Data Reconciliation on Complex Hydraulic System: Canal de Provence

Jean-Luc Deltour[1]; Eric Canivet[2]; Franck Sanfilippo[3]; and Jacques Sau[4]

Abstract: A data reconciliation module, based on the measurements from the hydraulic network, has been recently developed and implemented in the supervisory system of the Société du Canal de Provence (SCP). The software has initially been used daily to check the measured flow on the main canal. The data reconciliation occurs just after the measurement process. The measurement network on the hydraulic system includes many sensors subject to failure or deviation and is spread over a huge area. In addition, discharge and volume measurements in open-channel hydraulic networks are characterized by large uncertainties. The objective of the data reconciliation is to take advantage of information redundancy on a system to make a cross-check of real-time measurements. By using this information redundancy, a data reconciliation module allows detection of inconsistent measurements and measurement deviations and provides corrected values whether the initial measurements are valid, biased, or invalid. A derived consequence is better scheduling of the maintenance of sensors. The results are corrected values for measured variables and proposed values for nonmeasured quantities. A statistical analysis of the results is performed. This analysis allows evaluation of the uncertainties attached to the estimated flows and volume values. It allows also detecting invalid measurements and drift of sensors and making decesions about which maintenance operations to perform.

DOI: XXXX

CE Database subject headings: Canals; Hydraulic networks; Discharge measurements; Data analysis; France.

Introduction

The Canal de Provence is situated in the southeast of France. It supplies water to 80,000 ha of farmland, 110 towns and villages, and 400 industries. The water distribution strategy is user-oriented and resorts neither to rotations nor to any sort of priority allocation.

All the main structures are monitored and remotely controlled from the general center by a supervisory control and data acquisition (SCADA) system, including a module of dynamic regulation that provides automatic and permanent control of canal flows and safety systems.

The measurement network on the hydraulic system includes many sensors spread over a huge area, and they are subject to failure or deviation. In addition, discharge and volume measurements in open-channel hydraulic networks are characterized by large uncertainties. To overcome this kind of problem in process control industrial applications, data reconciliation is often used (Kratz-Bousghiri et al. 1996; Narasimham and Jordache 2000). The objective of the data reconciliation is to take advantage of information redundancy on a system to cross-check of real-time measurements.

By using this information redundancy, a data reconciliation module allows detection of inconsistent measurements and measurement deviations and provides corrected values whether the initial measurements are valid, biased, or invalid. A derived consequence is better scheduling of the maintenance of sensors.

A data reconciliation module, which is based on the measurements from the hydraulic network, has recently been developed and implemented in the supervisory system of the Société du Canal de Provence (SCP) (Canivet 2002). The application presented in this paper concerns daily volumes on the canal.

In the field of process control, the data reconciliation is a part of the general state estimation or reconstruction problem in dynamic systems. The usual basic tool when dealing with this problem is the Kalman filter (Chui and Cben 1998; Maquin and Ragot 2000). However, in our application, which is a daily based application, dynamic effects can be neglected; therefore we are led to a simplified version of the general approach (Ragot et al. 1992), applicable to static models.

The paper first presents the theory of the Canal de Provence data reconciliation application. The basic model is a hydraulic network with a series of nodes corresponding to balance equations (inflows, outflows, and storage). Constrained data reconciliation is used to satisfy the nonnegativity of the hydraulic variables and the mass-balance relations. The results are corrected values for measured variables and proposed values for nonmeasured quantities. A statistical analysis of the results is performed. This analysis allows evaluation of the uncertainties attached to the

[1]Water Control Engineer, Société du Canal de Provence, Le Tholonet BP 100, F-13603 Aix-en-Provence, cedex1, France. E-mail: jean-luc.deltour@canal-de-provence.com

[2]Consulting Engineer, Cogix, 68, Cours Albert Thomas, F-69371 Lyon cedex 08, France. E-mail: eric.canivet@cogix.fr

[3]Hydraulic Engineer, Société du Canal de Provence, Le Tholonet BP100, F-13603 Aix-en-Provence, cedex1, France. E-mail: franck.sanfilippo@canal-de-provence.com

[4]Professor, Université Claude Bernard Lyon 1, 43 Bd du 11 novembre, F-69622 Villeurbanne Cedex. E-mail: jacques.sau@univ-lyon1.fr

Note. Discussion open until November 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on January 16, 2004; approved on June 29, 2004. This paper is part of the ***Journal of Irrigation and Drainage Engineering***, Vol. 131, No. 3, June 1, 2005. ©ASCE, ISSN 0733-9437/2005/3-1–XXXX/$25.00.

