353–371
Elsevier Science Publishers B.V., Amsterdam

[3]

DETERMINATION IN REAL TIME OF THE RELIABILITY OF RADAR RAINFALL FORECASTS

T. DENOEUX[1]*, T. EINFALT[2] and G. JACQUET[2]

[1]*CERGRENE, Ecole Nationale des Ponts et Chaussées, La Courtine, F-93167 Noisy-le-Grand cedex (France)*

[2]*RHEA, 8d rue de la Ceinture, F-78000 Versailles (France)*

(Received October 12, 1989; accepted after revision May 7, 1990)

ABSTRACT

Denoeux, T., Einfalt, T. and Jacquet, G., 1991. Determination in real time of the reliability of radar rainfall forecasts. J. Hydrol., 122: 353–371.

Quantitative rainfall forecasts obtained from interpretation of radar data can be of great interest in urban hydrology, provided their reliability is known in real time. The aim of this study was to examine the feasibility of an a priori estimation of forecast reliability from characteristics of rainfall areas and atmospheric vertical structure. The first step has been to design a method to check the relevance of a criterion of forecasting quality to a particular application of the forecasts. This method was applied to the case of real-time control of a drainage network in a suburban area of Paris, and led to the definition of a new quality criterion, consistent with the user's utility function. Potential predictors of forecasting quality were then defined, to be calculated in real time from radar and rawinsonde data. In the final step, statistical and heuristic techniques, applied to a learning set of examples taken from 46 rainfall events, provided decision rules which can be used in real time to estimate the quality of radar forecasts. Although these rules are valid only in a specific operational context, the methodology is general, and can be transferred to other forecasting problems in hydrology, as well as in other domains.

INTRODUCTION

Since the 1950s, weather radar has increased in importance as a tool for precipitation nowcasting (i.e. forecasting with a lead time of up to 6 h). At the end of the 1960s the development in electronics and computers led to the operational availability of fast digital processing systems, which provided the possibility for the development of automatic rainfall forecasting systems, based on radar echo advection (see Collier (1978), Austin (1985) and Einfalt (1988) for reviews on these methods).

Users' needs in urban and rural areas have encouraged further development of these systems for hydrological use. In the U.K., for instance, river flooding

---

* Present address: Laboratoire d'Informatique Avancée de Compiègne, TECHNOPOLIS, Bât. 3, rue du Fonds Pernant, F-60200 Compiègne, France.

0022-1694/91/$03.50

problems were a source of collaboration between hydrologists and meteorologists (Dee Project Final Report, 1977; Collier et al., 1980). Elsewhere, increasing urbanization caused problems for sewer system managers, leading to considerably higher flow volumes for heavy rains. One possible solution to this problem is an optimized control of the sewer system, depending on the time and space distribution of rainfall (Schilling, 1989). The usefulness of a very short-term rainfall forecast in this context has already been demonstrated (e.g. Austin and Austin, 1974; Frérot, 1987; Schilling and Petersen, 1987; Frérot and Jacquet, 1989).

The importance of these needs and the availability of these solutions should have led to the operational use of the optimal forecasting methods in urban hydrology. However, in spite of very encouraging pilot projects (Huff et al., 1980; Damant et al., 1983), the number of operationally working systems in sewer management is extremely limited. In addition to the institutional problems caused by moving away from a 'static' management of sewer systems (Denoeux et al., 1987), a number of unsolved technical problems remain. Important points are, in particular, the lack of knowledge of forecast reliability and of the sensitivity of the hydrological system. Perfect forecasts are not an operational necessity, as long as the hydrological users know which types of errors may occur, and the possible consequences on their decisions (Einfalt and Denoeux, 1989). This holds for any application of weather radar, as the accuracy of a radar-based forecast is dependent on the features of the precipitation systems (Wilson, 1966; Austin and Bellon, 1974; Ciccione and Pircher, 1984). Frontal systems yield a higher reliability, as they are stable for a long time, whereas systems of a convective type may result in hardly any predictability, because of the very short life-time of the rainfall structures. For this reason, rainfall volumes provided by an automatic forecasting system should be used only if their accuracy is well known.

This paper presents a method to design rules for the estimation of the reliability of rainfall forecasts in real time. Each set of rules will be valid in given climatological conditions, and for a precisely defined application of the forecasts. The method itself consists of three steps, which will be described in detail. These steps are:

(1) the definition of an application-dependent criterion of forecast quality,

(2) the selection of meteorological parameters as reliability predictors,

(3) the generation of decision rules relating the predictors to the expected quality of the forecasts.

DEFINITION OF AN APPLICATION-DEPENDENT CRITERION OF FORECAST QUALITY

A criterion of forecast quality must be defined to compare different forecasting techniques (Elvander, 1976; Carpenter and Owens, 1981; Tsonis and Austin, 1981; Ciccione and Pircher, 1984; Einfalt et al., 1989), or the performance of one particular technique applied to different rainfall events (Austin and Bellon, 1974); the latter is the specific concern in this paper. The principal criteria used to assess the quality of radar forecasts (Denoeux, 1989; Denoeux et al., 1989) are based on

(1) the movement of rainfall areas, e.g. the absolute and relative difference between observed and forecast movement (Austin and Bellon, 1974);

(2) the rainfall field, e.g. the critical success index (Bellon and Austin, 1978);

(3) the hyetographs, such as the mean absolute error (ΔH) and the mean relative error ($\Delta H/H$), taken by comparing the measured and the forecast rainfall depths over particular catchments (Huff et al., 1980; Damant et al., 1983; Bellon and Austin, 1984).

The choice of any one of these criteria is based on hypotheses which are rarely specified. For example, the use of the ΔH and $\Delta H/H$ criteria, often preferred by hydrologists, relies on the three following conditions:

(1) the temporal rainfall distribution has not to be forecast precisely (thus, one can consider only rainfall depths, instead of complete hyetographs);

(2) the overestimation of a rainfall volume has the same impact as its underestimation (allowing for an error calculation regardless of sign);

(3) the average adequately represents, from the user's viewpoint, the error distribution for the catchment basins considered (i.e. an error of 50% on two basins is equivalent to an error of 0% on the first one and an error of 100% on the second one).

