Update (1.1) to ANDURIL—A MATLAB toolbox for ANalysis and Decisions with UnceRtaInty: Learning from expert judgments
ANDURYL

't Hart, Marcel; Leontaris, George; Morales Napoles, Oswaldo

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Software update

Update (1.1) to ANDURIL — A MATLAB toolbox for ANalysis and Decisions with UnceRtaInty: Learning from expert judgments ANDURYL

Cornelis Marcel Pieter 't Hart [a,b,*], Georgios Leontaris [a], Oswaldo Morales-Nápoles [a]

[a] *Civil Engineering and Geosciences, Delft University of Technology, The Netherlands*
[b] *Tunnel Engineering Consultants (TEC), Amersfoort, The Netherlands*

ARTICLE INFO

ABSTRACT

This is an update to PII: S2352711018300608

In this paper, we discuss ANDURYL, which is a Python-based open source successor of the MATLAB toolbox ANDURIL. The output of ANDURYL is in good agreement with the results obtained from ANDURIL and EXCALIBUR. Additional features available in ANDURYL, and not available in its predecessors, are discussed.

© 2018 The Authors. Published by Elsevier B.V. All rights reserved.

Code metadata

Current code version	Code: ANDURYL v1.0, Paper v1.1
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX_2019_237
Legal Code License	GNU General Public License
Code versioning system used	None
Software code languages, tools, and services used	Python, SCIPY, NUMPY, MATPLOTLIB
Compilation requirements, operating environments & dependencies	Python VERSION 3.6
If available Link to developer documentation/manual	PDF files containing material
Support email for questions	C.M.P.tHart@tudelft.nl

Software metadata

Current code version	ANDURYL v1.0
Permanent link to code/repository used for this code version	CodeOcean
Legal Code License	GNU General Public License
Code versioning system used	Code Ocean
Software code languages, tools, and services used	Python, SCIPY, NUMPY, MATPLOTLIB
Compilation requirements, operating environments & dependencies	Python VERSION 3.6
If available Link to developer documentation/manual	PDF files containing material
Support email for questions	C.M.P.tHart@tudelft.nl

Table 1
Overview of result comparison AI and AY against CC.

Software	Number of studies compared	Number of different scores in Table 2	Number of scores with approximation differences	Number of scores where AI = AY but different to CC	Relative agreement after correction for approximation and AI = AY
AI	18 (55%)	13 (96%)	4	9	100%
AY	33 (100%)	23 (96%)	8	9	99%

1. Motivation and significance

A MATLAB toolbox, named ANDURIL,[1] (AI), implementing Cooke's classical model [1] for structured expert judgment is presented in [2]. Until recently EXCALIBUR[2] (CC) was the only available software implementing Cooke's classical method. Though Eggstaff's studies were based on a MATLAB implementation[3] [3,4], the developed source code for these studies is not available for distribution.

In this paper we present ANDURYL (AY), which is a Python [5] implementation of Cooke's classical model [1]. The program name replacing the I with Y indicates that the AY source is based on Python instead of MATLAB. The program structure of AI has been retained in this implementation. The main obvious advantage of AY is that the MATLAB license required for AI is not required for AY. Other added features with respect to AI will be discussed along this paper.

2. Software description

AY is run from the command line with the Python function `main.py`, as it does not have a graphical user interface. Users can adapt the code to run their own studies in sequences as presented in `anduryl_example.py`. The program structure is setup in such a way that there is one main Python function `anduryl` which is used to run the full scope of AY. In this main script, the data obtained from expert judgments may be entered in order to conduct the desired analysis. The input variables are set as global variables and backed up. With 'restore' statements the variables can be reset to the original input values, which can be used in later calculations, but might also be useful in further developments of AY. In the current implementation, this is used in the process for investigating the robustness of the obtained Decision Makers (DM). The supported functionalities of Cooke's classical model in AY are:

1. Calculation of DM using global weights;
2. Calculation of DM using item weights;
3. Calculation of DM using equal or user defined weights;
4. Optimization of DM;
5. Robustness check itemwise;
6. Robustness check expertwise;
7. Plotting assessments itemwise;
8. Plotting robustness results.

