Description of the Expert elicitation Results for the ATC Performance Model

O Morales-Napoles, D Kurowicka, RM Cooke
EWI, TU Delft,
March, 2007

Quantification of the ATC Performance Model.

This document reports briefly on the final quantification of the ATC Error Model. The model is presented in Figure 1. A summary of the variables entered in the model is presented in table 1. The (un)conditional rank correlations shown in Figure 1 have been elicited through one conditional probability of exceedence [2] and ratios of unconditional rank correlations. These have been translated to conditional probabilities of exceedence later and combined through the classical method for expert judgment [1]. In total, six experts[1] have been elicited but only five used for the analyses. Expert number six presented estimates that were inconsistent and his/her estimates could not be used. Next, results of the expert judgment exercise will be discussed.

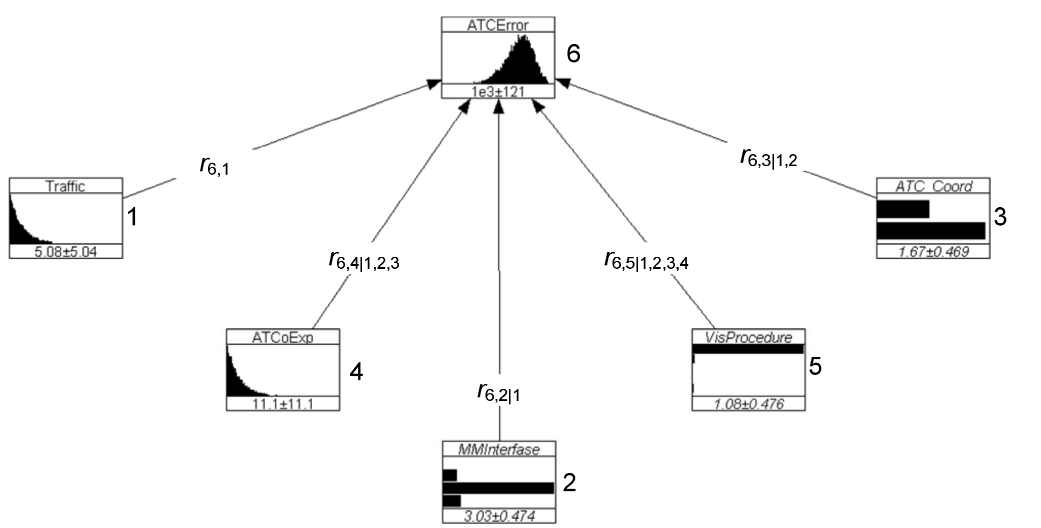

Figure 1. ATC Performance Model.

[1] All experts were area controllers and hence this model may be regarded only as an ATC-area control error probability model.

Name	Description	Source
Traffic	Number of aircraft (any type) simultaneously under control.	Data
MMInterface	Four states variable. From 1- using radio only to 4-using radio, primary and secondary radar and additional tools.	Data
ATC_Coord	1 - The communication with other ATCos takes place in the same room; 2 - The communication with other ATCos does not take place in the same room.	Data
AtCoExp	Number of years working as an ATCo in the same position.	Data
VisProcedure	Five states variable. From 1 - normal operations to 5 - operations below 200 meters visibility.	Data
ATCError	Likelihood that the ATC control will make an error of a given kind	Data, DNV

Table 1 ATC performance Model Variables.

The Classical Model For Expert Judgment.

The so-called Classical Model [1] for deriving uncertainty distributions over model parameters from expert judgements will be used. The Classical model is a performance based linear pooling or weighted averaging model. The weights are derived from experts' calibration and information performance, as measured on calibration or seed variables. These are variables from the experts' field whose values become known to the experts *post hoc*. Seed variables serve a threefold purpose: (i) to quantify experts' performance as subjective probability assessors, (ii) to enable performance-optimised combinations of expert distributions, and (iii) to evaluate and hopefully validate the combination of expert judgements.

Calibration measures the statistical likelihood that a set of experimental results corresponds, in a statistical sense, with the experts' assessments. In particular, the calibration score is the p-value of a standard Chi-square goodness of fit test. Loosely, the calibration score is the probability that the divergence between the expert's probabilities and the observed values of the seed variables might have arisen by chance. A high score (near one, but bigger than, say, 0.05) means that the expert's probabilities are statistically supported by the set of seed variables. Information represents the degree to which an expert's distribution is concentrated, relative to some user-selected background measure. The overall information score is the mean of information scores for each variable.

"Good expertise" corresponds to good calibration (high statistical likelihood) and high information. Highly informative distributions are desirable only if they are well calibrated. Information should discriminate between more or less equally well calibrated experts. The weights in the classical model are proportional to the product of statistical likelihood and information, and it turns out that this product is indeed dominated by calibration. When a combined expert has been formed, we can also measure the calibration and information of this combined expert. For more detail see [1]. Calculations are performed with the EXCALIBUR software developed at TUDelft.

Expert's Performance.

Table 2 below, shows the calibration and information scores for the five experts used in the study. The first and second columns give the expert number and id respectively; the third column gives the calibration score. The ratio of highest to lowest score is about 3.26E+7. It will be noted that only expert A had a score corresponding to a p-value above 5%. Scores of Experts D and E are marginal and for expert C is rather low. Calibration scores in the order 0.001 would fail to confer the requisite level of confidence in the results.

The information scores for all items and for calibrations items are shown in columns 4 and 5 respectively. It will be noted that the overall information scores are quite similar, within a factor 3. In this case the expert with the best calibration score (nr 1) also has the lowest information score for the calibration variables which is a recurrent pattern. The last column gives the "un-normlaized weight"; this is the product of columns 3 and 5. If this column were normalized and used to form weighted combinations, experts A, B and C would be influential with (52.07, 46.11 and 1.80 per cent respectively).