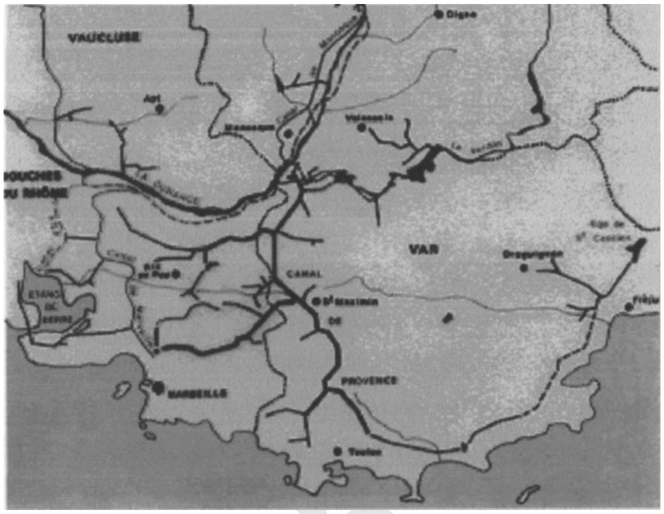

Fig. 1. Geographical location of Canal de Provence

estimated flows and volume values. It also allows detecting invalid measurements and drift of sensors, and helps make decisions about the maintenance operations to perform.

Second, field examples are presented, including measured and reestimated flow values with their standard deviations, detection of invalid sensors, and maintenance operations performed. The data reconciliation occurs just after the measurement process and takes place in the decision process for diagnosis, identification, and control.

Data Reconciliation Model

Canal Description

The geographic location of Canal de Provence is shown in Fig. 1. The water is taken from the Sainte-Croix Dam at the Boutre intake. The Canal de Provence supplies the Provence and Côte d'Azur (PACA) region. One can distinguish three main areas: the Aix-en-Provence, Marseille, and the French Riviera areas.

Fig. 2 shows the diagram of the network with the location of measurements. Different categories of sensors are used to measure the discharges. This equipment has been calibrated by the following appropriate methods: gates formula coefficient through gauging and flow meters through platform calibration. A procedure is required to maintain the initial good quality of measurements.

Since redundancy among measurements exists, a data reconciliation procedure can be implemented. This redundancy comes from system equations linking the measured variables because of the existence of a model of the system.

The first step is to establish a model linking the variables in which we are interested. In this paper, we are interested in the daily mean discharges. These discharges are obviously linked through mass conservation relations. The basic assumption is that volume variation in the canal is negligible compared with daily discharge passing through the cross structure. This result is verified by considering the mode operation of the canal. These relations lead to a static model.

The Basic Model

Let us call Q_i the (unknown) true discharge on the measurement point number i and call $\mathbf{Q}$ the vector of these discharges. Let us also call $\mathbf{Q}_m$ the vector of measured discharges. Uncertainties occur in the measurements, so we write

$$\mathbf{Q}_m = \mathbf{Q} + \boldsymbol{\varepsilon} \tag{1}$$

where ε=random part expressing the uncertainties.

Mass conservation relations exist between the discharges, elements of $\mathbf{Q}$. These relations, once collected, lead to a matrix relation

$$\mathbf{MQ} = 0 \tag{2}$$

The number n of rows of the matrix $\mathbf{M}$ is equal to the number of such independent relations (balancing equations), shown as nodes on the graph of the canal in Fig. 2 (here, n=9), and the number of columns (28 in our case) is the dimension of $\mathbf{Q}$ (Ragot 1992). The elements of a row of $\mathbf{M}$ are −1, +1, or 0, depending on whether the corresponding discharge is an inflow, an outflow, or is not involved in the node relation relative to this row. For example, for the third and fourth rows of matrix $\mathbf{M}$, relative to Nodes 3 and 4, the only nonnull elements are, respectively

$$\mathbf{M}_{3,3} = -\mathbf{M}_{3,4} = -\mathbf{M}_{3,9} = 1 \tag{3}$$

$$\mathbf{M}_{4,4} = -\mathbf{M}_{4,5} = -\mathbf{M}_{4,6} = -\mathbf{M}_{4,7} = -\mathbf{M}_{4,8} = -\mathbf{M}_{4,24} = 1 \tag{4}$$

The other rows are built in the same way.