If these conditions are not satisfied, it is not possible to state that criteria ΔH and $\Delta H/H$ are consistent with the user's conception of forecast quality, i.e. that a forecast f_1 which is superior to a forecast f_2 in terms of ΔH or $\Delta H/H$ would actually be preferred by the user. This lack of exactness is dangerous, because it has been shown (Denoeux, 1989; Denoeux et al., 1989) that different quality criteria usually are not equivalent: a forecast f_1 that is better than a forecast f_2 according to some criterion C_1, may, with a high probability, be worse according to some other criterion C_2.

These considerations show the need for a method to verify the relevance of a quality criterion to a given application.

General approach

If a forecast is used for decision-making, an initial hypothesis has to be verified:

Hypothesis 1. There exists a decision function f_d defining an action α, on the basis of a forecast p and n other decision criteria $(x_i)_{i=1,n}$:

$$f_d\colon (p, x_1,\ldots,x_n) \to \alpha$$

Such a function can be defined independently from the source of the decision, be it a human being, an optimization procedure, or an expert system.

Each action α implies, on the other hand, some consequences also dependent on the actual value r to be forecast and on m other influencing factors $(y_i)_{i=1,m}$. The q consequences of an action α, noted $(k_i)_{i=1,q}$, are related to α, r and $(y_i)_{i=1,m}$ by a function f_c:

f_c: $(\alpha, r, y_1, \ldots, y_m) \rightarrow (k_1, \ldots, k_q)$

This leads us to the second hypothesis.

Hypothesis 2. There is a user who is capable of deciding which of any two series of consequences $(k_i)_{i = 1, q}$ and $(k_i')_{i = 1, q}$ is preferable.

This second hypothesis actually describes the fact that there is a user able to decide a posteriori which of two decisions applied to the same situation is preferable in terms of their consequences. This corresponds to what Von Neumann and Morgenstern (1953) described as "the picture of an individual whose system of preferences is all-embracing and complete, i.e. who, for any two objects ... possesses a clear intuition of preference". This hypothesis immediately implies the existence (Montgolfier and Bertier, 1978) of an ordinal utility function f_u, associating a positive real number u (called a utility index) with any object, i.e. in the present case with any set of consequences $(k_1, \ldots, k_q)$:

f_u: $(k_1, \ldots, k_q) \rightarrow u \in \mathbb{R}$.

and such that, for any two sets of consequences $(k_i)_{i = 1,q}$ and $(k_i')_{i = 1,q}$:

$(k_1, \ldots, k_q)$ is preferred to $(k_1', \ldots, k_q') \Leftrightarrow f_u(k_1, \ldots, k_q) > f_u(k_1', \ldots, k_q')$

and

$(k_1, \ldots, k_q)$ is equivalent to $(k_1', \ldots, k'_q) \Leftrightarrow f_u(k_1, \ldots, k_q) = f_u(k'_1, \ldots, k'_q)$

It should be noted that:

(1) f_u is defined up to a monotonic transformation: consequently, there are an infinity of ordinal utility functions corresponding to the same system of preferences;

(2) the definition of a function f_u does not give any meaning to differences in utility numbers: in other terms, hypothesis 2 alone is not sufficient to guarantee the existence of a measurable, or cardinal utility function. The problem of the measurement of utility, which has been a major research area in economics (e.g. Ellsburg, 1954), will not be addressed here; it is assumed only that some ordinal utility function f_u can be approximated, providing a sufficient understanding of the user's preference structure.

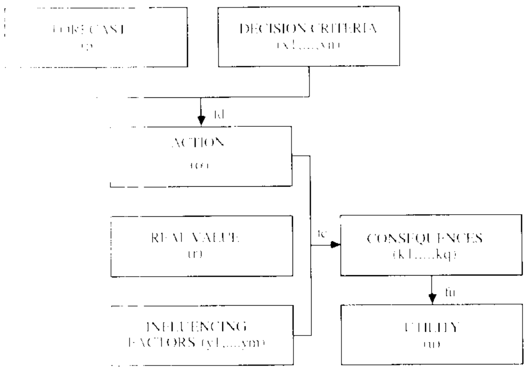

Fig. 1. Notations introduced in this section.

The definitions of functions f_d, f_c and f_u are summarized in Fig. 1. From these functions, it is possible to build a function g projecting a forecast p, a measured value r, the n decision criteria $(x_i)_{i=1,n}$ and the m influencing factors $(y_i)_{i=1,m}$ onto a real value u:

$$g: (p,x_1, ...,x_n,r,y_1, ...,y_m) \rightarrow f_u[f_c(f_d(p,x_1, ...,x_n),r,y_1, ...,y_m)] = u$$

If $(x_i)_{i=1,n}$ and $(y_i)_{i=1,m}$ are supposed to remain constant, it is therefore possible to use g to rank any pair of forecasts (p,p') corresponding to the same real value r:

$$p > p' \Leftrightarrow g(p,x_1, ...,x_n, r,y_1, ...,y_m) > g(p',x_1, ...,x_n, r,y_1, ...,y_m)$$

Thus, the supposition that the functions f_d, f_c and f_u exist, leads to a 'subjective' quality criterion based on the user's concerns, which can be compared with any 'objective' criterion. A method which allows comparison of different criteria is the random generation of a large number of forecasts for a number of known values of $(x_i)_{i=1,n}$ and $(y_i)_{i=1,m}$. For each situation, the correlation between any criterion and the function g can be evaluated; this leads to the choice of the most application-oriented criterion.

Application to the real-time control of an urban drainage system

The above-defined general approach can be applied to the problem of rainfall forecasting for the real-time control (RTC) of an urban drainage system. The problem is to forecast a set of hyetographs $(I_{i1}, ...,I_{ip})_{i=1,k}$, where I_{ij} is the intensity at time j on catchment i, p is the number of time steps and k is the number of catchments. The different actions to take are settings of control devices (valves, pumps, etc.) at certain time steps. The decision depends on some criteria in addition to the forecast itself, namely:

the previous rainfall,
the measured flow at different points of the sewer system,
the measured water level in the retention basins,
the known availability of the control devices.

