

Supplementary Online Material for Cross Validation of Classical Model for Structured Expert Judgment

Roger M. Cooke
Resources for the Future
Delft Univ. of Technology Dept Mathematics
November 18, 2016

Table of Contents

1.	Post-2006 Data and Applications Documentation	2
2.	Classical Model Performance Measures and Combination.....	4
2.1	Statistical accuracy	4
2.2	Information	5
2.3	Combination: Decision Maker	6
2.	Mathematical Pooling: Harmonic, Geometric and Arithmetic Means	9
3.	Strictly Proper Scoring Rules as Weights	11
4.1	Introduction.....	11
4.2	Scoring rules for individual variables	11
4.3	Scoring Rules for Average Probabilities / Expected Frequencies.....	12
4.3.1	Characterization of Strictly Proper Scoring Rules for Average Probabilities	12
4.4	Asymptotic Properties.....	16
4.5	Weights	18
5.	Additional Review of Expert Judgment Cross Validation Research.....	19
Discrepancies in Past Cross Validation Studies	19	
6.	Additional Classical Model Applications	22
6.1.	Nuclear reports Published as a result of the joint ec/usnrc project on uncertainty analysis of probabilistic accident consequence codes (under the Third EC-Framework Programme)	22
6.2.	Probabilistic accident consequence uncertainty analysis Reports published on the project uncertainty analysis of the probabilistic accident consequence code cosyra using expert judgement (under the fourth EC-framework programme)	23
6.3.	Ecosystems and public health	23
6.4.	Civil aviation and structural reliability.....	26
6.5.	Information security	28
6.6.	Natural hazards	28
6.7.	Climate.....	30
6.8.	Banking and finance.....	30
6.9.	Completed, publication in preparation	30
References	32	

1. Post-2006 Data and Applications Documentation

Expert judgment materials, including data from expert judgment studies, are available at (<http://rogermcooke.net/>). The studies can be read by the expert judgment software EXCALIBUR (Cooke and Solomatine 1992), which is a free downloadable at <http://www.lighttwist.net/wp/>. Summary information is presented in the following table. Full references are available at the end of the Supplementary Online Material.

Contracting Party	Performed by	Study name	Subject
Robert Wood Johnson Foundation	Center for Disease Dynamics, Economics & Policy	Arkansas	Grant effectiveness and child health insurance enrollment (Laxminarayan et al. 2012)
		Florida	
		Illinois	
		Nebraska	
		Washington	
		CoveringKids	Grant effectiveness and tobacco control (Colson, Cooke, and Laxminarayan 2016)
		Tobacco	Grant effectiveness, childhood obesity
		Obesity	Effectiveness of obstetric fistula repair (Colson, Adhikari, et al. 2015)
		Fistula	Effectiveness of surgical procedures
Disease Control Priority Project, 3 rd Edition		San_Diego	
Center for Disease Control and Prevention		CDC_ROI	Return on investment in public health tracking (Colson, Cohen, et al. 2015)
University of Wisconsin, CREATE	Vicki Bier	Create	Terrorism
National Oceanic and Atmospheric Administration and Environmental Protection Agency (EPA)	University of Notre Dame	Erie_Carp	Establishment of Asian Carp in Lake Erie (Cooke et al. 2014; Wittmann et al. 2015; Wittmann et al. 2014; Zhang et al. 2016)
		GL_NIS	Costs of invasive species in Great Lakes (Rothlisberger et al. 2010; Rothlisberger et al. 2012)
EPA and University of Maryland	University of Maryland	UMD_NREMO VAL	Nitrogen removal in Chesapeake Bay watershed (Koch, Febria, et al. 2015; Walsh 2015; Koch, Filoso, et al. 2015)
University Medical Center Utrecht	University Medical Center Utrecht	Hemophilia	Optimal treatment of patients with severe haemophilia (Fischer, Lewandowski, and Janssen 2013)
National Institute for Public Health and the Environment	TU Delft	ATCEP	Air traffic Controllers Human Error
Brand Preventie		FCEP	Flight Crew Human Error
Liander		Daniela	Fire prevention and control (Hanea 2009)
University of Cambridge		Liander	Underground cast iron gas-lines (Forys, Kurowicka, and Peppelman 2013)
Health Protection England	Willy Aspinall	Arsenic	Air quality levels for arsenic
University of Ottawa		Biol_Agent	Human dose-response curves for bioterror agents
		CWD	Infection transmission risks: chronic wasting disease (c wd) from deer to

			humans (Tyshenko et al. 2011; Tyshenko, ElSaadany, Oraby, Darshan, et al. 2012)
PrioNet		eBPP	Xenotropic murine leukemia virus-related virus (XMRV) blood/tissue infection transmission risks (Tyshenko, ElSaadany, Oraby, Laderoute, et al. 2012)
UK Government		Eff Erupt	Icelandic fissure eruptions: source characterization (Loughlin et al. 2013)
U Bristol / BAS / Ice2sea.eu		IceSheets	Contribution to sea level rise from ice sheets melting due to global warming (Bamber and Aspinall 2013; Bamber, Aspinall, and Cooke 2016)
University of Ottawa		PHAC_T4	Additional CWD factors
University of Bristol, Bristol Environmental Risk Research Centre		Sheep	Risk management policy for sheep scab control
Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy		SPEED	Volcano hazards at Vesuvius & Campi Flegrei, Italy (Bevilacqua et al. 2015; Neri et al. 2015)
Natural Environment Research Council, Economic & Social Research Council, and British Geological Survey		TdC	Volcano hazards (Tristan da Cunha) (Hicks et al. 2014)
Nuclear Waste Management Organization of Japan (NUMO) and Obayashi Corp		TOPAZ	Tectonic hazards for radwaste siting in Japan (Scourse, Aspinall, and Chapman 2015)
Natural Hazards Research Platform/GNS Science	Matthew Gerstenberger	Gerstenberger	Canterbury Seismic Hazard Model (Gerstenberger et al. 2014; Gerstenberger et al. 2013; Christophersen, Nicol, and Gerstenberger 2011)
Embry-Riddle	Benjamin Goodheart	Goodheart	Airport safety

Table 1: Details from 33 post-2006 applications of the Classical Model.

2. Classical Model Performance Measures and Combination

Similar expositions on the Classical Model can be found in the online supplementary material of other publications, including (Cooke et al. 2014; Wittmann et al. 2014; Cooke 2015; Koch, Febria, et al. 2015; Wittmann et al. 2015).

There are two generic, quantitative measures of expert performance, *calibration* or *statistical accuracy* and *information*. Loosely, statistical accuracy measures the statistical likelihood that a set of experimental results correspond, in a statistical sense, with an expert's assessments. More precisely, it is the p-value at which we would falsely reject the hypothesis that an expert's probability statements were accurate. Suppose the 5-, 50- and 95-percentiles, or quantiles, were elicited from each expert for each of N continuous variables. Each expert effectively divides the range of possible outcomes of each variable into four intervals: less than or equal to the 5 percent value, greater than the 5 percent value and less than or equal to the 50 percent value, etc. The probabilities for these intervals are expressed as a vector

$$p = (p_1, p_2, p_3, p_4) = (0.05, 0.45, 0.45, 0.05).$$

2.1 Statistical accuracy

If N quantities are assessed, each expert may be regarded as a statistical hypothesis, namely that each realization falls in one of the four inter-quantile intervals with probability vector p . Suppose we have realizations x_1, \dots, x_N of these quantities. We may then form the sample distribution of the expert's inter quantile intervals as:

$$\begin{aligned} s_1(e) &= \#\{ i \mid x_i \leq 5\% \text{ quantile} \} / N \\ s_2(e) &= \#\{ i \mid 5\% \text{ quantile} < x_i \leq 50\% \text{ quantile} \} / N \\ s_3(e) &= \#\{ i \mid 50\% \text{ quantile} < x_i \leq 95\% \text{ quantile} \} / N \\ s_4(e) &= \#\{ i \mid 95\% \text{ quantile} < x_i \} / N \\ s(e) &= (s_1, \dots, s_4) \end{aligned}$$

Note that the sample distribution depends on the expert e . If the realizations are indeed drawn independently from a distribution with quantiles as stated by the expert then the quantity

$$2N I(s(e) \mid p) = 2N \sum_{i=1..4} s_i \ln(s_i / p_i) \quad (1)$$

is asymptotically distributed as a chi-square variable with three degrees of freedom. This is the likelihood ratio statistic, and $I(s / p)$ is the relative information of distribution s with respect to p . Extracting the leading term of the logarithm yields the familiar chi-square test statistic for goodness of fit. There are advantages in using the form in (1) (Cooke 1991, and SOM 4).

If after a few realizations the expert were to see that all realization fell outside his/her 90 percent central confidence intervals, (s)he might conclude that these intervals were too narrow and might broaden them on subsequent assessments. This means that for this expert the uncertainty distributions are *not* independent, and (s)he learns from the realizations. Expert learning is not a goal of an expert judgment study. Rather, the problem owner wants experts who do not need to

learn from the elicitation. Independence is not an assumption about the expert's distribution but a desideratum of the problem owner. Hence the decision maker (see below) scores expert e as the statistical likelihood of the hypothesis

H_e : "the inter quantile interval containing the true value for each variable is drawn independently from probability vector p ."

A simple test for this hypothesis uses the test statistic (1), and the likelihood, or p-value, or *calibration score* of this hypothesis, is:

$$Cal(e) = p\text{-value}(e) = Prob\{2NI(s(e) | p) \geq r | H_e\}$$

where r is the value of (1) based on the observed values x_1, \dots, x_N . It is the probability under hypothesis H_e that a deviation at least as great as r should be observed on N realizations if H_e were true. Calibration scores are absolute and can be compared across studies. However, it is appropriate to equalize the power of the different hypothesis tests by equalizing the effective number of realizations. To compare scores on two data sets with N and N' realizations, we use the minimum of N and N' in (1), without changing the sample distribution s .

Although the calibration score uses the language of simple hypothesis testing, it must be emphasized that we are not rejecting expert-hypotheses; rather we are using this language to measure the degree to which the data supports the hypothesis that the expert's probabilities are accurate. Low scores, near zero, mean that it is unlikely that the expert's probabilities are correct. High scores, near 1, indicate good support.

2.2 Information

The second scoring variable is information. Loosely, the information in a distribution is the degree to which the distribution is concentrated. Information cannot be measured absolutely, but only with respect to a background measure. Being concentrated or "spread out" is measured relative to some other distribution.

Measuring information requires associating a density with each assessment of each expert. To do this, we use the unique density that complies with the experts' quantiles and is minimally informative with respect to the background measure. This density can easily be found with the method of Lagrange multipliers. For a uniform background measure, the density is constant between the assessed quantiles. The background measure is not elicited from experts as indeed it must be the same for all experts; instead it is chosen by the analyst.

The uniform and log-uniform background measures require an *intrinsic range* on which these measures are concentrated. The classical model implements the so-called " $k\%$ overshoot rule": for each item with uniform background we consider the smallest interval $I = [L, U]$ containing all the assessed quantiles of all experts and the realization, if known. This interval is extended to

$$I^* = [L^*, U^*]; L^* = L - k(U-L)/100; U^* = U + k(U-L)/100.$$

The value of k is chosen by the analyst. For a log uniform background, the same procedure is applied to logged values. A large value of k tends to make all experts look quite informative, and tends to suppress the relative differences in information scores. The *information score* of expert e on assessments for uncertain quantities $1 \dots N$ is

$$Inf(e) = \text{Average Relative information w.r.t. Background} = (1/N) \sum_{i=1..N} I(f_{e,i} | g_i)$$

where g_i is the background density for variable i and $f_{e,i}$ is expert e 's density for item i . This is proportional to the relative information of the expert's joint distribution given the background, under the assumption that the variables are independent. As with calibration, the assumption of independence here reflects a desideratum of the decision maker and not an elicited feature of the expert's joint distribution. The information score does not depend on the realizations. An expert can give her/himself a high information score by choosing quantiles very close together. The information score of e depends on the intrinsic range which depends on the assessments of the other experts. Hence, information scores cannot be exactly compared across studies.

The above information score is chosen because it is familiar, tail insensitive, scale invariant and "slow." The latter property means that relative information is a slow function; large changes in the expert assessments produce only modest changes in the information score. This contrasts with the likelihood function in the calibration score, which is a very "fast" function. This causes the normalized product of calibration and information to be driven by the calibration score.