The preceding matrix relation has the advantage of being easily generalized. The values of the elements of the matrix $\mathbf{M}$ can be other than 0, 1, or −1; and the second member vector can be a nonnull vector $\mathbf{R}$ if special relations take place in the model. For example, an $\mathbf{R}$ element value could be nonnull if the relation corresponding to that row is not balanced because of a fixed inflow or outflow. We then rewrite our model relations in Eq. (1)

$$\mathbf{Q}_m = \mathbf{Q} + \boldsymbol{\varepsilon}$$

$$\mathbf{R} = \mathbf{MQ} \tag{5}$$

A statistical hypothesis must next be made on the uncertainty vector $\boldsymbol{\varepsilon}$.

Statistical Hypothesis and Consequences

The errors that occur in the measurement process are assumed to follow a normal law characterized by a zero mean and a variance σ_i^2. The measurement equipment is independent, so the vector $\boldsymbol{\varepsilon}$, which characterizes the experimental uncertainties, is classically assumed to be a Gaussian vector, of zero mean and having statistical independent components. Its covariance matrix $\boldsymbol{\Gamma}_\varepsilon$ is therefore diagonal but is not a multiple of the unity matrix, since the measurement points refer to different equipment with different accuracies. A variance σ_i^2 is then assigned to each measurement point, depending on its accuracy. We then have

$$\boldsymbol{\Gamma}_\varepsilon = \mathrm{diag}(\sigma_i^2) \tag{6}$$

The problem now amounts to finding the best estimate $\mathbf{Q}_e$ of $\mathbf{Q}$, knowing the measurements, in Eq. (1) under the equality constraint Eq. (2). This best estimate is the one that maximizes the likelihood which, in the Gaussian case, is also the one that minimizes the weighted least-square expression (Chui and Chen 1998; Freund and Wilson 2003)

