Data and Information Management for Integrated Hydrological and Meteorological Networks

Adarsha P. Pokhrel[1] and Dilip K. Gautam[2]
[1]Director General, [2]Hydrologist Engineer, HMGN Department of Hydrology and Meteorology, Kathmandu, Nepal

Abstract
Hydrological and meteorological data are widely used in the development of water resources projects (such as hydropower, irrigation, water supply, navigation, and flood control) as well as for agricultural planning and infrastructure development. Moreover, forecasts are essential to the day-to-day operations of civil aviation, the safety of mountaineering expeditions, and the mitigation of natural disasters such as floods and landslides. These applications require that the data are not only reliable but that they are available in a user-friendly format that implies effective management of data and information. The implementation of data and information management cycles consists of first assessing the data and information requirements of the users and then delivering the information to them in a useable format. The latter implies the acquisition of data and information, transfer of data, quality control, the building and maintenance of databases, the retrieval and analysis of data, the dissemination of data and information, and the evaluation of data and information application. This paper describes in brief how these different steps can be implemented for the effective management of data and information in integrated hydrological and meteorological networks.

Introduction

Water is one of the most important natural resources and is one of global concern. Any change in its availability, quality, or quantity impacts living beings all over the world. The dynamics of water on a global scale can be monitored through the hydrological cycle, which contains complementary hydrological and meteorological information – a thorough knowledge of both of which is required for proper water management. Water management is the interaction of human beings with the natural water cycle; thus effective water management can only be achieved by taking an integrated approach. Integration at the national level needs to balance the country's requirements, that is the demands for and uses of water, on the one hand, with its scientific and technical capability for physical water management on the other. Technical capability implies the combined use of hydrological and meteorological information systems. Different approaches are possible, and while some countries choose to maintain separate hydrological and meteorological institutions others choose to combine these two disciplines under one roof, as in the case of Nepal's Department of Hydrology and Meteorology (DHM).

Hydrological and meteorological data are the basis for the development of water resource projects, infrastructure development, and agricultural planning. For example, hydrological data such as river flow forecasts are essential for the management of flood plains, for the operation of reservoirs and for the mitigation of flood disasters. Meteorological forecasts are essential for civil aviation, for agricultural planning, and for the safety of mountaineering expeditions, among others. An integrated network of hydrological and meteorological stations provides the essential primary data; once collected this data needs to be properly stored, archived, processed, analysed, and made available to the end users in easily digestible formats by an effective data management process.

An effective data and information management process consists of the following (Figure 1):

1. assessment of data and information requirements,
2. acquisition of data and information,
3. transfer of data,
4. quality control,
5. construction and maintenance of databases,
6. data retrieval and analysis,
7. dissemination and application of data and information,
8. evaluation of information application, and
9. re-assessment of data requirements.