The functions of AY are similar to the functions presented for AI. AY keeps its architecture as similar as possible to that of AI. The main difference however is in the function `calculate_weights`, which merges AI's functions `global_weights` and `item_weights`. A more detailed explanation of the program is presented in the Supplement. The remaining differences will be further discussed in Section 4. Next we present results of comparing AY's output to both CC and the MATLAB implementation AI.

---

3. Comparing output of ANDURYL with previous expert judgment studies

In [4], 33 post-2006 studies using Cooke's classical method are presented using CC. We use these data to compare output from AY to both CC and the MATLAB implementation AI of the previous paper [2]. Table 2 presents the results reported in Table 1 of [4] (the study name followed by CC) extended with calculations from AI (AI) and AY (AY). Table 2 includes the statistical accuracy (SA), information (In) and the combined scores (Co).

Equal weight, Global weights without optimization (Global No Op.), Global weights optimized (PW Global), Item weights optimized (PW Item) and the expert with highest combined score (Best Expert) are presented. In the supplement, an extended table including Item weights without optimization (Item No Op.) and the expert with the lowest combined score is presented.

From the 33 studies reported [4], 14 were performed using 5 quantiles, 3 with quantiles other than the 5th, 50th and 95th or contained missing items for some experts. These results cannot be compared with AI and are marked by (*). On the EBPP study, a software error appeared in the MATLAB code. This error will be resolved in a future update of AI. Hence, a total 18 studies were compared with AI. Each study in Table 2 presents 17 numbers. Differences between the calculations reported in [4] and AI are highlighted in blue. There are a total of 153 blue numbers in Table 2 and hence an agreement of $(1 - \frac{13}{306}) \times 100 \approx 96\%$ between AI and the calculations reported in [4] for the studies that can be compared. From the 13 numbers 4 are clearly approximation differences. Notice that though the numbers in CC are MATLAB-based we compare our results to the published results in [4] and no way to investigate further the approximation used in [4] is available to the authors. Additionally, 9 numbers are equal to the results obtained with AY. These two observations would bring the agreement to 100%.

Differences between the calculations reported in [4] and AY are highlighted in red in the same table. There are a total of 23 red numbers in Table 2 and hence an agreement of $(1 - \frac{23}{561}) \times 100 \approx 96\%$ between AY and the calculations reported in [4]. From the 23 red numbers 8 are clearly approximation differences. Additionally, 9 AY results are equal to those obtained with AI which would bring the agreement to ≈99%. This result indicate that both AI and AY may be used with enough confidence by interested users.

The results of the comparison are summarized in 1.

In Table 2, 9 values are equal for AI and AY but different compared to CC. The authors checked the input files of the "Icesheets" study. It was found that the realization file (*.rls) and the file with assessments (*.dtt) presented inconsistencies in the labeling of assessment questions. We speculate that this could be the source of this misalignment of both AI and AY with CC.

The differences found in the "Gerstenberger", "Goodheart" and "Hemopilia" study are related to the optimization process. For example, the optimization process for "Goodheart" data shows in CC 1 expert as the optimal combination. For both AI and AY the optimal combination consists of 3 experts. Without the source code of CC the authors cannot investigate further this source of misalignment.
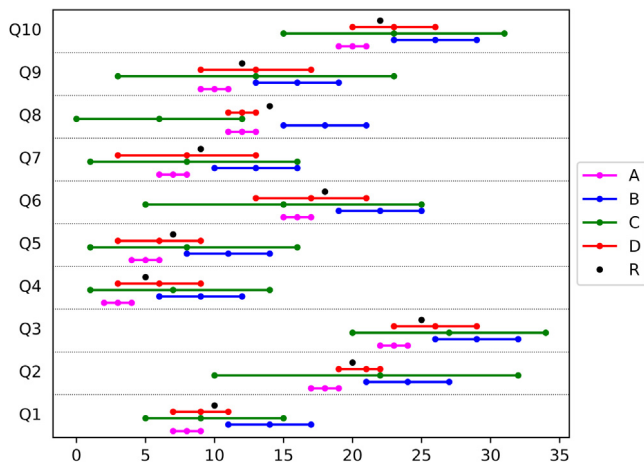
Table 2

Comparison of results presented in Table 1 of [4] (CC) and calculations with AI (AI) and AY (AY).