Results of scoring experts
Bayesian Updates: no Weights: global DM Optimisation: yes
Significance Level: 0.00131 Calibration Power: 1

Nr.	Id	Calibr.	Mean relati total	Mean relati realizatioo	Numb real	UnNormalize weight
1	A	0.1012	0.5633	0.5034	10	0.05095
2	B	0.04706	1.03	0.9588	10	0.04512
3	C	0.00131	1.423	1.349	10	0.001767
4	D	2.795E-009	1.669	1.655	10	0
5	E	2.501E-006	1.017	0.9624	10	0

Table 2. Expert's Performance.

Combinations Schemes.

In this exercise, experts give their uncertainty assessments on calibration variables in the form of 5%, 50%, and 95% quantiles. To combine all experts' assessments into one uncertainty assessment there are three combination schemes. The combined distributions are weighted sums of the individual experts' distributions, with non-negative weights adding to one. Different combination schemes are distinguished by the method according to which the weights are assigned to densities. These schemes are designated "Decision Makers". Three kinds of decision makers are described below.

Equal Weight Decision Maker.

The equal weight decision maker (EWDM) results by assigning equal weight to each density. In Table 3 the 6th expert is identified as "EWDM". As it may be seen in table 3 the EWDM is better calibrated than each expert individually. However information scores derived from the EWDM are poor. They are the lowest amongst the 6 experts (that is including the EWDM as an expert) in both all variables and calibration questions alone. Weights for each expert and for the decision maker are shown in the last three columns. Since the equal weight combination has been chosen in this case, column 8 displays the same weight for all experts except the decision maker, and finally column 9 shows that when normalizing the numbers in column 7 the decision maker's weight is 1.66 times smaller than the best calibrated expert.

Results of scoring experts
Bayesian Updates: no Weights: equal DM Optimisation: no
Significance Level: 0 Calibration Power: 1

Nr.	Id	Calibr.	Mean relati total	Mean relati realizatioo	Numb real	UnNormalize weight	Normaliz.we without DM	Normaliz.we with DM
1	A	0.1012	0.5633	0.5034	10	0.05095	0.2	0.3964
2	B	0.04706	1.03	0.9588	10	0.04512	0.2	0.351
3	C	0.00131	1.423	1.349	10	0.001767	0.2	0.01375
4	D	2.795E-009	1.669	1.655	10	4.625E-009	0.2	3.598E-008
5	E	2.501E-006	1.017	0.9624	10	2.407E-006	0.2	1.872E-005
6	EWDM	0.1242	0.2662	0.2472	10	0.0307		0.2388

Table 3. Expert's and EW Decision Maker's Performance.

Item and Global Weight Decision Maker.

This kind of decision maker is from the class of performance based decision makers that are those where the weights are based on the experts' performance on calibration variables. Two performance based decision makers are supported in the software EXCALIBUR. The "global weight" decision maker and the "item weight" decision maker.

The global weight decision maker (Table 4) uses average information over all calibration variables and computes one set of weights for all items. It computes performance based weights which are defined, per expert, by the product of expert's calibration score and his(her) overall information score on calibration variables, and by an optimization procedure[2]. In this case, all experts with a calibration score less than the significance level found by the optimization procedure are outweighed as reflected by the zeros in columns 7, 8 and 9 in Table 4.

In table 4 one can see that after the optimization procedure is applied, 3 experts have non-zero weight. Comparing tables 3 and 4 one can see that the calibration score of the GWDM is about 3 times higher than the EWDM. The information score is comparable for both decision makers in both all variables and calibration variables alone.

[2] The weights used in the classical model are *proper scores*. A score is called *strictly proper* if an expert can achieve his highest expected score by and only by stating his true opinion. The measures for information and calibration described above must be combined in such a way that the result is (in the long run) a strictly proper scoring rule. This requires that the measurement of calibration be combined with "classical significance tests". Briefly, there must be some value $\alpha > 0$, such that if the expert's statistical likelihood drops below α, his/her weight becomes zero. For each value of α it is defined a decision maker dm_α computed as a weighted linear combination of the experts whose calibration score exceeds α. dm_α is scored with respect to calibration and information. The weight which this dm_α would receive if he were added as a "virtual expert" is called the "virtual weight" of dm_α. The value of α for which the virtual weight of dm_α is the greatest is chosen as the cut-off value for determining the unweighted expert.

Results of scoring experts and Relative Information to the DM
Bayesian Updates: no Weights: global DM Optimisation: yes
Significance Level: 0.00131 Calibration Power: 1

Nr.	Id	Calibr.	Mean relati total	Mean relati realizatioo	Numb real	UnNormalize weight	Normaliz.we without DM	Normaliz.we with DM	Rel.Inf to total	Rel.Inf to realiz.
1	A	0.1012	0.5633	0.5034	10	0.05095	0.5208	0.2015	0.379	0.4075
2	B	0.04706	1.03	0.9588	10	0.04512	0.4612	0.1784	0.578	0.5823
3	C	0.00131	1.423	1.349	10	0.001767	0.01806	0.006987	1.135	1.156
4	D	2.795E-009	1.669	1.655	10	0	0	0	1.458	1.55
5	E	2.501E-006	1.017	0.9624	10	0	0	0	1.229	0.9386
6	GWDM	0.6827	0.3094	0.2271	10	0.1551		0.6131	0	0

Table 4. Expert’s and GW Decision Maker’s Performance.

