



EC DG-TREN
Transport Programme
(2.3.1/2)
Contract : 2000-AM.10011



MAEVA

A Master ATM European VAlidation Plan

D3.3: Measurement and Analysis Specification for Validation Exercises



Document Change Log

Release	Author	Edition Date	Affected Sections / Comments
0.1	A. BARBER	17/4/01	All
0.2	A. BARBER	26/04/01	All, incorporating NATS Real time comments
0.3	A. BARBER	11/05/01	All, incorporating NATS Fast time comments
0.4	A. BARBER	21/5/01	All, incorporating R. A. Whitaker's comments.
1.0	R EVELEIGH	10/07/01	Incorporating comments from Sofreavia and AENA

Document Distribution

To/cc	Code	Organisation	Name
To		Project Manager	J. REVUELTA
cc.	AEN	Aena	M. HERNÁNDEZ
cc.	AVT	AVTECH	L. LINDBERG
cc.	NAT	NATS	R. WHITAKER
cc.	SOF	SOFREAVIA	J.R. VELTEN
cc.	AEN	Aena	J.M. de PABLO

Review and Approval of the Document

Organisation Responsible for Review	Reference of comment documents	Date
AENA	33RA0101	29/06/01
Organisation Responsible for Approval	Name of person approving the document	Date
Work Package Leader	J.R. VELTEN	10/07/2001
Project Manager	J. REVUELTA	17/07/2001
EC Official	P. BERNARD	27/09/2001

Table of Contents

0.	EXECUTIVE SUMMARY	1
1.	INTRODUCTION	1
1.1.	Purpose	1
1.2.	Scope	1
1.3.	Background.....	1
1.4.	Document Structure.....	1
1.5.	Glossary.....	1
1.6.	Reference Documents	2
2.	RELEVANCE.....	2
3.	MEASUREMENT AND ANALYSIS SPECIFICATION	2
3.1.	Validation Aims.....	2
3.2.	Operational Participant Requirements	3
3.3.	Training Issues.....	4
3.3.1.	Specification of Training Methods	4
3.4.	Measurements.....	5
3.5.	Analytical Methods.....	6
3.5.1.	Selection of Analytical Approach	6
3.5.2.	Exploration.....	6
3.5.3.	Inference	7
3.5.4.	Synthesis	7
4.	CONCLUSION	8



Index of Tables

Table 2-1	Relevant Subsections.....	2
Table 3-1	Example of a Definition of a Measure	5



0. Executive Summary

This document provides a template from which an MAS for real-time, fast-time, shadow mode and operational trials may be produced. Such an MAS will have the following sections:

- *a restatement of the validation aims;*
- *the requirements placed on operational participants, if they are to participate in the exercise;*
- *training issues for participants;*
- *measurements to be made;*
- *recommended analysis methods.*

The document gives advice on the scope of each of these sections and provides example wording that can be used in an MAS. References of MAS developed for real-time simulations are provided.

1. Introduction

1.1. Purpose

The purpose of this document is to provide a template to assist in the production of the Measurement and Analysis Specification (MAS) for each validation exercise. An MAS is required for each validation exercise to support the planning and execution of the measured runs. It draws together all the relevant information on the conduct of the measured runs and the subsequent analysis of the results. In this way any gaps or inconsistencies in these aspects of the validation exercise can be identified and addressed. The issues that should be taken into account include:

- the validation aims;
- the requirements placed on operational participants;
- training issues;
- measurements;
- recommended analysis methods.

1.2. Scope

This document is written with particular regard to how an MAS should be written for Operational and Shadow-mode trials, and Real-time and Fast-time simulations. It does not attempt to cover Analytical methods, Literature studies or Judgmental techniques.

It is inevitable that each MAS will contain specific detail relevant only to the trial for which it is written. Therefore this document should not be seen as the definitive MAS, but as a basis for building exercise specific MAS for each validation exercise (see VGH Sub-activity 2.5.1).

The MAS will eventually be incorporated within the VGH. While it is intended to be useful as a standalone document, it is best read in close reference to the VGH.