Study	♯E	♯S	Equal Weight			Global No Op.			PW Global			PW Item			Best Expert		
			Sa	In	Co	Sa	In	Co	Sa	In	Co	Sa	In	Co	Sa	In	Co
Arkansas (CC)	4	10	0.39	0.20	0.08	0.50	0.34	0.17	0.50	0.34	0.17	0.50	0.52	0.26	0.07	0.41	0.03
Arkansas* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Arkansas (AY)	4	10	0.39	0.20	0.08	0.50	0.34	0.17	0.50	0.34	0.17	0.50	0.52	0.26	0.07	0.41	0.03
Arsenic (CC)	9	10	0.06	1.10	0.07	0.04	1.68	0.06	0.04	2.74	0.10	0.04	2.74	0.10	0.04	2.74	0.10
Arsenic (AI)	9	10	0.06	1.10	0.07	0.04	1.68	0.06	0.04	2.74	0.10	0.04	2.74	0.10	0.04	2.74	0.10
Arsenic (AY)	9	10	0.06	1.10	0.07	0.04	1.68	0.06	0.04	2.74	0.10	0.04	2.74	0.10	0.04	2.74	0.10
ATCEP (CC)	5	10	0.12	0.25	0.03	0.68	0.23	0.16	0.68	0.23	0.16	0.24	0.38	0.09	0.10	0.50	0.05
ATCEP (AI)	5	10	0.12	0.25	0.03	0.68	0.23	0.16	0.68	0.23	0.16	0.24	0.38	0.09	0.10	0.50	0.05
ATCEP (AY)	5	10	0.12	0.25	0.03	0.68	0.23	0.16	0.68	0.23	0.16	0.24	0.38	0.09	0.10	0.50	0.05
Biol Agent (CC)	12	12	0.41	0.24	0.10	0.41	0.43	0.18	0.68	0.61	0.41	0.68	0.66	0.45	0.31	1.00	0.31
Biol Agent (AI)	12	12	0.41	0.24	0.10	0.41	0.43	0.18	0.68	0.61	0.41	0.68	0.66	0.45	0.31	1.00	0.31
Biol Agent (AY)	12	12	0.41	0.24	0.10	0.41	0.43	0.18	0.68	0.61	0.41	0.68	0.66	0.45	0.31	1.00	0.31
CDC ROI (CC)	20	10	0.23	1.23	0.29	0.39	1.35	0.52	0.72	2.31	1.66	0.72	2.31	1.66	0.72	2.31	1.66
CDC ROI* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
CDC ROI (AY)	20	10	0.23	1.23	0.29	0.39	1.35	0.52	0.72	2.30	1.66	0.72	2.30	1.66	0.72	2.30	1.66
CoveringKids (CC)	5	10	0.63	0.27	0.17	0.72	0.38	0.28	0.72	0.43	0.31	0.72	0.51	0.36	0.62	0.89	0.55
CoveringKids* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
CoveringKids (AY)	5	10	0.63	0.27	0.17	0.72	0.38	0.28	0.72	0.43	0.31	0.72	0.51	0.36	0.62	0.89	0.55
create-vicki (CC)	7	10	0.06	0.21	0.01	0.19	0.27	0.05	0.39	0.28	0.11	0.31	0.30	0.09	0.02	0.25	0.00
create-vicki (AI)	7	10	0.06	0.21	0.01	0.19	0.27	0.05	0.39	0.28	0.11	0.31	0.30	0.09	0.02	0.25	0.00
create-vicki (AY)	7	10	0.06	0.21	0.01	0.19	0.27	0.05	0.39	0.28	0.11	0.31	0.30	0.09	0.02	0.25	0.00
CWD (CC)	14	10	0.47	0.93	0.44	0.47	0.94	0.45	0.49	1.22	0.60	0.68	1.33	0.90	0.31	2.19	0.69
CWD (AI)	14	10	0.47	0.93	0.44	0.47	0.94	0.45	0.49	1.21	0.60	0.68	1.33	0.90	0.31	2.19	0.69