Results of scoring experts and Relative Information to the DM
Bayesian Updates: no Weights: item DM Optimisation: yes
Significance Level: 0.00131 Calibration Power: 1

Nr.	Id	Calibr.	Mean relati total	Mean relati realizatioo	Numb real	UnNormalize weight	Normaliz.we without DM	Normaliz.we with DM
1	A	0.1012	0.5633	0.5034	10	0.05095		0.2688
2	B	0.04706	1.03	0.9588	10	0.04512		0.238
3	C	0.00131	1.423	1.349	10	0.001767		0.009322
4	D	2.795E-009	1.669	1.655	10	0		0
5	E	2.501E-006	1.017	0.9624	10	0		0
6	IWDM	0.2441	0.4441	0.3757	10	0.0917		0.4838

Table 5. Expert’s and IW Decision Maker’s Performance.

The "item weight" decision maker constructs weights for each item separately, using the experts' information scores for the given item, rather than the average information score. The calibration score of the GWDM is about 2.8 times larger than the IWDM. The IWDM is slightly more informative than the GWDM. However the gain in information is not a sufficient argument to justify a preference of the IWDM over the GWDM.

Next, Figure 2 presents range graphs for experts' assessments in calibration variables. The GWDM is also shown for comparison. The 5[th] and 95[th] percentile of each expert's uncertainty distribution is represented by '[' and ']' respectively. The median is shown as '*' and the realization as '#'.

```
                          Range graph of input data
Item no.:   1 Item name: Airprox_rep Scale: LOG
Experts
  1 [-------------------------------*---------------------]
  2                                                  [--------*---]
  3                                           [------*------------]
  4                                       [-------*--]
  5                                           [------------------------*------]
GWDM  [===============================================*===========]
Real::::::::::::::::::::::::::::::::::::::::::::::::::::::::::#::::::::::::::::
                                                              25
     1                                                                      60

Item no.:   2 Item name: Missed_rate Scale: LOG
Experts
  1 [--------------------------------------------*----------]
  2                                                                [---*----]
  3                                                                [-*---]
  4                                                                    [----*-]
  5                                                         [--------*---]
GWDM   [====================================================*==============]
Real::::::::::::::::::::::::::::::::::::::::::::::::::::::#::::::::::::::::::::
                                                        0.0011
    1E-009                                                                 0.2

Item no.:   3 Item name: Infringements Scale: LOG
Experts
  1    [----------------------------------------------*-----------------------]
  2    [-----------*------------]
  3 [-----------*-------------------------------]
  4                                                          [----*---]
  5    [-------------------------------------*--------]
GWDM  [====================*================================================]
Real:::::::::::::::::::::::::::::::::::::::::::::::::::::::#:::::::::::::::::::
                                                           214
     4                                                                     900

Item no.:   4 Item name: Fleet Scale: LOG
Experts
  1 [-------------------------------*--------------------------]
  2                       [---------*----------------------]
  3                [---------*-]
  4                                       [---*-----]
  5                                                 [--------*----------------]
GWDM  [=============================*=========================]
Real:::::::::::::::::::::::::::::::::#:::::::::::::::::::::::::::::::::::::::::
                                  1.084E004
    1000                                                                 2E005
```

```
Item no.:   5 Item name: Readback Scale: LOG
Experts
  1                           [-----------------------------------------*-----]
  2                                                         [------------*----]
  3                                                   [------*-------]
  4                                                              [------*-]
  5                           [---------------*--------------]
GWDM                            [=======================================*=====]
Real#::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
      1
     1                                                                      95

Item no.:   6 Item name: Overrun_fatal Scale: LOG
Experts
  1 [-------------------------------------------------*-----------------------]
  2                            [--------------------*------------------------]
  3                                           [-------------*----]
  4                             [--------*------------------]
  5                                      [-----------------------*------------]
GWDM   [=============================================*========================]
Real::::::::::::::::::::::::::::::::::::::::::::::::::#::::::::::::::::::::::::
                                                      6
    0.1                                                                     50

Item no.:   7 Item name: BadTO Scale: LOG
Experts
  1 [---------------------------------------------------------*---------------]
  2                                [--------------------------*-------]
  3                                [--------------------------------*----]
  4                                                         [-------*-----]
  5                                                      [------*-----]
GWDM     [====================================================*===============]
Real:::::::::::::::::::::::::::::::::::::::::::::#:::::::::::::::::::::::::::::
                                                20
     2                                                                      99

Item no.:   8 Item name: FirstGen Scale: LOG
Experts
  1 [------------------------------------------------*------------------------]
  2                                                  [--------*-----------]
  3                                       [----------*-----]
  4                                                           [----*------]
  5                                            [-----------*----]
GWDM   [====================================================*================]
Real::::::::::#::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
             0.23
    0.1                                                                     40

Item no.:   9 Item name: Overrun_TO Scale: LOG
Experts
  1                               [--------*---------]
  2 [-------------------------------------------------------------------*-----]
  3                               [---*----]
  4                                   [----*--]
  5                                                  [---------*--------------]
GWDM     [====================================*===============================]
Real:::::::::::::::::#:::::::::::::::::::::::::::::::::::::::::::::::::::::::::
                  6.5E-008
    1E-009                                                                0.05

Item no.:  10 Item name: Incursion Scale: LOG
Experts
  1                         [---*-]
  2 [-----------------------------------------------------------------*------]
  3                                 [-----*---]
  4                                                                  [------*-]
  5                                       [--------*---------------]
GWDM     [========================*=========================================]
Real::::::::::::::::::::::::::::::::::::::::#::::::::::::::::::::::::::::::::::
                                        1.6E-005
    1E-009                                                                 0.1
```

Figure 2. Range Graphs for expert's assessments in calibration variables.