2.3 Combination: Decision Maker

The **combined score** of expert e will serve as an unnormalized weight for e :

$$w_\alpha(e) = Cal(e) \times Inf(e) \times \mathbb{1}_\alpha(Cal(e) \geq \alpha), \quad (2)$$

where $\mathbb{1}_\alpha(Cal(e) \geq \alpha) = 1$ if $Cal(e) \geq \alpha$, and is zero otherwise. The combined score thus depends on α ; if $Cal(e)$ falls below cut-off level α , expert e is unweighted. The presence of a cut-off level is imposed by the requirement that the combined score be an asymptotically strictly proper scoring rule. That is, an expert maximizes his/her long run expected score by and only by ensuring that his probabilities $p = (0.05, 0.45, 0.45, 0.05)$ correspond to his true beliefs (Cooke, 1991, section 6). α is similar to a significance level in simple hypothesis testing, but its origin is to measure "goodness" and not to reject hypotheses.

A combination of expert assessments is called a "decision maker" (*DM*). All decision makers discussed here are examples of linear pooling; the classical model is essentially a method for deriving weights in a linear pool. "Good expertise" corresponds to good calibration (high statistical likelihood, high p-value) and high information. Weights that reward good expertise and pass these virtues on to the decision maker are desired.

The reward aspect of weights is very important. We could simply solve the following optimization problem: find a set of weights such that the linear pool under these weights maximizes the product of calibration and information. Solving this problem on real data, one finds that the weights do not generally reflect the performance of the individual experts. As an expert's influence on the decision maker should not appear haphazard, and "gaming" the system

with assessments tilted to achieve a desired outcome should be discouraged, we must impose a strictly scoring rule constraint on the weighting scheme.

The scoring rule constraint requires the term $\mathbb{1}_\alpha(\text{Cal}(e) \geq \alpha)$ in eq (2), but does not indicate what value of α we should choose. Therefore, we choose α to maximize the combined score of the resulting decision maker. Let $DM_\alpha(i)$ be the result of linear pooling for any item i with weights proportional to (2):

$$DM_\alpha(i) = \sum_{e=1,..E} w_\alpha(e) f_{e,i} / \sum_{e=1,..E} w_\alpha(e) \quad (3)$$

The *optimized global weight DM* is DM_{α^*} where α^* maximizes

$$\text{calibration score}(DM_{\alpha^*}) \times \text{information score}(DM_{\alpha^*}). \quad (4)$$

This weight is termed global because the information score is based on all the assessed calibration variables. A variation on this scheme allows a different set of weights to be used for each item. This is accomplished by using information scores for each item rather than the average information score:

$$w_\alpha(e,i) = I_\alpha(\text{calibration score}) \times \text{calibration score}(e) \times I(f_{e,i} / g_i) \quad (5)$$

For each α we define the *item weight DM* for item i as

$$IDM_\alpha(i) = \sum_{e=1,..E} w_\alpha(e,i) f_{e,i} / \sum_{e=1,..E} w_\alpha(e,i) \quad (6)$$

The *optimized item weight DM* is IDM_{α^*} where α^* maximizes

$$\text{calibration score}(IDM_{\alpha^*}) \times \text{information score}(IDM_{\alpha^*}). \quad (7)$$

The non-optimized versions of the global and item weight DM's are obtained by setting $\alpha = 0$.

The optimization in (5) and (7) often causes experts to be unweighted, even experts with good scores. Such experts are not "rejected;" unweighting simply means that their input is already captured by a smaller subset of experts. Their value to the whole study is brought out in studying the robustness of the optimal DM under loss of experts.

Item weights are potentially more attractive as they allow an expert to up- or down- weight her/himself for individual items according to how much (s)he feels (s)he knows about that item. "Knowing less" means choosing quantiles farther apart and lowering the information score for that item. Of course, good performance of item weights requires that experts successfully perform this up-down weighting. Anecdotal evidence suggests that item weights improve over global weights as the experts receive more training in probabilistic assessment. Both item and global weights can be described as optimal weights under a strictly proper scoring rule constraint. With both global and item weights, calibration strongly dominates over information, and information serves to modulate between more or less equally well calibrated experts.

Since any combination of expert distributions yields assessments for the seed variables, any combination can be evaluated on the seed variables. In particular, we can compute the calibration and the information of any proposed decision maker. We should hope that the "performance based decision maker" would perform better than the result of simple averaging, and we should also hope that the proposed DM is not worse than the best expert in the panel.

2. Mathematical Pooling: Harmonic, Geometric and Arithmetic Means

This analysis builds on the material in Bamber et al. (2016) and the main manuscript. Using the 33 professional expert judgment studies performed since 2006, it is possible to compare *HW*, *EW* and performance weighting (*PW*). To facilitate third party checks of the results, for this comparison *PW* is based on global weights, and experts who assessed less than the full set of calibration variables are excluded. This causes the *PW* and *EW* solutions used here to differ slightly from the solutions published elsewhere, but the integrity of the present comparison is not affected.

	PW			EW			HW			#seeds	#experts
	P-value PW	inf PW	comb	P-value EW	inf EW	comb	P-value HW	inf HW	comb		
Arkansas	0.499	0.337	0.168	0.386	0.198	0.076	5.55E-02	0.640	3.55E-02	10	4
Arsenic D-R	0.036	2.739	0.098	0.061	1.095	0.067	7.99E-04	1.324	1.06E-03	10	9
ATCEP Error	0.683	0.227	0.155	0.124	0.247	0.031	5.99E-04	1.066	6.38E-04	10	5
Biol agents	0.678	0.610	0.414	0.413	0.244	0.101	3.60E-02	0.884	3.18E-02	12	12
CDC ROI	0.720	2.305	1.660	0.233	1.230	0.286	7.56E-01	1.565	1.18E+00	10	20
CoveringKid	0.720	0.431	0.310	0.628	0.274	0.172	9.03E-01	0.595	5.38E-01	10	5
CREATE	0.394	0.276	0.109	0.061	0.207	0.013	2.77E-04	0.52	1.44E-04	10	7
CWD	0.493	1.215	0.598	0.474	0.930	0.441	7.07E-01	1.494	1.06E+00	10	14
Daniela	0.554	0.634	0.351	0.533	0.168	0.089	1.82E-01	0.520	9.48E-02	7	4
dcpn_fistula	0.119	1.309	0.156	0.059	0.622	0.037	8.78E-08	1.125	9.88E-08	10	8
eBBP	0.833	1.406	1.172	0.358	0.316	0.113	8.04E-02	0.954	7.67E-02	15	14
EffusiveErup	0.664	1.123	0.745	0.286	0.796	0.228	2.65E-02	1.505	3.99E-02	8	14
Erie Carps	0.661	0.856	0.566	0.182	0.281	0.051	3.87E-01	0.754	2.92E-01	15	10
FCPEP Error	0.664	0.574	0.381	0.222	0.099	0.022	1.75E-05	0.771	1.35E-05	8	5
Florida	0.756	1.133	0.857	0.756	0.455	0.344	6.98E-02	0.880	6.15E-02	10	7
GL-NIS	0.928	0.209	0.194	0.044	0.307	0.014	5.53E-02	0.842	4.66E-02	13	9
Gerstenberge	0.9302	1.095	1.018	0.6439	0.4815	0.31	8.10E-02	0.966	7.82E-02	14	12
Goodheart	0.707	0.959	0.678	0.550	0.277	0.153	6.83E-01	0.888	6.07E-01	10	6
Hemophilia	0.312	0.494	0.154	0.254	0.202	0.051	3.12E-01	0.779	2.43E-01	8	18
IceSheet2012	0.399	1.552	0.620	0.492	0.517	0.254	7.96E-02	1.201	9.56E-02	11	10
Illinois	0.337	0.647	0.218	0.620	0.264	0.163	2.37E-03	0.793	1.88E-03	10	5
Liander	0.228	0.524	0.120	0.228	0.484	0.111	2.81E-03	1.198	3.36E-03	10	11
Nebraska	0.033	1.447	0.048	0.368	0.695	0.256	2.40E-05	1.192	2.86E-05	10	4
Obesity	0.440	0.507	0.223	0.070	0.243	0.017	6.68E-04	0.745	4.98E-04	10	4
PHAC T4	0.178	0.351	0.062	0.298	0.211	0.063	1.64E-02	0.640	0.01048	12	10
San Diego	0.155	0.758	0.117	0.147	1.012	0.148	3.02E-03	1.583	3.32E-02	10	8
Sheep Scab	0.643	1.310	0.843	0.661	0.780	0.516	1.15E-02	1.411	1.63E-02	15	14
SPEED	0.676	0.777	0.525	0.517	0.751	0.389	2.97E-02	1.165	3.46E-02	16	14
T4C	0.989	1.256	1.242	0.166	0.364	0.060	1.24E-02	1.079	1.34E-02	17	18
Tobacco	0.688	1.062	0.730	0.200	0.451	0.090	2.11E-01	0.708	1.49E-01	10	7
Topaz	0.411	1.455	0.598	0.629	0.922	0.580	8.66E-05	1.528	1.32E-04	16	21
umd_nremov	0.706	1.988	1.404	0.068	0.804	0.054	2.40E-03	1.219	2.93E-03	11	9
Washington	0.200	0.724	0.145	0.155	0.529	0.082	4.21E-01	0.862	3.63E-01	10	5
nr < 0.05	2			1			18				
nr best			26			3			4		
Ave Inf		1.042			0.531			1.077			

Table 2: Performance of *PW*, *EW* and *HW*. “#seeds” denotes the number of calibration variables used in each study, “#experts” denotes the number of experts who assessed all calibration variables in each study.

The performance of *HW*, *EW* and *PW* are compared with regard to statistical accuracy, informativeness and the combined score (the product of the former two) (Table 2). *HW* is the best (as determined by the combined score) in four of the thirty-three cases. *HW*'s informativeness is slightly higher than that of *PW* and substantially higher than *EW*. The statistical accuracy of *HW* is substantially below that of *EW* and *PW*. As reported in Bamber et al. (2016), in 18 cases (55 percent) the hypothesis that *HW* is statistically accurate would be rejected at the 5 percent level. In nine cases rejection would be at the 0.001 level.

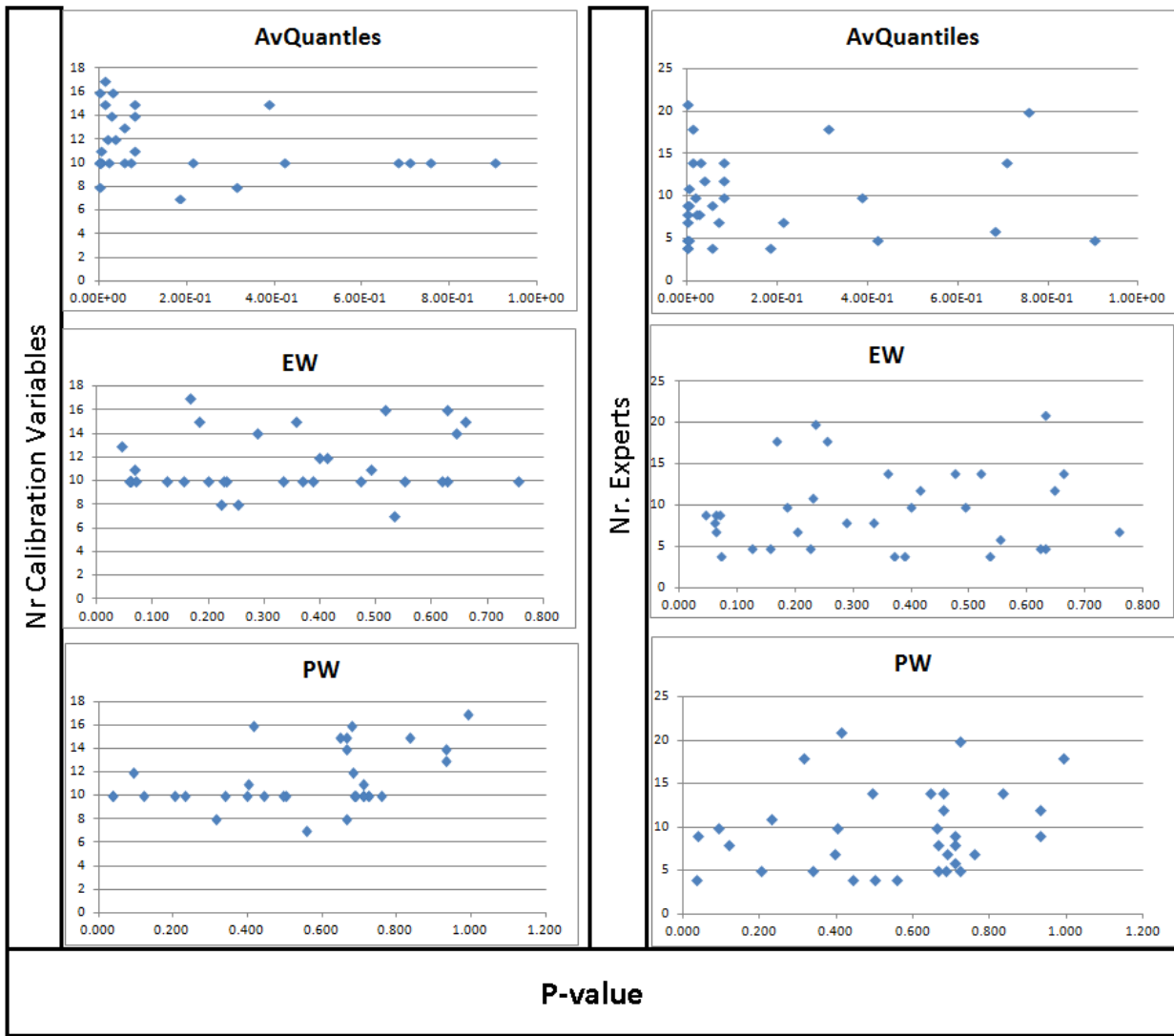


Figure 1: Number of calibration variables and number of experts against P-values for *HW*, *EW* and *PW*.

