Capacity building at the National Meteorological Institute: Training personnel to apply quality control tools.

National Meteorological Institute

P.O. Box 5583-1000 San Jose, Costa Rica

Tel. (506) 22-22-56-16

E-mail: jlaraya@imn.ac.cr

José Luis Araya López.

Abstract

As part of an initiative to improve the quality control procedures at the Costa Rican Meteorological Institute, new operational tools to check large amounts of data were developed. These tools rely on open software technology. In order to allow this scientific computing software to be properly used by technicians, a capacity building effort was undertaken. Training on the new tools was a must, so that personnel without academic background could understand and apply them properly. The stages of this capacity building program involved independent learning from operational meteorologists, in particular the use of scientific computing software such as Scilab and Python. The second stage included development of the quality control programs, which also implied a lot of team work and cooperation among scientists and technicians alike, so that it were possible to determine what errors are likely to occur in an operational context. Finally, after a period of testing and implementing operational scripts, the need arises for transferring the new tools to other technicians. Nowadays, scientific computing tools are used on a daily basis at the data processing office. At present, technicians who were resilient to new technological approaches have tested them and become keen on the initiative. Apart from that, the new quality control tools have allowed personnel to improve their ability to detect anomalous situations in meteorological data. It is expected that new and improved technological approaches will be implemented, as long as additional experience is attained.

Introduction

Quality control of meteorological data is a highly deemed area of expertise whose purpose is to guarantee reasonable quality standards in meteorological information. It is clear the process of automation has raised the need for new ways of looking at the way data errors may come up. This is important in those institutions whose main purpose is making sure that data networks do not only generate data, but to guarantee that those data are being generated in agreement to the goals that motivated the whole project. Traditionally, the way things were dealt with conventional weather stations at the National Meteorological Institute (NMI) was a different approach; technicians were in charge of processing data generated by this network, so that the former quality control procedures and the errors generated were manually detected. The former procedures were lacking effectiveness as long as new automatic observing systems were added to the data network. In fact, there were reasons to believe that this former manual quality control approach was lacking some important features to cope with relatively large volumes of data, so the need for a different approach for quality controlling data became a primary need at NMI.

Considering the need of more effective ways to revise large amounts of data, a project to develop operational methods to deal with this information was proposed. The present paper delves into the different stages of this capacity building initiative to improve the efficiency of the former quality control protocol.

The data network at NMI is traced back to 1995, with a massive endeavor to cover the national territory at large. However, the concept of real-time quality control was not applied in former stages, so that data have been collected by manual methods, namely, technicians who visit the placements at regular intervals, download data and take them to the data processing office for check and final storage. Checks made to such information have depended on the technician ability to detect and deal with potential errors, normally there has always been a single technician that deals with the massive amounts of information that is generated by the data network. Because of this situation, large amounts of data lacked an objective quality control. As a matter of fact, training was compulsory in the different stages of development. The goals of this endeavor are:

1) Generate capacity building on the most well-known quality control techniques, as well as computing tools that facilitate their application in an operational context.
2) Develop operational tools that can deal with large amounts of data.
3) Generate confidence on the effectiveness of these quality control tools.
4) Develop quality control tools that can be applied by data network personnel, so that their capacity to detect anomalous information is enhanced. A diagram of the objectives proposed is seen in Fig. 1. This figure shows the feedback that training has in building capacity in the subsequent stages of the project.
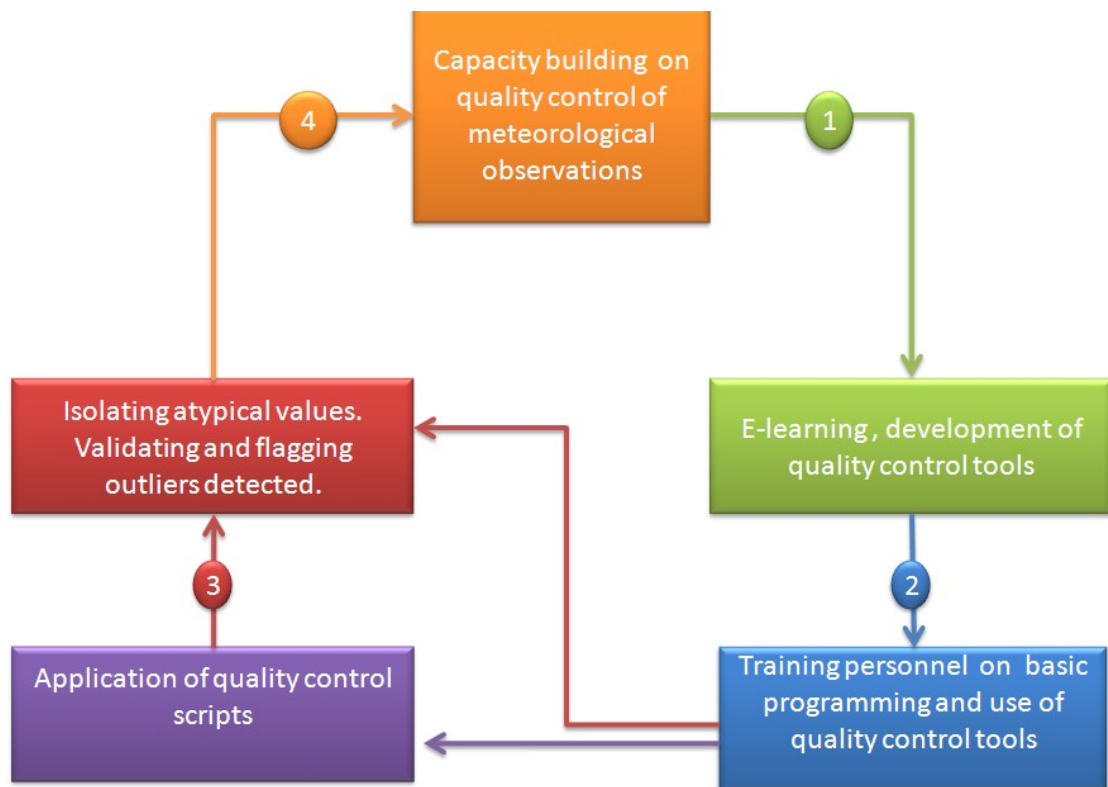
Figure 1: Different stages of development. Red and purple arrows show how training has had an impact in later stages of the process.

The quality control project involves the following steps:

a) Learning scientific computing programming.
b) Development of quality control scripts to cope with large amounts of hourly and daily data.
c) Testing of quality control applications.
d) Application of these programs to data sets generated by the data network.
e) Assessment of results generated by the quality control algorithms.
f) Detection of atypical values and validation.
g) Flagging atypical data.
h) Feedback to the data network personnel. This allows people in charge to take measures in order to point towards possible technical vulnerabilities that have permitted the outliers detected to occur.
i) Documentation of results and procedures.

Previous research

Once the need for better tools to quality control data was presented, it was compulsory to undertake a research on the topic of quality control of meteorological observations. It was necessary to go over the quality control bibliography in order to determine the most efficient methods applied to an operational level by National Meteorological and Hidrological Services(NMHS). It was found that very useful information had been delivered on the internet, so

that it was straightforward to compile all the material that may be useful to undertake this endeavor.

Once this massive revision was finished, it was blatant there is enough sources available to get to grips with the mean aspects of modern quality control. An exhaustive study on the application of quality control algorithms lead to the establishment of a very basic real-time quality control protocol. This was important to analyze the viability of such a proposal and its advantages at a practical level. Discussions and presentations on how this quality control algorithms work and the way they can be interpreted were made to spread the word about the possibility to departmental level to count on computing software, which could allow the personnel to understand and apply such concepts on a daily basis.

Learning programming languages

Fig. 2 shows the programming tools that were learned in this capacity building program. The second stage of this quality control required experience on programming languages, so that personnel from NMI could develop quality control applications. This was important due to the difficulties that meteorological services such as NMI face with payment of licenses of commercial software. Capacity building on programming was a reasonable choice because those who develop such expertise are capable of customizing applications; so that they can solve problems. Apart from that, these skills were useful to put to the test quality control algorithms in a straightforward manner.

