Comparing Causal Discovery Methods using Synthetic and Real Data

Christoph Käding and Jakob Runge

DLR German Aerospace Center
Institute of Data Science, Jena, Germany

May 7, 2020

DLR

Motivation

- Unveiling causal structures, *i.e.,* distinguishing cause from effect, from observational data plays a key role in climate science.
- Various techniques are available to approach this while each comes with own assumptions about the data.
  $\rightarrow$ restricted applicability
- **Goal:** Evaluate and compare a number of state-of-the-art methods in a joint benchmark.
  - Synthetic data allows for controlling dataset conditions in detail.
  - Real data gives performance indicators for applications.
- **Disclaimer:**
  - We concentrate on the case with two uni-variate variables.
  - This is an intermediate report of an ongoing study.

Methods I

Adaptive noise model (**ANM**) by [Hoyer et al., 2009]

- decision according to regression residuals

Causal additive models (**CAM**) by [Bühlmann et al., 2014]

- relies on fitting Gaussian processes

Concave penalized coordinate descent with reparametrization (**CCDr**) by [Aragam and Zhou, 2015]

- utilizes a score-based structure learning approach

Conditional distribution similarity (**CDS**) by [Fonollosa, 2016]

- decision according to the standard deviation of conditional distributions

Information geometric causal inference (**IGCI**) by [Daniusis et al., 2012]

- relies on differential entropies

Methods II

Kernel conditional deviance for causal inference (**KCDC**) by [Mitrovic et al., 2018]

- decision according to smaller deviance in the norms of conditional kernel mean embeddings

Linear non-Gaussian acyclic model (**LiNGAM**) by [Shimizu et al., 2006]

- utilizes independent component analysis to resolve directions

Non-combinatorial optimization via trace exponential and augmented lagrangian for structure learning (**noTEARS**) by [Zheng et al., 2018]

- reformulation of the problem as continuous optimization task

Randomized causation coefficient (**RCC**) by [Lopez-Paz et al., 2015]

- transforms data into a feature representation and decides with the help of trained classifiers

Regression error based causal inference (**RECI**) by [Blöbaum et al., 2018]

- decision according to least-squares errors of regressors fitted for both directions

Synthetic Data

Synthetic data allows for constructing datasets with specific conditions. Therefore, we extend the functions used by [Mitrovic et al., 2018]:

lin_a: $Y = X + 10 + \epsilon$
lin_b: $Y = (X + 10) \cdot \epsilon$
lin_c: $Y = (X + 10)^{\epsilon}$

add_a: $Y = X^3 + X + \epsilon$
add_b: $Y = log(X + 10) + X^6 + \epsilon$
add_c: $Y = sin(10 \cdot X) + e^{3 \cdot X} + \epsilon$

mul_a: $Y = (X^3 + X) \cdot e^{\epsilon}$
mul_b: $Y = (sin(10 \cdot X) + e^{3 \cdot X}) \cdot e^{\epsilon}$
mul_c: $Y = (log(X + 10) + X^6) \cdot e^{\epsilon}$

com_a: $Y = (log(X + 10) + X^2)^{\epsilon}$
com_b: $Y = log(X + 10) + |X|^{2 \cdot |\epsilon|}$
com_c: $Y = log(|X|^7 + 5) + X^5 - sin(X^2 \cdot |\epsilon|)$

With:

- cause $X \sim \mathcal{N}(0, 1)$
- noise $\epsilon \sim \mathcal{N}(0, 1)$, or $\epsilon \sim \mathcal{U}(0, 1)$, or $\epsilon \sim EXP(1)$
- 100 or 1000 data points per realization
- 100 realizations each

Goal: Predict whether X causes Y or Y causes X.

Please note: Due to implementation details of the methods, we actually deal with a three class problem ($X \rightarrow Y$, $X \leftarrow Y$, and $X \perp\!\!\!\perp Y$) while predicting the third class is always wrong.