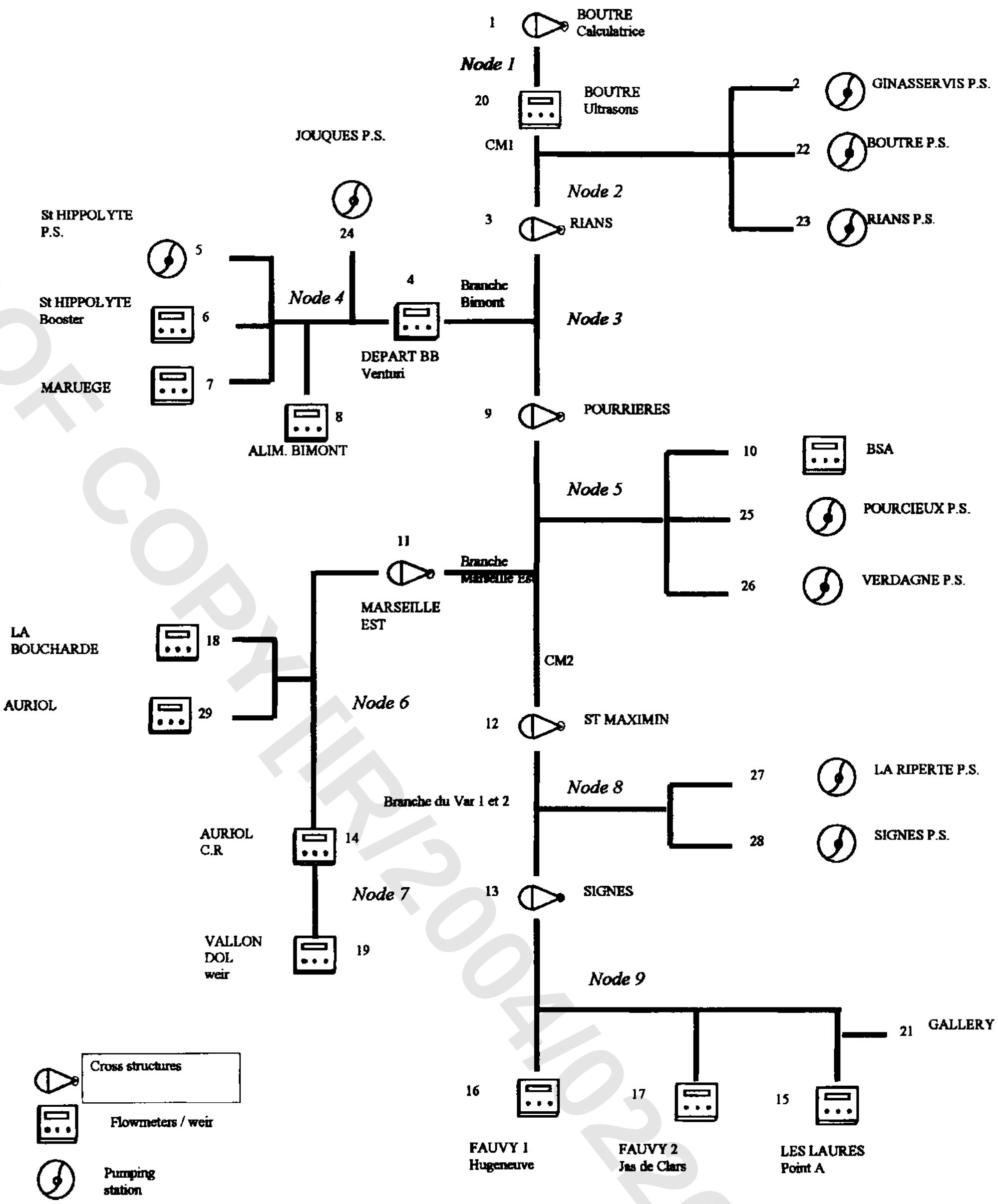

Fig. 2. Hydraulic diagram

$$\frac{1}{2}(\mathbf{Q}_m-\mathbf{Q})^t\boldsymbol{\Gamma}_\varepsilon^{-1}(\mathbf{Q}_m-\mathbf{Q}) \tag{7}$$

under the equality constraint in Eq. (2). This problem is a well-known problem of quadratic optimization (Gill and Murray 1981; Lawson and Hanson 1995), the solution of which takes the form

$$\mathbf{Q}_e=\mathbf{T}\mathbf{Q}_m+\mathbf{S}\mathbf{R} \tag{8}$$

with matrices

$$\mathbf{T}=\mathbf{I}-\boldsymbol{\Gamma}_\varepsilon\mathbf{M}^t(\mathbf{M}\boldsymbol{\Gamma}_\varepsilon\mathbf{M}^t)^{-1}\mathbf{M} \tag{9}$$

$$\mathbf{S}=\boldsymbol{\Gamma}_\varepsilon\mathbf{M}^t(\mathbf{M}\boldsymbol{\Gamma}_\varepsilon\mathbf{M}^t)^{-1} \tag{10}$$

In the following, we also need the deviations

$$d=\mathbf{Q}_m-\mathbf{Q}_e, \tag{11}$$

the covariance matrix of which is

$$\boldsymbol{\Gamma}_d=(\mathbf{I}-\mathbf{T})\boldsymbol{\Gamma}_\varepsilon(\mathbf{I}-\mathbf{T})^t, \tag{12}$$

and the reduced deviations

$$e_i=\frac{d_i}{\sqrt{(\boldsymbol{\Gamma}_d)_{ii}}}. \tag{13}$$

Hypothesis of Normality

The next section describes various supervising procedures. All of them are based on the normality of the variables on which the statistical tests are performed. This hypothesis is classically the usual one in this kind of problem. Verifying that these variables are, at least approximately, normal is interesting. However, the variables on which the tests are performed are not the discharge measurements themselves but are variables obtained after a number of calculations have been performed on the measurement set; and they are a linear function of the deviations d, or, equivalently,

Table 1. Normality Test Results for Reduced Deviations

Measurement points	Test result (95% level)	Critical probability
1	Yes	0.654
2	Yes	0.625
3	Yes	0.852
4	Yes	0.312
5	Yes	0.971
6	Yes	0.971
7	Yes	0.971
8	Yes	0.452
9	Yes	0.059
10	Yes	0.964
11	Yes	0.289
12	Yes	0.802
13	Yes	0.687
14	Yes	0.567
15	Yes	0.686
16	Yes	0.686
17	Yes	0.686
18	Yes	0.319
19	Yes	0.582
20	Yes	0.892
21	Yes	0.686
22	Yes	0.586
23	Yes	0.705
24	No	0.000
25	Yes	0.449
26	Yes	0.502
27	Yes	0.196
28	Yes	0.122
29	Yes	0.319

the reduced deviations e. One can then expect that, because of the central limit theorem (Wonnacott and Wonnacott 1990), the resulting variables are approximately normal, even if the measurements are not.