These data provide evidence on how performance is affected by the number of experts and number of calibration variables. Figure 1 graphs the number of calibration variables and number of experts against the statistical accuracy scores for *HW*, *EW*, and *PW*. *HW* degrades as the number of calibration variables increases, whereas *EW* is unaffected, and *PW* actually improves. The statistical power of the measure of statistical accuracy increases with the number of calibration variables, and this would tend to suppress statistical accuracy scores of all experts and combinations alike. However, no such tendency is observed for *EW* or *PW* (the effect would presumably be observed if greater numbers of calibration variables were available). The number of experts does not have a marked effect on any of the combinations.

3. Strictly Proper Scoring Rules as Weights

This builds on the original material from (Cooke 1991).

4.1 Introduction

This chapter covers the mathematics underlying the classical model in full detail. This material was published in 1991, and the original exposition has been simplified and improved for courses given to NASA. The original numbering of theorems and propositions from (Cooke 1991) is preserved for reference. The main simplification concerns the characterization of scoring rules for average probabilities. This improved rendition is published here for the first time.

4.2 Scoring rules for individual variables

Scoring rules were introduced by L. J. Savage and B. DeFinetti for *elicitation*.

Where S^n is the simplex on n dimensions, i.e. the set of non negative n -vectors whose components sum to one, suppose expert is asked to state $p \in S^n$. When outcome i is known, the expert is paid $R(p,i)$.

Definition: $R(p,i)$ is strictly proper positive (negative) sensed if for all $q \in S^n$,
 $Argmax_p (Argmin_p) E_q(R(p,i)) = q$.

Direct Rule: $R_d(p,i) = Kp_i$, (reward proportional to probability of observed outcome)

$$E_q(R_d(p,i)) = \sum q_i R_d(p,i) = K \sum q_i p_i.$$

Lemma: $Argmax_p E_q(R_d(p,i)) = (0,0\dots 0, 1, 0\dots 0)$, where "1" is at the position of the largest component of q .

Quadratic rule: $R_Q(p,i) = 2p_i - \sum_j p_j^2$.

Spherical rule: $R_S(p,i) = p_i / (\sum_j p_j^2)^{1/2}$

Logarithmic rule: $R_L(p,i) = \ln(p_i)$

Brier score $R_B(p,i) = (p_i - i)^2$; $i \in \{0,1\}$ ($= (1 - R_Q)/2$) (used in rain forecasting)

Shuford, Albert and Massengil (1966) prove that R_L is the only rule which depends only on the probability of the observed outcome, for $n > 2$.

DeGroot Feinberg decomposition

Imagine that an expert assigns variables to different "probability bins" where the events assigned to the same bin are assessed to have the same distribution. The variables are subsequently observed and a sample distribution is found for each bin:

Let $\mathbf{p} = p_{1..}, \dots, p_{B..} \in (S^n)^B$ (a vector of probability vectors);

$\mathbf{s} = s_{1..}, \dots, s_{B..} \in (S^n)^B$ (vector of sample distributions)

$\mathbf{n} = n_{1..}, \dots, n_{B..} \in \mathbb{R}^B$ (occupation vector, $n_i = \#$ vbls assigned to bin i).

(Expert assigns variables to probability bins, thus \mathbf{n} and \mathbf{s} depend on the expert.)

For $r, t \in S^n$, write: $E_t(R(r)) = \sum_{i=1..n} t_i R(r, i)$. The DeGroot, Feinberg decomposition of scoring rule $\mathbf{R}(\mathbf{p}, \mathbf{n}, \mathbf{s})$ summed over bins $i = 1 \dots B$ is (note that $s_i, p_i \in S^n, i = 1 \dots B$):

$$\mathbf{R}(\mathbf{p}, \mathbf{n}, \mathbf{s}) = \sum_{i=1..B} n_i [E_{s_i}(R(p_i)) - E_{s_i}(R(s_i))] + \sum_{i=1..B} n_i E_{s_i}(R(s_i)).$$

“Calibration term”
“Resolution term”

Lemma: For the logarithmic rule,

$$\mathbf{R}(\mathbf{p}, \mathbf{n}, \mathbf{s}) = - \sum_{i=1..B} n_i [I(s_i | p_i) + H(s_i)]$$

where $H(p) = - \sum p_i \ln p_i$ is called the entropy of $p \in S^n$.

4.3 Scoring Rules for Average Probabilities / Expected Frequencies

$\mathbf{O} = \{1, \dots, m\}$; *outcomes*

$M(\mathbf{O}) = \{p \in S^m \mid p_i > 0\}$ *non-degenerate probability vectors*

$X_i : \Omega \rightarrow \mathbf{O}$: *Uncertain quantity (finite valued)*

$X = X_1, \dots$

$K_i^n = \sum_{j=1..n} \mathbb{1}_{\{X_j=i\}}$: *frequency counter for outcome i in variables 1...n.*

$S_i(n) = K_i^n/n$ *relative frequency for outcome i.*

$S(n) = S_1(n), \dots, S_m(n)$: *sample distribution over $X_1 \dots X_n$.*

$M(X) =$ *non-degenerate measures on X; $P \in M(X) \Rightarrow \forall \text{words } W, P(W) > 0$*

$Q \in M(X), q_i(n) = (1/n) \sum_{j=1..n} Q(X_j = i)$: *Assessed distribution.*

$q(n) = q_1(n) \dots q_m(n)$. : *vector of average probabilities.*

Definition: For $M \subseteq M(X)$, $R(p, s(n))$ is a positively sensed M strictly proper scoring rule for average probabilities (SPSRAP) if

$$\forall Q \in M, \arg \max_{p \in S^m} E_Q(R(p, s(n))) = q(n).$$

Note, the quantifier is over a different (and potentially much larger) set than the argmax.

4.3.1 Characterization of Strictly Proper Scoring Rules for Average Probabilities

Numbered as in (Cooke 1991, chap. 9).

THEOREM 9.1 (Cooke 1991): If $R(p, s(n))$ is differentiable in p , then the following are equivalent:

1. $\forall Q \in M(X), \quad \nabla_p E_Q(R(p, s(n)))|_{p=q(n)} = 0$
2. $\forall i, j, k \in \{1, \dots, m-1\} \exists$ integrable functions g_i, g_{ijk} such that

$$\partial R / \partial p_i = g_i(p, n)(p_i - s_i(n)) + \sum_{k < j} g_{ijk}(p, n)(s_j(n)p_k - s_k(n)p_j).$$

PROOF: We use the following lemmata:

Lemma 9.1 Where F is a field of events, $A_1 \dots A_N \in F : (1/N) \sum_{j=1 \dots N} Q(I_{A_j} = 1) = (1/N) \sum_{k=1 \dots N} k Q\{k \text{ of } A_1 \dots A_N \text{ occur}\}$ (average probability = expected relative frequency).

Proof: The expected relative frequency of occurrence of A_1, \dots, A_N , $(1/N) \sum_k k Q\{k \text{ of } A_1 \dots A_N \text{ occur}\}$, may be written $E((1/N) \sum_{j=1 \dots N} I_{A_j}) = (1/N) \sum_{j=1 \dots N} E(I_{A_j}) = (1/N) \sum_{j=1 \dots N} Q(I_{A_j} = 1)$. \square

Lemma 9.2 $\forall A \in M(L \times N)$ (set of $L \times N$ matrices) with $\text{Rank}(A) = L < N$; and

$$V = \{x \in \mathbb{R}^N \mid Ax = b: b \neq 0; x_i > 0, i = 1, \dots, N\}$$

$$\text{If } \dim V > 0, \text{ then } \dim V = N - L + 1.$$

Proof: The null space of A has dimension $N - L$, if $V \neq \emptyset$, then there is $x \in V$ which is not in the null space of A . The difference of any two vectors in V is in the null space of A . Hence, $\dim V = N - L + 1$. \square

Proof strategy:

Statement (1) says that for all i , the following two vectors are orthogonal:

$$\partial_i (R(p, k^{N,1})) \dots \partial_i (R(p, k^{N,\alpha})) \perp Q(k^{N,1}) \dots Q(k^{N,\alpha});$$

where $k^{N,1} \dots k^{N,\alpha}$ is the set of all possible frequency vectors on $X_1 \dots X_N$. We will count the dimensions of the subspaces generated by these vectors, and show that scoring rules satisfying statement (2) of the theorem span the subspace of vectors orthogonal to $Q(k^{N,1}) \dots Q(k^{N,\alpha})$. We simplify this notation in the following definition:

Define $W = \{k \in \mathbb{N}^{m-1} \mid k_i \geq 0; i=1 \dots m-1, \sum_{i=1 \dots m-1} k_i \leq N\}$ (set of frequency vectors over outcomes $1 \dots m$ $\boxtimes 1$)

Use $k \in W$ to index $\mathbb{R}^{|W|}$: $k \in W, Q_k = Q(k)$ (assessed prob. of frequency vector $k \in W$)

$$Q = (Q_1, \dots, Q_{|W|}) \in \mathbb{R}^{|W|}.$$

$A(p) =$ subspace generated by $\{Q \in \mathbb{R}^{|W|} \mid q(N) = p\}$; ($q(N)$ is the vector of average probabilities associated with Q)

$$R(p, i)_k = (\partial/\partial p_i) R(p, k/N);$$

$B(p, i)$ = subspace of $\mathbb{R}^{|W|}$ generated by vectors $R(p, i)$ where scoring rule $R(p, k/N)$ is differentiable in p and satisfies statement (2) of the theorem.