The result of a control strategy can be measured, for instance, in overflow volume or quality change in the receiving waters. These consequences also depend on such influencing factors as the actual availability and reliability of the control and measuring devices during the event.

These considerations have been applied to the RTC of the Morée sewer system, situated in the northern part of Seine-Saint-Denis, a surburban area of Paris. Two retention basins are controlled: Blanc-Mesnil and Pont-Yblon, with capacities of 95 000 and 65 000 m^3, respectively. The Pont-Yblon basin comprises two parts, one of which is considered as a recreational area to be used in extreme cases. Three overflow points are situated just downstream of the inlet points of the Garonor, Bourget and Croult catchments (Fig. 2).

A computer program for the optimization of the control actions has been operational since 1987 (Frérot et al., 1986; Frérot, 1987). This program, based on

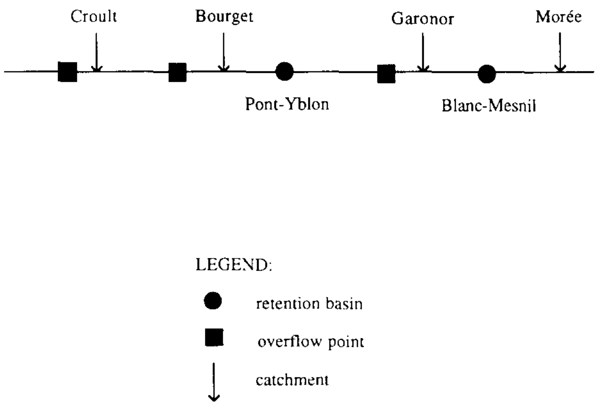

Fig. 2. Schematic description of the part of the drainage system controlled in real time in the Seine-Saint-Denis project.

forecast inflows calculated by rainfall–runoff models (using the measured and forecast hyetographs on each catchment), generates control strategies for the next 6 h with a time step of 15 min.

The cost function to be minimized expresses the following three hierarchical control goals:

(1) overflow reduction,

(2) preservation of water quality in the recreational part of the Pont-Yblon basin,

(3) limitation of control device operations.

The cost C of a control strategy was defined as

$$\begin{aligned} C = {} & \sum_{i=1}^{i=3} \alpha_i \int_0^T \mathrm{deb}_i^2\,(t)\mathrm{d}t \\ & + \alpha_4 \sum_{k=1}^{k=N} [\sup(V_{\mathrm{PY},k}-15\,000,\ 0)]^2 \\ & + \beta_1 \sum_{k=1}^{k=N} (Q_{\mathrm{BM},k}-Q'_{\mathrm{BM}})^2 + \beta_2 \sum_{k=1}^{k=N} (Q_{\mathrm{PY},k}-Q'_{\mathrm{PY}})^2 \end{aligned}$$

where $\mathrm{deb}_i(t)$ is the overflow at time t ($t \in [0, T]$) and at overflow point i ($i = 1{,}3$); $V_{\mathrm{PY},k}$ is the water volume in the Pont-Yblon basin, at time step k; $Q_{\mathrm{BM},k}$ and $Q_{\mathrm{PY},k}$ are the outflows at time step k from the Blanc-Mesnil and Pont-Yblon basins, respectively; Q'_{BM} and Q'_{PY} are the default maximum values for the outflows from these two basins; N is the total number of time steps considered by the decision procedure; and $(\alpha_i)_{i=1,4}$ and $(\beta_i)_{i=1,2}$ are weights reflecting the relative importance of the different goals.

The decision procedure chooses the strategy for which C, calculated with the forecast inflow, is minimum. Once the real inflow is known, the value of C,

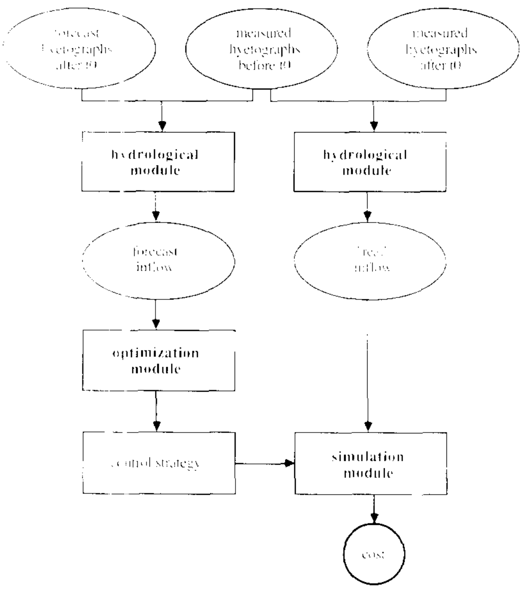

Fig. 3. Sequence of computer programs used to compute the cost of a forecast, in terms of the user-defined objective function.

calculated with the observed inflow, can serve to evaluate the applied control strategy.

As it is based on the sewer managers' concerns, the cost function (with changed sign) can be considered as an approximation of the user's utility function. It can therefore be used to compare rainfall forecasts, applied to that particular RTC scheme.

Figure 3 shows the links between the different programs that were used, in this particular case, to compute the function g introduced in the previous section. The hydrological module calculates a forecast inflow for the decision time t_0, taking into account historical rainfall (before t_0) and forecast rainfall. On this basis, the optimization module determines an optimal control strategy for the following 6 h. A flow simulation module then calculates the cost of this strategy applied to the 'real' inflow, i.e. the simulated inflow corresponding to the measured hyetographs. The final result of these calculations is the cost of the chosen strategy, which, by definition, is high for bad forecasts, and low for good forecasts, in terms of the users' concerns.

