CNR-UNICA-USTO

Stay Report Document

„Quality Control of Agro-alimentary Products“

Document Ref.: CUU-US-RD-002

Issue: 1

Prepared by: **Ahmed Nabil BELBACHIR** Date: **18 July 2003**

Contents

Index of Figures

Index of Tables

1 Introduction

1.1 Purpose of the document

The purpose of this document is to summarize the topics discussed between Ahmed Nabil BELBACHIR and researchers from the University of Cagliari under the supervision of Prof. Alessandra FANNI.

This document is split into four parts:

1. Section 2 describes a proposition to how to develop the project „Quality Control of Food“
2. In Section 3, preliminary results for the British competition regarding the Flood Forecast Simualtion are given
3. In Section 4, Minutes of the attended Meetings are provided.

1.2 Definitions, Acronyms and Abbreviations

Acronyms and Abbreviations	Description
CCA	Canonical Correlation Analysis
CNR	Centro Nazionale di Ricerca
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
HW	HardWare
ICA	Independent Component Analysis
LDA	Linear Discriminant Analysis
MSE	Mean Square Error
PCA	Principal Component Analysis
RMSE	Root Mean Square Error
SW	SoftWare
TBD	To Be Defined
UNICA	UNIversità degli studi di Cagliari
USTO	University of Science and Technology of Oran

Definition	Description
NN	Neural Network
QC	Quality Criteria

1.3 References

1.3.1 Applicable Documents

AD001	Yanbo Huang et al	Food Quality Quantiyation and Process Control
AD002	J. Jackson	A User's Guide to Principal Components
AD003	R.O. Duda et al	Pattern Classification and Scene Analysis
AD004	R. Gittins	Canonical Analysis-Review with Applications in Ecology
AD005	Aapo Hyvärinen et al	Independent Component Analysis
AD006	B. Schölkopf.	Support Vector Machine and kernel methods
AD007	M. Vetterli & J. Kovacevic	Wavelets and Subband Coding

1.3.2 Reference Documents

RD001	Technical Brief	http://www.itdg.org/html/technical_enquiries/docs/quality%20control.pdf
RD002	W. Sweldens	The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets
RD003	J.Shapiro	Embedded Image Coding Using Zerotrees of Wavelet Coefficients, ”

2 Quality Control of Agro-alimentary Products

2.1 Motivation

The production of high quality food products has the widest scope for scientific research and development[AD001]. Research can encompass everything from biotechnology and statistics to engineering and computing. Although the industry has come up with advanced technology and techniques for any food processing business such as online bio-sensors and inspections systems, these techniques still suffer from expensive time consuming or expensive and the results of quality control tests does not really help to save money in the long run. In general, the quality control procedures used should be as simple as possible and only give the required amount of information (too little information means the test has not done its job, too much information and management decisions may be delayed or confused).

The quality of foods or ingredients can be measured in different ways but one popular method is to describe 'quality attributes', see Table 1. A specification can then be written and agreed with the supplier or seller, which lists the quality attributes that are required in a food. An example of a quality specification for tomatoes intended for processing into a paste or leather is shown in Table 1[RD002].

Attribute	Accept	Reject
Colour	Orange/red	More than 10% green
Size	Any	-
Shape	Any	-
Damage - splitting - insect - mould	Less than 5% Less than 5% None	More than 5% More than 5% Any evidence of mould
Hardness	Soft to oversoft	More than 10% hard

Table 1. Quality attributes for tomatoes

2.2 Basic Quality Control Flow

Figure 1 depicts the basic concept for a quality control of an agro-alimentary product.