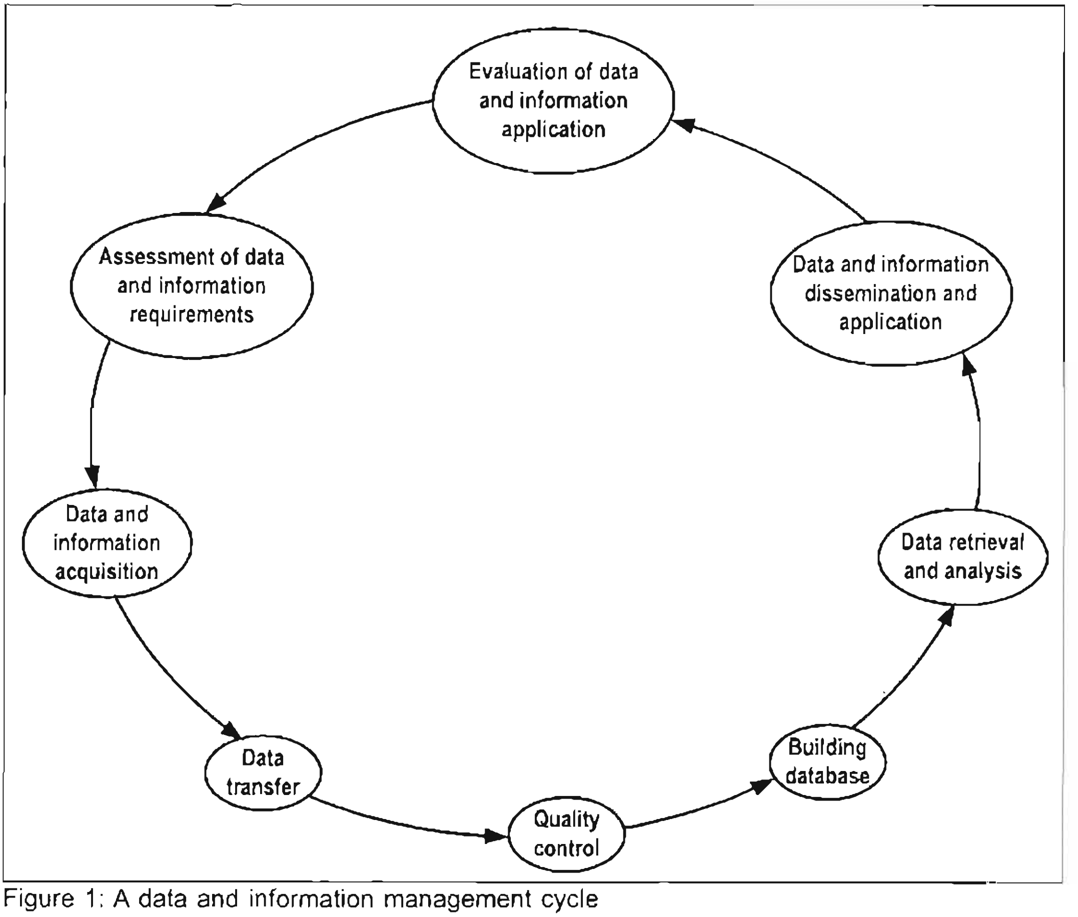

Figure 1: A data and information management cycle

Assessing Data and Information Requirements

River stage (water level), discharge, and sediment data are required for hydropower, water supply, and irrigation projects. The data requirements for flood forecasting depend on the type of forecasting required and the forecasting models used. Either water level or river flow can be forecast; however, water level forecasting is relatively easier than flow forecasting. Water level forecasting makes use of water level time series data from upstream stations and tributaries, together with rainfall data to establish gauge-to-gauge correlations; this, combined with calculations of time of travel, allows forecasts to be made of the water level at downstream stations.

Flow forecasts are usually produced by models. The simplest method of forecasting flow is to use an equation for the stage-discharge relationship after the river stages (water levels of different sections) have been forecast. There are three types of modelling paradigms in hydrology, namely, physically-based distributed models, lumped conceptual models, and empirical models.

The physically-based distributed models and lumped conceptual models use data for the catchment characteristics, which include information on the river section, initial soil moisture content of the upper and lower zone of soil, river stages, discharges, rainfall, temperature, and evapotranspiration among others. Empirical models need only rainfall and runoff/water level time series data.

All of these models have in common that they use rainfall data as a model input; hence, rainfall forecasting is intimately linked with flood forecasting. Rainfall forecasting is a complex process that includes the tracking and forecasting of frontal systems, disturbances, waves, cyclones, and storms. It is also part of the more general weather forecasting that uses observations of wind speed and direction, atmospheric pressure, cloud cover, and visibility.

Data Acquisition

Hydrological and meteorological data are collected using land-based observation systems; meteorological data are also collected from space-based observation systems such as satellites and radar. Since flood forecasting models use both hydrological and meteorological data, an integrated network of hydrological and meteorological stations is needed to record regular observations.

River stage (water level) data are collected by constructing simple staff gauges or using water level recorders. Data loggers can be used to acquire data in electronic form. Automatic weather stations equipped with data loggers provide meteorological data in electronic form.

Historical data and other information about the stations can also be obtained from periodic publications. The data on catchment characteristics, such as catchment area, length of hillslopes, and main channels, can be obtained from maps.

Exchange and Transfer of Data and Information

The conventional methods of transferring observational data from recording stations to forecasting centres were wireless sets, telephone, fax, mail, and meetings (face-to-face).

With the development of the Internet and the World Wide Web (www), data and information can be exchanged through electronic networks. Internet tools include e-mail, mailing lists, newsgroups, and chat rooms; www-based information sharing tools include file servers (based on file transfer protocol (ftp) and network file system (nfs)), document servers (based on hypertext transfer protocol (http)), newsgroups (to share e-mails), and bulletin boards. Other technologies include tele or video conferencing tools such as Net Meeting, CU-SeeMe, and MBone based systems.

Internet based tools provide the opportunity for the real-time data transfer needed for real-time forecasting; whereas historical data can be transferred to the forecasting centre by slower means. Historical data are needed to calibrate forecasting models.