To verify their normality, reduced deviations have been calculated on a number of days with no measurement anomalies. Statistical Bera-Jarque normality tests (Judge et al. 1998; Mathworks 2002) have then been applied on the values obtained for each measurement point. Table 1 gives the results of the tests, at the 95% significance level, together with the critical probability. As can be seen, all measurement points except Measurement Point 14 (Jouques station) fulfill the requirements of the test. The Jouques station was almost not operating during that period. A normal probability plot (Mathworks 2002) is classically used to verify graphically the normality of a distribution. The points of the graph must be approximately aligned. Fig. 3 shows an example of a normal probability plot of the reduced deviations for a given measurement point. Our hypothesis of approximate normality is therefore confirmed.

Supervising Procedures

To supervise the discharge measurements in the canal system, five procedures have been implemented. The first one deals with global consistency of measurements, and the remaining four deal with the consistency of each measurement point taken individually. The threshold used in these tests were tuned during the initial implementation period to be error-sensitive but to avoid false alarms:

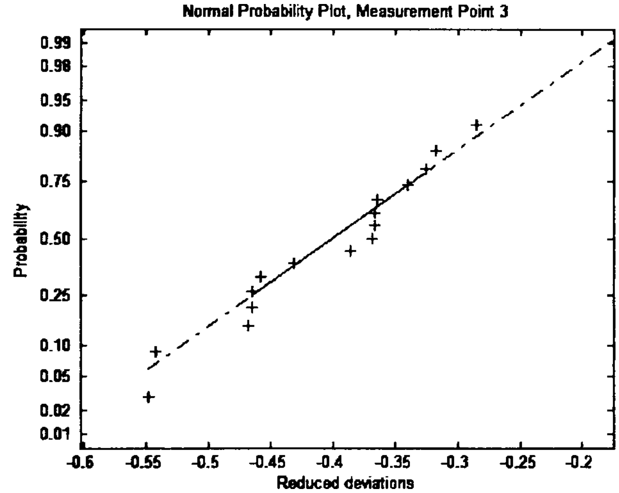

Fig. 3. Normal probability plot of reduced deviations

Global Consistency

For the global consistency of measurements, the vector of balance residues, **r**, is calculated:

$$\mathbf{r} = \mathbf{R} - \mathbf{M}\mathbf{Q}_m \tag{14}$$

This vector is an indicator of the overall statistical consistency of the measurements $\mathbf{Q}_m$, which, of course, do not fulfill the constraint relation in Eq. (2).

The covariance matrix of **r** is

$$\boldsymbol{\Gamma}_r = \mathbf{M}\boldsymbol{\Gamma}_\varepsilon\mathbf{M}^t$$

The overall consistency of the measurements is then characterized by the critical probability of the n degrees of freedom $\chi_n{}^2$ quantity (Freund and Wilson 2003):

$$\chi_n{}^2 = \mathbf{r}^t\boldsymbol{\Gamma}_r^{-1}\mathbf{r} \tag{15}$$

We have considered here a critical probability threshold of 0.05, which leads to a $\chi_n{}^2$ threshold of 17.

Individual Consistency of Measurement

For individual consistency of measurement, we investigate the difference, d, between measurements $\mathbf{Q}_m$ and estimations $\mathbf{Q}_e$ given by the relation in Eq. (11). In addition, the drift that may occur in the measurement between two tunings of equipment is estimated through the mean value of d ($\bar{d}$) calculated on an $n_d = 30$ days sliding window.

Four tests have been implemented:

- Detection of a value outside its confidence interval: Every day, the field deviation is calculated at each point, considering for each measurement the standard deviation σ_i:

$$d_c = \frac{(\mathbf{Q}_m - \mathbf{Q}_e) - \bar{\mathbf{d}}}{\sigma_i} \tag{16}$$

We consider here a critical probability of 0.006, which leads us to consider the measured value outside its confidence interval if the following condition is not fulfilled:

PROOF COPY [IR/2004/022644] 009503QIR

$$-2.8 < d_c < 2.8 \quad (17)$$

This test is able to detect sensor default for a given day.