Consequently, the $\Delta H/H$ criterion (averaged over the four catchments) can be compared with the value of cost C. For this purpose, two historical events have been selected, which have caused flooding in the observed area. For each event, two forecast situations have been chosen: one at the beginning and the other in the middle of the event. Two hundred forecasts have been randomly and independently generated for each forecast situation. The results of

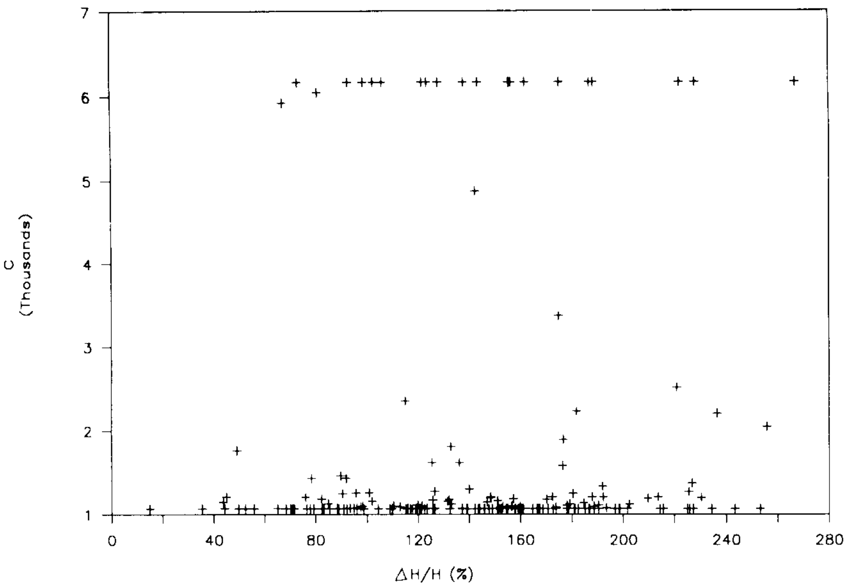

Fig. 4. Plot of $\Delta H/H$ vs. C for 200 randomly generated rainfall forecasts.

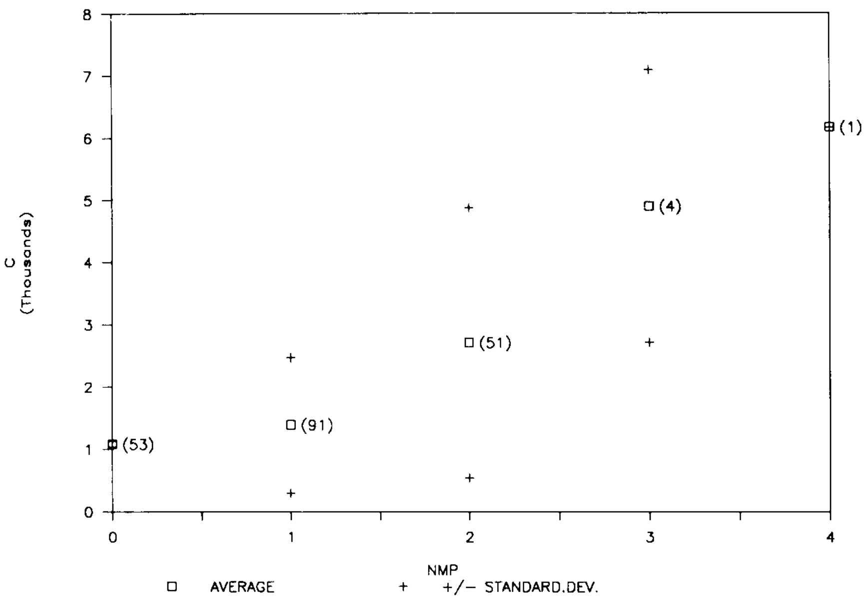

Fig. 5. Average of cost C for each NMP value, $\pm$ 1 SD (the numbers of cases are indicated in parentheses).

comparing $\Delta H/H$ and C for one set of 200 forecasts (Fig. 4) show no obvious relationship between the two criteria. Similar results have been observed for the other three sets of 200 forecasts.

A more thorough analysis of these results showed that only overestimations of $> 150\%$ for the two upstream catchments (Morée and Garonor), and underestimations of $> 50\%$, for the two downstream catchments (Bourget and Croult) result in 'bad' control strategies, i.e. produce costs which are $> 20\%$ higher than that produced by a perfect forecast. Thus, two of the three hypotheses for the use of the $\Delta H/H$ criterion mentioned above are not verified in this case, namely the equivalence of over- and underestimations, and the representativity of the mean error for several catchments.

This finding has led to the definition of the application-specific criterion NMP ('Nombre de Mauvaises Prévisions' — number of bad forecasts), as the number of cases including either an overestimation of $> 150\%$ upstream, or an underestimation of $> 50\%$ downstream. As four catchments are considered in our example, the values of NMP range from zero to four for excellent and very bad forecasts, respectively.

Figure 5 shows, for the same situation as in Fig. 4, the mean cost, $\pm$ 1 SD, corresponding to each NMP value. For that situation as well as for the three others, the NMP appears to be far better correlated than $\Delta H/H$ with the costs of control strategies. Hence, it can reasonably be considered as a better criterion for this application of the forecasts.

DEFINITION OF METEOROLOGICAL PARAMETERS

From convective cells to large frontal rainbands, precipitation systems show a great variety of sizes, durations, and behaviours in development and motion (e.g. Harrold and Austin, 1974; Browning, 1985). These factors can be expected to determine the predictability of the rainfall systems, i.e. the possibility of forecasting the amount of precipitation for a specific area and a specific time interval. Many arguments support this assertion:

(1) Point forecasts are more sensitive to errors arising from the determination of advection, as rainfall areas are smaller (Newton and Frankhauser, 1964).

(2) If T is the maximum life-time of the precipitation systems perceived by radar at time t_0, no forecast can be made beyond $t_0 + T$. Moreover, the deformation of a rainfall area in time, which can be expected to be related to its duration, has been recognized as a major source of forecast error (Bellon and Austin, 1984; Denoeux et al., 1989).