The predominant picture is that the experts' central 90% confidence, generally show considerable overlap. To appreciate Figure 2 numerical table 4 also compares each expert to the GWDM. In discrepancy analysis, the relative information of each expert with respect to the Decision Maker is computed per item. These scores are averaged over all variables (column 10) and calibration variables only (column 11). It can be observed that experts tend to agree with the GWDM. The ratio of columns 10 to 11 is close to one in each case.

In summary the recommended choice of the decision maker is the Global Weight Decision Maker as it achieves better performance than the Equal weight and Item Weight combinations. Future analysis will be performed also based on the GWDM.

Combining Experts' Dependence Opinions.

As explained before, in the classical method for expert judgment, the marginal distributions of experts $e = 1,\ldots,N$ are combined through a weighted average. In equation (1) $f_{Y,DM}$ is the density of elicitation variable Y of the DM, w_e the weight for expert e and $f_{Y,e}$ the density of expert e on variable Y. The weights can be chosen as equal weights or derived from performance measures.

$$f_{Y,DM} = \sum_{e=1}^{N} w_e f_{Y,e} \tag{1}$$

From previous analysis it was concluded that the GWDM would be the combination choice. The combination of experts' dependence estimates derives from equation (1) above with some modifications. A questionnaire was designed to elicit the dependence information from figure 1. This questionnaire included 12 calibration variables from which only 10 could be used for the performance measures. Additionally 6 questions were asked to obtain the rank and conditional rank correlations required by the model in figure 1. These 6 questions were essentially the following:

1. A ranking of variables assigning 1 to the one with the largest influence (highest absolute rank correlation coefficient) on ATC errors until 5 for the one with the smallest absolute rank correlation. For each expert variables were labeled as $v_1,\ldots, v_5$. Variable v_6 would denote ATC errors for all experts. Notice that $v_1,\ldots, v_5$ could denote different variables for each expert according to their own beliefs.
2. A Probability of exceedence for the '*most influential*' variable was asked to each expert in the usual way. For example: suppose variable v_1 was found to lie above its median value[3] what is your probability that also v_6 would be above its median value. This would give us the rank correlation r_{v_6,v_1}. This rank correlation is maximal in absolute value over all 5 rank correlations required in the correlation matrix.
3. Next the following 4 ratios were elicited: $r_{v_6,v_j} / r_{v_6,v_1}$ for $j = 2,\ldots,5$. Observe that the ratios should be decreasing as j increases.

[3] Or some other percentile if the variable is discrete

This information together with the marginal distributions for each node is enough to compute the (un)conditional rank correlations required by the model. The results of asking the 6 questions above are presented in table A1 in Appendix 1. All (un)conditional rank correlations were computed from conditional probabilities of exceedence as in [2] assuming a normal copula to construct the joint distribution.

After the estimates from each expert were obtained, they were combined with the procedure described in Appendix 2. The general idea consists basically in taking a weighted average of each expert's individual assessments for a given probabilistic statement concerning the joint distribution. The weights were taken from the GWDM in table 4. This procedure is described in Appendix 2.

Probability[4]		(Un)conditional rank correlation	
P_1	0.48	$r_{v_6,v_4 \mid v_1,v_2,v_3}$	-0.060
P_2	0.44	$r_{v_6, \mid v_1}$	-0.179
P_3	0.47	$r_{v_6,v_2 \mid v_1}$	-0.210
P_4	0.58	$r_{v_6,v_3 \mid v_1,v_2}$	0.180
P_5	0.52	$r_{v_6,v_5 \mid v_1,v_2,v_3,v_4}$	0.020

Table 6. GWDM Dependence Information

Table 6 presents the final results of applying the procedure described in Appendix 2 for combining dependence information to the expert's individual estimates in table A1. The determinant of the correlation matrix of the BBN in Figure 1 with the estimates in Table 6 is equal to 0.89 showing that the GWDM considers a very small linear relationship of ATC error with other variables in table 1. This may be explained by observing that experts A and B are dominant in the performance based combination. For expert A the highest rank correlation with ATC error was communication and coordination equal to 0.22 ($r_{6,1}$ = 0.22 in table A1). For expert B the highest was -0.34 with man-machine interface ($r_{6,1}$ = -0.34 in table A1). This explains the high value of the determinant of the correlation matrix of the GWDM. Expert C has the lowest determinant for the correlation matrix amongst all 5 experts but his/her contribution is marginal to the performance based combination. Experts D and E do not contribute to the GWDM.

Conclusions.

This report shows that the elicitation and optimal combination of expert judgments in the form of rank and conditional rank correlations for real applications is possible. The optimal combination regards variables 1 – 5 as showing strong non-monotonic relationships with ATC error. Large differences with respect to tower controllers may be expected.

---

[4] $P_1 = P$(ATC error $\geq$ median | Experience $\geq$ median), $P_2 = P$(ATC error $\geq$ median | Traffic $\geq$ median), $P_1 = P$(ATC error $\geq$ median | MMI $\geq$ 4), $P_1 = P$(ATC error $\geq$ median | Communication & Coordination = 2), $P_1 = P$(ATC error $\geq$ median | Visibility Procedure $\geq$2)

Basic bibliography.

[1] Cooke R.M. "*Experts in Uncertainty: Opinion and Subjective Probability in Science*" in Environmental Ethics and Science Policy Series, Oxford University Press, June 1991.

[2] Morales Napoles, O. "*Eliciting conditional and unconditional rank correlations from conditional probabilities*" Reliability Engineering & System Safety on press doi:10.1016/j.ress.2007.03.020.