1.3. Background

The major objective of MAEVA is to provide the EC with an overall validation strategy for the TORCH concept, as represented by the 5th Framework Programme projects selected for MAEVA. The first phase of MAEVA is to provide a framework for the validation within these projects. The framework comprises:

- a VGH for the conduct of individual validation exercises;
- a Validation Master Plan that schedules all of the anticipated validation exercises to be conducted by these projects to ensure an efficient and complete validation of the overall ATM concept envisaged by the EATMP.

MAEVA deliverable D1.3 (Ref.1) provides a basic framework for the conduct of validation exercises. It was based on MAEVA work packages, WP1.1, WP1.2, WP2.2 and WP2.3. The work conducted in WP3.4, providing an outline MAS, completes the initial consolidation of guidance material. This information will be incorporated in the VGH for dissemination to the projects and other interested parties.

1.4. Document Structure

This document should be read in conjunction with the Validation Guideline Handbook (VGH) (Ref.1), the activities and steps mentioned may be found in the VGH. Section 2 indicates which sections within the template are relevant for each type of trial covered by this document. Section 3 provides a template from which an MAS may be produced. Examples are written in italics. Small-scale, Real-time simulations have been used for examples as this type of exercise illustrates most of the aspects of interest.

1.5. Glossary

Term:	Description:
ATC	Air Traffic Control.
ATCO	Air Traffic Control Officer.
ATM	Air Traffic Management.
CBT	Computer Based Trainer.
EATCHIP	European ATC Harmonisation and Integration Programme.
EATMP	European Air Traffic Management Programme.
EC	European Commission.
EEC	Eurocontrol Experimental Centre.
HMI	Human Machine Interface.
ISA	Instantaneous Self Assessment.
MAEVA	Master ATM European Validation Plan.
MAS	Measurement and Analysis Specification.
MTCD	Medium Term Conflict Detection.
NATS	National Air Traffic Services.

TORCH Technical, EcOnomical and OpeRational Assessment of an ATM Concept AchIEvable from the year 2005.
VGH Validation Guidelines Handbook.
WP Work Package.

1.6. Reference Documents

LIST OF REFERENCE DOCUMENTS	
Reference	Author / Organisation, Title, Edition and Date
1.	T. Kiebalá / Isdefe, Validation Guideline Handbook, 25/05/01, 1.0, MVA/ISD/WP1/13DI__10
2.	S. D. Kay, NATS, PD/3 Measurement and Analysis Specification, 1,02/98, PHARE/NATS/VAL-4.4.2/WP005
3.	K. Jackson & D. Marsh, NATS, A Methodology for Validating EATCHIP III Medium-Term Conflict Detection using Shadow-mode Trials, Issue 1, 25.4.01, ASA.01.MTCD.DEL01.VAL1
4.	R. S. Eveleigh, NATS, Validation Methodology, Issue 1, 29/03/01, MVA/NAT/WP22DN__10
5.	M. R. Phillips & D. T. Marsh, The Validation of Fast-time Air Traffic Simulations in Practice, 2000, Journal of the Operational Research Society Volume 51 No. 4
6.	A. Harvey & C. Costello, EATCHIP III Evaluation and Demonstration PHASE 3 Project Experiment 3Abis: MTCD Final Report, 12/00, EEC Report No. 355

2. Relevance

This Section indicates which parts of the template are relevant for each type of trial covered by this document. Each validation technique is suited to validating different stages in the development cycle of ATM Concepts. Consequently there are different requirements for each validation exercise depending on the technique used. For example, generally Fast-time simulations do not require a controller or controllers to be present on the day of simulation; accordingly, the section on operational participant requirements (Section 3.2) is not relevant for Fast-time simulation. This does not, however, imply that controllers have no input into Fast-time simulations (see Ref. 5). Table 2-1 indicates which sections are relevant to the various validation techniques.