- A jump of mean value of d on a sliding window: This second test compares the mean value of d calculated on an $n_1 = 10$ days sliding window $(\bar{d}_1)$ with the mean value of the same d calculated on an $n_2 = 30$ days sliding windows $(\bar{d}_2)$. The diagnosis indicator is

$$t = \frac{(\overline{d_1} - \overline{d_2})}{\sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2}{(n_1 + n_2 - 2)} + \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (18)$$

t follows an $n_1 + n_2 - 2$ (28 in our case) degrees of freedom Student's t distribution law (Wonnacott and Wonnacott 1990). An abnormal jump of the mean value has occurred with a 5% critical probability if the following condition is not fulfilled:

$$-2.05 < t < 2.05 \quad (19)$$

This test is able to detect a drift in sensor.

- A too large variance of d: This test compares the supposed variance of each measurement σ_i^2 with the variance σ_e^2 estimated over an $n_e = 12$ days period with

$$\sigma_e^2 = \frac{1}{n_e - 1}\sum_{j=1}^{n_e} (d_j - \bar{d})^2 \quad (20)$$

The indicator shown hereafter follows a chi-square distribution with $n_e - 1$ degrees of freedom (Wonnacott and Wonnacott 1990):

$$\chi_i^2 = \frac{(n_e - 1)\sigma_e^2}{\sigma_i^2} \quad (21)$$

The variance of d is then considered too large, with a 5% critical probability if the following condition is not fulfilled: $\chi_i^2 < 19.7$.
This test is able to detect an incoherency between the sensor behavior and its supposed accuracy.
- For some measurement points, we check also that the value of $\mathbf{Q}_e$ does not exceed known physical thresholds for these points.

The operator in charge of supervision at the General Control Center is alerted if one of these tests shows that measurements are not consistent. The operator then analyzes the results and requires, if necessary, that one technician go to the field to check out the equipment.

Result Analysis

Saint Maximin Cross Structure

The Saint Maximin cross structure consists of two gates that are used alternately on a monthly basis, for maintenance reasons. The discharge value is obtained from the gate opening and the upstream level, since the structure works in a free-flow condition. The software detects inconsistencies depending on the gate used. Fig. 4(a) displays the differences between reconciled and measured discharge values, together with the gates' status. Inconsistencies that are correlated with the gate used appear very clearly. A field investigation pointed out a slight error on the position measurement of the right-side gate. This error was small in comparison with the total gate opening. However, since the structure operates at a low rate, it had a large effect on the discharge calculation. Fig. 4(b) shows the actual situation after correction.

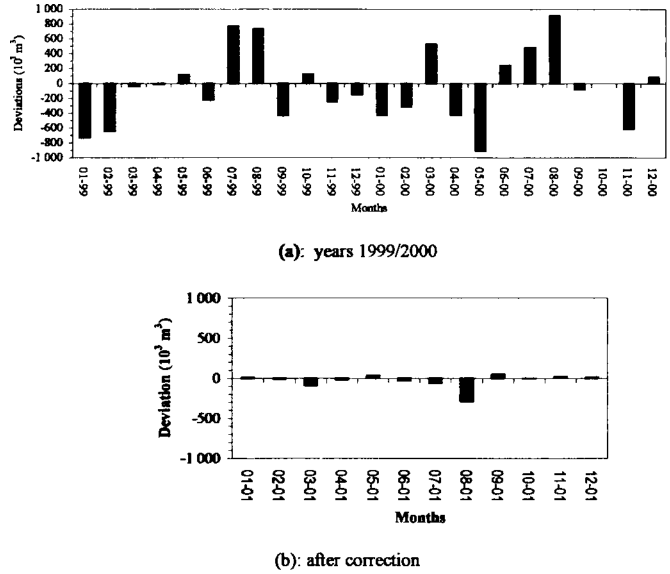

Fig. 4. Deviation between reconciled and measured values at St. Maximin cross structure: (a) years 1999–2000; (b) after correction

Marseille Est Cross Structure

At Marseille Est, the structure was initially designed for free-flow-condition operations. The software detected that the discharge calculated at that point was too low in comparison with the reconciled value. This outcome was confirmed by a gauging that was based on flow velocity measurement. We then diagnosed that the canal downstream had an effect on the gate, although the canal slope was high. The discharge calculation formula now takes into account the submerged condition.