Appendix 1

A

Variable	Rank	Ratio	Probability[5]		(Un)conditional rank correlation	
Experience	4	0.20	P_1	0.515	$r_{v_6,v_4\|v_1,v_2,v_3}$	0.05
Traffic	2	-0.40	P_2	0.471	$r_{v_6,v_2\|v_1}$	-0.09
MMI	3	-0.20	P_3	0.494	$r_{v_6,v_3\|v_1,v_2}$	-0.05
Comm. & Coord.	1		P_4	0.6	r_{v_6,v_1}	0.22
Visibility Procedure	5	0.20	P_5	0.539	$r_{v_6,v_5\|v_1,v_2,v_3,v_4}$	0.05

B

Variable	Rank	Ratio	Probability		(Un)conditional rank correlation	
Experience	3	-0.50	P_1	0.443	$r_{v_6,v_3\|v_1,v_2}$	-0.19
Traffic	2	-0.75	P_2	0.414	$r_{v_6,v_2\|v_1}$	-0.27
MMI	1		P_3	0.45	r_{v_6,v_1}	-0.34
Comm. & Coord.	4	0.40	P_4	0.562	$r_{v_6,v_4\|v_1,v_2,v_3}$	0.16
Visibility Procedure	5	0.00	P_5	0.5	$r_{v_6,v_5\|v_1,v_2,v_3,v_4}$	0.00

C

Variable	Rank	Ratio	Probability		(Un)conditional rank correlation	
Experience	3	-0.65	P_1	0.374	$r_{v_6,v_3\|v_1,v_2}$	-0.54
Traffic	1		P_2	0.3	r_{v_6,v_1}	-0.57
MMI	2	-0.70	P_3	0.442	$r_{v_6,v_2\|v_1}$	-0.50
Comm. & Coord.	4	-0.60	P_4	0.341	$r_{v_6,v_4\|v_1,v_2,v_3}$	-0.60
Visibility Procedure	5	0.00	P_5	0.5	$r_{v_6,v_5\|v_1,v_2,v_3,v_4}$	0.00

D

Variable	Rank	Ratio	Probability		(Un)conditional rank correlation	
Experience	4	-0.25	P_1	0.442	$r_{v_6,v_4\|v_1,v_2,v_3}$	-0.28
Traffic	3	0.26	P_2	0.56	$r_{v_6,v_3\|v_1,v_2}$	0.28
MMI	1		P_3	0.4	r_{v_6,v_1}	-0.69
Comm. & Coord.	2	0.40	P_4	0.628	$r_{v_6,v_2\|v_1}$	0.39
Visibility Procedure	5	0.10	P_5	0.562	$r_{v_6,v_5\|v_1,v_2,v_3,v_4}$	0.12

E

Variable	Rank	Ratio	Probability		(Un)conditional rank correlation	
Experience	1		P_1	0.4	r_{v_6,v_1}	-0.30
Traffic	3	0.20	P_2	0.52	$r_{v_6,v_3\|v_1,v_2}$	0.07
MMI	2	0.70	P_3	0.531	$r_{v_6,v_2\|v_1}$	0.22
Comm. & Coord.	4	0.10	P_4	0.514	$r_{v_6,v_4\|v_1,v_2,v_3}$	0.03
Visibility Procedure	5	0.03	P_5	0.508	$r_{v_6,v_5\|v_1,v_2,v_3,v_4}$	0.01

GWDM

Variable	Rank	Ratio	Probability		(Un)conditional rank correlation	
Experience		-	P_1	0.48	$r_{v_6,v_4\|v_1,v_2,v_3}$	-0.060
Traffic		-	P_2	0.44	r_{v_6,v_1}	-0.179
MMI		-	P_3	0.47	$r_{v_6,v_2\|v_1}$	-0.210
Comm. & Coord.		-	P_4	0.58	$r_{v_6,v_3\|v_1,v_2}$	0.180
Visibility Procedure		-	P_5	0.52	$r_{v_6,v_5\|v_1,v_2,v_3,v_4}$	0.020

Table A1. Expert elicitation results for dependence information of the ATC model in Figure 1.

[5] $P_1 = P$(ATC error ≥ median | Experience ≥ median), $P_2 = P$(ATC error ≥ median | Traffic ≥ median), $P_3 = P$(ATC error ≥ median | MMI ≥ 4), $P_4 = P$(ATC error ≥ median | Communication & Coordination = 2), $P_5 = P$(ATC error ≥ median | Visibility Procedure ≥2)

Appendix 2

To combine the dependence information elicited from expert's via conditional probabilities it would be tempting to pool the conditional probabilities linearly to determine the conditional probability of the decision maker. This strategy would work well in general if the medians off all experts were the same which is not typically the case. In order to combine the experts' dependence information a different strategy has to be taken. An example is presented with the BBN in figure A2.

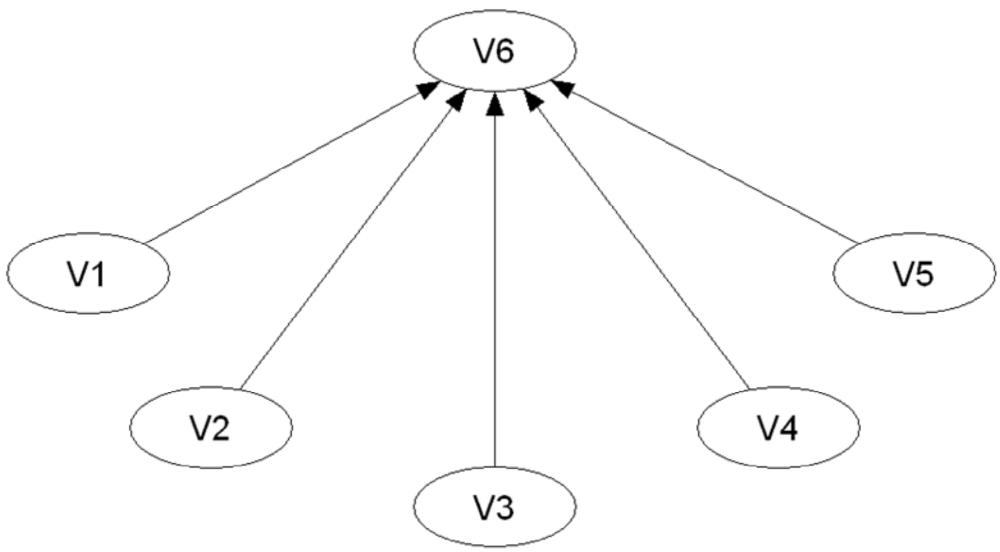

Figure A2. BBN example for the elicitation procedure.