RELEVANCE		
	Real-time ¹ Techniques	Fast-time techniques
Validation Aims	✓	✓
Operational Participant Requirements	✓	✗
Training Issues	✓	✗
Measurements	✓	✓
Analysis methods	✓	✓

Table 2-1 Relevant Subsections

3. Measurement and Analysis Specification

3.1. Validation Aims

The validation aims to be detailed within the validation aims section of the MAS will be those identified in Step 1 of the Validation Guideline Handbook (VGH). The validation aims identified by the Validation Team in VGH Activity 1.3 should be stated along with the high-level validation objectives identified in VGH Activity 1.4. These objectives provide the basis for identifying the measurements for the validation exercise. These measurements are described in Section 3.4 of this document, the analysis of which will follow the guidelines specified in Section 3.5.

Including aims and objectives within the MAS ensures that the document can be read in isolation.

Examples

The following examples are included to illustrate how a validation aim and its related objectives could appear within the MAS. For simplicity only one of the validation aims has been presented.

Example 1 is from a small scale real time trial of a prototype HMI and reflects a more general approach to Real-time simulation.

Example 2 is from Experiment 3Abis:MTCD of the EATCHIP III Evaluation and Demonstration programme (Ref. 6). It is described as being "small scale" and is designed to be just detailed enough to provide initial answers to the questions posed. Therefore the aims are specific to the questions the EUROCONTROL customer wished to answer.

Example 3 is from a fast time simulation of future traffic sector loadings.

¹ Real-Time Techniques encompass Real-Time simulations, and Shadow-Mode and Operational Trials.



Example 1

Validation Aim

- explore the impact of the prototype in terms of capacity.

Related High Level Objectives

- Determine whether capacity is changed by the introduction of the prototype;
- Identify any limitations of the prototype that may limit capacity;
- Identify any features of the prototype that may improve capacity.

Example 2

Validation Aim

- Demonstrate and evaluate the EATCHIP III operational concept 'hands on' to obtain feedback from controllers. An evaluation should include the identification of the strengths and weaknesses of the concept from the users' viewpoint. Where weaknesses are found, workable solutions shall be proposed.

Related High Level Objectives

- Evaluate the level of confidence felt by the controllers in using the MTCD tools;
- Determine the effect on the workload of each controller (Executive & Planning) caused by the enhanced planning capabilities provided by the MTCD tools;
- Evaluate controller roles, tasks and working methods when using MTCD tools to identify, analyse and resolve problems in a stripless environment;
- Evaluate the different characteristics of the two HMI provided for the Potential Problem Display.

Example 3

Validation Aim

- Estimate the sector loadings in terms of the number and duration of flights in each sector and the airspace as a whole.

Related High Level Objectives

- Obtain the peak number of flights within a controller's accept bay
- Obtain the peak number of flights in the ACC airspace;
- Obtain the number of flights in an area of airspace equivalent to the display screen centred on the busiest areas;
- Obtain the mean and maximum number of flights per hour in each sector and the whole ACC airspace;

- Obtain the average time that flights spend in a sector and in the ACC airspace;
- Obtain the average number of sectors that a flight passes through.

3.2. Operational Participant Requirements

The section on the requirements for operational participants, typically refers only to controller requirements, so here for clarity only controllers requirements will be described, but the specification applies equally to pilots where the exercise involves operational aircraft or high fidelity cockpit simulations.

This section of the MAS should contain a description of the number of controllers required and the sectors on which they are valid. The requirements placed on the controllers will differ depending on:

- the validation aims and objectives;
- the type of trial;
- the length of the trial;
- the number of sectors.

The following qualities should be considered when requesting controllers:

- familiarity with the airspace that they are asked to control;
- general ATC experience (i.e. length in years or any other sectors on which they are valid);
- participation in previous simulations/working groups involved in developing the prototype (this may or may not be desirable);
- use of computer based ATC systems;
- general interest and skills in using computers;
- familiarity with the baseline system where relevant;
- other qualities of interest such as height; eyesight; left/right handed; etc.