In order for the personnel to be able to apply such tools, a training period on scientific programming and application was a must. In this case, two scientific computing packages were applied: Scilab and Python. Scilab is high-level scientific computing software, which provides an interpreted programming environment. It is very similar to other commercial programs such as Matlab and IDL. Python is a general-purpose high-level scientific computing language which points out the importance of code readability. Due to the general character of this computing language, it is important to emphasize that these tools have been applied using the modules Numpy, Scipy and Matplotlib to develop quality-control scripts in a Matlab-like environment.

In order to cope with the training process, online resources were actively used and applied to real problems at the data processing office, so that it were possible to determine the advantages and disadvantages of any application. The websites of these computing languages contain tutorials and examples on how to apply the basics of programming (Fig. 2); these resources were actively applied to boost the learning process.
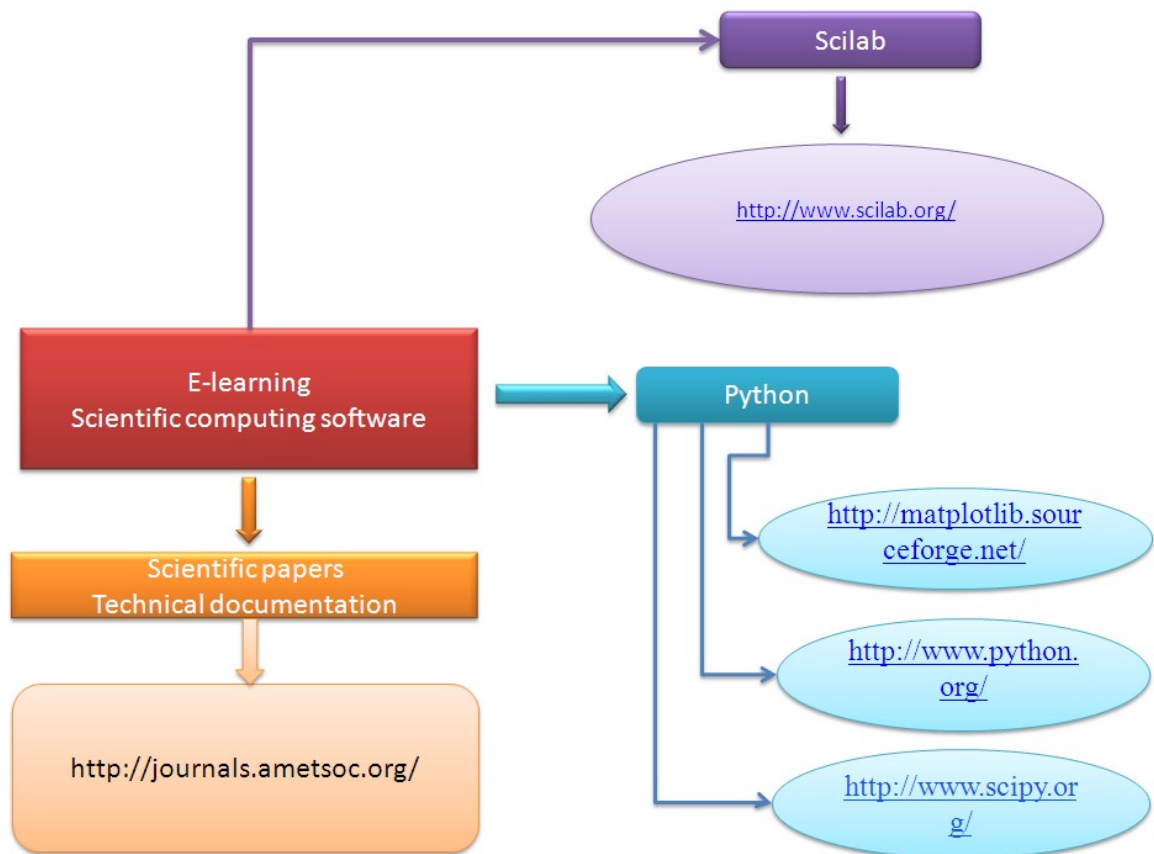
Figure 2: E-learning roll in the development of this quality control protocol.