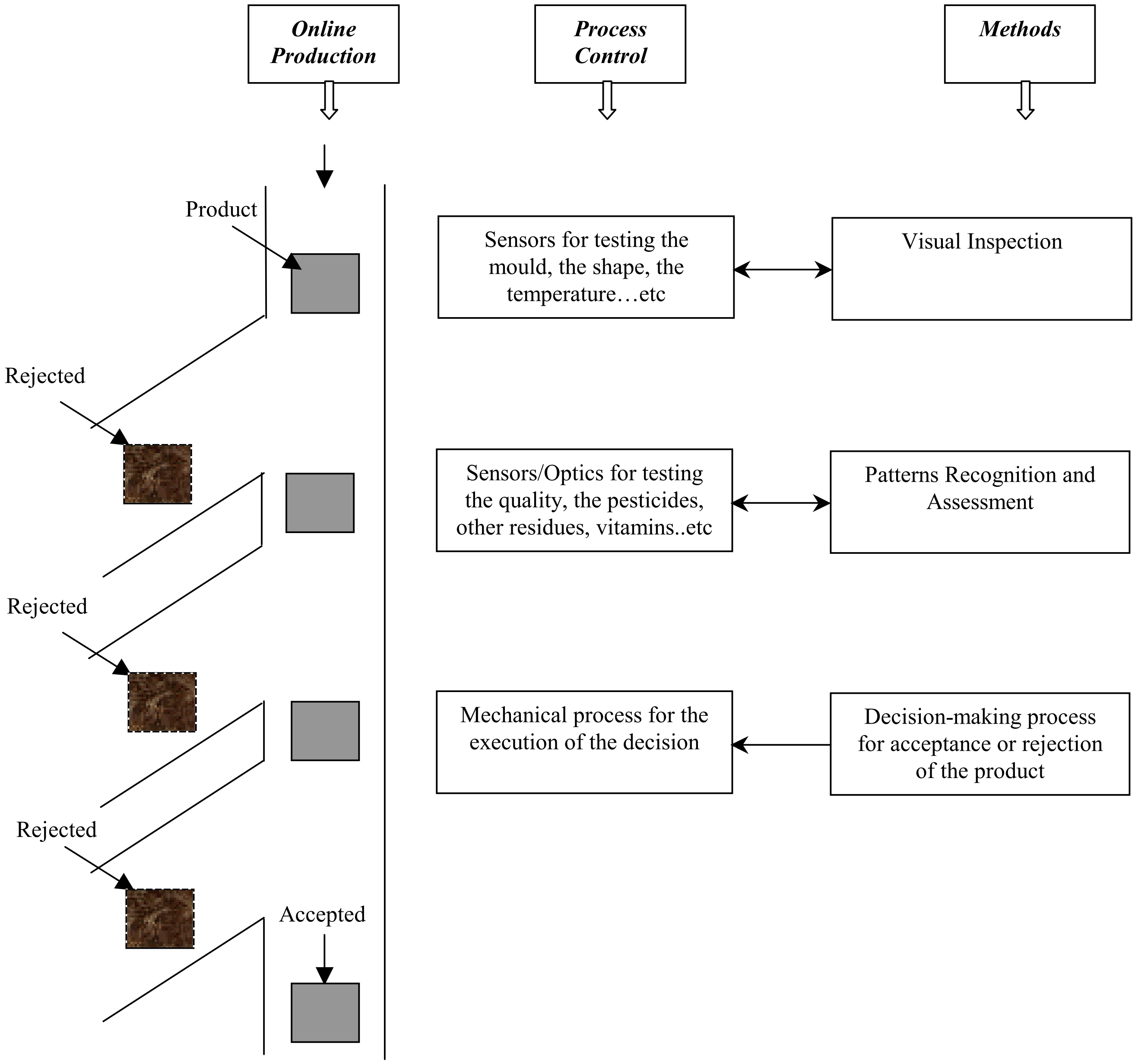


Figure 1. Basic Concept of Food Quality Control

2.3 Online Production

It consists of the agro-alimentary products whom the quality should be evaluated. The speed of the production should be compatible with the control process such that the product testing and analysis could be accurately performed.

It is clear that the parameters required to assess the quality of the product has to be provided by the expert of the product(e.g. enzymes, vitamins, shape, mould…etc). In the following sections, we call these parameters the quality criteria

2.4 Process Control

This part represents the instrumentation part where a set of measurements has to be performed and to be transmitted to the control PC (analysis machine) for the diagnosis of the measurements used one of the methods described in Section 2.5.

It consists of the a set of sensors and vision systems (at different wavelength) to measure the predefined parameters required for qualtity assessment of the agro-alimentary product. Indeed, the parameters to measure depend on the type of product to control the quality. Usually, the sensors required for the quality control fall under the categories listed below.

Sensors for process conditions: Temperature; Pressure, mass and level; Flow; Density, viscosity and particle size. Moisture, fat and protein content: Equilibrium moisture; Microwave sensors; Infrared sensors.