Data Quality Control

The quality of rainfall data depends on the instrument used, its installation, its site characteristics, and its operation by a trained observer. Common errors include, incorrect readings, misplaced decimal points, copying errors, arithmetic errors, correct readings entered on wrong days, and systematic errors such as leaky gauges. The quality of river flow data depends on the quality of field data collection and proper establishment of stage-discharge relationships.

Technicians **assure data quality** using various methods including screening, graphical plotting, double-mass analysis, comparison with nearest-neighbour stations, and regional mapping and correlation analysis.

Data screening can be achieved in one of two ways. The first method uses maximum and minimum limits to identify outliers that do not fall within given maximum and minimum limits; the second identifies outliers that don't fall within the range of 'mean ± n * standard deviation', where n = 1, 2, 3. Simple graphical plots can also help to identify any abnormalities in time series data. Double mass analysis is performed using one 'test station' and a number of 'base stations'. The cumulative value of the mean of the base stations is plotted on the X-axis and cumulative value of the test station is plotted on the Y-axis. The plot is useful for checking the consistency of the data. Correlation analysis and the plot of the relation curve can be employed to compare data with neighbouring stations. Cross-correlation function plots of rainfall and water level time series also provide valuable information about the time lag, which is important for estimating the forecast lead time.

Continuity of data flow is achieved by filling in any missing time series data using various methods such as arithmetic averaging, the normal ratio method, the inverse distance method, and the relation curve method.

The arithmetic mean method estimates missing data from a particular station by calculating the arithmetic mean of the data from selected neighbouring stations for the same period. The formula used is

$$P_x = \left(\frac{1}{M}\right)\sum_{i=1}^{M} P_i$$

where, P_i is the data for the neighbouring station i for the duration of missing data at station x, and M is the number of selected neighbouring stations.

The formula used in the normal ratio method is

$$P_x = \left(\frac{N_x}{M}\right)\sum_{i=1}^{m} P_i N_i$$

where, P_i = the data of the neighbouring station i for the duration of missing data at station x, N_i is the normal value of the data from station i for the duration of missing data at station x, and M is the number of selected neighbouring stations.

In the inverse distance method, a weight value is assigned to the data from selected neighbouring stations. The weight value is inversely proportional to the distance between the station with missing data and the selected station. The formula used is:

$$P_x = \sum_{i=1}^{m} P_i W_i \Big/ \sum_{i=1}^{m} W_i$$

where, $W_i = 1/D_i^2$ and D_i is the distance from the missing data station to the selected station i.

The relation curve method is a statistical method for filling in missing data in which the missing data is correlated with the same data from a selected station or various combinations of selected stations. The equation for the relation curve that provides the best fit, as indicated by the highest value of r^2, is used to generate values for the missing data.

Construction and Maintenance of Databases

Observational data and other information about the catchment can be stored in a database. There are many database programmes and decision support systems for flood forecasting, including the **WINHYDRO** program that DHM has used written in Visual Basic for the purpose of storing, analysing, and publishing hydrological data (Figure 2). WINHYDRO is essentially a hydrological database program and has very limited data analysis facilities. Recently, DHM has developed another program, the **Hydrometeorological Information System** (HIS), to store, analyse, and publish both hydrological and meteorological data (Figure 3). The HIS program is an application for data management, time series analysis, spatial analysis, and production of hydrological and meteorological atlases. It uses the Environmental Systems Research Institute's (ESRI) geodatabase concept, together with the ORACLE program as the relational database management system (RDBMS), which can store both the time series and spatial data. ESRI's Spatial Database Engine (SDE) plays the role of middleware between the data repository and the front-end application that has been developed in Visual Basic. ESRI's desktop products (ArcView 8.1 and ArcInfo 8.1) have been heavily customised with ArcObjects programming to develop the interface for the spatial data. Other ArcGIS extensions such as 3D Analyst, Spatial Analyst, and Geostatistical Analyst have also been used in HIS. STATISTICA, another software package, has been used for a variety of statistical analyses (for details see Kumar et al. 2002).

The HIS consists of four major modules: the System Administration, Master Data Definition, Time Series, and Spatial modules. The System Administration module is used to define different levels of access rights to users and thus assures data integrity and security. The Master Data Definition module allows the user to define master data and