CWD (AY)	14	10	0.47	0.93	0.44	0.47	0.94	0.45	0.49	1.21	0.60	0.68	1.33	0.90	0.31	2.19	0.69
Daniela (CC)	4	7	0.53	0.17	0.09	0.68	0.23	0.16	0.55	0.63	0.35	0.55	0.63	0.35	0.55	0.63	0.35
Daniela (AI)	4	7	0.53	0.17	0.09	0.68	0.23	0.16	0.55	0.63	0.35	0.55	0.63	0.35	0.55	0.63	0.35
Daniela (AY)	4	7	0.53	0.17	0.09	0.68	0.23	0.16	0.55	0.63	0.35	0.55	0.63	0.35	0.55	0.63	0.35
DCPN Fistula (CC)	8	10	0.06	0.62	0.04	0.12	1.14	0.14	0.12	1.31	0.16	0.27	1.34	0.36	0.01	1.92	0.01
DCPN Fistula* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
DCPN Fistula (AY)	8	10	0.06	0.62	0.04	0.12	1.14	0.14	0.12	1.31	0.16	0.27	1.34	0.36	0.01	1.92	0.01
eBPP (CC)	14	15	0.36	0.32	0.11	0.36	0.43	0.15	0.83	1.41	1.17	0.83	1.41	1.17	0.83	1.41	1.17
eBPP (AI)[a]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
eBPP (AY)	14	15	0.36	0.32	0.11	0.36	0.43	0.15	0.83	1.41	1.17	0.83	1.41	1.17	0.83	1.41	1.17
Eff Erup (CC)	14	8	0.29	0.80	0.23	0.29	1.02	0.29	0.66	1.12	0.75	0.66	1.24	0.82	0.19	1.80	0.33
Eff Erup (AI)	14	8	0.29	0.80	0.23	0.29	1.02	0.29	0.66	1.12	0.75	0.66	1.24	0.82	0.19	1.80	0.33
Eff Erup (AY)	14	8	0.29	0.80	0.23	0.29	1.02	0.29	0.66	1.12	0.75	0.66	1.24	0.82	0.19	1.80	0.33
Erie Carp (CC)	11	15	0.31	0.29	0.09	0.57	0.45	0.25	0.76	0.86	0.65	0.76	0.86	0.65	0.53	1.29	0.68
Erie Carp* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Erie Carp (AY)	11	15	0.31	0.29	0.09	0.57	0.45	0.25	0.76	0.86	0.65	0.76	0.86	0.65	0.53	1.29	0.68
FCEP (CC)	5	8	0.22	0.10	0.02	0.14	0.39	0.06	0.66	0.57	0.38	0.66	0.57	0.38	0.66	0.57	0.38
FCEP (AI)	5	8	0.22	0.10	0.02	0.14	0.39	0.06	0.66	0.57	0.38	0.66	0.57	0.38	0.66	0.57	0.38
FCEP (AY)	5	8	0.22	0.10	0.02	0.14	0.39	0.06	0.66	0.57	0.38	0.66	0.57	0.38	0.66	0.57	0.38
Florida (CC)	7	10	0.76	0.46	0.34	0.56	0.80	0.45	0.76	1.13	0.86	0.76	1.15	0.87	0.12	1.74	0.22
Florida* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Florida (AY)	7	10	0.76	0.46	0.34	0.56	0.80	0.45	0.76	1.13	0.86	0.76	1.15	0.87	0.12	1.74	0.22
Gerstenberger (CC)	12	14	0.64	0.48	0.31	0.35	0.61	0.21	0.93	1.10	1.02	0.76	1.20	0.91	0.54	1.74	0.93
Gerstenberger* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Gerstenberger (AY)	12	14	0.64	0.48	0.31	0.35	0.61	0.21	0.93	1.09	1.02	0.76	1.09	0.82	0.54	1.74	0.93
GL NIS (CC)	9	13	0.04	0.31	0.01	0.93	0.21	0.19	0.93	0.21	0.19	0.93	0.26	0.24	0.45	0.27	0.12