(3) All radar rainfall forecasting techniques are based on the estimation of the motion of rainfall areas. This motion may be too slow to be measured accurately; it may also differ greatly from the traditional linearity assumption, or contain a more complex propagation term which may not easily be modelled (Boucher and Wexler, 1961).

(4) The dynamics of the formation of precipitation systems determine not only all their characteristics, but also the speed at which they grow and decay.

For example, the presence of wind shear has been shown to play an important role in the evolution of a convective cell into the form of an ordinary cell, or a supercell (Chalon, 1978).

These remarks guided us in the search, among the many possible candidates, for potential predictors of the quality of radar rainfall forecasts. Thirty-three parameters were defined to be calculated from either radar data or rawinsonde data.

Definition of parameters from radar data

Precipitation systems appear on a PPI radar image as groups of precipitation areas. These areas can be described mathematically as 'echoes', i.e. sets of connected pixels with a reflectivity above some threshold t (Einfalt et al., 1989). Two categories of echoes have been defined:

(1) lower echoes, composed of pixels with a reflectivity > 13 dBZ; dBZ is the unit of the radar reflectivity Z, defined as the summation per unit volume of the sixth power of the diameter of spherical water drops

$$\mathrm{dBZ} = 10 \log_{10} \left(\frac{Z\,\mathrm{mm}^6\,\mathrm{m}^{-3}}{1\,\mathrm{mm}^6\,\mathrm{m}^{-3}} \right)$$

(2) upper echoes, for which the threshold t has been set so that there remain at least 1500 pixels of reflectivity $\geqslant t$.

These definitions had been introduced previously in the SCOUT forecasting method (Einfalt, 1988; Einfalt et al., 1989).

TABLE 1

Parameters defined from radar data

Names	Units	Definitions
N (or N')	—	Number of lower (or upper) echoes
S (or S')	km^2	Average area of lower (or upper) echoes
I (or I')	mm h^{-1}	Average intensity of lower (or upper) echoes
σ (or σ')	mm h^{-1}	Standard deviation of intensities, for lower (or upper) echoes
ST	km^2	Total area covered by lower (or upper) echoes
E	—	Surface-weighted mean of elongations (quotients of the principal moments of inertia), for lower (or upper) echoes
ΔN (or $\Delta N'$)	—	Absolute variation of N (or N')
ΔST (or $\Delta ST'$)	km^2	Absolute variation of ST (or ST')
ΔI (or $\Delta I'$)	mm h^{-1}	Absolute variation of I (or I')
ΔN_R(or $\Delta N_R'$)	%	Relative variation of N (or N')
ΔST_R (resp. $\Delta ST_R'$)	%	Relative variation of ST (or ST')
ΔI_R (resp. $\Delta I_R'$)	%	Relative variation of I (or I')
C	—	Maximum cross-correlation coefficient between the two images used in the forecasting process
VA	km h^{-1}	Advection speed of lower echoes, estimated by the cross-correlation method

The echo concept served as a basis for the definition of 26 parameters (see Table 1), which can be classified in two categories.

(1) geometrical parameters (N, S, I, σ, ST, E, and their equivalents for upper echoes), which describe echo structure, both in space and reflectivity;

(2) evolution parameters (ΔN, ΔST, ΔI, ΔN_R, ΔST_R, ΔI_R and their equivalents for upper echoes, plus C and VA), which are related to the transformation of the rainfall field.

Definition of parameters from rawinsonde data

The influence of temperature, humidity and wind profiles in the troposphere on the formation of precipitation systems has been studied by meteorologists for a long time (see e.g. Triplet and Roche, 1977; Wallace and Hobbs, 1977). These profiles can be obtained from the upper-air measurements regularly performed by meteorological services. Interpretation of the full profiles is often necessary to make accurate weather forecasts, but other parameters can also be calculated and provide some valuable information. Table 2 shows the definition of seven of these parameters, which have been chosen as features to predict forecast reliability. Wind speed (V) in the 3500–5000-m layer is well correlated with the advection speed of convective cells (Battan, 1973). The Showalter index (SI) is used by meteorologists to evaluate the probability of storm occurrence (Jarmuzynski, 1978), and the Convective Available Potential Energy ($CAPE$), the Convective Inhibition (CIN) and the Energy Index (EI) can be used to measure the degree of instability, as well as to discriminate between different types of convective systems (Bluestein and Jain, 1985). Lastly, the vertical wind shear (WS) and the bulk Richardson number (Ri) have

TABLE 2

Parameters defined from rawinsonde data

Names	Units	Definitions
V	km h^{-1}	Pressure-weighted mean of wind in the 3500–5000-m layer
WS	km h^{-1}	Pressure-weighted vertical wind shear
SI	K	Difference between the temperature at 500 hPa and the temperature of an air parcel originally at 850 hPa and lifted at 500 hPa
CIN	J kg^{-1}	Net work per unit mass required to lift a negatively buoyant air parcel from the surface to the level of free convection
$CAPE$	J kg^{-1}	Energy per unit mass gained by an air parcel which rises from the level of free convection to the equilibrium level
EI	J kg^{-1}	Change in kinetic energy per unit mass of an air parcel moving from the level of greatest wet bulb potential temperature in the lowest 150 hPa of the sounding, to the 400-hPa level
Ri	—	Ratio of the total energy available because of buoyancy to the total energy available from vertical wind shear

been related to the number and size of convective cells (Chalon, 1978; Bluestein and Jain, 1985).

All of these predictors can easily be calculated from sounding data, for example using the algorithms described by Stackpole (1967).

GENERATION OF DECISION RULES

Having defined a criterion of forecast quality and a list of predictors, the next step was to assemble a data set large enough for the use of statistical (discriminant analysis) or heuristic (machine learning) methods, to generate decision rules relating the values of the predictors to forecast reliability.

Such a data set has been constituted from archive data of the C-band radar in Trappes (near Paris), operated by the French National Weather Service. Among the 46 available rainfall events, 619 30-min periods were selected, for which there were at least 10 $1.6 \times 1.6\,km^2$ areas, < 100 km away from the radar, where the rainfall depth, measured by radar, was > 1 mm.