The rank correlations from each expert are elicited as in [2]. Denote each expert as e_i for i = 1,…,N. The rank correlations will be denoted as $r_{6,1}^{e_i}$ for the rank correlation between V6 and V1 in figure A2 for each i. Similarly the conditional rank correlation between V6 and V2 given V1 will be denoted as $r_{6,2|1}^{e_i}$ for expert e_i. All other (un)conditional rank correlations in the BBN will be denoted similarly. The median value of variable V_j for expert e_i is denoted as $v_{j,q_{50}}^{e_i}$. Similarly the k^{th} percentile of variable V_j is denoted as $v_{j,q_k}^{e_i}$. Proceed as follows.

1. Query from each expert e_i some probabilistic statement as in relation (A1) and translate them to the corresponding rank correlations. $F_{V_j}^{e_i}$ is the cumulative distribution function of V_j for expert e_i. Observe that $v_{j,q_{50}}^{e_i}$ will in general be different for each variable and for each expert when the marginal distribution comes from expert judgment. When the marginal distributions come from data then the median value across experts and the decision maker will be the same.

$$
\begin{aligned}
P_1^{e_i} &= P\left(V_6 \geq v_{6,q_{50}}^{e_i} \mid V_1 \geq v_{1,q_{50}}^{e_i}\right) \\
&= P\left(F_{V_6}^{e_i}(V_6) \geq 0.5 \mid F_{V_1}^{e_i}(V_1) \geq 0.5\right) \\
P_2^{e_i} &= P\left(V_6 \geq v_{6,q_{50}}^{e_i} \mid V_1 \geq v_{1,q_{50}}^{e_i}, V_2 \geq v_{2,q_{50}}^{e_i}\right) \\
&= P\left(F_{V_6}^{e_i}(V_6) \geq 0.5 \mid F_{V_1}^{e_i}(V_1) \geq 0.5, F_{V_2}^{e_i}(V_2) \geq 0.5\right) \\
P_3^{e_i} &= P\left(V_6 \geq v_{6,q_{50}}^{e_i} \mid V_1 \geq v_{1,q_{50}}^{e_i}, V_2 \geq v_{2,q_{50}}^{e_i}, V_3 \geq v_{3,q_{50}}^{e_i}\right) \\
&= P\left(F_{V_6}^{e_i}(V_6) \geq 0.5 \mid F_{V_1}^{e_i}(V_1) \geq 0.5, F_{V_2}^{e_i}(V_2) \geq 0.5, F_{V_3}^{e_i}(V_3) \geq 0.5\right) \\
&\;\;\vdots \\
P_5^{e_i} &= P\left(V_6 \geq v_{6,q_{50}}^{e_i} \mid V_1 \geq v_{1,q_{50}}^{e_i}, V_2 \geq v_{2,q_{50}}^{e_i}, \ldots, V_5 \geq v_{5,q_{50}}^{e_i}\right) \\
&= P\left(F_{V_6}^{e_i}(V_6) \geq 0.5 \mid F_{V_1}^{e_i}(V_1) \geq 0.5, F_{V_2}^{e_i}(V_2) \geq 0.5, \ldots, F_{V_5}^{e_i}(V_5) \geq 0.5\right)
\end{aligned}
\tag{A1}
$$

$$
\begin{aligned}
P_1^{e_i} &\rightarrow r_{6,1}^{e_i} \\
P_2^{e_i} &\rightarrow r_{6,2|1}^{e_i} \\
P_3^{e_i} &\rightarrow r_{6,3|1,2}^{e_i} \\
P_4^{e_i} &\rightarrow r_{6,4|1,2,3}^{e_i} \\
P_5^{e_i} &\rightarrow r_{6,5|1,2,3,4}^{e_i}
\end{aligned}
$$

Notice that other probabilistic statements could be elicited in (A1) according to the analysts preference. For example shorter conditioning sets might be considered: $P_1^{e_i} = P\left(V_6 \geq v_{6,q_{50}}^{e_i} \mid V_1 \geq v_{1,q_{50}}^{e_i}\right), \ldots, P_5^{e_i} = P\left(V_6 \geq v_{6,q_{50}}^{e_i} \mid V_5 \geq v_{5,q_{50}}^{e_i}\right)$. Another option would be to elicit a joint distribution $P_1^{e_i} = P\left(V_6 \geq v_{6,q_{50}}^{e_i}, V_1 \geq v_{1,q_{50}}^{e_i}\right), \ldots, P_5^{e_i} = P\left(V_6 \geq v_{6,q_{50}}^{e_i}, \ldots, V_5 \geq v_{5,q_{50}}^{e_i}\right)$ instead of a conditional probability of exceedence[6]. If the medians of each expert differ then the probabilities above are taken over different events and a correction needs to be computed. Notice also that the recommended choice for the percentile used in the probabilities stated is the median however any other percentile $v_{j,q_k}^{e_i}$ may be used. In particular other percentiles are necessary for discrete variables.