Biosensors which use an immuno-reaction of molecular recognition and photrometric sensing for transduction to an electron signal. Enyzme sensors for online measurement urea for improving the dairy herd nutritional management. Biosensor for the detection of pesticide residues; hardware may include a fluorescent immunoassay, an infrared laser diode and a high sensitivity photodiode.

Chemical and biochemical sensors: Spectroscopy; Nuclear magnetic resonance; pH and ion-selective electrodes; Immobilised enzyme sensors. Size, color and turgidity: Height and length; Color; Turbidity. Sorting, vision systems and foreign body detection: Sorting by color; Vision systems such as CCD camera or far infrared camera for the foreign body detection and the reliability of measurement.

2.5 Methods

In this Section a set of methods which may be used for the processing of the taken measurements are listed. The choise of one of these methods depends on the specifications of the product to contrtol the quality

2.5.1 Visual Inspection

This part consists mainly of the evaluation of the visual aspects of the product. Classication methods may be used such as backroung canceling method to separate the product from the background. Then, the visual aspect of the object may be assessed using a dynamic comparison (no absolute such as a tolerence table may be asserted) between the actual object and a golden object (template)

2.5.2 Pattern Recognition

In the following subsections, several image processing method for the analysis of the agro-alimentary products are listed. One can use one method or several method in serial depending on the complexity of target product to analyze.

2.5.2.1 Colour Inspection

The objective is to analyse the colouring of any food, particularly baked products. Visual appearance of food is a major factor in the judgement of quality. The system provides reliable relative quantitative measurement data to decide whether a product meets customer acceptable colour standards. The system may consist of colour camera, lighting system, frame grabber and control software.

To keep the light level constant, a light feedback function should be added to the lightsource. Otherwise, the coulour correctness may be affected.

2.5.2.2 Principal Component Analysis

Principal Component Analysis (PCA)[AD002] involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The objective of PCA is:

- To discover or to reduce the dimensionality of the data set.
- To identify new meaningful underlying variables.

PCA may used for the quality control of the product such the set of components (transformed image) could be matched to templates. It is useful to use a small number of patterns from the image instead of the whole image for a rapid assessment of the quality of the product. However, the number of components depends on the variability in the image. The PCA is suitable if the original image has to be reconstructed.

2.5.2.3 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) [AD003]classifies a sample object into one of two categories based on certain object properties.

This method is well suited for the classification (Clustering) of different objects in a image. If the agro-alimentary product consists of different elements, this method may separate the image component into several classes according to the number of elements in the product.

2.5.2.4 Canonical Correlation Analysis

In Canonical Correlation Analysis (CCA) [AD004], it required to find the correlations between two data sets. One data set is called the *dependent* set, the other the *independent* set.

The optimal correlation is obtained using this method however, this method is not well-suited if the product contains several element as the algorithmic complexity may increase

2.5.2.5 Independent Component Analysis

Independent Component Analysis (ICA) [AD005] is a powerful technique from the signal processing for the blind source separation. It can be seen as an extemsion to PCA. While PCA takes into account only statistics up to 2nd order, the ICA finds components that are statistically independent (as independent as possible).

This method may be a powerful method for the separation of the different elements consisting the agro-alimentary product. Each element may be processed independently afterwards.

The algorithmic complexity has to be evaluated (TBD).

2.5.2.6 Kernel Methods

All methods presented above are linear. Kernel methods [AD006] are used to generalize to non-linear data in a computational efficient manner. The basic idea is:

1. Non-linear mapping of data in high dimensional space
2. Perform linear method in high-dimensional space
3. Non-linear method in original space

The problem is that the processing in high dimensional space is not realizable due to the high complexity. Therefore, it is mandatory to avoid the computation of the non linear mapping directly, if the algorithm can be specified in terms of dot products and non-linearity satisfies the Mercers condition. Then, we can use the Kernel trick.

This method may be used for the quality control of agro-alimentary products, as an extension of the previous listed methods (e.g. Kernel PCA or Kernel CCA), if a non-linear separation between the product elements has to be performed.