GL NIS (AI)	9	13	0.04	0.31	0.01	0.93	0.21	0.19	0.93	0.21	0.19	0.93	0.26	0.24	0.45	0.27	0.12
GL NIS (AY)	9	13	0.04	0.31	0.01	0.93	0.21	0.19	0.93	0.21	0.19	0.93	0.26	0.24	0.45	0.27	0.12
Goodheart (CC)	6	10	0.55	0.28	0.15	0.47	0.35	0.16	0.71	0.96	0.68	0.71	0.96	0.68	0.71	0.96	0.68
Goodheart (AI)	6	10	0.55	0.28	0.15	0.47	0.35	0.16	0.47	0.35	0.17	0.68	0.64	0.43	0.71	0.96	0.68
Goodheart (AY)	6	10	0.55	0.28	0.15	0.47	0.35	0.16	0.47	0.35	0.17	0.68	0.64	0.43	0.71	0.96	0.68
Hemopilia (CC)	18	8	0.25	0.20	0.05	0.31	0.27	0.08	0.31	0.49	0.15	0.31	0.46	0.14	0.85	1.07	0.91
Hemopilia* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Hemopilia (AY)	18	8	0.25	0.20	0.05	0.31	0.27	0.08	0.31	0.30	0.09	0.31	0.15	0.05	0.85	1.07	0.91
IceSheets (CC)	10	11	0.49	0.52	0.25	0.62	0.70	0.43	0.40	1.55	0.62	0.62	1.04	0.64	0.40	1.55	0.62
IceSheets (AI)	10	11	0.49	0.52	0.25	0.37	0.66	0.25	0.40	1.55	0.62	0.62	1.04	0.64	0.40	1.55	0.62
IceSheets (AY)	10	11	0.49	0.52	0.25	0.37	0.66	0.25	0.40	1.55	0.62	0.62	1.04	0.64	0.40	1.55	0.62
Illinois (CC)	5	10	0.62	0.26	0.16	0.39	0.51	0.20	0.34	0.65	0.22	0.39	0.60	0.23	0.13	0.97	0.13
Illinois* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Illinois (AY)	5	10	0.62	0.26	0.16	0.39	0.51	0.20	0.34	0.65	0.22	0.39	0.60	0.23	0.13	0.97	0.13
Liander (CC)	11	10	0.23	0.48	0.11	0.23	0.50	0.11	0.23	0.52	0.12	0.68	0.75	0.51	0.00	0.86	0.00
Liander (AI)	11	10	0.23	0.48	0.11	0.23	0.50	0.11	0.23	0.52	0.12	0.68	0.75	0.51	0.00	0.86	0.00
Liander (AY)	11	10	0.23	0.48	0.11	0.23	0.50	0.11	0.23	0.52	0.12	0.68	0.75	0.51	0.00	0.86	0.00
Nebraska (CC)	4	10	0.37	0.70	0.26	0.03	1.25	0.04	0.03	1.45	0.05	0.03	1.45	0.05	0.03	1.45	0.05
Nebraska* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Nebraska (AY)	4	10	0.37	0.70	0.26	0.03	1.25	0.04	0.03	1.45	0.05	0.03	1.45	0.05	0.03	1.45	0.05
Obesity (CC)	4	10	0.07	0.24	0.02	0.50	0.23	0.12	0.44	0.51	0.22	0.78	0.49	0.38	0.44	0.51	0.22
Obesity* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Obesity (AY)	4	10	0.07	0.24	0.02	0.50	0.23	0.12	0.44	0.51	0.22	0.78	0.49	0.38	0.44	0.51	0.22
PHAC T4 (CC)	10	13	0.27	0.20	0.05	0.09	0.26	0.02	0.14	0.40	0.06	0.10	0.49	0.05	0.01	1.25	0.01
PHAC T4 (AI)	10	13	0.27	0.20	0.05	0.09	0.26	0.02	0.14	0.40	0.06	0.10	0.49	0.05	0.01	1.25	0.01