2. Take linear pooling of expert's assessments for V_i to determine $F_{V_j}^{DM}$ (The cumulative distribution function of the decision maker) and compute the median value $v_{j,q_{50}}^{DM}$ for the Decision Maker.
3. Compute the probabilities that each expert "*would have stated*" if he/she had been asked about the same quintile (median in the example above) as the Decision Maker (A2) such that his/her estimates for the rank correlations remain unchanged. These probability will be denoted as $P_{1^*}^{e_i}, \ldots, P_{5^*}^{e_i}$

[6] Probabilities of concordance or discordance might be another option.

$$
\begin{array}{lll}
r_{6,1}^{e_i} & \rightarrow & P_{1^*}^{e_i} \\
r_{6,2|1}^{e_i} & \rightarrow & P_{2^*}^{e_i} \\
r_{6,3|1,2}^{e_i} & \rightarrow & P_{3^*}^{e_i} \\
r_{6,4|1,2,3}^{e_i} & \rightarrow & P_{4^*}^{e_i} \\
r_{6,5|1,2,3,4}^{e_i} & \rightarrow & P_{5^*}^{e_i}
\end{array}
$$

$$
\begin{aligned}
P_{1*}^{e_i} &= P\left(F_{V_6}^{e_i}(V_6) \geq F_{V_6}^{e_i}(v_{6,q_{50}}^{DM}) \mid F_{V_1}^{e_i}(V_1) \geq F_{V_1}^{e_i}(v_{1,q_{50}}^{DM})\right) \\
P_{2*}^{e_i} &= P\left(F_{V_6}^{e_i}(V_6) \geq F_{V_6}^{e_i}(v_{6,q_{50}}^{DM}) \mid F_{V_1}^{e_i}(V_1) \geq F_{V_1}^{e_i}(v_{1,q_{50}}^{DM}), F_{V_2}^{e_i}(V_2) \geq F_{V_2}^{e_i}(v_{2,q_{50}}^{DM})\right) \\
P_{3*}^{e_i} &= P\left(F_{V_6}^{e_i}(V_6) \geq F_{V_6}^{e_i}(v_{6,q_{50}}^{DM}) \mid F_{V_1}^{e_i}(V_1) \geq F_{V_1}^{e_i}(v_{1,q_{50}}^{DM}), \ldots, F_{V_3}^{e_i}(V_3) \geq F_{V_3}^{e_i}(v_{3,q_{50}}^{DM})\right) \\
&\;\;\vdots \\
P_{5*}^{e_i} &= P\left(F_{V_6}^{e_i}(V_6) \geq F_{V_6}^{e_i}(v_{6,q_{50}}^{DM}) \mid F_{V_1}^{e_i}(V_1) \geq F_{V_1}^{e_i}(v_{1,q_{50}}^{DM}), \ldots, F_{V_5}^{e_i}(V_5) \geq F_{V_5}^{e_i}(v_{5,q_{50}}^{DM})\right)
\end{aligned} \quad \text{(A2)}
$$

An example of the situation where the expert's estimate and the Decision Maker's median differ is presented in figure A2. Observe that expert's e_1 median for V_1 is 50 while the DM's median for the same variable is 100. In this case the DM's median realizes the 66% percentile of expert e_1 distribution.

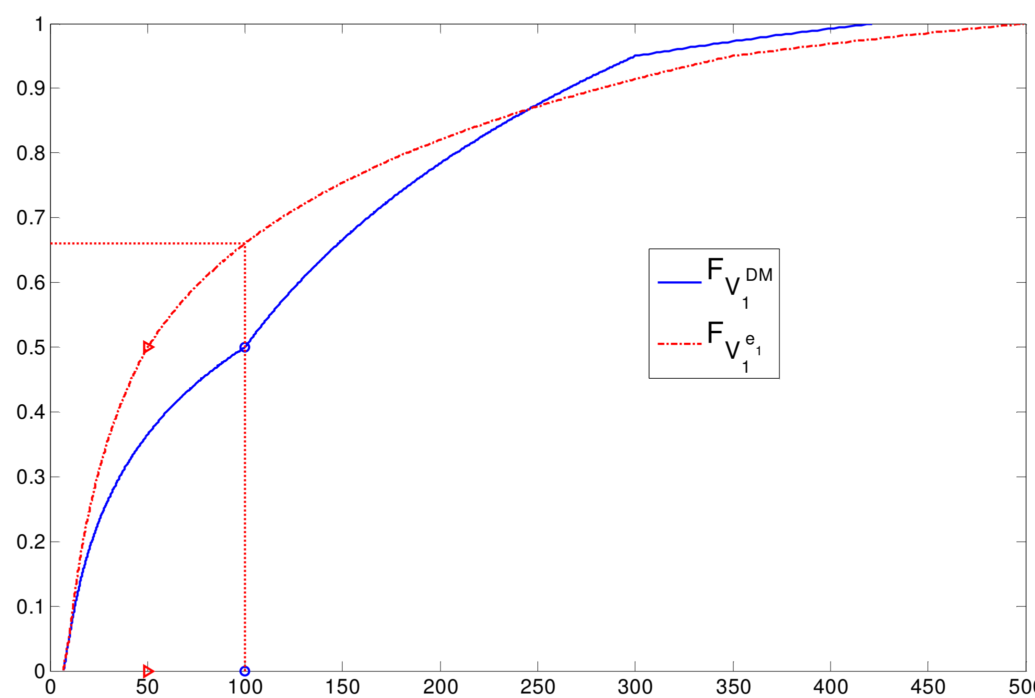

Figure A2. Expert 1 and DM disagreement in the median for V_1.