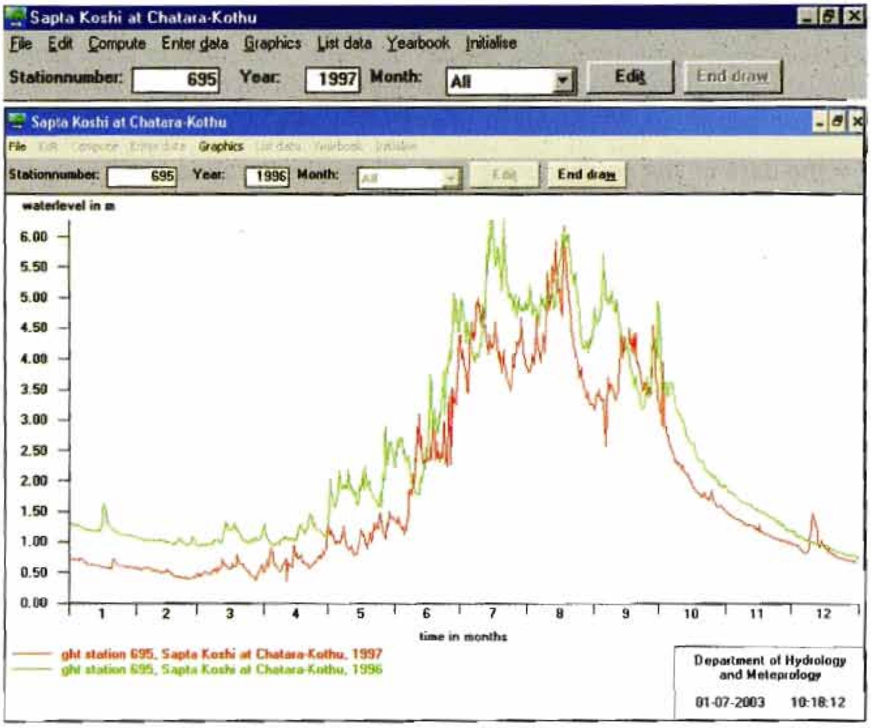

Figure 2: WINHYDRO, main menu (top), and water level hydrograph plot (bottom)

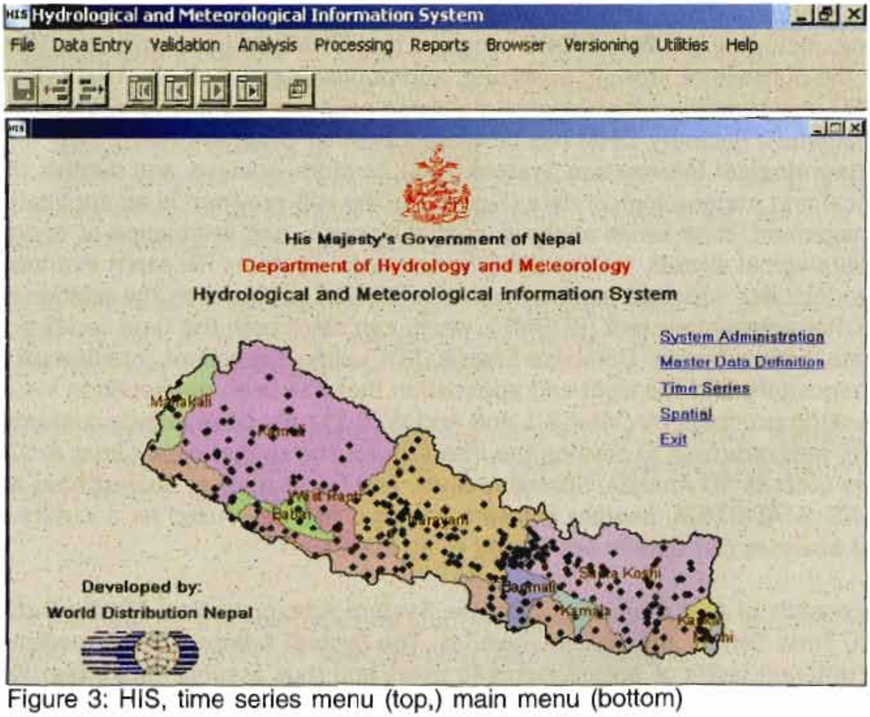

Figure 3: HIS, time series menu (top,) main menu (bottom)

their management. The Time Series module consists of a variety of tools for time series data processing and analysis. It has a menu-driven system for data entry, validation, data processing, analysis, and data browsing. The Spatial module is used to capture, analyse, and visualise hydrological and meteorological phenomena with simulated geography. It provides the tools for basin analysis as well as drainage analysis. Tools are also available to delineate watershed boundaries, to calculate slope, aspect, basin lag, and peak flow using various methods. Similarly, isohyete, isotherms, and isobars can also be generated and the average basin precipitation can be calculated.