Lemma 9.3 $\text{Dim } A(p)^\perp \leq \text{Dim } B(p, i)$:

Proof of lemma 9.3

$$Q \in A(p) \text{ satisfies: } \sum_{k \in W} Q(k) = 1, \sum_{k \in W} k_i Q_k/N = p_i, i = 1, \dots, m \text{ } \forall 1.$$

These are m independent equations so $\text{Dim}(A(p)) = |W| - (m - 1)$ (Lemma 9.2)

$$\Rightarrow \text{Dim}(A(p)^\perp) = m - 1.$$

We must show $\text{Dim } B(p, i) \geq m - 1$. It suffices to find $m - 1$ linearly independent vectors satisfying (2). In fact, it suffices to find $m - 1$ vectors of the form (2) which are independent on the $m - 1$ components of $|W|$:

$$K^{(1)} = N, 0, 0 \dots ;$$

$$K^{(2)} = 0, N, 0, \dots ;$$

...

$$K^{(m-1)} = 0, 0, \dots, N.$$

Choose:

$$R^{(i)}(p, s(N)) = 1/2 p_i^2 - p_i s_i$$

For $j = 1 \dots m \text{ } \forall 1; j \neq i$, choose:

$$R^{(j)}(p, s(N)) = -s_i \ln p_i - s_j \ln p_j - (1 - s_i - s_j) \ln (1 - p_i - p_j).$$

Verify:

$$\partial_i R^{(i)} = p_i - s_i$$

$$\partial_i R^{(j)} = \frac{p_i - s_i + s_i p_j - p_i s_j}{p_i (1 - p_i - p_j)}$$

So; filling in $s = K^{(h)}/N$:

$j \neq i$:

$$\partial_i R^{(j)}(p, k^{(h)}/N) = \frac{p_i}{p_i(1-p_i-p_j)}; \text{ if } h \neq i, h \neq j;$$

$$\partial_i R^{(j)}(p, k^{(h)}/N) = 0; \text{ if } h = j;$$

$$\partial_i R^{(j)}(p, k^{(h)}/N) = \frac{p_i - 1 + p_j}{p_i(1-p_i-p_j)}; \text{ if } h = i;$$

$$\partial_i R^{(i)}(p, k^{(h)}/N) = p_i; h \neq i$$

$$\partial_i R^{(i)}(p, k^{(h)}/N) = p_i - 1; h = i$$

Multiply row j by $p_i(1-p_i-p_j)$, $j \neq i$:

	$k(1)$	$k(2)$	$k(i)$		$k(i+1)$	$k(m-1)$
$\partial_i R^{(1)}$	0	p_i		p_i+p_1-1		p_i	p_i	p_i
$\partial_i R^{(2)}$	p_i	0		p_i+p_2-1		p_i	p_i	p_i
$\partial_i R^{(3)}$	p_i	p_i		.		.	.	p_i
.	p_i
$\partial_i R^{(i)}$	p_i	p_i	$p_i - 1$		p_i	p_i	p_i
.	p_i
.	p_i	p_i	p_i	.		0	p_i	p_i
.	p_i	p_i	p_i	.		p_i	0	p_i
$\partial_i R^{(m-1)}$	p_i	p_i	p_i	$p_i+p_{m-1}-1$		p_i	p_i	0

Subtract the i -th row from each row:

	$k(1)$	$k(2)$	$K(i)$		$k(i+1)$	$K(m-1)$
$\partial_i R^{(1)}$	$-p_i$	0	0	p_1		0	0	0
$\partial_i R^{(2)}$	0	$-p_i$	0	p_2		0	0	0
$\partial_i R^{(3)}$	0	0	$-p_i$	p_3		.	.	0
.	0
$\partial_i R^{(i)}$	p_i	p_i	p_i-1		p_i	p_i	0
.	0
.	0	0	0	.		$-p_i$	0	0
.	0	0	0	.		0	$-p_i$	0
$\partial_i R^{(m-1)}$	0	0	0	p_{m-1}		0	0	$-p_i$

These are linearly dependent only if some linear combination of the rows = $(0, 0, \dots, 0)$. That can only be the sum of all rows. However, for column k^i the sum is $p_1 + p_2 + \dots + p_{m-1} - 1$. This equals 0 only if $p_m = 0$, but we assume $p_m > 0$. \square

Proof of Theorem

(2) \Rightarrow (1)

$$\begin{aligned}
 \partial_i E_Q R &= \sum_{k \in W} Q(k) \partial_i R \\
 &= \sum_{k \in W} Q(k) g_i(p, N)(p_i - k_i/N) + \sum Q(k) \sum_{j < h} g_{ijh}(p, N)(p_h k_j/N - p_j k_h/N) \\
 &= g_i(p, n) \times (p_i - q_i(N)) + g_{ijh}(p, n) \times (p_i q_h(N) - p_h q_i(N)) \\
 &= 0 \text{ if } p = q(N).
 \end{aligned}$$

(1) \Rightarrow (2):

From (2) \Rightarrow (1) we have that $B(q(N), i) \subseteq A(p)^\perp$.

But also $\text{Dim } A(p)^\perp \leq \text{Dim } B(q, i)$ (Lemma 9.3)

Therefore $A(p)^\perp = B(q, i)$, and any R satisfying (1) satisfies (2). \square

Examples

where $s, p \in S^n$, strictly positive, and $I(s / p) = \sum_{i=1..n} s_i \ln(s_i / p_i)$:

- $I(s(n) / p)$ is $M(X)$ *spsrap*
- $I(p / s(n))$ is *NOT* $M(X)$ *spsrap*
- $\sum_{i=1..n} (s_i - p_i)^2$ is $M(X)$ *spsrap* (quadratic loss)
- $\sum_{i=1..n} (s_i - p_i)^2 / p_i$ is *NOT* $M(X)$ *spsrap* (chi-square goodness of fit, also leading term of $I(s / p)$).

4.4 Asymptotic Properties

Definition:

For $M \subseteq M(X)$, $R(p, s(n))$ is *strongly asymptotic M – strictly proper* scoring rule for average probabilities if

$$\forall Q \in M$$

$$\text{Argmax}_p E_Q(R(p, s(n))) = p(n), \text{ and } q(n) \rightarrow r \text{ as } n \rightarrow \infty \Rightarrow p(n) \rightarrow r.$$

For $M \subseteq M(X)$, $R(p, s(n))$ is *weakly asymptotic M – strictly proper* scoring rule for average probabilities (*WAM SPSRAP*) if

$$\forall Q \in M \text{ whenever } q(n) \rightarrow r \text{ and } r' \in M(O), r' \neq r,$$

then there exists $N' \in \mathcal{N}$ such that for all $n > N'$,

$$E_Q(R(r, s(n))) > E_Q(R(r', s(n))).$$

Relative information score: $RI(p, s(n)) = 2n(I(s | p))$.

$$\text{Hypothesis test score } w_t(p, s(n)) = \begin{cases} 1 & \text{if } RI(p, s(n)) \leq t \\ 0 & \text{if } RI(p, s(n)) > t \end{cases}$$

Calibration Score: $W_t(p, s(n)) = (1 - \chi^2_{m-1}(RI)) \times w_t$

χ^2_{m-1} is the cdf of a chi square variable with $m-1$ degrees of freedom.

Proposition 9.4 Let $M = \{ \prod p \mid p \in M(O) \}$ (set of product measures on X) For any $t \in (0, \infty)$, w_t is WAM SPSRAP.

Proof: choose $Q \in M$, Then $\lim_{n \rightarrow \infty} E_Q w_t(q, s(n)) = \chi^2_{m-1}(t)$. Choose $r \in S^m$, $r \neq q$, By the strong law of large numbers, $s(n) \rightarrow q$, Q a.s. and by Egoroff's theorem, for every $d > 0$ the convergence is uniform on a set of probability greater than $1 - d$. Choose $d < \chi^2_{m-1}(t)$, for some $k_d > 0$ we can find N_d such that $\forall n > N_d$, $I(s(n) | r) > k_d$, with probability $1 - d$. For $n > \max\{N_d, t/k_d\}$, on this set

$$2nI(s(n) | r) > 2(t/k_d)k_d = 2t > t.$$

Hence, for n sufficiently large, $Q\{w_t(r, s(n)) = 0\} > 1 - d$ and

$$E_Q w_t(r, s(n)) < d < \chi^2_{m-1}(t). \quad \square$$

Remark This also shows that $E_Q w_t(r, s(n)) \rightarrow 0$ as $n \rightarrow \infty$ if $r \neq q$.

Lemma: For any CDF, F :

$$0 \leq \int_{-\infty}^z F(x) dF(x) - F(z)^2/2 \leq \max_x [F(x) - F(x)]$$

Proposition 9.5: Let $M = \{ \prod p \mid p \in M(O) \}$ (set of product measures on X) For any $t \in (0, \infty)$, W_t is WAM SPSRAP. For any function

$f: M(O) \times N \rightarrow [a, b] ; 0 < a < b < \infty$

$W_t f(p, n)$ is WAM SPSRAP

Proof: Let Q_n denote expert's CDF for $RI(q, s(n))$.

$$E_Q W_t(q, s(n)) = \int_0^t [1 - \chi_{m-1}^2(x)] d Q_n(x)$$

$$= E_Q W_t(q, s(n)) - \int_0^t \chi_{m-1}^2(x) d Q_n(x).$$

$Q_n \rightarrow \chi_{m-1}^2$. χ_{m-1}^2 is continuous and bounded so the Helly Bray theorem together with the above Lemma gives, as $n \rightarrow \infty$

$$= \int_0^t \chi_{m-1}^2(x) d Q_n(x) \rightarrow \chi_{m-1}^2(t)^2/2 > 0.$$

Suppose $r \in M(O)$, with $r \neq q$. Since $1 - \chi_{m-1}^2(x) < 1$, it follows from the proof of 9.4 that $\lim_{n \rightarrow \infty} E_Q W_t(r, s(n)) = 0$. \square

4.5 Weights

Collecting all desiderata: Weights should

1. reward good statistical likelihood and high relative information
2. be asymptotically strictly proper for average probabilities
3. be meaningful, familiar, easy to explain
4. allow likelihood to dominate over informativeness

The weights in SOM 2 satisfy these desiderata.

5. Additional Review of Expert Judgment Cross Validation Research

Discrepancies in Past Cross Validation Studies

EXCALIBUR (Cooke and Solomatine 1992) is the standard software for implementing the Classical Model. It is not possible to use EXCALIBUR for cross validation, but it can be used to spot-check the results of code that implements cross validation. Lin and Cheng (2008; 2009), Flandoli et al. (2011), and Eggstaff et al. (2014) all developed their own code for cross validation, but only Eggstaff et al. report their code and results have been vetted against EXCALIBUR, a process that required extensive coordination.

Large differences exist between the values reported in Lin and Cheng (2008) and Cooke and Goossens (2008). Table 3 details those differences and provides a strong argument for communicating with the authors of the data set before publishing results. The numbers of Lin and Cheng bear little resemblance to those of Cooke and Goossens. In 6 of the 28 studies, the study name and number of experts and seed variables are so divergent that it is not possible to determine which study from the TU Delft database is referenced.

The values in Cooke and Goossens (2008) are published values, some of which were computed with archaic MS-DOS code. That code had a crude method for estimating the tail of a chi square distribution, leading to poor resolution below $1E-4$. For large numbers of calibration variables (e.g., as in study 24), this problem could be acute. It was addressed in Cooke and Goossens (2008) by reducing the statistical power to a default value of 10. This might explain part of the discrepancy in study 24. In their later cross validation analysis of more studies, Lin and Cheng (2009) report that they deleted questions that were not answered by many experts and experts who did not answer many questions. If they also did this for their 2008 analysis, that could explain differences in the number of seed variables reported in Lin and Cheng (2008) and Cooke and Goossens (2008). For the other studies, however, no explanation suggests itself for the observed differences in *EW* and *PW* combined scores.

	Study name (from Lin and Cheng)	# of experts	# of seed variables	Lin and Cheng 2008, Table 1 "within sample"		Cooke and Goossens 2008 Table 1	
				PWComb	EWComb	PWComb	EWComb
1	Acrylonitrile	7	10	0.47	0.44	0.76	0.42
2	Option trading ¹	5	34				
3	Dike ring	17	47	0.42	0.03	0.25	0.038
4	Flanges	10	8	0.6	0.2	0.91	0.43
5	Crane risk ²	8	10	0.93	0.28	1.15	0.35
6	Groundwater	7	10	0.95	0.05	2.11	0.16
7	Space debris ³	7	26	6.0E-06	0.13	0.25	0.14
8	Composite materials	6	12	0.55	0.21	0.39	0.11

9	Radiation in food ¹	7	6				
10	Dry deposition	8	14	0.48	0.003	0.70	0.001
11	Atmospheric dispersion	8	23	0.38	0.18	0.98	0.13
12	Early health effects	7/9	15	0.06	0.01	0.050	0.012
13	Radiation dosimetry	5	38				
14	Soil transfer	4	31	1.0E-06	1.0E-07	1.0E-04	9.7E-05
15	Wet deposition	7	19	0.11	0.002	0.11	0.00073
16	Gas pipelines ¹	16	14				
17	MONTSE 1 ¹	9	8				
18	MOTHER5 ¹	5	34				
19	Montserrat ¹	10	8				
20	Movable barriers	8	14	0.06	0.13	0.54	0.13
21	Real estate	5	31	0.7	0.001	0.63	0.0009
22	River dredging	6	8	0.54	0.18	0.45	0.19
23	Sulphur trioxide	4	7	2.53	0.3	0.547	0.29
24	Building temperature ⁴	6	48	0.002	2.0E-10	0.2005	0.0035
25	Atmospheric dispersion (TNO)	7	36	0.09	0.002	0.60	0.24
26	Radioactive deposition (Delft) ⁵	4	24	0.3	0.22	0.74	0.42
27	Atmospheric dispersion (Delft)	11	36	0.14	0.06	0.56	0.51
28	Water pollution ⁶	11	9	0.62	0.4	0.66	0.48

Table 3: Comparison of combined scores for performance based and equal weight of Lin and Cheng (2008) and Cooke and Goossens (2008).