However, while it is possible to make requests there are usually only a limited number of available controllers who cannot be guaranteed to meet the requirements. When the controllers and their qualities are known, the trial design should be re-examined (see VGH Activities 1.8 and 1.9) in order to assess the impact on the analysis and to identify any steps that could be taken to ensure that the aims of the trial are not comprised.

Experience from previous validation exercises suggests that the views of the controllers will be influenced by many factors, including their hopes and expectations of the system under validation. It has previously been shown that controller opinion is influenced by the level of traffic loading and technology that they are used to. For example, controllers used to low traffic loading with only basic radar

support, may be more impressed by the technology than controllers used to more electronic support with high traffic loading. On the other hand, they may also be more wary of technology; therefore, appropriate training should be given in which both of these issues are addressed. (Training issues are discussed in Section 3.3.).

One vital "view" of the controller is that related to subjective workload (i.e. their assessment of how busy they think they are). This is likely to be influenced by many diverse factors, such as:

- control experience relevant to their controller working position;
- experience in using computer-based ATC systems;
- the traffic levels that they are used to handling;

These issues should be taken into account during the synthesis stage of the analysis, which is discussed in more detail in Section 3.5.4.5.4.

Example

The following text is an example of how the controller requirements may appear within the MAS. Depending on the type of trial and the expected influence of the factors mentioned above this section of the MAS could and should be lengthened or shortened.

'A request has been made for five controllers familiar with the baseline operational concept interface and sectors 1 and 2, although not necessarily validated on them, to participate in the trial. Two controllers familiar with a variant of this interface and operational concept are already confirmed for the trial.'

3.3. Training Issues

A detailed description of the training issues will be included here. The training issues will vary depending on:

- whether controllers meet the requirements set out in Section 3.2;
- the language the training is to be conducted in;
- the understanding of the concept by the controller;
- the scale of the difference between current day systems and the advanced system;
- the level of knowledge of the current and advanced systems required;
- familiarity with the simulation environment;
- the aims and objectives of the trial.

It is likely that the training will not be sufficient for controllers to have learnt everything there is to know about a system before measurements begin to be taken. It is further expected that the controllers will become familiar with the traffic samples, although the level of familiarity can

be minimised by careful experimental design (see VGH Activities 1.8 and 1.9). Learning effects should therefore be considered during the analysis. The amount of time dedicated specifically to training during the trial should be minimised. As far as possible, training should be undertaken a few weeks prior to the trial, with some time spent during the trial to refresh the controllers' memory. For example in a two-week trial approximately 2 days should be dedicated to reinforcing the previous training.

In a trial where a significant amount of training is necessary it is prudent to produce a training plan. The first stage in developing a training plan is to consider the objectives of the training. It is unlikely that the controllers will be familiar with everything before the start of the validation exercise, however, it is useful to define exactly what knowledge each controller should have. This will enable decisions to be made on the best way in which to train the controllers and the extent to which the training has achieved its aim will also be clear.

Example

The following example training objectives are taken from Appendix D of Ref. 6.

- *To be aware of the study context: experiment objectives and organisation, what is expected from the ATCOs' participation,*
- *To be aware of the operational context: time horizon, traffic characteristics, airspace management, sectors configuration, ATC procedures,*
- *To understand the controllers' roles and tasks: the system mission and all information associated with describing the new working method;*
- *To know the proposed system services: functionalities provided by the new system and all information concerning its functioning;*
- *To master the HMI principles and functions: applications/use of the functionalities via the HMI, coding of the information, allowed actions.*

Next the various training methods need to be considered in order to decide which ones will best meet the objectives. These training methods include:

- Literature;
- Computer Based Trainer (CBT);
- Presentations;
- Hands on experience including:
 - Illustration runs;
 - Mentored runs.

3.3.1. Specification of Training Methods

Training literature can take the form of an aide memoire, user guide and/or mock Temporary Operating Instructions. Temporary Operating Instructions are used at operational sites to communicate temporary changes in the standing agreements, routings or the method of operation. In the



same way they may be used to communicate any differences between current operations and the simulation environment.