Recently, DHM has adopted a conceptual rainfall-runoff model, ADM (A Distributed Model), for flood forecasting and the simulation of rainfall-runoff processes (Figure 4). ADM is effectively a combination of the ARNO model (Todini 1996) and the XINANJIANG model (Zhao et al.1980). ADM has two distinct parts: the first calculates the soil-level water balance and expresses the balance between the moisture content of the soil and the incoming precipitation; the second calculates the transfer to the basin outlet by estimating the outgoing quantities (evapotranspiration, surface runoff, interflow, and baseflow). The runoff is then subjected to a transfer operation which represents the transfer to the network channels along the hillslopes, and transfer to the basin outlet along the channel network. The model is characterised by 11 parameters, 7 of which identify the water balance component at soil level and 4 of which identify

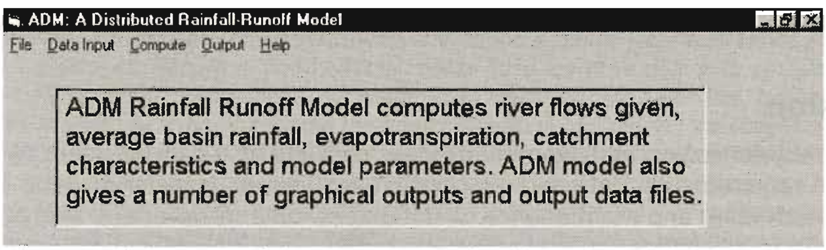

Figure 4: Main menu of the conceptual rainfall-runoff model ADM

the transfer component to the basin outlet. The water balance component consists of two parts representing the upper and lower zones of the soil respectively.
One program commonly used as a decision support system for real-time flood forecasting is **MIKE Flood Watch** (see Paudyal G.N., this volume). This program combines an advanced time series database with the MIKE 11 hydrodynamic modelling and real-time forecasting system (MIKE 11 FF), and works in the ArcView geographical information system (GIS) environment. MIKE Flood Watch is applied operationally at flood forecasting centres throughout the world – it consists of three modules: a database management module, a map-oriented data editing and presentation module, and a module for model preparation, simulation, and post-processing of results.

The MIKE Flood Watch management module is used to establish and georeference the real-time stations in ArcView GIS display. Having established the database, a data conversion module links the data blocks within MIKE 11 FF, the ArcView GIS environment, and the database. Results can be imported into the MIKE 11 GIS flood-mapping module and overlaid with depth/area inundation maps for further analysis.

The ArcView GIS display of the different stations enables the user to select stations for data editing and presentation directly from the GIS map. By direct access to MIKE 11 FF, results can be visualised in combination with other stations and maps. The simulation module includes tools to set up and execute MIKE 11 FF. A bulletin-editor is used to formulate and prepare flood warnings and other flood information for dissemination, for example on the Internet.

Data and Information Dissemination

The analysed data and information are disseminated to the public and users in various forms such as the following.

1. **Bulletins and graphs:** Daily weather forecasts are issued by television and radio transmissions throughout the country.
2. **Reports:** Other information is disseminated in the form of publications and reports, such as the Hydrological Records of Nepal, Precipitation Records of Nepal, Climatological Records of Nepal, and others.
3. **Maps**: Inundation maps, flood hazard maps, hydrological atlases, and meteorological atlases are published and made available.
4. **Internet:** www based dynamic interactive documents can be used for publishing data and information.
5. **Diskettes and CD-ROMs:** Data can also be obtained for research, insurance claims, and other purposes in electronic form on diskettes and CDs from DHM for a nominal fee.

Conclusion

The data and information management process consists of the assessment of data and information requirements, data and information acquisition, transfer of data, quality control, construction and maintenance of databases, data retrieval and analysis, data and information dissemination and application, and evaluation of the data and information application process. This paper has discussed each of these steps and described how they are carried out for hydrological and meteorological data at the Department of Hydrology and Meteorology of Nepal.

References

Kumar, M.; Budhathoki, N.R.; Kansakar, S. (2002) *Hydro-meteorological Information System: Implementation of Multi-user Geodatabase Concept in Nepal.* Paper presented at the 23rd Asian Conference on Remote Sensing, November 25-29, 2002, Kathmandu, Nepal

Todini, E. (1996) 'The ARNO Rainfall-Runoff Model'. In *Journal of Hydrology*, 175: 339-382

Zhao, R.J.; Zuang, Y.L.; Fang, L.R.; Liu, X.R; Zhang, Q.S. (1980) *The XINANJIANG model, Hydrological forecasting.* Proc. of the Oxford Symposium, April 1980, IAHS Publ. No. 129. Wallingford (UK): International Association of Hydrological Sciences