PHAC T4 (AY)	10	13	0.27	0.20	0.05	0.09	0.26	0.02	0.14	0.40	0.06	0.10	0.49	0.05	0.01	1.25	0.01
San Diego (CC)	8	10	0.33	1.07	0.36	0.78	0.69	0.54	0.88	0.69	0.61	0.35	1.19	0.41	0.03	1.12	0.04
San Diego* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
San Diego (AY)	8	10	0.33	1.07	0.36	0.78	0.69	0.54	0.88	0.69	0.61	0.35	1.19	0.41	0.03	1.12	0.04
Sheep (CC)	14	15	0.66	0.78	0.52	0.36	0.98	0.35	0.64	1.31	0.84	0.64	1.31	0.84	0.64	1.31	0.84
Sheep (AI)	14	15	0.66	0.78	0.52	0.36	0.98	0.35	0.64	1.31	0.84	0.64	1.31	0.84	0.64	1.31	0.84
Sheep (AY)	14	15	0.66	0.78	0.52	0.36	0.98	0.35	0.64	1.31	0.84	0.64	1.31	0.84	0.64	1.31	0.84
SPEED (CC)	14	16	0.52	0.75	0.39	0.63	0.71	0.45	0.68	0.78	0.53	0.99	0.78	0.78	0.23	0.84	0.19
SPEED (AI)	14	16	0.52	0.75	0.39	0.63	0.71	0.45	0.68	0.78	0.53	0.99	0.78	0.78	0.23	0.84	0.19
SPEED (AY)	14	16	0.52	0.75	0.39	0.63	0.71	0.45	0.68	0.78	0.53	0.99	0.78	0.78	0.23	0.84	0.19
TDC (CC)	18	17	0.17	0.36	0.06	0.30	0.55	0.17	0.99	1.26	1.24	0.99	1.26	1.24	0.99	1.26	1.24
TDC (AI)	18	17	0.17	0.36	0.06	0.30	0.55	0.17	0.99	1.26	1.24	0.99	1.26	1.24	0.99	1.26	1.24
TDC (AY)	18	17	0.17	0.36	0.06	0.30	0.55	0.17	0.99	1.26	1.24	0.99	1.26	1.24	0.99	1.26	1.24
Tobacco (CC)	7	10	0.20	0.45	0.09	0.66	0.53	0.35	0.69	1.06	0.73	0.69	1.06	0.73	0.69	1.06	0.73
Tobacco* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Tobacco (AY)	7	10	0.20	0.45	0.09	0.66	0.53	0.35	0.69	1.06	0.73	0.69	1.06	0.73	0.69	1.06	0.73
Topaz (CC)	21	16	0.63	0.92	0.58	0.31	1.12	0.34	0.41	1.46	0.60	0.41	1.46	0.60	0.41	1.46	0.60
Topaz (AI)	21	16	0.63	0.92	0.58	0.31	1.12	0.34	0.41	1.45	0.60	0.41	1.45	0.60	0.41	1.45	0.60
Topaz (AY)	21	16	0.63	0.92	0.58	0.31	1.12	0.34	0.41	1.45	0.60	0.41	1.45	0.60	0.41	1.45	0.60
UMD NREMOVAL (CC)	9	11	0.07	0.80	0.05	0.49	1.43	0.70	0.71	1.99	1.40	0.71	1.99	1.40	0.71	1.99	1.40
UMD NREMOVAL (AI)	9	11	0.07	0.80	0.05	0.49	1.43	0.70	0.71	1.99	1.40	0.71	1.99	1.40	0.71	1.99	1.40
UMD NREMOVAL (AY)	9	11	0.07	0.80	0.05	0.49	1.43	0.70	0.71	1.99	1.40	0.71	1.99	1.40	0.71	1.99	1.40
Washington (CC)	5	10	0.15	0.53	0.08	0.20	0.65	0.13	0.20	0.72	0.14	0.50	0.99	0.49	0.06	1.29	0.08
Washington* (AI)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Washington (AY)	5	10	0.15	0.53	0.08	0.20	0.65	0.13	0.20	0.72	0.14	0.50	0.99	0.49	0.06	1.29	0.08