Suppose for simplicity that both the DM and expert e_1 agree in the median value for variable V_6. Suppose that expert 1 answered:

$$\begin{aligned} P_1^{e_1} &= P\left(V_6 \geq v_{6,q_{50}}^{e_1} \mid V_1 \geq 50\right) \\ &= P\left(F_{V_6}^{e_1}(V_6) \geq 0.5 \mid F_{V_1}^{e_1}(V_1) \geq 0.5\right) = 0.75 \rightarrow r_{6,1} = 0.7 \end{aligned}$$

Then according to Figure A3 $r_{6,1} = 0.7$. This may also be observed in figure A4 in the left hand side. Both V_1 and V_6 have been transformed to uniform and 10,000 samples with correlation 0.7 shown in a scatter plot. The expert's probability is approximately $P_1^{e_1} \approx \#\text{ of points in B}/\#\text{ of points in (B + D)}$. According to the method described above and figure A3, we have:

$$\begin{aligned} P_{1^*}^{e_1} &= P\left(V_6 \geq v_{6,q_{50}}^{e_1} \mid V_1 \geq 100\right) \\ &= P\left(F_{V_6}^{e_1}(V_6) \geq 0.5 \mid F_{V_1}^{e_1}(V_1) \geq 0.66\right) = 0.83 \rightarrow r_{6,1} = 0.7 \end{aligned}$$

In other words the expert 'would have stated' a probability of 0.83 for his/her estimate of dependence to remain constant if asked about the percentile corresponding to the Decision Maker's median. This may be observes in Figure A4 in the right hand side. As before, V_1 and V_6 have been transformed to uniform and the same 10,000 samples with correlation 0.7 shown in a scatter plot. Again $P_{1^*}^{e_1} \approx \#\text{ of points in B}_1/\#\text{ of points in (B}_1 + \text{D}_1)$. A similar kind of correction needs to be applied to all probabilistic statements of all experts in relation (A1) by applying relation (A2).

4. Take linear pooling (A3) of $P_{j^*}^{e_i}$ $i = 1, \ldots, N$ for all j to determine (A4) as explained in [2].

$$P_j^{DM} = \sum_i w_{e_i} P_{j^*}^{e_i} \tag{A3}$$

$$\begin{aligned} P_1^{DM} &\rightarrow r_{6,1}^{DM} \\ P_2^{DM} &\rightarrow r_{6,2|1}^{DM} \\ P_3^{DM} &\rightarrow r_{6,3|1,2}^{DM} \\ P_4^{DM} &\rightarrow r_{6,4|1,2,3}^{DM} \\ P_5^{DM} &\rightarrow r_{6,5|1,2,3,4}^{DM} \end{aligned} \tag{A4}$$

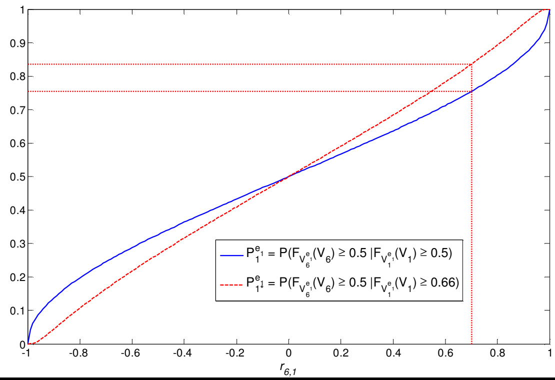

Figure A3. Expert 1 and DM disagreement in the median for V_1.

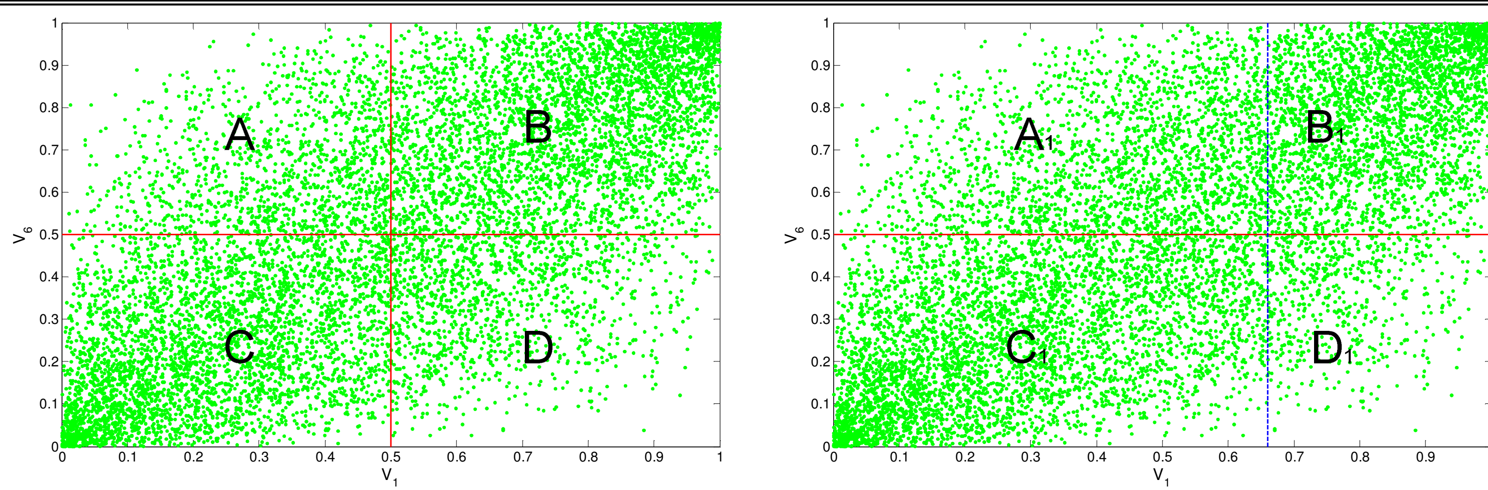

Figure A4. Expert 1 and DM disagreement in the median for V_1 (Rank correlation 0.7)