Notes:

The effective number of seed variables is the smallest number assessed by an expert in the study. The number of total and effective seeds reported in Lin and Cheng do not match Cooke and Goossens for several studies (see below).

The statistical accuracy is powered to the effective number of seed variables by EXCALIBUR. In some cases, (e.g., study 24) scores are powered down because of numerical limitations of archaic code.

1. It is not clear which study Cooke and Goossens study this corresponds to.
2. Cooke and Goossens report this study had 12 total seeds and 11 effective seeds.
3. Cooke and Goossens report this study had 18 effective seeds.
4. Cooke and Goossens report this study had 10 effective seeds.
5. Cooke and Goossens report this study had 22 effective seeds.
6. Cooke and Goossens report this study had 11 total and 10 effective seeds.

The out-of-sample code of Flandoli et al. (2011) has been reviewed and found to optimize incorrectly and to conflate uniform and loguniform background measures. Two of the four cases reported had 15 and 16 calibration variables, enabling direct comparison with results from the verified Eggstaff et al. code (2014). The other cases are too large, as Eggstaff split studies with large numbers of seed variables into two separate sets for analysis. Flandoli et al. draw 500 random samples from training sets of fixed size and compute the scores on the complementary test set. Table 4 compares the results from the Flandoli sampling with the complete set using the Eggstaff code.

		PW			EW		
		SA	Inf	Comb	SA	Inf	Comb
Pbearl 7 training, 8 test	Eggstaff	0.149	0.617	0.072	0.271	0.167	0.046
	Flandoli Table 8	0.229	0.407	0.093	0.273	0.167	0.046
Vesuvius 8 training 8 test	Eggstaff	0.277	1.176	0.231	0.520	0.756	0.380
	Flandoli Table 4	0.449	0.896	0.377	0.519	0.720	0.365

Table 4: Results of Flandoli et al. (2011) based on 500 samples compared with the vetted code of Eggstaff based on the complete set.

6. Additional Classical Model Applications

This list is current as of November 2016.

- 6.1. **Nuclear reports** Published as a result of the joint ec/usnrc project on uncertainty analysis of probabilistic accident consequence codes (under the Third EC-Framework Programme)
- 6.1.1. F.T. Harper, L.H.J. Goossens, R.M. Cooke, S.C. Hora, M.L. Young, J. Päsler-Sauer, L.A. Miller, B. Kraan, C. Lui, M.D. McKay, J.C. Helton and J.A. Jones *Probabilistic accident consequence uncertainty study: Dispersion and deposition uncertainty assessment* (1994) Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities NUREG/CR-6244, EUR 15855 EN, SAND94-1453, Washington/USA, and Brussels-Luxembourg, November 1994, published January 1995. Volume I: Main report, Volume II: Appendices A and B, Volume III: Appendices C, D, E, F, G, H
 - 6.1.2. R.M. Cooke, L.H.J. Goossens and B.C.P. Kraan *Methods for CEC\USNRC accident consequence uncertainty analysis of dispersion - Performance based aggregating of expert judgements and PARFUM method for capturing modeling uncertainty* (1995) Prepared for the Commission of European Communities, EUR 15856, Brussels-Luxembourg, June 1994, published 1995
 - 6.1.3. J. Brown, L.H.J. Goossens, F.T. Harper, B.C.P. Kraan, F.E. Haskin, M.L. Abbott, R.M. Cooke, M.L. Young, J.A. Jones S.C. Hora, A. Rood and J. Randall (1997) *Probabilistic accident consequence uncertainty study: Food chain uncertainty assessment* Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities NUREG/CR-6523, EUR 16771, SAND97-0335 Washington/USA, and Brussels-Luxembourg, March 1997, published June 1997. Volume 1: Main report, Volume 2: Appendices.
 - 6.1.4. L.H.J. Goossens, J. Boardman, F.T. Harper, B.C.P. Kraan, R.M. Cooke, M.L. Young, J.A. Jones and S.C. Hora *Probabilistic accident consequence uncertainty study: Uncertainty assessment for deposited material and external doses* (1997) Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities, NUREG/CR-6526, EUR 16772, SAND97-2323 Washington/USA, and Brussels-Luxembourg, September 1997, published December 1997. Volume 1: Main report, Volume 2: Appendices
 - 6.1.5. F.E. Haskin, F.T. Harper, L.H.J. Goossens, B.C.P. Kraan, J.B. Grupa and J. Randall (1997) *Probabilistic accident consequence uncertainty study: Early health effects uncertainty assessment* Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities NUREG/CR-6545, EUR 16775, SAND97-2689 Washington/USA, and Brussels-Luxembourg, November 1997, published December 1997. Volume 1: Main report, Volume 2: Appendices
 - 6.1.6. M. Little, C.M. Muirhead, L.H.J. Goossens, F.T. Harper, B.C.P. Kraan, R.M. Cooke and S.C. Hora (1997) *Probabilistic accident consequence uncertainty study: Late health effects uncertainty assessment* Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities, NUREG/CR-6555, EUR 16774, SAND97-2322 Washington/USA, and Brussels-Luxembourg, September 1997, published December 1997. Volume 1: Main report, Volume 2: Appendices

6.1.7. L.H.J. Goossens, J.D. Harrison, F.T. Harper, B.C.P. Kraan, R.M. Cooke and S.C. Hora (1998) *Probabilistic accident consequence uncertainty study: Uncertainty assessment for internal dosimetry* Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities, NUREG/CR-6571, EUR 16773, SAND98-0119 Washington/USA, and Brussels-Luxembourg, February 1998, published April 1998 Volume 1: Main report, Volume 2: Appendices

6.2. Probabilistic accident consequence uncertainty analysis Reports published on the project uncertainty analysis of the probabilistic accident consequence code cosyma using expert judgement (under the fourth EC-framework programme)

6.2.1. *Radiation Protection and Dosimetry Special Issue* , vol. 90 no 3, 2000.

6.2.2. R.M. Cooke, L.H.J. Goossens, B.C.P. Kraan (2000) Probabilistic Accident Consequence Uncertainty Assessment *Procedures Guide Using Expert Judgement* EUR 18820EN European Commission. Luxembourg 2000, Euratom.

6.2.3. L.H.J. Goossens, J.A. Jones, J. Ehrhardt, B.C.P. Kraan (2001) *Probabilistic Accident Consequence Uncertainty Assessment Countermeasures Uncertainty Assessment* EUR 18821EN European Commission. Luxembourg 2001, Euratom.

6.2.4. J.A. Jones, J. Ehrhardt, F. Fischer, I. Hasemann, L.H.J. Goossens, B.C.P. Kraan, R.M. Cooke (2001) *Probabilistic Accident Consequence Uncertainty Assessment Using COSYMA Uncertainty from the Atmospheric Dispersion and Deposition Module* EUR 18822EN European Commission. Luxembourg 2001, Euratom.

6.2.5. J.A. Jones, J. Brown, F. Fischer, I. Hasemann, L.H.J. Goossens, B.C.P. Kraan, R.M. Cooke (2001) *Probabilistic Accident Consequence Uncertainty Assessment Using COSYMA: Uncertainty from the Food Chain Module* EUR 18823EN European Commission. Luxembourg 2001, Euratom.

6.2.6. J.A. Jones, F. Fischer, I. Hasemann, L.H.J. Goossens, B.C.P. Kraan, R.M. Cooke (2001) *Probabilistic Accident Consequence Uncertainty Assessment Using COSYMA Uncertainty from the Health Effects Module* EUR 18824EN European Commission. Luxembourg 2001, Euratom.

6.2.7. J.A. Jones, F. Fischer, I. Hasemann, L.H.J. Goossens, B.C.P. Kraan, R.M. Cooke (2001) *Probabilistic Accident Consequence Uncertainty Assessment Using COSYMA Uncertainty from the Dose Module* EUR 18825EN European Commission. Luxembourg 2001, Euratom.

6.2.8. J.A. Jones, J. Ehrhardt, L.H.J. Goossens, F. Fischer, I. Hasemann, B.C.P. Kraan, R.M. Cooke (2001) *Probabilistic Accident Consequence Uncertainty Assessment Using COSYMA Uncertainty from the Complete System* EUR 18826EN European Commission. Luxembourg 2001, Euratom

6.2.9. J.A. Jones, B.C.P. Kraan, R.M. Cooke, L.H.J. Goossens, F. Fischer, I. Hasemann (2001) *Probabilistic Accident Consequence Uncertainty Assessment Using COSYMA Methodology and Processing Techniques* EUR 18827EN European Commission. Luxembourg 2001, Euratom.

6.2.10. Klügel, J-U., (2008) “Seismic Hazard Analysis — Quo vadis?” *Earth-Science Reviews* 88, 1–32

6.3. Ecosystems and public health

- 6.3.1. Tuomisto J.T, Wilson A., Cooke R.M., Tainio M., Evans J.S. (2005) "Mortality in Kuwait due to PM from oil fires after the Gulf War: Combining expert elicitation assessments" (2005) *Epidemiology*, Volume 16 (5) September 2005 p. S74-S75
- 6.3.2. Evans J.S., Wilson A, Tuomisto JT, Tainio M, Cooke RM (2005) "What risk assessment can tell us about the mortality impacts of the Kuwaiti oil fires" *Epidemiology*, Volume 16 {5} September 2005 p.S137-S138
- 6.3.3. Burgman, M. (2005) *Risk and Decisions for Conservation and Environmental Management*, Cambridge University Press, Cambridge, UK.
- 6.3.4. Van der Fels-Klerx, H.J., Cooke, R.M., Nauta, M.J., Goossens, L.H.J., Havelaar, A.H. (2005) "A Structured Expert Judgement Study For A Model of *Campylobacter* Transmission During Broiler Chicken Processing" *Risk Analysis* 25 No. 1, 2005, pp 109-124.
- 6.3.5. Cooke, R.M. Wilson,A.M., Tuomisto,J.T. Morales,O. Tainio,M. and Evans, J.S.. (2007) A Probabilistic Characterization of the Relationship Between Fine Particulate Matter and Mortality: Elicitation of European Experts. *Environmental Science and Technology* 2007 Sep 15;41(18): pp 6598-6605.
- 6.3.6. Kurowicka, D., Cooke, R.M. Goossens, L. and Ale, B.J.M, (2008) Expert Judgment study for Placement Ladder Bowtie, *Safety Science*, Volume 46, Issue 6, July 2008, Pages 921–93.
- 6.3.7. Radboud J. Duintjer Tebbens, Kimberly M. Thompson, M.G. Myriam Hunink, M.D., Thomas M Mazzuchi, Daniel Lewandowski, Dorota Kurowicka, Roger M. Cooke, (2008) "Uncertainty And Sensitivity Analyses Of A Dynamic Economic Evaluation Model For Vaccination Programs" *Medical Decision Making* 2008
- 6.3.8. Arie H. Havelaar, Ángela Vargas Galindo, Dorota Kurowicka, Roger M. Cooke (2008) Attribution of Foodborne Pathogens Using Structured Expert Elicitation, *Foodborne Pathogens and Disease*, October 2008, 5(5): 649-659. doi:10.1089/fpd.2008.0115
- 6.3.9. Rothlisberger, J.D., Lodge, D.M. Cooke, R.M. and Finnoff, D.C. (2009) "Future declines of the binational Laurentian Great Lakes fisheries: recognizing the importance of environmental and cultural change" *Frontiers in Ecology and the Environment*; doi:10.1890/090002
- 6.3.10. Neslo R.E.J, and Cooke, R.M.,(2011) "Modeling and Validating Stakeholder Preferences with Probabilistic Inversion" *Applied Stochastic Models in Business and Industry*.
- 6.3.11. Kurowicka, D. Bucura, C. Cooke, R.M. and Havelaar, A. (2010) Probabilistic Inversion in Priority Setting of Emerging Zoonoses, *Risk Analysis*, Vol. 30, No. 5, 2010 DOI: 10.1111/j.1539-6924.2010.01378.x
- 6.3.12. Kurowicka, D. Nauta, M. Jozwiak,, K. 1 and Cooke, R.M. (2010) " Updating Parameters of the Chicken Processing Line Model" *Risk Analysis*, Vol. 30, No. 6, 2010 DOI: 10.1111/j.1539-6924.2010.01379.x
- 6.3.13. Sarah J. Teck, Benjamin S. Halpern, Carrie V. Kappel, Fiorenza Micheli, Kimberly A. Selkoe, Caitlin M. Crain, Rebecca Martone, Christine Shearer, Joe Arvai, Baruch Fischhoff, Grant Murray, Rabin Neslo, and Roger Cooke (2010) Using expert judgment to estimate marine ecosystem vulnerability in the