[a] The authors found a software error in AI, this particular study has not been validated to AI. In a future update of AI the software error will be solved.

Fig. 1. Hypothetical example of 4 experts assessing 10 seed variables.

Table 3

Statistical accuracy and Informativeness computed with AY and CC for the hypothetical example presented in Fig. 1 assuming experts elicited 10th, 50th and 90th percentiles of their uncertainty distribution.

Expert ID	Calibration (CC)	Calibration (AY)	Information (CC)	Information (AY)
Expert A	5.529E−10	5.530E−10	1.371	1.371
Expert B	5.529E−10	5.530E−10	0.571	0.571
Expert C	0.371	0.371	0.039	0.039
Expert D	0.526	0.526	0.629	0.629
Global (non-opt.)	0.526	0.526	0.431	0.431

4. Impact

The advantages of AI, discussed in [2], with respect to CC are inherited by AY. A number of limitations of AI were discussed in the supplement of [2]. Besides the full open source character using Python as a programming language, two other advantages were implemented in comparison with CC and/or AI. These are elaborated further next.

4.1. User defined quantiles

From Table 2 it may be observed that AY presents good agreement with the 11 studies reported in [4] where 5 quantiles (5th, 25th 50th, 75th and 95th) were used to elicit expert judgments, hence we do not elaborate further on this issue.

As stated earlier, AY provides the option of user defined quantiles. CC allows for the use of 3, 4 or 5 user defined quantiles. Fig. 1 presents a hypothetical example of 4 experts: A, B C and D, assessing 10 calibration or seed variables. The realization (R) is also shown.

Intuitively, the reader may already appreciate that expert A will be informative but with low SA. Expert B will be less informative and also present low SA. The SA for C and D will be equal, however, D will be more informative than C. Table 3 presents a comparison of the calculations of SA and informativeness between AY and CC assuming experts elicited 10th, 50th and 90th percentiles of their uncertainty distribution. The reader may appreciate that the agreement between the calculations performed by CC and AY is almost exact.

Because the source code of AY is available and extended with respect to CC, practitioners may use more that 3, 4 or 5 user defined quantiles to elicit expert judgments. The same hypothetical example with four experts as in Table 3 is used but with experts assessing 7 quantiles (10th, 25th, 35th, 50th, 65th, 75th

Table 4

Statistical accuracy and Informativeness computed with AY with 7 quantiles for the hypothetical example presented in Section 4.1 assuming experts elicited 10th, 25th, 35th 50th, 65th, 75th and 90th percentiles of their uncertainty distribution.

Expert ID	Calibration score	Information score	Un-normalized weights	Normalized weights
Expert A	8.542E−08	1.3738	1.173E−07	9.403E−07
Expert B	8.542E−08	0.5710	4.877E−08	3.908E−07
Expert C	0.0041	0.0393	0.0002	0.0013
Expert D	0.1004	0.6302	0.0633	0.5069
Global (non-opt.)	0.1004	0.6114	0.0614	0.4918

and 90th) is presented in Table 4 (intermediate assessments have been obtained by interpolating linearly the estimates summarized in Fig. 1).