The training literature, along with the Computer Based Trainer (CBT), are useful for providing an initial introduction to the system and the simulation environment. As part of the CBT some of the system components can be presented with which the users can interact. These components will have limited functionality but can be used to present examples that the user can work through. However, CBTs can be costly to produce, they are also poor at explaining things such as the Method of Operations, procedures or any changes in the current sectorisation to be used for the validation exercise. Presentations or 'classroom' sessions where controllers may ask questions are much better at this. However, presentations should be kept short and with breaks so that the controllers can consider what they have learned.

Illustration runs are where members of the Validation Team use the system while the controllers watch and ask questions. Illustration runs should be followed by mentored runs. Mentored runs are where the controller uses the system and a member of the Validation Team, or someone experienced in the use of the system, sits with the controller(s) and explains how to use the system. As the controllers learn the system, the mentors can then 'sit back' and allow the controllers to explore the system. As far as possible, the time spent on the system should be maximised within the allotted training time and shakedown trials (see VGH Activity 2.5) should be used as a further opportunity to train the participants.

In order to minimise the time spent during the validation exercise on training, the training literature and CBTs should be sent to participants before the trial begins. (It is recommended that they be sent out approximately 3 weeks before the start of the trial). The Validation Team should check whether the controllers have had time to complete the training, and have understood the material. This may be done through observation, debriefs or questionnaires. These checks will help to ensure that the controllers arriving for the trial, have a common level of knowledge. The training should be followed by a training debrief and/or questionnaire so that any lessons can be learnt for future validation exercises. It is also useful to have a record of how well the controllers feel they have learnt the system before the validation exercise begins. This may then be fed into the analysis.

Example

The following example of a training plan is taken from a previous trial run by NATS.

The shakedown trial will be used as training. Of the 11 controllers participating in the trial 7 are available for the shakedown and 3 have been involved in previous simulations. A CBT will be produced, this and training literature including mock Temporary Operating Instructions will be sent out to controllers three weeks prior to the start of trial. During the main trial there will be two days set-aside for training prior to the start of the validation exercise. The training will include:

- briefings on:

- the background to the trial;
- the measurements to be taken;

- the method of operations;
- two illustration runs;
- mentored runs;
- a refresher CBT session.

3.4. Measurements

The measurements section of the MAS will describe the measurements to be used to assess each low-level validation objective. A number of measurements could contribute to more than one objective. The process of analysing these measurements is described in Section 3.5.

The low-level and subsidiary validation objectives, identified in VGH Activity 1.4, should be stated under the heading of the related high-level validation objective or research aim. In a fast-time simulation, there may be no need to specify low-level objectives. The hypotheses developed in VGH Activity 1.7 should also be noted along with a description of the measurements that are to be recorded (VGH Activity 1.6). For detail on measurements that can be made in fast-time and real-time trials, Appendix 5 of the VGH should be read.

Example

The following example contains two low-level objectives, a corresponding hypothesis and one of the measurements that may be taken in order to meet the objective.

Low-level objective 1: Determine whether the overall level of planner workload is changed by the introduction of the prototype.

Low-level objective 2: Determine whether the overall level of tactical workload is changed by the introduction of the prototype.

Hypothesis: H0 The workload is the same in each of the organisations.

H1 The workload is different in each of the organisations.

Measure	Definition of measure	Role	Notes
ISA	The mean ISA value per role per sector per organisation and run.	Planner and Tactical	This will show how each role differs in terms of workload (ISA) between the organisations. ISA is recorded every two minutes and these values will then be averaged to produce a mean ISA per exercise per role.

Table 3-1 Example of a Definition of a Measure

In the example shown in Table 2-1 the data will be aggregated over three factors: role; sector; organisation.

Factors are independent variables (ie those that can be controlled), which have two or more levels, for instance, in the example above, organisation may have two levels one for the baseline and another for the advanced system. The level of each factor in each run will be predetermined and specified in the experimental design (VGH Activity 2.1).