At the beginning of each of these 30-min periods, a forecast 30 min ahead has been simulated, using two different forecasting methods:

(1) a cross-correlation method (CROS), very similar to that described by Austin and Bellon (1974), and

(2) an echo-tracking method (SCOUT), described by Einfalt (1988) and Einfalt et al. (1989).

Because it is only applicable to cases of heavy rainfall over the Seine-Saint-Denis catchments, the above-defined NMP criterion could not be applied to evaluate all of these forecasts. It was therefore generalized so as to take into account the forecast rainfall depths over 10 areas, chosen at random among the $1.6 \times 1.6\,km^2$ areas where the measured rainfall depth was > 1 mm. The generalized NMP criterion (denoted NMP in this section) was defined as the number of cases, out of 10, including either an overestimation of $> 150\%$, or an underestimation of $> 50\%$. This new index (the possible values of which range from zero for very good forecasts to 10 for very poor forecasts) keeps the main properties that were found in the first part of the study, to be desirable, i.e.

(1) it concentrates on the evaluation of forecasts over areas where rainfall was hydrologically significant;

(2) it does not take into account errors below certain thresholds, and

(3) it distinguishes between over- and underestimations, by applying different thresholds.

Because point rain-gauge measurements cannot be easily compared with the areal measurements provided by radar (e.g. Collier, 1986), and considering the sparseness of rain-gauge data, radar measurements were used as a reference to assess the quality of the forecasts. The bias introduced by this choice can reasonably be expected to affect only the absolute performance values, and not the relative values which are needed in this study.

For each 30-min period, the above-defined parameters have been calculated from radar data, and from rawinsonde measurements performed in Trappes two to four times a day by the French National Weather Service.

The final data set thus consisted of 619 'examples', each composed of the values of the 33 predictors, and of the NMP values of the forecasts performed by CROS (NMPCR) and SCOUT (NMPSC). This data set has been randomly divided into a learning set, and a test set of equivalent size.

To make the learning process easier, only three levels of forecast quality have been considered, corresponding to

(1) NMP $= 0$
(2) $0 <$ NMP ≤ 2
(3) NMP ≥ 3

Two learning methods have been used:

(1) a statistical method: Bayes discriminant analysis, with hypotheses of normality and of equality of the covariance matrices in the different classes (Anderson, 1958);

(2) an heuristic method: the generation of production rules from decision trees (Quinlan, 1986, 1987).

Let us suppose that an observation is described by a vector of attributes $X = (X_1, ..., X_n)$, and that it has to be classified in one of two classes C_1 and C_2. The first method is based on the Bayes rule:

$$d(X) = C_1 \Leftrightarrow P(1 \mid X) \geq P(2 \mid X)$$

where $d(X)$ is the predicted class, and $P(i \mid X)$ is the conditional probability that an observation belongs to class C_i, if its representation is X.

If classes C_1 and C_2 are assumed to contain multivariate normal populations with equal covariance matrices, the Bayes rule becomes

$$d(X) = C_1 \Leftrightarrow -(X - \mu_1)' \sum{}^{-1} (X - \mu_1) + (X - \mu_2)' \sum{}^{-1} (X - \mu_2) \geq 2 \ln\left(\frac{p_2}{p_1}\right)$$

where $\sum$ is the matrix of variances and covariances of each population, μ_i is the vector of means of the ith population ($i = 1,2$), and p_i is the prior probability of class $C_i (i = 1,2)$.

The second learning procedure that was used in this study is a classical technique in Artificial Intelligence. Compared with the former method, it has the advantages of requiring no particular hypotheses, and of producing more intuitive results. The basic algorithm starts by partitioning the learning set with respect to the most discriminatory variable (or test). Following Quinlan (1986), a measure of entropy was used to assess the discriminatory power of each variable. Each subset can then be partitioned in a similar way, unless it is too small or contains only examples of one class. The process repeats iteratively until no subset can be divided. The result is a recursive structure called a decision tree, where each leaf denotes a class, and each interior node denotes a test. Such a tree can be transformed into a set of production rules (Quinlan, 1987).

TABLE 3

Decision rules for the recognition of good forecasts (NMP = 0) and bad forecasts (NMP > 2), generated by discriminant analysis (DA), and by induction of decision trees (IDT)

Rule	e (%)	e' (%)	r (%)	Method
R1: NMPCR = 0 $\Leftrightarrow$ 0.958 $\ln(S)$ + 1.611 10^{2} VA − 6.779 $\geq$ 0	32.3	44.5	27.4	DA
R2: NMPCR = 0 $\Leftrightarrow$ ($I \leq$ 0.725) and ($C >$ 0.769)	32.3	44.5	27.4	IDT
R3: NMPCR > 2 $\Leftrightarrow$ 4.608 C + 0.794 $\ln(ST)$ + 2.238 10^{-2} VA − 11.656 < 0	25.8	42.6	39.4	DA
R4: NMPCR > 2 $\Leftrightarrow$ ($C \leq$ 0.599) or ($EI \leq -$ 516.995) or ($S \leq$162.795)	18.4	42.6	56.8	IDT
R5: NMPCR > 2 $\Leftrightarrow$ $C \geq$ 0.599	23.9	42.6	43.9	IDT
R6: NMPSC = 0 $\Leftrightarrow$ 0.728 $\ln(ST)$ − 7.278 $\geq$ 0	40.7	40.7	0	DA
R7: NMPSC = 0 $\Leftrightarrow$ ($C >$ 0.771) and ($SI >$ 2.67)	24.8	40.7	39.1	IDT
R8: NMPSC > 2 $\Leftrightarrow$ 2.750 C + 0.690 $\ln(ST)$ − 0.625 $\ln(S')$ − 5.1 < 0	37.7	48.4	22.1	AD
R9: NMPSC > 2 $\Leftrightarrow$ ($ST \leq$ 35445) and [($VA \leq$ 40.729) or ($VA >$ 57.422)]	31.6	48.4	34.7	IDT
R10: NMPSC > 2 $\Leftrightarrow$ ($ST \leq$ 35445) and [($SI >$ 7.585) or ($Ri >$ 128.445)]	33.9	48.4	30.0	IDT

e: error rate; e': error rate of random classification; r: error rate reduction.