2.5.2.7 Wavelet Decomposition

The Discrete Wavelet Transform (DWT) can be seen from two complementary perspectives [AD007]. In the first case the DWT is seen as a linear projection of the signal into the wavelet domain. In the second case the DWT is seen as an operation separating the different frequency components of the signal using a Filter Bank (FB). The principal idea was to divide the input into different frequency bands. This division was performed by filtering the signal, generally with FIR filters. The simplest case is when the signal is filtered using a low and a high pass filter, respectively hl(k) and hh(k). However, this procedure has the drawback that the size of the data is doubled. To overcome this problem the low and high-pass components are downsampled by a factor 2.

This approach may be used to reduce the image size before its processing. It is recommended to combine the Swelden implementation of the DWT (the Lifting Scheme)[RD002] and the Shapiro approach [RD003] to have a lower algorithmic complexity.

2.5.3 Decision-Making Process

This step consists of taking the decision of either rejecting or accepting the product using the analysis results and according to the quality criteria. Depending on the complexity of the product to control, two kind of decision- making process may be used.

2.5.3.1 Classical Methods

A statistical decision tree may be used to decide whether to accept or reject the product. The decision may consist of a set and if-else setps and/or several lookup tables which may contain the acceptance tolerances (i.e. for every quality parameters a tolerance range is stored).

The advantage of this method is the easy implementation and the fast decision. The maintenance of the method is very easy. It consists of the update of the decision tree and the lookup tables

The disadvantage is the non-flexibility as the decision is a part from the decision tree resp. the lookup tables

2.5.3.2 Neural Network

The decision to accept or reject the product may be performed using an intelligent system based on Neural Network NN.

The advantage of this method is the flexible decision which depends on the performance of the NN.

The disadvantage is the complexity of the implementation and the maintenance which requires a training of the NN each time the quality parameters are updated.

3 Flood Forecasting -Results from 2003 British Competition-

3.1 Background

The intention of this work is to assess our estimation method compared to Neural Network method by participating to the British competition „ANNEXG“. This competition intends to initiate a project to evaluate the effectiveness of artificial neural neworks in rainfall-runoff modelling/flow forecasting. In this Section we present the results of flood forecasting method not using the Neural Network .

There exists 3 data sets from measurements in a region in UK. Two sets are provided for calibration (Train1.xls and Train2.xls) and one set for final testing of your models (Test.xls). The first calibration set (Train1.xls) covers the period 1 October 1993 to 31 March 1994 and the second calibration set (Train2.xls - which can be used for further training or validation/selection of your 'best' model) covers the period 1 October 1995 - 31 March 1996. The ultimate test set (Test.xls) covers the period 1 October 1994 - 31 March 1995.

These sets contain flow data (level, m) at three upstream sites (US1, US2, US3), rainfall data at five catchment rain gauges (mm) (RG1, RG2, RG3, RG4, RG5) and flow (level, m) at the target site (Q). The data are sampled at 15 minute intervals.

Note, the peak stage in the Train1.xls file is 5.04m, in the Train2.xls file it is 4.13m and in the final test data set the peak stage is recorded as 5.78m. In these files any missing flow data are represented by *-1* and missing rainfall data are represented by blanks - it is up to you how you deal with these.

There are a limited number of unusually high readings in the rainfall data and it is up to you to decide how to deal with these. For information, the highest ever recorded rainfall in the UK for a 15 minute period is 50mm, so any values over 40mm should be treated with scepticism.

These data should be used to create models for three forecast horizons (t+8 hours, t+16 hours and t+24 hours ahead). For each of these forecast horizons you should create two neural network models - one which can use upstream flow data (predictors US1, US2, US3) and one which does not (although it can use antecedent flow at the target site with appropriate lead times of 8, 16, or 24 hours respectively).

3.2 Methodology

In general, the relationship between the water flows could given like this:

$$\sum_{t=0}^{t=t1} In_t(flow) + Er(Input\_Meas.) = \sum_{t=0}^{t=t2} Out_t(flow) + Er(Output\_Meas.) + Er(Water\_Loss)$$

where:
In_t is the measured water level for several upstream sections at instant t
Out_t is the measured water level for target section Q at instant t
Er is the error due either to the accuracy of the measument or to the water loss due to evaporation..etc
t2 and t1 are time constant wher t2>t1

We can fusion the errors such that

$$\sum_{t=0}^{t=t1} In_t(flow) = \sum_{t=0}^{t=t2} Out_t(flow) + Er(total)$$

Where

Er(total) = Er(Output_Meas.)+ Er(Water_Loss) - Er(Input_Meas.)