Though this option is available in AY, it is unclear to the authors its applicability in practice since the complexity of eliciting expert judgments grows significantly with the number of quantiles to be elicited from experts. It is also unclear to the authors if no study considered the elicitation of more than 5 quantiles because this feature was not available in any software implementation.

4.2. Missing items for some experts

In [6] two panels of 9 experts were gathered in order to assess uncertainty over economic growth and oil prices for Mexico in 2020 and 2030. In the panel corresponding to international gas and oil prices, expert A did not answer 10 of 26 calibration variables. No answer for expert D was recorded for 5 calibration variables. Similarly, no answer to 1 calibration variable was observed for expert G. The results of calculations obtained with missing items for both AY and CC are presented in Table 5. Similarly as in Table 3, the agreement between the calculations obtained with CC and AY is almost exact.

5. Conclusions

The MATLAB toolbox named AI for combining expert judgments applying Cooke's classical model for structured expert judgment has been extended. The new software is called AN-DURYL. The main purpose for developing these toolboxes is to create open source solutions that can be used by practitioners and researchers who are interested in applying or developing further Cooke's method. In comparison with AI and/or CC, AY presents the following new features:

AY has inherited all advantages of AI discussed in [2]. Additionally, AY is fully open source and allows for user defined quantiles (see 4.1) and missing items (see 4.2).

The software tool presented in this paper validates Cooke's classical model successfully with a range of studies presented in [4]. Despite the limitations of the current version of AY, it is to the authors belief that similarly as AI the developed toolbox will be valuable to those who are interested in developing and further applying the method. It is the ambition of the authors to extend AI and AY with more features than those currently available in CC and with the more recent techniques of elicitation of multivariate dependence [7].

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Table 5
Comparison of calculations from AY and CC for the expert panel presented in [6].

Expert ID	Calibration (CC)	Calibration (AY)	Information (CC)	Information (AY)	Information (CC)	Information (AY)
Expert A	1.634E−7	1.635E−7	1.347	1.347	1.235	1.235
Expert D	0.07205	0.07209	1.045	1.045	1.004	1.004
Expert G	0.0004775	0.0004774	1.075	1.0745	1.262	1.262
Global (non-opt.)	0.1512	0.1512	0.8549	0.8549	0.8683	0.8684

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.softx.2019.100295.

References

[1] Cooke R. Experts in uncertainty: Opinion and subjective probability in science. Environmental ethics and science policy, Oxford University Press; 1991.

[2] Leontaris G, Morales-Nápoles O. ANDURIL - a MATLAB toolbox for ANal-ysis and Decisions with UnceRtaInty: Learning from expert judgments. SoftwareX 2018;7:313–7. http://dx.doi.org/10.1016/j.softx.2018.07.001.

[3] Eggstaff JW, Mazzuchi TA, Sarkani S. The effect of the number of seed variables on the performance of Cooke's classical model. Reliab Eng Syst Saf 2014;121:72–82.

[4] Colson AR, Cooke RM. Cross validation for the classical model of structured expert judgment. Reliab Eng Syst Saf 2017;163:109–20.

[5] Oliphant T. Python for scientific computing. Comput Sci Eng 2007;9:10–20. http://dx.doi.org/10.1109/MCSE.2007.58.

[6] Puig D, Morales-Nápoles O, Bakhtiari F, Landa G. The accountability im-perative for quantifying the uncertainty of emission forecasts: evidence from Mexico. Climate Policy 2018;18(6):742–51. http://dx.doi.org/10.1080/14693062.2017.1373623.

[7] Werner C, Bedford T, Cooke RM, Hanea AM, Morales-Nápoles O. Ex-pert judgement for dependence in probabilistic modelling: A systematic literature review and future research directions. European J Oper Res 2017;258(3):801–19.