Aggregating data is then a technique that can be utilised to ease comparison between levels of factors; it may also reduce autocorrelation and have the effect of 'normalising' the data, which will further ease inference (Section 3.5.3). However, summarising the data in this way means that there is loss of information about the variability of the data. Striking a balance between having as many data points as possible and satisfying any requirements the analysis may have for independence or normality, can be a matter of statistical judgement.

3.5. Analytical Methods

The experimental design defined in VGH Activity 1.8 of Step 1 and refined in VGH Activity 2.1 of Step 2 will impact on the analysis methods used. The operational and statistical significance levels determined in VGH Activity 1.9 will be stated here, with an explanation of the reasons why these levels were chosen.

Example

The level of significance to be used in the statistical tests has been chosen to be 95%. This level was decided, based on previous experience of simulations, in order to have an acceptable risk of a false positive.

3.5.1. Selection of Analytical Approach

The selection of analytical approach section of the MAS will describe the structure of the analysis process. It is assumed that the team conducting the trial has experience of data analysis methods. The aim of this section is to give an idea of the type of tasks that this part of the exercise will involve, what decisions are involved and what the output should be.

There are three principle phases of the data analysis:

- *exploration of the data;*
- *inference about the objectives;*
- *synthesis of conclusions concerning the objectives.*

Before these three areas are explored in detail there are a number of considerations that should be kept in mind when preparing for the trial and when performing the analysis:

- *Statistical tests:* The type of data that can be gathered from the trial will determine the types of exploratory and analysis technique that are available. The analysis methods should be chosen to suit the experimental design and consequently the data; it is not possible to force the data to fit a particular analysis technique.
- *Quantitative versus qualitative data:* Although there are formal, quantitative approaches to analysing qualitative data, such as questionnaire responses, it can often be difficult to obtain the wide range of respondents that

these methods may require. So, whilst qualitative data contain important information, especially about the acceptability of the prototype to controllers, there will always be a need for quantitative data to be gathered if a quantitative conclusion about validity is to be drawn.

- *Objectivity:* Although much of the data gathered is subjective, and the analysis process involves the exercise of judgement, the process can still have a degree of objectivity. This is achieved by taking a systematic approach to analysing all of the data, by presenting results which may or may not be consistent with the final conclusion and by presenting the arguments for reaching the final conclusion from this (possibly self-inconsistent) set of results.
- *Auditability:* The whole analysis process must be auditable, so that, in particular, key conclusions can be traced back to the raw data. This allows the possibility of errors in the analysis to be checked and the sensitivity of the conclusions to assumptions in the analysis to be tested.

3.5.2. Exploration

Exploratory data analysis, or descriptive statistics, is the process of understanding the data that have been gathered:

- looking for gaps and outliers;
- looking for relationships that might be of interest;
- looking for potential distributions which might explain the data;
- categorising and grouping observations and comments.

At its simplest, it involves gathering and plotting the data and looking at the results. The two key tools for exploration are graphs and contingency tables (ie tables cross classifying frequencies of occurrence). Contingency tables apply just as well to questionnaire responses (eg number in favour, neutral, or against a statement), as to traffic analysis (eg number of cruise, climbing, descending aircraft against meeting or not the accuracy requirement, or re-coordination).

Many other methods of categorising and breaking down the data will be available. For example, the data could be separated for individual controllers, controller role or for each sector if there is more than one. Data from some of these categories may require preliminary analysis to find suitable groupings, eg different traffic patterns or weather conditions. Trial to trial tabulations might also prove useful.

The types of decision to be made in this phase are:

- Which of the data sets or data points should be included in the analysis and which excluded? Are there sound reasons for actually excluding some data rather than merely a wish to exclude any measurements that would refute the desired outcome of the trial?
- Are there unexpected results that need further investigation?



- Is there more data which needs to be extracted from the trial records that would help the analysis?
- And principally
- What are the main messages coming from the quantitative data?