Er(total) is minimal whenever the accuracy of the measurements is high (e.g. high sampling) and the water loss is negligible.

In this case, to calculate the water level (in meter) at a target site Q (Target_Stream) depending on the measurement of the water level for 3 other sites USx(UpStreamx) we can use this general formula.

$$Q(t+T) = \sqrt{(w_1.US1_t(t))^2 + (w_2.US2_t(t))^2 + (w_3.US3_t(t))^2}$$

Where:
T is the average time constant between the Upstreams site and the target site (depending on the distance between the site and the speed of the water flow)
Wi is the weight of every upstream (i=1:3) which consists of the contribution of every US water flow to the target site water flow.

1.) As we Dont Know the Distance between the Upstream Site and the Target Site, we Have to Calculate the Time Constants (T) bewteen them by finding the mean difference of the Peaks

2.) Estimation of the Non-Weighted Water Flow

3.) Calculation the Mean Absolute Error between the Estimation and the Upstreams

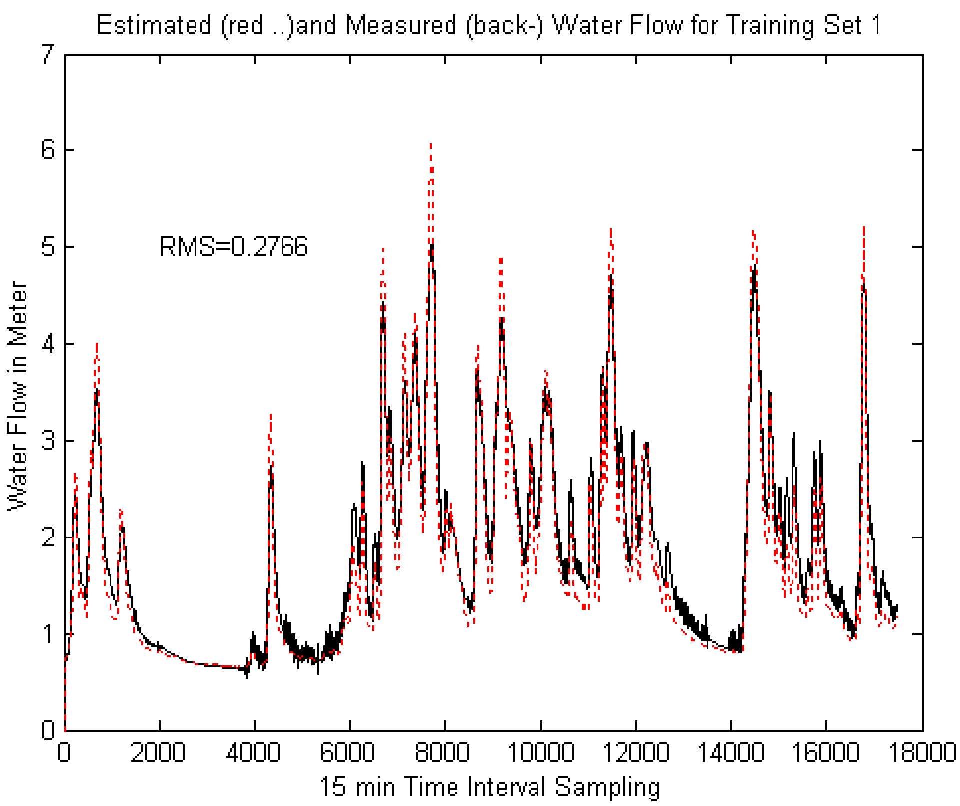
4.) Calculate the Weight for every Upstream Site Using the Mean Absolute Error Calculated Above