Development and Application of operational tools

Once experience on scientific computing was gained, a project on quality control of meteorological observations took place. Quality control scripts were developed to analyze hourly and daily data generated by automatic weather stations. These scripts normally include a set of quality control algorithms, as well as numerical display, plotting and reports on the results of the quality-control tests. These computer programs were run on the datasets generated by the data network, which includes 10 years of hourly data for all meteorological parameters (air surface temperature, relative humidity, radiance, precipitation, wind speed and direction, surface atmospheric pressure). Some parameters also included daily data sets that had to be quality controlled. After filling temporal gaps in the datasets the quality control programs were run. Depending on the parameter analyzed, the quality control script included algorithms, plots and reports that allowed a qualified technician to isolate suspicious information. Fig. 3 and Fig. 4 show examples of two quality control scripts that were applied. In particular, the goal was to apply this quality control tools on the data generated by automatic weather stations. The methodology took into consideration the experience obtained by technicians and meteorologists on the most common sources of error that they have detected. These errors had not been properly characterized due to lack of more sophisticated tools to deal with burdensome amounts of data.
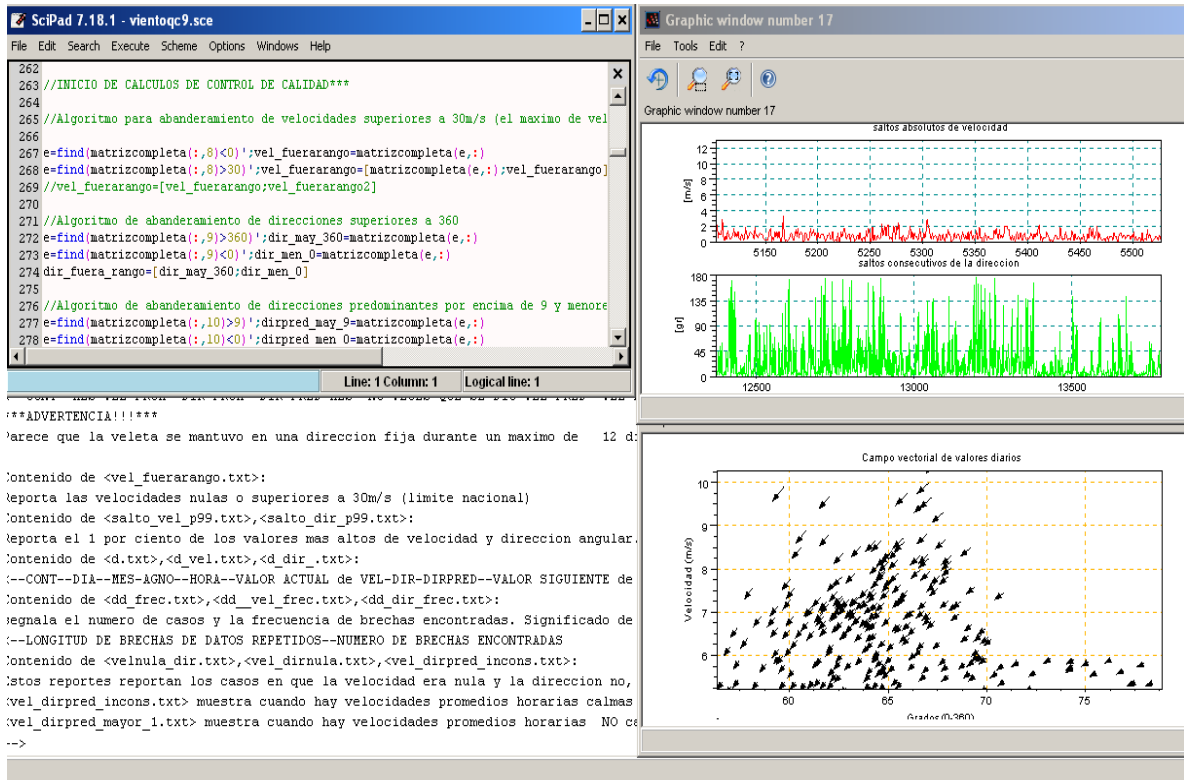
Figure 3: Scilab quality control script for detecting outliers in hourly wind speed and direction
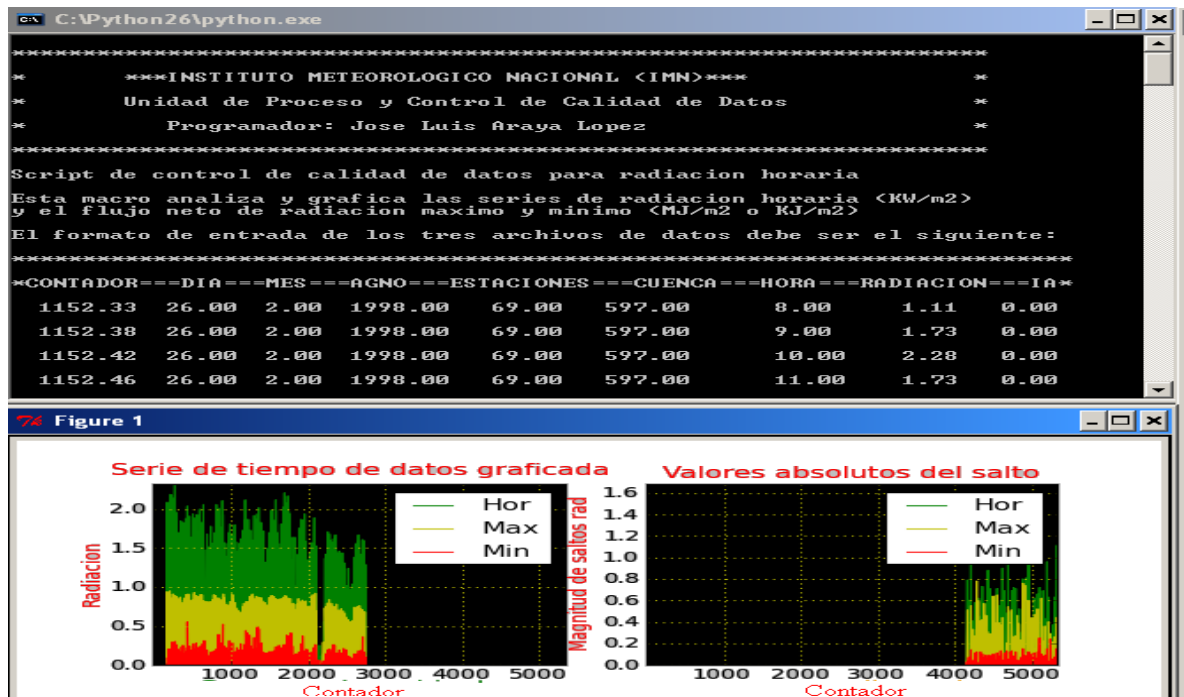