The emphasis in this analysis process on picturing and tabulating the data should mean that the analysis at an early stage is being conducted in terms of operational significance. Appendix 6 of the VGH and statistical text books provide further detail on the methods that can be specified in this part of the MAS

Example

The MAS should state that exploratory analysis will be undertaken and the types of methods that will be used, for example:

For each measure, descriptive statistics, usually in the form of a graph or histogram will be produced prior to the statistical tests being carried out, giving an indication of the scale and range of each measure.

3.5.3. Inference

Statistical inference is concerned with two types of problem: estimation of population parameters and tests of hypotheses. It allows the analyst to draw conclusions about a large group of subjects based on measurements taken from a small sample. In particular measurements taken under baseline experimental conditions can be compared to measurements taken under advanced experimental conditions and conclusions drawn about how and to what extent they differ. The level of confidence about that conclusion may also be calculated and stated. The level of confidence relates to the fact that there is always the possibility that a difference between the measurements may simply occur by chance.

The emphasis placed on the inference stage of the analysis will depend on the type of trial and data gathered. For simulations where no baseline is simulated, i.e. shadow-mode and operational trials, or where there is a lack of data, it is likely that the exploratory stage will be more important than the inferential one in developing ideas as to what are the quantitative results from the trial.

Ref. 3 discusses how shadow-mode trials may be thought of as a survey rather than an 'experiment', because "the experimenter cannot control the conditions in which the trial takes place, eg in terms of traffic patterns and weather etc, but also in terms of the baseline". The trial should then be designed (see VGH Activities 1.8 and 2.1) in terms of a statistical survey. Statistical inference may then be carried out on contingency tables produced from the data.

Where there is a lack of data two effects will be observed, either:

- the probability that any observed differences could have occurred by chance will be high;

- or the probability of observing a statistically significant difference is reduced.

Using the sample mean and variance it is possible to work out the size of the sample that would have been needed in order to produce a result at the level of precision required. This level should then be stated as part of the analysis. Further detail on this aspect of the MAS may be found in statistical text books.

Example

The example below shows how the inference stage of statistical analysis may be described in an MAS.

Two-sided tests will be used throughout the analysis. If significance testing determines a significant difference between the two systems, the direction and scale of the difference will be reported.

The statistical tests used are the Friedman and Kruskal-Wallis tests. These are standard statistical tests for use with ordinal and non-parametric data, (i.e. data for which no a priori assumption about the distribution can be made). The Friedman test requires matched data, but Kruskal-Wallis does not.

Where no statistically significant difference is found, the 'power' of the test will be estimated. The power provides an estimate of the confidence that can be attached to the hypothesis that two samples do in fact come from the same population. Additionally, the percentage difference between the medians of the samples will be calculated and compared with the percentage difference that would have been required for a statistically significant difference to be concluded.

3.5.4. Synthesis

The final stage of analysis, that of synthesis, may not always be specified in the MAS, however it is an important part of the analysis process so is described here for completeness.

It is only at the synthesis stage that the individual measurements are brought together to address each objective. Synthesis is then the iterative process of finding the consensus amongst the quantitative and qualitative results for each objective. Therefore this is the stage in which the analyst tries to pull together messages from the various measures/metrics into a coherent account of what actually happened during the trial and what it means to the trialled system. At first glance the data may appear contradictory; however, on closer examination a deeper understanding will evolve enabling the data to be better explained. The analyst should be careful to build the whole picture and not just focus on the one or two outspoken individuals. The effects of the following issues on the data gathered should also be considered:

- stability of the system;
- learning effects;
- controller experience;
- unexpected use of the system;
- unexpected problems during the exercise.

Example

If a statement on synthesis is to be included in the MAS, it could be described using the wording below.

After the initial analysis covered in the previous sections, there will be a stage of synthesis where a coherent assessment is generated from the individual results relating to each objective.

4. CONCLUSION

This document has provided a template from which an MAS may be produced. It covers advice on the restatement of the validation aims, with further specification of controller requirements and associated training issues, measurements and analytical methods. These analytical methods include exploration of the data, inference about the data and synthesis of results. If the reader requires a full example of an MAS for a validation exercise, it is recommended that Reference 2 is used.