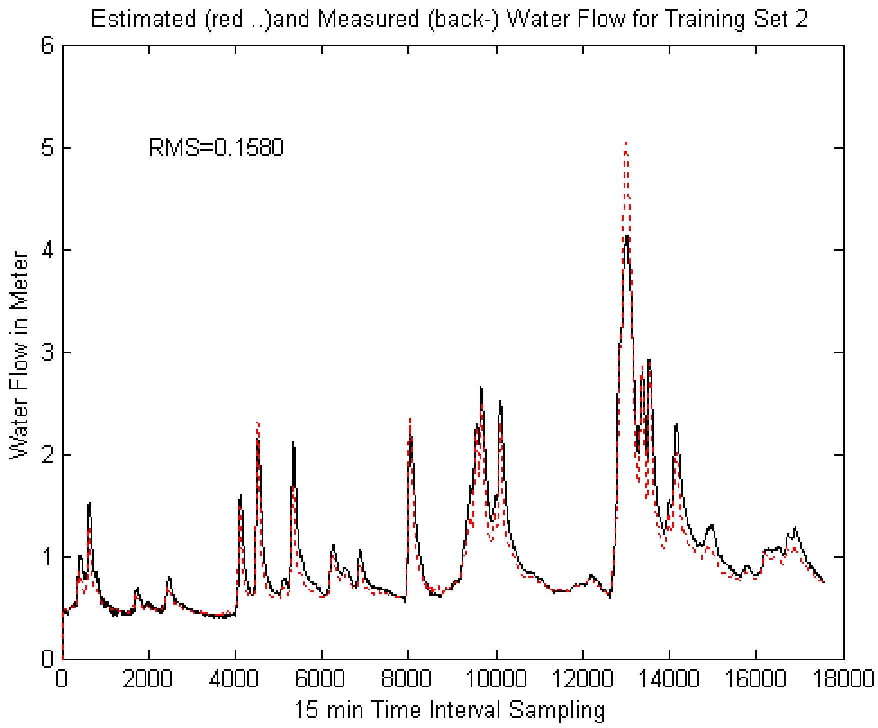
5.) Estimation of the Weighted Water Flow

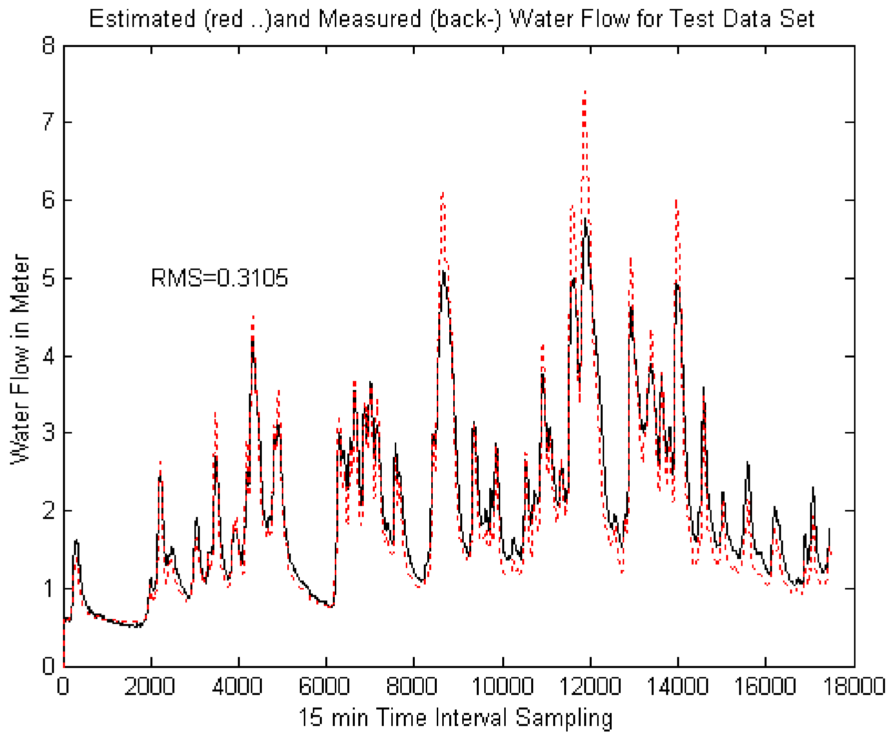
6.) Store the results

3.3 Results

Preliminary Results when Estimating the Water Flow Q with Only using the 3 Upstreams information US1, US2 and US3.