Figure 4: Example of a Python script to quality control radiance data.

Training personnel to apply quality control tools

The beginning of this initiative can be traced back to 2007, when the whole project was officially proposed. Personnel who have participated include 2 meteorologists and a technician. The new tools were also passed on to other technicians, this in order to be used by them to quality control the information they analyzed. These technicians had never got in touch with scientific computing programming, so they were resilient and skeptical in the matter. However, it was possible to devise a training plan in which they applied the quality control tools, so that they could assess them.

This training plan involved the following steps:

1. **Selected tutorials on basic programming:** The goal was to allow the quality control technicians to learn the basics of some applications. The plan involved active hands-on tutorials, which was a good introduction to these tools, so that they could get some experience on the basics of the software and possible applications to real data. Technicians were instructed in the know-how of these applications: installing, running and interpreting.

2. **Training on application of quality control scripts:** Technicians were trained to apply the quality control scripts. These technicians had prior experience on quality control of meteorological observations, either generated by conventional or automatic weather stations. It is important to point out that there was a previous discussion with these operational experts, so that important feedback could be obtained. Comments and suggestions based on their experience of a lifetime on quality control made possible to program quality control applications that could fit their needs.

3. **Application of quality control tools:** Once the resources to undertake the revision were set, a complete check on the whole data base was undertaken. Atypical values were found, which were shown to have gone undetected by the former manual quality control. Evidence and confidence at an operational level was gained on the viability to make effective quality control tools using scientific computing programs.

Conclusions and perspectives

After two years of sustained effort in quality control activities, it is clear that personnel at the data processing office have benefited from this training to tackle the problem of outlier detection. The former lack of quality control tools that could be applied to large data sets limited the capacity to analyze vulnerabilities in the current quality control protocol. Nowadays, quality control algorithms can be programmed, tested, implemented and applied on a daily basis. We strongly believe it is a good start to catch up with basic modern quality control methodologies normally quoted in peer-reviewed literature. Capacity to use the internet resources and free software technology is making a difference, because it allows to develop customized programs that can solve specific needs. Scientific literature is available on the internet; it can be accessed and studied by trained scientists. Operational expertise at NMI increases due to modern approaches given by the use of quality control literature and scientific computing tools, which is important to achieve some level of modernization, chiefly in the operational practices from developing countries like Costa Rica. An important aspect of this endeavor is that a quality control protocol has been developed using

limited resources. The usefulness of operational developments of this kind is that NMHS from developing countries may benefit from the discussion of these approaches, which may objectively evaluate if it is viable for them to implement them, if not already available. Some level of cooperation among institutions may encourage training and application of approaches of this type for supporting operational quality control in developing countries.

Challenges that have to be tackled in the near future are:

1) **Follow-up of quality control operational research:** Due to the very special field of activities of NMI and the need that this institution has for automating procedures, it is expected that more test and research on quality control strategies will carry on being applied in the future. Former experiences suggest that this kind of research is fundamental to be able to cope with new needs and demands.
2) **Development of a real-time quality control protocol:** One of the goals of the project was to determine to what extent errors could be found, and how to use this experience to point towards potential methodological vulnerabilities in the way things are done. It was desirable to build capacity on operational tools that could ease this task. Once basic operational knowledge is set, the possibility to develop customized tools for the small NMI real-time data network is real. It has to be recognized the real-time data network's density is not high, so that deferred quality control may continue being compulsory.
3) **Improvement of quality control tools:** it involves development of GUIs and other goodies that will facilitate interaction, either with not very specialized users or with the data base. It includes development of spatial quality control algorithms, as well as introduction of more techniques to improve the outlier validation process.

Acknowledgements

Bibliography

Baker, N.L., 1992: Quality Control for the Navy Operational Atmospheric Database. *Wea. Forcast.*. 7(2). 250-261.

Eischeid, J.K., Baker, C.B., Karl,T.R. y Diaz, H.F., 1995:The Quality Control of Long-Term Climatological Data Using Objective Data Analysis. *J.Appl. Meteor.*, **34,** 2787-2795.

Feng, S., Hu, Q., Qian, W., 2004: Quality Control of Daily Meteorological Data in China, 1951-2000: A new dataset. *Int. J. Climatol.*,**24**: 853-870.

Gandin, L. S., 1988: Complex Quality Control of Meteorological Data. *Mon. Wea. Rev.*, **116,** 1137-1156.

Shafer, M.A., Friebrich, C.A., Arndt, D.S., Fredrickson,S.E. y Hughes, T.W.,2000: Quality Assurance Procedures in the Oklahoma Mesonetwork. *J. Atmos. Oceanic Technol.*,**17**,474-494.

Vejen, F., Jacobson, C., Fredrikson, U., Moe, M., Andresen, L., Hellsten, E., Rissanen, P., Pálsdóttir, T. y Arason, T., 2002: Quality Control of Meteorological Observations. Automatic Methods Used in the Nordic Countries. Nordklim, Nordic Co-operation within Climate Activities, Report Nº 8 KLIMA. 109 pp.

Øgland, P.,1993: Theoretical Analysis of the Dip Test in Quality Control of Geophysical Observations. Report Nº 24 KLIMA. 18 pp.