- California Current, . *Ecological Applications*, 20(5), 2010, pp. 1402–1416 2010 by the Ecological Society of America
- 6.3.14. Burgman, M. Carr, A. Godden, L. Gregory, R. McBride, M. Flander, L. Maguire, L.(2010) “Redefining expertise and improving ecological judgement” under review *Conservation Letters*.
- 6.3.15. Burgman, M. McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., Fidler, F., Rumpff, L. and Twardy, C., (2011) “Perception of expertise and performance of experts: How do they compare?”
- 6.3.16. Villie Flari, Qasim Chaudhry, Rabin Neslo and Roger Cooke (2011) Expert judgment based multi-criteria decision model to address uncertainties in risk assessment of nanotechnology-enabled food products *J Nanopart Res (2011) 13: 1813–1831; DOI 10.1007/s11051-011-0335-x*
- 6.3.17. Martin, T.G. Burgman, M.A., Fidler, F.F. , Kuhnert, P.M. , Low-Choy, S., McBride, M. , and Mengersen, K., (2011) eliciting expert knowledge in conservation science *Conservation biology*, volume 26, no. 1, 29–38 2011 *Conservation biology c* _ 2011 society for conservation biology Doi: 10.1111/j.1523-1739.2011.01806.x
- 6.3.18. Tyshenko, M.G., S. ElSaadany, T. Oraby, M. Laderoute, J. Wu, W. Aspinall and D. Krewski (2011) Risk Assessment and Management of Emerging Blood-Borne Pathogens in Canada: Xenotropic Murine Leukaemia Virus-Related Virus as a Case Study for the Use of a Precautionary Approach. Chapter in: *Risk Assessment* (ISBN 979-953-307-765-8).
- 6.3.19. Tyshenko, M.G., ElSaadany, S., Oraby, T., Darshan, S., Catford, A., Aspinall, W., Cooke, R. and Krewski, D. (2012) Expert judgement and re-elicitation for prion disease risk uncertainties. *International Journal of Risk Assessment and Management*, 16(1-3), 48-77. doi:10.1504/IJRAM.2012.047552
- 6.3.20. Tyshenko, M.G., S. ElSaadany, T. Oraby, S. Darshan, W. Aspinall, R. Cooke, A. Catford, and D. Krewski (2011) Expert elicitation for the judgment of prion disease risk uncertainties. *J Toxicol Environ Health A.*; 74(2-4):261-285.
- 6.3.21. Oraby, T., Tyshenko, M.G., Westphal, M., Darshan, S., Croteau, M., Aspinall, W., Elsaadany, S., Cashman, N. and Krewski, D. (2011) Using Expert Judgments to Improve Chronic Wasting Disease Risk Management in Canada. *Journal of Toxicology and Environmental Health*, in press. Volume 74, Issue 2-4, 2011 Special Issue: Prion Research in Perspective 2010
- 6.3.22. Rothlisberger, J.D. Finnoff, D.C. Cooke, R.M. and Lodge, D.M. (2012) “Ship-borne nonindigenous species diminish Great Lakes ecosystem services” *Ecosystems* (2012) 15: 462–476 DOI: 10.1007/s10021-012-9522-6
- 6.3.23. Halpern, B.S., Longo, C McLeod. C.L., Cooke, Roger M., Fischhoff, B., Samhuri, J.F., Scarborough, C.(2013) Elicited preferences for components of ocean health in the California Current, *Marine Policy* 42 (2013) 68–73.
- 6.3.24. Cashman, N.R., Cheung, R., Aspinall, W., Wong, M. and Krewski, D. (2014) Expert Elicitation for the Judgment of Prion Disease Risk Uncertainties associated with Urine-derived and Recombinant Fertility Drugs. Submitted to: *Journal of Toxicology and Environmental Health*.
- 6.3.25. Colson, Abigail R., Sweta Adhikari, Ambereen Sleemi, and Ramanan Laxminarayan. (2015) “Quantifying Uncertainty in Intervention Effectiveness with

- Structured Expert Judgment: An Application to Obstetric Fistula.” *BMJ Open*, 1–8. doi:10.1136/bmjopen-2014-007233.
- 6.3.26. Fischer K, Lewandowski D, Janssen MP. Estimating unknown parameters in haemophilia using expert judgement elicitation. *Haemophilia*. 2013 Sep;19(5):e282-e288.
- 6.3.27. Cooke, R.M., Wittmann, M.E., Lodge, D.M., Rothlisberger, J.D., Rutherford E.S., Zhang, H. and Mason, D.M. (2014) "Out-of-Sample Validation for Structured Expert Judgment of Asian Carp Establishment in Lake Erie", *Integrated Environmental Assessment and Management*, open access. DOI: 10.1002/ieam.1559 <http://onlinelibrary.wiley.com/doi/10.1002/ieam.1559/abstract> Hald, T.
- 6.3.28. Koch, Benjamin J., Filoso, S., Cooke, R. M. Hosen, J. D., Colson, A.R. Febria, Catherine M., Palmer, M. A. , (2015) Nitrogen in stormwater runoff from Coastal Plain watersheds: The need for empirical data, reply to Walsh , *Elementa* DOI 10.12952/journal.elementa.000079. <https://www.elementascience.org/articles/79>
- 6.3.29. Aspinall,W. Devleesschauwer, B. Cooke, R.M., Corrigan,T., Havelaar, A.H., Gibb, H., Torgerson, P., Kirk, M., Angulo, F., Lake, R., Speybroeck, N., and Hoffmann, S. (2015) World Health Organization estimates of the relative contributions of food to the burden of disease due to selected foodborne hazards: a structured expert elicitation. *PLOS ONE*, : January 19, 2016 DOI: 10.1371/journal.pone.0145839, <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0145839>
- 6.3.30. Zhang, H, Rutherford E.S., Mason, D.M., Breck, J,T,, Wittmann M.E., Cooke R.M., Lodge D.M., Rothlisberger J.D., Zhu X., and Johnson, T B., (2015) Forecasting the Impacts of Silver and Bighead Carp on the Lake Erie Food Web, *Transactions of the American Fisheries Society*, Volume 145, Issue 1, pp 136-162, DOI:10.1080/00028487.2015.1069211
- 6.3.31. Koch, Benjamin J., Febria, Catherine M. , Cooke, Roger M. Hosen, Jacob D. , Baker, Matthew E. , Colson, Abigail R. Filoso, Solange, Hayhoe, Katharine, Loperfido, J.V. , Stoner, Anne M.K. , Palmer, Margaret A. , (2015) Suburban watershed nitrogen retention: Estimating the effectiveness of storm water management structures, *Elementa*, DOI 10.12952/journal.elementa.000063 <https://www.elementascience.org/articles/63>
- 6.3.32. Hoffmann, S. Aspinall,W. Cooke, R.M., Cawthorne,A., Corrigan,T. Havelaar,A. Gibb H., Torgerson,P. Kirk,M. Angulo,F., Lake,R. Speyboeck,N. Devleesschauwer, B. Hald,T. (2015) Research Synthesis Methods in an Age of Globalized Risks: Lessons from the Global Burden of Foodborne Disease Expert Elicitation, accepted for publication in *Risk Analysis*.
- 6.3.33. Hald, T. Aspinall,W. Devleesschauwer, B. Cooke, R.M., Corrigan,T., Havelaar, A.H., Gibb, H., Torgerson, P., Kirk, M., Angulo, F., Lake, R., Speybroeck, N., and Hoffmann, S. (2015) World Health Organization estimates of the relative contributions of food to the burden of disease due to selected foodborne hazards: a structured expert elicitation. *PLOS ONE*, : January 19, 2016 DOI: 10.1371/journal.pone.0145839

6.4. Civil aviation and structural reliability

- 6.4.1. Cooke, R.M. & Jager, E. (1998) "Failure Frequency of Underground Gas Pipelines", *Risk Analysis*, vol. 1, no 4, 511-527, 1998.
- 6.4.2. Cooke, R.M., E. Jager, D. Lewandowski (002) "Reliability model for underground gas pipelines" *Probabilistic Safety Assessment and Management* E.J. Bonano, A.L. Camp, M.J. Majors, R.A. Thompson (eds), Elsevier, 2002; 1045-1050.
- 6.4.3. Cooke, R.M., Eric Jager and D. Lewandowski "Reliability Model for Underground Gas Pipelines" (2003) *Case Studies in Reliability and Maintenance*. Edited by Wallace R. Blischke, D.N. Prabhakar Murthy; p. 423-446, ISBN: 0-471-41373-9, 2003, John Wiley and Sons, Inc.
- 6.4.4. Cooke, R.M. and Slijkhuis, Karen A. (2003) "Expert Judgment in the Uncertainty Analysis of Dike Ring Failure Frequency" *Case Studies in Reliability and Maintenance*. Edited by Wallace R. Blischke, D.N. Prabhakar Murthy; p. 331-352, ISBN: 0-471-41373-9, 2003, John Wiley and Sons, Inc.
- 6.4.5. Brown A.J. and Aspinall W.P. (2004) Use of expert opinion elicitation to quantify the internal erosion process in dams. In *Proc: The 13th Biennial British Dams Society Conference*: University of Kent, Canterbury, 22-26th June 2004; 16pp (<http://www.britishdams.org/2004conf/synopses/brown.pdf>).
- 6.4.6. Cooke, R.M. and Goossens L.H.G.(2004) "Expert judgement elicitation for risk assessments of critical infrastructures" *Journal of Risk Research* vol. 7 issue 6, 2004, ISSN 1366-9877, pp 643-657.
- 6.4.7. B.J.M. Ale, L.J. Bellamy, A.L.C. Roelen, R.M. Cooke, L.H.J. Goossens A.R. Hale, D. Kurowicka, E. Smith, (2005) Development of a causal model for air transport safety, *Proceedings of IMECE2005, 2005 ASME International Mechanical Engineering Congress and Exposition* November 5-11, 2005, Orlando, Florida USA IMECE2005-79374
- 6.4.8. Ale, B.J.M. Bellamy, L.J., Boom, R. van der, Cooper, J., Cooke, R.M., Goossens, L.H.J., Hale, A.R., Kurowicka, D., Morales, O. Roelen, A.L.C. and Spouge, J.(2009). "Further development of a Causal model for Air Transport Safety (CATS); Building the mathematical heart." *Reliability Engineering & System Safety*, Vol. 94, No. 9. (September 2009), pp. 1433-1441. doi:10.1016/j.ress.2009.02.024 Key: citeulike:5143231
- 6.4.9. Ale, B.J.M. (2006), L.J. Bellamy, R.M. Cooke, L.H.J. Goossens, A.R. Hale, A.L.C. Roelen, E. Smith, Towards a causal model for air transport safety – an ongoing research project, *SAFETY SCIENCE*, Volume 44, Issue 8, October 2006, Pages 657-673.
- 6.4.10. Ale, B., Bellamy, L.J., Cooke, R.M., Duyvis, M., Kurowicka, D. Lin, P.H., Morales, O., Roelen, A. Spouge, J. (2009) "Causal Model for Air Transport Safety, Final Report" Ministerie van Verkeer en Waterstaat, Directoraat-Generaal Luchtvaart en Maritieme Zaken, ISBN 10:90 369 1724-7; ISB N 13: 978 90 369 1724-7, The Hague, The Netherlands
- 6.4.11. Forsys, M.B., Kurowicka, D., Peppelman, B.(2013) "A probabilistic model for a gas explosion due to leakages in the grey cast iron gas mains" *Reliability Engineering & System Safety* volume 119, issue , year 2013, pp. 270 - 279.
- 6.4.12. Jaiswal, K.S., Wald, D.J., Perkins, D., Aspinall, W.P. and Kiremidjian, A.S. (2014) Estimating structural collapse fragility of generic building typologies using expert judgment. Chap 117 in: *Safety, Reliability, Risk and Life-Cycle*

Performance of Structures and Infrastructures (eds: Deodatis, G., Ellingwood, B.R., Frangopol, D.M.), CRC Press; 879-886.

- 6.4.13. Goodheart, B. (2013). Identification of causal paths and prediction of runway incursion risk by means of Bayesian belief networks. *Transportation Research Record: Journal of the Transportation Research Board*, (2400), 9-20.