Estimated (red ..)and Measured (back-) Water Flow for Training Set 2
RMS=0.1580
Water Flow in Meter
15 min Time Interval Sampling
0
1
2
3
4
5
6
0
2000
4000
6000
8000
10000
12000
14000
16000
18000


Estimated (red ..)and Measured (back-) Water Flow for Test Data Set
RMS=0.3105
Water Flow in Meter
15 min Time Interval Sampling
0
1
2
3
4
5
6
7
8
0
2000
4000
6000
8000
10000
12000
14000
16000
18000

4 Meetings

4.1 Meeting 1. “University of Cagliari 11.07.2003 at 16h00”

Attendees: A.N. Belbachir, B.Cannas, A. Fanni, M. Lera, A. Montisci, A. Sechi, L. See and Antonino Serri

The aim of this Meeting was to understand the environment of the flood forecasting in Sardinia. A set of data measurements (Rainfalls and Runoffs) has been collected for the new Dam in Tirso. Two data set are available: one from the period Nov-2000 till Jan-2001 and the second from Jun-2001 till Jun-2002.

The data sampling intervall is either hours or days. There exists also montly data which consist of an averaging of the daily data. A. Sechi said that accuracy of the data is not high such that the measurement does not cover the whole sampling interval (i.e. the measurement is performed during the first 10 min but affected to an hour measurement).

A. Sechi recommend to not use a high resolution data „ hours data“ for building the flood forecasting model. Experience shows that daily data is enough to have a robust flood forecasting model.

L. See asked to have the monthly data for the Tirso in order to her method results with the results of Sardinian working group.

4.2 Meeting 2. “University of Cagliari 15.07.2003 at 16h00”

Attendees: A.N. Belbachir, B.Cannas, A. Montisci, L. See, Antonini Serri and Michele ?

The aim of this Meeting was to prepare the Sardinian-British proposal for a possible cooperation in the flood forecasting. Administrative forms have to be filled and sent to the appropriate organization before 18.07.2003.

A.N. Belbachir asked B. Cannas about her experience while participating into the british competition for testing the Neural Network flood forecasting methods. It is possible to apply for the actual british competition (deadline is 30.11.2003) together taking into account the results of Cannas of the competition of last year

4.3 Meeting 3. "University of Cagliari 17.07.2003 at 16h30"

Attendees: A.N. Belbachir, A. Fanni, M. Lera and A. Montisci

The aim of this Meeting was to discuss about a new project „quality control of agro-alimentary products“. As the technique to use is highly dependent of the product to qualify, the specifications of the target product has to be known in advance to focus on the work to do.

A Meeting with the compagnies consultant „Effisio?“ has been organized for 21.07.2003 to discuss about practical product quality problems with agro-alimentary products in Sardinia.

M. Lera takes the action to contact the consultant.

A.N. Belbachir takes the actions to write a report (included in this document) to summariye the possible methods to use for quality control of some agro-alimentary products.

4.4 Meeting 4. "University of Cagliari 21.07.2003 at 15h45"

Attendees: A.N. Belbachir, A. Fanni, M. Lera and A. Montisci.

The items discussed during this Meeting was :

1. To forthgo with the discussion about „quality control of agro-alimentary products“. The compagnies consultant „Effisio?“ was not able to attend the Meeting for TBD reasons. M. Lera collected some inputs from the consultant about some possible product issues for the quality control. Two kind of products has been maintained because of their relevance for the algerian and Sardinian industry „Cheese and Pasta“. While Pasta is not a common product for Algerian industry, the cheese could be for a high interest of both parts „Algeria and Sardinia“. The next step to do is to define precisely which kind of cheese (product) and at which stage of production this product could be qualified in optimal manner. M. Lera takes the action to meet Effisio to have more details. A.N. Belbachir takes the action to contact Algerian industries who produce cheese for more information. The collaboration will be followed by both partners in Algeria and Sardinia.

2. To discuss about the first draft document from A.N. Belbachir for the possible methods to use for agro-alimentary product qualification.

3. A possible visit of Sardinian partner to Algeria or to Austria has to be discussed and organized.

4. A.N. Belbachir talks about his actual field of reasearch in Vienna. They consist of:

   a. Entropy Coding (data Compression, Noise Estimation\Modelisation, Resampling techniques)

   b. DSP Programming (DSP Implementation, Real Time systems)

   c. Complexity Evaluation and Minimization (Parallel Processing, Efficient Implementation Structure)