The rules obtained for the recognition of 'good' forecasts (NMP = 0) and 'bad' forecasts (NMP $\geq$ 3) are presented in Table 3, together with the corresponding error rate (e) and, for comparison, the error rate (e') of a random estimation of reliability, based only on the frequencies of the different classes in the learning set. From these results, it is apparent that:

(1) except in one case, the error rate reduction owing to the application of the decision rules, compared with chance, is significant (22.1–56.8%);

(2) the best rule for the recognition of good forecasts by SCOUT (R7) is more efficient than the best rule for the recognition of bad forecasts by this same method (R9), whereas the inverse tendency is observed in the case of CROS (R4 is better than R1 and R2);

(3) only three rules, out of 10, make use of parameters calculated from upper-air measurements; a closer analysis of the results tends to show that these parameters do provide some valuable predictive information, but that this information is, to some extent, duplicated by the information provided by the radar data alone;

(4) the heuristic method gave consistently better results than the statistical one, probably because the very restrictive hypotheses imposed by this last method were only approximately verified in this case.

Additionally, the combination of two rules R and R′, for the recognition of 'good' and 'bad' forecasts, respectively, allows for a more precise determination of forecast reliability using the trivial rules contained in Table 4. Tables 5 and 6 show the results of such a classification in three classes, obtained by the combination of R2/R4 and R7/R10, respectively. What should be noticed is the very small percentage of cases (4.5 and 3.2%) where a 'bad' forecast has been classified as 'good', or a 'good' forecast has been classified as 'bad'.

The best rules obtained in this study will soon be integrated in the real-time control system in the Seine-Saint-Denis project, where the SCOUT forecasting method has been operational since May 1988. They will be refined progressively, as new radar data and forecast results become available and are analysed by the learning algorithm.

TABLE 4

Combination of two decision rules R and R′ for a more precise determination of forecast reliability

	R	
	NMP = 0	NMP > 0
R′		
NMP $\leq$ 2	NMP = 0	0 < NMP $\leq$ 2
NMP > 2	?	NMP > 2

TABLE 5

Performance table of rules R2 and R4 combined to determine the reliability of CROS forecasts

	Forecast			
	NMP = 0	0 < NMP ≤ 2	NMP > 3	?
Observed				
NMP = 0	47	51	7	0
0 < NMP≤ 2	35	58	17	0
NMP > 3	7	26	62	0

TABLE 6

Performance table of rules R7 and R10 combined to determine the reliability of SCOUT forecasts

	Forecast			
	NMP = 0	0 < NMP ≤ 2	NMP > 3	?
Observed				
NMP = 0	40	42	0	0
0 < NMP ≤ 2	19	76	1	2
NMP > 3	10	92	26	2

CONCLUSIONS

The study presented in this paper has led, on the one hand, to very specific decision rules, and, on the other hand, to a general methodology.

The decision rules obtained are very specific because they are linked to:

(1) a given region (~ 150 km around Paris), with particular climatic and microclimatic conditions (influence of urbanized areas, of the Seine river, etc.), and

(2) a given operational context: the real-time control of a large urban drainage system, with well-defined objectives.

Another main interest of the study is nevertheless the generality of the methodology which has been designed; it can be transferred to other hydrological applications of radar rainfall forecasts, provided two conditions are met:

(1) the decision process must have been formalized and the user's utility function approximated, so that a criterion of forecast quality can be defined rigorously;

(2) radar data corresponding to a representative set of rainfall events must be available, to allow for the identification of statistically significant relations between the quality of the forecasts and characteristics of rainfall.

This approach can be extended in many directions. For example, cases where no utility function can be defined (because the user is not able to choose between any two sets of consequences) could be studied. New meteorological parameters could also be defined, e.g. from results of simulations performed by mesoscale atmospheric models.

The transfer of this approach to a completely different context, even outside hydrology, is also an interesting possibility.

ACKNOWLEDGEMENTS

This study was supported by the French Ministry of Construction (Plan Urbain contract 84-31395-00-223-75-01), and by the Seine-Saint-Denis regional authority. The authors thank Pr A. Frérot for providing the simulation and optimization programs, as well as Mr. V. Pircher and Mr. P. David, of the French National Weather Service, for many discussions that helped to develop some of the ideas expressed in this paper.

REFERENCES

Anderson, T.W., 1958. An Introduction to Multivariate Statistical Analysis. Wiley, New York, NY.

Austin, G.L., 1985. Application of pattern recognition and extrapolation techniques to forecasting. ESA J., 9: 147–155.

Austin, G.L. and Austin, L.B., 1974. The use of radar in urban hydrology. J. Hydrol., 22: 131–142.

Austin, G.L. and Bellon, A., 1974. The use of digital weather radar records for short-term precipitation forecasting. Q. J.R. Meteorol. Soc., 100: 658–664.

Battan, L.J., 1973. Radar Observation of the Atmosphere. University of Chicago Press, Chicago, IL.

Bellon, A. and Austin, G.L., 1978. The evaluation of two years of real-time operation of a short-term precipitation forecasting procedure (SHARP). J. Appl. Meteorol., 17 (12): 1778–1787.

Bellon, A. and Austin, G.L., 1984. The accuracy of short-term radar rainfall forecasts. J. Hydrol., 70: 35–49.

Bluestein, H.B. and Jain, M.H., 1985. Formation of mesoscale lines of precipitation: severe squall lines in Oklahoma during the spring. J. Atmos. Sci., 42 (16): 1711–1732.

Boucher, R.J. and Wexler, R., 1961. The motion and predictability of precipitation lines. J. Meteorol., 18: 160–171.

Browning, K.A., 1985. Conceptual models of precipitation systems. ESA J., 9: 157–180.

Carpenter, K.M. and Owens, R.G., 1981. Use of radar data for forecasting rain. COST 72 Workshop, Seminar on Weather Radar, pp. 161–181.