Boutre Intake

Boutre is the main intake of the canal. The discharge at that point had previously been calculated from a gate formula established from a scale model 30 years ago. The accuracy of this system has

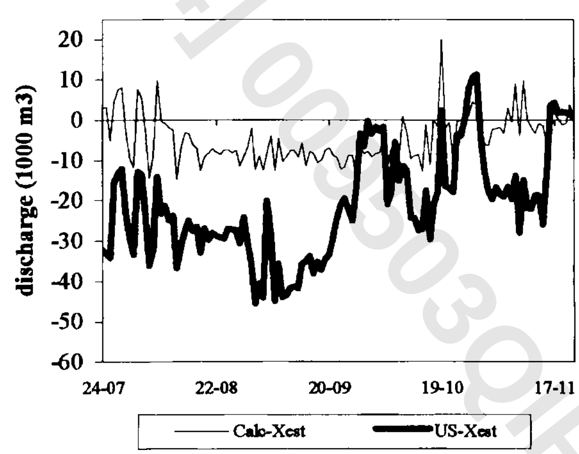

Fig. 5. Deviation between measured and reconciled values at Boutre cross structure

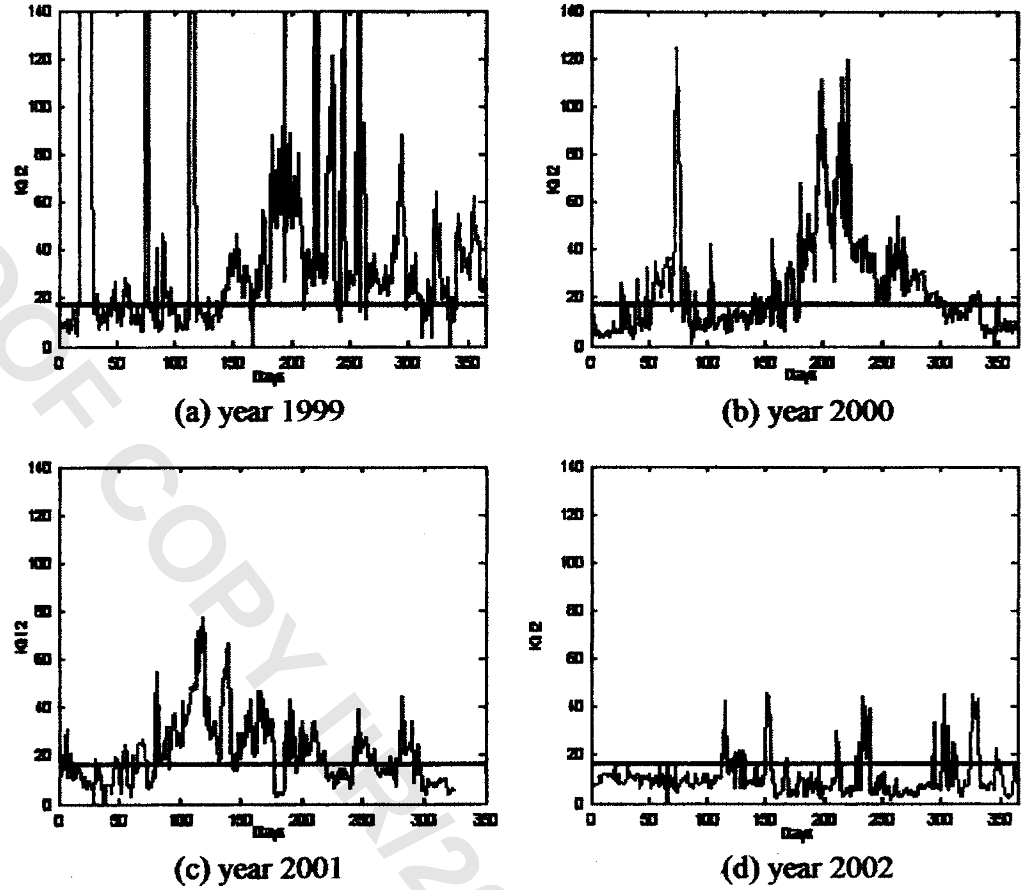

Fig. 6. Evolution of the chi-squared value

been improved by replacing the gate-opening measurement. The operating staff decided to add an ultrasonic flow meter located 150 meters (492 ft) downstream from that point.