6.5. Information security

- 6.5.1. Ryan, J.C.H. Mazzuchi, T.A. Ryan, D.J., Lopez de la Cruz, J. and Cooke, R.M. (2012) "Quantifying information security risks using expert judgment elicitation", *Computers & Operations Research Journal* 39, 774-784.
doi:10.1016/j.cor.2010.11.013

6.6. Natural hazards

- 6.6.1. Aspinall, W.P. (1997) Elicitation of expert opinions for decision-making: concepts, experience and issues. Paper presented to The Royal Society IDNDR conference "Extreme Natural Disasters - Mitigating Strategies for the 21st Century", March 10, 1997; London.
- 6.6.2. Aspinall W. and Cooke R.M. (1998) Expert judgement and the Montserrat Volcano eruption. *Proceedings of the 4th International Conference on Probabilistic Safety Assessment and Management PSAM4*, September 13th -18th 1998, New York City, USA (eds. Ali Mosleh and Robert A. Bari), Vol.3, 2113-2118.
- 6.6.3. Aspinall W.P., Loughlin S.C., Michael F.V., Miller A.D., Norton G.E., Rowley K.C., Sparks R.S.J. and Young S.R. (2002) The Montserrat Volcano Observatory: its evolution, organisation, role and activities. In: Druitt, T.H. & Kokelaar, B.P. (eds) *The eruption of Soufrière Hills Volcano, Montserrat, from 1995 to 1999*. Geological Society, London, Memoir.
- 6.6.4. Aspinall W.P., Woo G., Voight B. and Baxter P.J. (2003) Evidence-based volcanology: application to eruption crises. *Journal of Volcanology & Geothermal Research*, 123, 273-285.
- 6.6.5. Aspinall, W.P. (2006) Structured elicitation of expert judgment for probabilistic hazard and risk assessment in volcanic eruptions. In: Mader, H.M., Coles, S.G., Connor, C.B. & Connor, L.J. (eds) *Statistics in Volcanology*. Special Publications of IAVCEI, 1. Geological Society, London, 15-
- 6.6.6. Aspinall, W.P. (2006) Structured elicitation of expert judgment for probabilistic hazard and risk assessment in volcanic eruptions. Chapter 2, in *IAVCEI Volume Statistics in Volcanology*, 31pp.
- 6.6.7. Sparks, R.S.J., Aspinall, W.P. et al. (2013) Risk and uncertainty assessment of volcanic hazards. In "Risk and Uncertainty assessment in Natural Hazards." Rougier, J.C., Sparks R.S.J., Hill, L. (eds). Cambridge University Press, Chapter 11, 364-397.
- 6.6.8. Martí, J., W. P. Aspinall, R. Sobradelo, A. Felpeto, A. Geyer, R. Ortiz, P. Baxter, P. Cole, J. Pacheco, M.J. Blanco, C. Lopez (2008). A long-term volcanic hazard event tree for Teide-Pico Viejo stratovolcanoes (Tenerife, Canary Islands). *Journal of Volcanology and Geothermal Research*, 178, 543-552.
[doi:10.1016/j.jvolgeores.2008.09.023](https://doi.org/10.1016/j.jvolgeores.2008.09.023)

- 6.6.9. Queiroz, G., J. M. Pacheco, J.L. Gaspar, W.P. Aspinall, J.E. Guest and T. Ferreira. (2008). The last 5000 years of activity at Sete Cidades volcano (São Miguel Island, Azores): Implications for hazard assessment. *Journal of Volcanology and Geothermal Research*, 178, 562-573. doi:10.1016/j.jvolgeores.2008.03.001
- 6.6.10. Neri, A., Aspinall, W.P., Cioni, R., Bertagnini, A., Baxter, P.J., Zuccaro, G., Andronico, D., Barsotti, S., Cole, P.D., Esposti Ongaro, T., Hincks, T.K., Macedonio, G., Papale, P., Rosi, M., Santacroce, R., Woo, G. (2008) Developing an Event Tree for Probabilistic Hazard and Risk Assessment at Vesuvius. *Journal of Volcanology & Geothermal Research*, 178, 397-415. doi:10.1016/j.jvolgeores.2008.05.014
- 6.6.11. Aspinall, W.P. (2012) Comment on “Social studies of volcanology: knowledge generation and expert advice on active volcanoes” by Amy Donovan, Clive Oppenheimer and Michael Bravo [*Bull Volcanol* (2012) 74:677-689] *Bulletin of Volcanology*, 74, 1569-1570. doi: 10.1007/s00445-012-0625-x
- 6.6.12. Loughlin, S. C., Aspinall, W. P., Vye-Brown, C., Baxter, P. J., Braban, C., Hort, M., Schmidt, A., Thordarson, T., Witham, C. (2012) Large-magnitude fissure eruptions in Iceland: source characterisation. BGS Open File Report, OR/12/098; 231pp (available at: <http://www.bgs.ac.uk/research/volcanoes/LakiEruptionScenarioPlanning.html>)
- 6.6.13. Aspinall, W.P., Cooke, R.M., (2013) Expert Elicitation and Judgement. In “Risk and Uncertainty assessment in Natural Hazards.” Rougier, J.C., Sparks R.S.J., Hill, L. (eds). Cambridge University Press, Chapter 4, 64-99.
- 6.6.14. Hincks, T.K., Komorowski, J.-C., Sparks, S.R., and Aspinall, (2014) W.P., Retrospective analysis of uncertain eruption precursors at La Soufrière volcano, Guadeloupe, 1975-77: volcanic hazard assessment using a Bayesian Belief Network approach *Journal of Applied Volcanology* 2014, 3:3 doi:10.1186/2191-5040-3-3
- 6.6.15. Wadge, G. and Aspinall, W.P. (2014) A Review of Volcanic Hazard and Risk Assessments at the Soufrière Hills Volcano, Montserrat from 1997 to 2011. Ch. 24 in: *The Eruption of Soufrière Hills Volcano, Montserrat, from 2000 to 2010: Geological Society Memoirs*, Vol. 39, (edited by G. Wadge, R.E.A. Robertson and B. Voight). Geological Society: London; 439-456
- 6.6.16. Scourse, E., Aspinall, W.P. and Chapman, N. (2014) Using expert elicitation to characterise long-term tectonic risks to radioactive waste repositories in Japan. *Journal of Risk Research*, doi: 10.1080/13669877.2014.971334
- 6.6.17. Hicks, A., Hicks, A., Barclay, J., Simmons, P. and Loughlin, S. (2014). "An interdisciplinary approach to volcanic risk reduction under conditions of uncertainty: a case study of Tristan da Cunha." *Nat. Hazards Earth Syst. Sci.* 14(7): 1871-1887. Doi: 10.5194/nhess-14-1871-2014. www.nat-hazards-earth-syst-sci-discuss.net/1/7779/2013/
- 6.6.18. Bevilacqua, A., Isaia, R., Neri, A., Vitale, S., Aspinall, W.P. and eight others (2015) Quantifying volcanic hazard at Campi Flegrei caldera (Italy) with uncertainty assessment: I. Vent opening maps. *Journal of Geophysical Research - Solid Earth*; AGU. doi:10.1002/2014JB011775
- 6.6.19. Neri, A., Bevilacqua, A., Esposti Ongaro, T., Isaia, R., Aspinall, W.P. and nine others (2015) Quantifying volcanic hazard at Campi Flegrei caldera (Italy) with

uncertainty assessment: II. Pyroclastic density current invasion maps. *Journal of Geophysical Research - Solid Earth*; AGU. doi:10.1002/2014JB011776

6.7. Climate

- 6.7.1. Bamber, J.L., and Aspinall, W.P., (2012) An expert judgement assessment of future sea level rise from the ice sheets, *Nature Climate Change*, PUBLISHED ONLINE: January 6, 2012 | DOI: 10.1038/NCLIMATE1778. <http://www.nature.com/nclimate/journal/vaop/ncurrent/full/nclimate1778.html>
- 6.7.2. Gerstenberger, M.C.; McVerry, G.H.; Rhoades, D.A.; Stirling, M.W. (2014) Seismic hazard modeling for the recovery of Christchurch, New Zealand. *Earthquake Spectra*, 30(1): 17-29; doi: 10.1193/021913EQS037M
- 6.7.3. Christophersen, A.; Nicol, A.; Gerstenberger, M.C. (2011) The feasibility of using seed questions for weighting expert opinion in CCS risk assessment. Canberra: Cooperative Research Centre for Greenhouse Gas Technologies. *CO2CRC Report RPT11-2868*. 14 p.
- 6.7.4. Gerstenberger, M.C.; Christophersen, A.; Buxton, R.; Allinson, G.; Hou, W.; Leamon, G.; Nicol, A. (2013) Integrated risk assessment for CCS. p. 2775-2782; doi: 10.1016/j.egypro.2013.06.162 IN: Dixon, T.; Yamaji, K. (eds) *11th International Conference on Greenhouse Gas Control Technologies, 18th-22nd November 2012, Kyoto International Conference Center, Japan*. Elsevier. *Energy procedia* 37
- 6.7.5. Hanea, D. (2009) Human Risk of Fire: Building a decision support tool using Bayesian networks, PhD Thesis, Dept. Technical Administrative Science, T.U. Delft

6.8. Banking and finance

- 6.8.1. Van Overbeek, F.N.A. (1999) Financial experts in Uncertainty, Masters Thesis, Department of Mathematics, Delft University of Technology, Delft.
- 6.8.2. Qing, X. (2002) Risk analysis for real estate investment, PhD thesis, Dept. of Architecture, TU Delft.
- 6.8.3. Bakker, M. (2004) "Quantifying Operational Risks within Banks According to Basel II" Masters Thesis, Delft University of Technology, Dept. of Mathematics.

6.9. Completed, publication in preparation

- 6.9.1. Attribution study for Robert Wood Johnson Covering Kids & Families in Pennsylvania, Washington, Nebraska, Illinois, Arkansas, and Florida, conducted by Center for Disease Dynamics, Economics & Policy, 2012.
- 6.9.2. Colson, Abigail R., R.M. Cooke, R. Laxminarayan. (2015) "Attributing Impact to a Charitable Foundation's Programs with Structured Expert Judgment." Working Paper. Center for Disease Dynamics, Economics & Policy.
- 6.9.3. Colson, Abigail R., Mark A. Cohen, Sudha Regmi, Arindam Nandi, Ramanan Laxminarayan, and Molly K. Macauley. 2015. "Structured Expert Judgment for Informing the Return on Investment in Surveillance: The Case of Environmental Public Health Tracking." SSRN Scholarly Paper ID 2704189. Rochester, NY: Social Science Research Network. <http://papers.ssrn.com/abstract=2704189>.

- 6.9.4. Cashman, N.R., Cheung, R., Aspinall, W., Wong, M. and Krewski, D. (2014) Expert Elicitation for the Judgment of Prion Disease Risk Uncertainties associated with Urine-derived and Recombinant Fertility Drugs. Submitted to: Journal of Toxicology and Environmental Health.
- 6.9.5. Ioannou I., Aspinall W., Bisby L., Rush D. and Rossetto T. (2015) Judgment-Based Fragility Assessment of Reinforced Concrete Buildings Exposed to Fire (submitted to Engineering Structures Journal).
- 6.9.6. Bevilacqua, A., Isaia, R., Neri, A., Vitale, S., Aspinall, W.P. and eight others (2015) Quantifying volcanic hazard at Campi Flegrei caldera (Italy) with uncertainty assessment: I. Vent opening maps. Journal of Geophysical Research - Solid Earth; AGU. doi:10.1002/2014JB011775
- 6.9.7. Neri, A., Bevilacqua, A., Esposti Ongaro, T., Isaia, R., Aspinall, W.P. and nine others (2015) Quantifying volcanic hazard at Campi Flegrei caldera (Italy) with uncertainty assessment: II. Pyroclastic density current invasion maps. Journal of Geophysical Research - Solid Earth; AGU. doi:10.1002/2014JB011776.
- 6.9.8. Aspinall, W. and Blong, R. (2015) Volcanic Risk Management. Chapter 70 in: The Encyclopedia of Volcanoes, Second Edition (eds H. Sigurdsson & four others), Academic Press ISBN 978-0-12-385938-9; pp1215-1234.
- 6.9.9. Jaiswal, K.S., Wald, D.J., Perkins, D., Aspinall, W.P. and Kiremidjian, A.S. (2014) Estimating structural collapse fragility of generic building typologies using expert judgment. Chap 117 in: Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures (eds: Deodatis, G., Ellingwood, B.R., Frangopol, D.M.), CRC Press; 879-886.