Chalon, J.P., 1978. Dynamique des nuages convectifs; influence du cisaillement de vent sur les cumulonimbus. La Météorologie, VIe série, No. 13: 17—26.

Ciccione, M. and Pircher, V., 1984. Preliminary assessment of very short-term forecasting of rain from single radar data. Proc. Nowcasting Symposium, Norrköping, Sweden.

Collier, C.G., 1978. Objective forecasting using radar data: a review. Research Rep. 9, Meteorological Office Research Laboratory, Malvern, U.K., 16 pp.

Collier, C.G., 1986. Accuracy of rainfall estimates by radar. Part II: Comparison with raingauge network. J. Hydrol., 83: 225–235.

Collier, C.G., Cole, J.A. and Robertson, R.B., 1980. The North-West Weather Radar Project; the establishment of a weather radar system for hydrological forecasting. Proc. Int. Symp. Hydrological Forecasting, Oxford, IAHS Publ. 129: 31–40.

Damant, C., Austin, G.L., Bellon, A., Osseyrane, M. and Nguyen, N., 1983. Radar rain forecasting for wastewater control. J. Hydraul. Div., ASCE, 109(2): 293–297.

Dee Weather Radar and Real-Time Hydrological Forecasting Project, 1977. Final Rep., Central Water Planning Unit, Dept. of the Environment, Reading.

Denoeux, T., 1989. Fiabilité de la prévision de pluie par radar en hydrologie urbaine. Ph.D. Thesis, Ecole Nationale des Ponts et Chaussées, Paris, 246 pp.

Denoeux, T., Einfalt, T. and Jacquet, G., 1987. Introduction of a weather radar as an operational tool. Proc. 4th Conf. Urban Storm Drainage, Topics in Urban Storm Water Quality, Planning and Management, IAHR, Lausanne, pp. 289–290.

Denoeux, T., Einfalt, T. and Jacquet, G., 1989. On the evaluation of radar rainfall forecasts. Proc. Int. Symp. Hydrological Applications of Weather Radar, University of Salford.

Einfalt, T., 1988. Recherche d'une méthode optimale de Prévision de pluie par radar en hydrologie urbaine. Ph.D. thesis, Ecole Nationale des Ponts et Chaussées, Paris, 189 pp.

Einfalt, T. and Denoeux, T., 1989. Never expect a perfect forecast. Proc. Int. Symp. Hydrological Applications of Weather Radar, University of Salford.

Einfalt, T., Denoeux, T. and Jacquet, G., 1990. A radar rainfall forecasting method designed for hydrological purposes. J. Hydrol., 114: 229–244.

Ellsburg, D., 1954. Classic and current notions of "measurable utility". Econ. J., LXIV: 528–556.

Elvander, R.C., 1976. An evaluation of the relative performance of three radar echo forecasting techniques. Proc. 17th Conf. Radar Meteorology, Seattle, WA, pp. 526–532.

Frérot, A., 1987. Procédures d'optimisation des consignes de gestion d'un réseau d'assainissement automatisé. Ph.D. Thesis, Ecole Nationale des Ponts et Chaussées, Paris, 296 pp.

Frérot, A. and Jacquet, G., 1989. Contrôle-commande et gestion en temps réel des réseaux d'assainissement. Ann. Ponts et Chaussées, No. 49: 3–25.

Frérot, A., Jacquet, G. and Delattre, J.M., 1986. Elaboration des consignes optimales de gestion d'un réseau d'assainissement. In: Eau et Informatique, Actes du Colloque Organisé par l'ENPC. Presses de l'ENPC, Paris.

Harrold, T.W. and Austin, P.M., 1974. The structure of precipitation systems; a review. J. Rech. Atm., 8: 41–57.

Huff, F.A., Changnon, S.A. and Vogel, J.L., 1980. Convective rain monitoring and forecasting system for an urban area. 19th Conf. Radar Meteorology, Miami, FL, pp. 56–61.

Jarmuzynski, M., 1978. Etude comparative de quelques critères de stabilité pour une situation convective particulière. La Météorologie, VIe série, No. 14: 61–68.

Montgolfier (de), J. and Bertier, P., 1978. Approche Multicritère des Problèmes de Décision. Editions Hommes et Techniques, Coll. AFCET, Suresne.

Newton, C.W. and Fankhauser, J.C., 1964. On the movement of convective storms, with emphasis on size discrimination in relation to water-budget requirements. J. Appl. Meteorol., 3 (6): 651–668.

Quinlan, J.R., 1986. Induction of decision trees. Machine Learning, 1: 81–106.

Quinlan, J.R., 1987. Generating production rules from decision trees. Proc. 10th Int. Joint Conf. Artificial Intelligence, pp. 304–307.

Schilling, W. (Editor), 1989. Real time control of urban drainage systems, the state-of-the-art. IAWPRC Task Group on Real Time Control of Drainage Systems. IAWPRC, London, 84 pp.

Schilling, W. and Petersen, S.O., 1987. Real time control of urban drainage system; validity and sensitivity of optimization techniques. 1st IAWPRC Int. Symp. Systems Analysis in Water Quality Management, London.

Stackpole, J.D., 1967. Numerical analysis of atmospheric soundings. J. Appl. Meteorol., 6: 464–467.

Triplet, J.P. and Roche, G., 1977. Météorologie Générale. Ecole Nationale de la Météorologie, Paris, 317 pp.

Tsonis, A.A. and Austin, G.L., 1981. An evaluation of extrapolation techniques for the short-term prediction of rain amounts. Atmosphere–Ocean, 19(1): 54–65.

Von Neumann, J. and Morgenstern, O., 1953. The Theory of Games and Economic Behaviour. Princeton University Press, Princeton, NJ.

Wallace and Hobbs, 1977. Atmospheric Science; An Introductory Survey. Academic Press, New York, NY.

Wilson, J.W., 1966. Movement and Predictability of Radar Echoes. Technical Memorandum 28, NSSL, Norman, OK, 30 pp.