The calibration of this sensor happened to be difficult because of a backwater effect. The software unexpectedly established that the older method gives values closer to the reconciled ones. Fig. 5 shows the ultrasonic-reconciled (US-Xest) and formula–based-reconciled (Calc-Xest) discharge deviations. A new calibration of the ultrasonic sensor is in progress.

Evolution of Global Consistency

The examples have illustrated the various improvements that have been undertaken on the canal. As previously indicated, the global chi-square distribution is a valid indicator of consistency of measurement. Figs. 6(a–d) display the evolution of the chi-squared values for the last four years; evidence of consistency enhancement is clearly seen in the consecutive graphs.

Conclusion

In this paper we have presented a data-reconciliation application on the Canal de Provence system. A canal is clearly a system where data reconciliation is a necessary and very helpful process. Indeed, discharge and volume measurements are spread over a large area and are characterized by large uncertainties. However, because of relations existing among the hydraulic variables within the system, redundancy exists among these measured values. Therefore, we can take advantage of this last characteristic to improve the reliability of the hydraulic measurements. After an initial calibration of the measurement points, the use of data reconciliation associated with appropriate tests represents a good day-to-day tool to help operating staff maintain the measurement quality and availability. After an initial implementation period, the daily application has been running on the Canal de Provence system for more than two years, giving good results and confirming the robustness of the approach. It allowed the operation and maintenance department to enhance the quality of measurement. The global chi-squared test represents a good indicator that assures that the quality of archived values is sufficient independently of initial default in some measurement points. The next step would clearly be the insertion of data reconciliation in the control software. Because of the command time step of the control process, a quarter of an hour, this would imply using a dynamic model of the canal. This project is actually in progress at the Canal de Provence.

Notation

The following symbols are used in this paper:

- $\mathbf{d}$ = vector of deviation between measured and estimated discharge;
- e_i = reduced deviation;
- $\mathbf{M}$ = balancing equations matrix;
- n = number of nodes in the hydraulic network;

$\mathbf{Q}$ = true discharge vector (vector of all Q_i);
$\mathbf{Q}_e$ = best estimated discharge vector;
Q_i = true discharge on measurement point i;
$\mathbf{Q}_m$ = measured discharges vector;
$\mathbf{R}$ = vector of fixed intakes at nodes;
$\mathbf{r}$ = vector of balance residues;
$\mathbf{\Gamma}_*$ = covariance matrix of subscript variable *;
$\boldsymbol{\varepsilon}$ = measurement uncertainties vector;
σ_e^2 = estimated variance at one point; and
σ_i^2 = variance of measurement error at point i.

References

Canivet, ■. (2002). "Réconciliation et validation des données sur un système hydraulique complexe, le Canal de Provence." PhD thesis, Université Lyon 1.

Chui, C. K., and Chen, G. (1999). *Kalman filtering with real time applications*, Springer-Verlag, Berlin.

Freund, R., and Wilson, W. (2003). *Statistical methods*, Elsevier, New York.

Gill, E., Murray, W., and Wright, M. H. (1999). *Practical optimization*, Academic Press, New York.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lutkepohl, H., and Lee, T.-C. (1998). *Introduction to the theory and practice of econometrics*, Wiley, New York.

Kratz-Bousghiri, ■., Nuninger, W., and Kratz, F. (1996). "Fault detection in stochastic dynamic systems by data reconciliation." *Engineering Simulation*, 13, 837.

Lawson, C. L., and Hanson, R. J. (1995). *Solving least-square problems*, SIAM, Philadelphia.

Maquin, D., and Ragot, J. (2000). *Diagnostic des systèmes linéaires*, Hermès, Paris.

Mathworks. (2002). *Statistical toolbox user guide*, Mathworks, Natick, Mass.

Narasimham, S., and Jordache, C. (2000). *Data reconciliation and gross error detection: An intelligent use of process data*, Gulf Professional Publishing, New York.

Ragot, D., et al. (1992). *Validation de données et diagnostic*, Hermès, Paris.

Wonnacott, T., and Wonacott, R. (1990). *Introductory statistics*, Wiley, New York.

PROOF COPY [IR/2004/022644] 009503QIR