References

- Bamber, J. L., and W. P. Aspinall. 2013. "An Expert Judgement Assessment of Future Sea Level Rise from the Ice Sheets." *Nature Climate Change* 3 (4): 424–27. doi:10.1038/nclimate1778.
- Bamber, J. L., W. P. Aspinall, and R. M. Cooke. 2016. "A Commentary on 'how to Interpret Expert Judgment Assessments of Twenty-First Century Sea-Level Rise' by Hylke de Vries and Roderik SW van de Wal." *Climatic Change* 137 (3–4): 321–28. doi:10.1007/s10584-016-1672-7.
- Bevilacqua, Andrea, Roberto Isaia, Augusto Neri, Stefano Vitale, Willy P. Aspinall, Marina Bisson, Franco Flandoli, et al. 2015. "Quantifying Volcanic Hazard at Campi Flegrei Caldera (Italy) with Uncertainty Assessment: 1. Vent Opening Maps." *Journal of Geophysical Research: Solid Earth* 120 (4): 2309–2329.
- Christophersen, A., A. Nicol, and Matthew C. Gerstenberger. 2011. "The Feasibility of Using Seed Questions for Weighting Expert Opinion in CCS Risk Assessment." RPT11-2868. CO2CRC. Canberra, Australia: Cooperative Research Centre for Greenhouse Gas Technologies.
- Colson, Abigail R., Sweta Adhikari, Ambereen Sleemi, and Ramanan Laxminarayan. 2015. "Quantifying Uncertainty in Intervention Effectiveness with Structured Expert Judgement: An Application to Obstetric Fistula." *BMJ Open* 5 (6): e007233. doi:10.1136/bmjopen-2014-007233.
- Colson, Abigail R., Mark A. Cohen, Sudha Regmi, Arindam Nandi, Ramanan Laxminarayan, and Molly K. Macauley. 2015. "Structured Expert Judgment for Informing the Return on Investment in Surveillance: The Case of Environmental Public Health Tracking." SSRN Scholarly Paper ID 2704189. Rochester, NY: Social Science Research Network. <http://papers.ssrn.com/abstract=2704189>.
- Colson, Abigail R., Roger M. Cooke, and Ramanan Laxminarayan. 2016. "Estimating the Attributable Impact of a Charitable Foundation's Programs Using Structured Expert Judgment." Working Paper.
- Cooke, Roger M. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. New York: Oxford University Press.
- . 2015. "Messaging Climate Change Uncertainty." *Nature Climate Change* 5 (1): 8–10. doi:10.1038/nclimate2466.
- Cooke, Roger M., and Louis L. H. J. Goossens. 2008. "TU Delft Expert Judgment Data Base." *Reliability Engineering & System Safety*, Expert Judgement, 93 (5): 657–74. doi:10.1016/j.ress.2007.03.005.
- Cooke, Roger M., and D. Solomatine. 1992. "EXCALIBUR--Integrated System for Processing Expert Judgments, User's Manual Version 3.0." Delft, The Netherlands: Delft University of Technology and SoLogic Delft.
- Cooke, Roger M., Marion E Wittmann, David M Lodge, John D Rothlisberger, Edward S Rutherford, Hongyan Zhang, and Doran M Mason. 2014. "Out-of-Sample Validation for Structured Expert Judgment of Asian Carp Establishment in Lake Erie." *Integrated Environmental Assessment and Management* 10 (4): 522–28. doi:10.1002/ieam.1559.
- Eggstaff, Justin W., Thomas A. Mazzuchi, and Shahram Sarkani. 2014. "The Effect of the Number of Seed Variables on the Performance of Cooke's Classical Model." *Reliability Engineering & System Safety* 121 (January): 72–82. doi:10.1016/j.ress.2013.07.015.

- Fischer, K., D. Lewandowski, and M. P. Janssen. 2013. "Estimating Unknown Parameters in Haemophilia Using Expert Judgement Elicitation." *Haemophilia: The Official Journal of the World Federation of Hemophilia* 19 (5): e282-288. doi:10.1111/hae.12166.
- Flandoli, F., E. Giorgi, Willy P. Aspinall, and A. Neri. 2011. "Comparison of a New Expert Elicitation Model with the Classical Model, Equal Weights and Single Experts, Using a Cross-Validation Technique." *Reliability Engineering & System Safety* 96 (10): 1292–1310. doi:10.1016/j.ress.2011.05.012.
- Forys, Monika B., Dorota Kurowicka, and Bas Peppelman. 2013. "A Probabilistic Model for a Gas Explosion due to Leakages in the Grey Cast Iron Gas Mains." *Reliability Engineering & System Safety* 119 (November): 270–79. doi:10.1016/j.ress.2013.06.034.
- Gerstenberger, Matthew C., A. Christophersen, R. Buxton, G. Allinson, W. Hou, G. Leamon, and A. Nicol. 2013. "Integrated Risk Assessment for CCS." *Energy Procedia*, GHGT-11 Proceedings of the 11th International Conference on Greenhouse Gas Control Technologies, 18-22 November 2012, Kyoto, Japan, 37: 2775–82. doi:10.1016/j.egypro.2013.06.162.
- Gerstenberger, Matthew C., Graeme McVerry, David Rhoades, and Mark Stirling. 2014. "Seismic Hazard Modeling for the Recovery of Christchurch." *Earthquake Spectra* 30 (1): 17–29. doi:10.1193/021913EQS037M.
- Hanea, D. M. 2009. "Human Risk of Fire: Building a Decision Support Tool Using Bayesian Networks." Dissertation, Delft, The Netherlands: TU Delft. <http://repository.tudelft.nl/view/ir/uuid%3A8c7e167e-ee61-488e-afef-9ad7723f4959/>.
- Hicks, A., J. Barclay, P. Simmons, and S. Loughlin. 2014. "An Interdisciplinary Approach to Volcanic Risk Reduction under Conditions of Uncertainty: A Case Study of Tristan Da Cunha." *Nat. Hazards Earth Syst. Sci.* 14 (7): 1871–87. doi:10.5194/nhess-14-1871-2014.
- Koch, Benjamin J., Catherine M. Febria, Roger M. Cooke, Jacob D. Hosen, Matthew E. Baker, Abigail R. Colson, Solange Filoso, et al. 2015. "Suburban Watershed Nitrogen Retention: Estimating the Effectiveness of Stormwater Management Structures." *Elementa: Science of the Anthropocene* 3 (July): 63. doi:10.12952/journal.elementa.000063.
- Koch, Benjamin J., Solange Filoso, Roger M. Cooke, Jacob D. Hosen, Abigail R. Colson, Catherine M. Febria, and Margaret A. Palmer. 2015. "Reply to Comment on 'Suburban Watershed Nitrogen Retention: Estimating the Effectiveness of Stormwater Management Structures' by Koch et Al. (Elem Sci Anth 3:000063, July 2015)." *Elementa: Science of the Anthropocene* 3 (December): 79. doi:10.12952/journal.elementa.000079.
- Laxminarayan, Ramanan, Abigail R. Colson, Sweta Adhikari, and Roger M. Cooke. 2012. "The Value of Giving, Part 2." Princeton, NJ: Princeton University.
- Lin, Shi-Woei, and Chih-Hsing Cheng. 2008. "Can Cooke's Model Sift out Better Experts and Produce Well-Calibrated Aggregated Probabilities?" In *IEEE International Conference on Industrial Engineering and Engineering Management, 2008. IEEM 2008*, 425–29. doi:10.1109/IEEM.2008.4737904.
- . 2009. "The Reliability of Aggregated Probability Judgments Obtained through Cooke's Classical Model." *Journal of Modelling in Management* 4 (2): 149–61. doi:10.1108/17465660910973961.
- Loughlin, S. C., Willy P. Aspinall, C. Vye-Brown, P. J. Baxter, C. Braban, M. Hort, A. Schmidt, T. Thordarson, and C. Witham. 2013. "Large-Magnitude Fissure Eruptions in Iceland: Source Characterisation." BGS Open File Report OR/12/098. British Geological Survey. <http://www.bgs.ac.uk/research/volcanoes/LakiEruptionScenarioPlanning.html>.

- Neri, Augusto, Andrea Bevilacqua, Tomaso Esposti Ongaro, Roberto Isaia, Willy P. Aspinall, Marina Bisson, Franco Flandoli, et al. 2015. "Quantifying Volcanic Hazard at Campi Flegrei Caldera (Italy) with Uncertainty Assessment: 2. Pyroclastic Density Current Invasion Maps." *Journal of Geophysical Research: Solid Earth* 120 (4): 2014JB011776. doi:10.1002/2014JB011776.
- Rothlisberger, John D., David C. Finnoff, Roger M. Cooke, and David M. Lodge. 2012. "Ship-Borne Nonindigenous Species Diminish Great Lakes Ecosystem Services." *Ecosystems* 15 (3): 1–15. doi:10.1007/s10021-012-9522-6.
- Rothlisberger, John D, David M Lodge, Roger M. Cooke, and David C Finnoff. 2010. "Future Declines of the Binational Laurentian Great Lakes Fisheries: The Importance of Environmental and Cultural Change." *Frontiers in Ecology and the Environment* 8 (5): 239–44. doi:10.1890/090002.
- Scourse, E., W. P. Aspinall, and N. Chapman. 2015. "Using Expert Elicitation to Characterise Long-Term Tectonic Risks to Radioactive Waste Repositories in Japan." *Journal of Risk Research* 18 (3): 364–77. doi:10.1080/13669877.2014.971334.
- Tyshenko, Michael G, Susie ElSaadany, Tamer Oraby, Shalu Darshan, Willy P. Aspinall, Roger M. Cooke, Angela Catford, and Daniel Krewski. 2011. "Expert Elicitation for the Judgment of Prion Disease Risk Uncertainties." *Journal of Toxicology and Environmental Health. Part A* 74 (2–4): 261–85. doi:10.1080/15287394.2011.529783.
- Tyshenko, Michael G., Susie ElSaadany, Tamer Oraby, Shalu Darshan, Angela Catford, Willy Aspinall, Roger Cooke, and Daniel Krewski. 2012. "Expert Judgement and Re-Elicitation for Prion Disease Risk Uncertainties." *International Journal of Risk Assessment and Management* 16 (1–3): 48–77. doi:10.1504/IJRAM.2012.047552.
- Tyshenko, Michael G., Susie ElSaadany, Tamer Oraby, Marian Laderoute, Jun Wu, Willy Aspinall, Daniel Krewski, and Peter R. 2012. "Xenotropic Murine Leukemia Virus-Related Virus as a Case Study: Using a Precautionary Risk Management Approach for Emerging Blood-Borne Pathogens in Canada." In *The Continuum of Health Risk Assessments*, edited by Michael G. Tyshenko. InTech. <http://www.intechopen.com/books/the-continuum-of-health-risk-assessments/xenotropic-murine-leukemia-virus-related-virus-as-a-case-study-using-a-precautionary-risk-managemen>.
- Walsh, Christopher J. 2015. "Comment on 'Suburban Watershed Nitrogen Retention: Estimating the Effectiveness of Stormwater Management Structures' by Koch et Al. (Elem Sci Anth 3:000063, July 2015)." *Elementa: Science of the Anthropocene* 3 (December): 77. doi:10.12952/journal.elementa.000077.
- Wittmann, Marion E., Roger M. Cooke, John D. Rothlisberger, and David M. Lodge. 2014. "Using Structured Expert Judgment to Assess Invasive Species Prevention: Asian Carp and the Mississippi—Great Lakes Hydrologic Connection." *Environmental Science & Technology* 48 (4): 2150–56. doi:10.1021/es4043098.
- Wittmann, Marion E., Roger M. Cooke, John D. Rothlisberger, Edward S. Rutherford, Hongyan Zhang, Doran M. Mason, and David M. Lodge. 2015. "Use of Structured Expert Judgment to Forecast Invasions by Bighead and Silver Carp in Lake Erie." *Conservation Biology* 29 (1): 187–97. doi:10.1111/cobi.12369.
- Zhang, Hongyan, Edward S. Rutherford, Doran M. Mason, Jason T. Breck, Marion E. Wittmann, Roger M. Cooke, David M. Lodge, John D. Rothlisberger, Xinhua Zhu, and Timothy B. Johnson. 2016. "Forecasting the Impacts of Silver and Bighead Carp on the Lake Erie

Food Web.” *Transactions of the American Fisheries Society* 145 (1): 136–62.
doi:10.1080/00028487.2015.1069211.