ANM Results for Synthetic (100)

Correct estimates overall: 48.361 % (1741 / 3600, bootstrap err. 0.815 %)

Correct predictions for individual setups:

lin_a + normal	54.000 % (54 / 100)
lin_a + uniform	81.000 % (81 / 100)
lin_a + exponential	90.000 % (90 / 100)
lin_b + normal	47.000 % (47 / 100)
lin_b + uniform	28.000 % (28 / 100)
lin_b + exponential	17.000 % (17 / 100)
lin_c + normal	1.000 % (1 / 100)
lin_c + uniform	8.000 % (8 / 100)
lin_c + exponential	0.000 % (0 / 100)
add_a + normal	100.000 % (100 / 100)
add_a + uniform	100.000 % (100 / 100)
add_a + exponential	98.000 % (98 / 100)
add_b + normal	100.000 % (100 / 100)
add_b + uniform	99.000 % (99 / 100)
add_b + exponential	97.000 % (97 / 100)
add_c + normal	100.000 % (100 / 100)
add_c + uniform	100.000 % (100 / 100)
add_c + exponential	100.000 % (100 / 100)
mul_a + normal	1.000 % (1 / 100)
mul_a + uniform	0.000 % (0 / 100)
mul_a + exponential	10.000 % (10 / 100)
mul_b + normal	0.000 % (0 / 100)
mul_b + uniform	4.000 % (4 / 100)
mul_b + exponential	5.000 % (5 / 100)
mul_c + normal	10.000 % (10 / 100)
mul_c + uniform	35.000 % (35 / 100)
mul_c + exponential	8.000 % (8 / 100)
com_a + normal	3.000 % (3 / 100)
com_a + uniform	25.000 % (25 / 100)
com_a + exponential	1.000 % (1 / 100)
com_b + normal	100.000 % (100 / 100)
com_b + uniform	100.000 % (100 / 100)
com_b + exponential	94.000 % (94 / 100)
com_c + normal	41.000 % (41 / 100)
com_c + uniform	37.000 % (37 / 100)
com_c + exponential	47.000 % (47 / 100)

Correct predictions averaged for function type:

lin_a	75.000 % (225 / 300)
lin_b	30.667 % (92 / 300)
lin_c	3.000 % (9 / 300)
add_a	99.333 % (298 / 300)
add_b	98.667 % (296 / 300)
add_c	100.000 % (300 / 300)
mul_a	3.667 % (11 / 300)
mul_b	3.000 % (9 / 300)
mul_c	17.667 % (53 / 300)
com_a	9.667 % (29 / 300)
com_b	98.000 % (294 / 300)
com_c	41.667 % (125 / 300)

Correct predictions averaged for noise type:

normal	46.417 % (557 / 1200)
uniform	51.417 % (617 / 1200)
exponential	47.250 % (567 / 1200)

DLR

ANM Results for Synthetic (1000)

Correct estimates overall: 48.139 % (1733 / 3600, bootstrap err. 0.836 %)

Correct predictions for individual setups:

lin_a + normal	48.000 % (48 / 100)
lin_a + uniform	99.000 % (99 / 100)
lin_a + exponential	100.000 % (100 / 100)
lin_b + normal	30.000 % (30 / 100)
lin_b + uniform	30.000 % (30 / 100)
lin_b + exponential	19.000 % (19 / 100)
lin_c + normal	0.000 % (0 / 100)
lin_c + uniform	10.000 % (10 / 100)
lin_c + exponential	0.000 % (0 / 100)
add_a + normal	100.000 % (100 / 100)
add_a + uniform	100.000 % (100 / 100)
add_a + exponential	100.000 % (100 / 100)
add_b + normal	100.000 % (100 / 100)
add_b + uniform	100.000 % (100 / 100)
add_b + exponential	100.000 % (100 / 100)
add_c + normal	100.000 % (100 / 100)
add_c + uniform	100.000 % (100 / 100)
add_c + exponential	100.000 % (100 / 100)
mul_a + normal	0.000 % (0 / 100)
mul_a + uniform	0.000 % (0 / 100)
mul_a + exponential	27.000 % (27 / 100)
mul_b + normal	8.000 % (8 / 100)
mul_b + uniform	29.000 % (29 / 100)
mul_b + exponential	22.000 % (22 / 100)
mul_c + normal	15.000 % (15 / 100)
mul_c + uniform	76.000 % (76 / 100)
mul_c + exponential	14.000 % (14 / 100)
com_a + normal	0.000 % (0 / 100)
com_a + uniform	2.000 % (2 / 100)
com_a + exponential	0.000 % (0 / 100)
com_b + normal	100.000 % (100 / 100)
com_b + uniform	100.000 % (100 / 100)
com_b + exponential	98.000 % (98 / 100)
com_c + normal	3.000 % (3 / 100)
com_c + uniform	0.000 % (0 / 100)
com_c + exponential	3.000 % (3 / 100)

Correct predictions averaged for function type:

lin_a	82.333 % (247 / 300)
lin_b	26.333 % (79 / 300)
lin_c	3.333 % (10 / 300)
add_a	100.000 % (300 / 300)
add_b	100.000 % (300 / 300)
add_c	100.000 % (300 / 300)
mul_a	9.000 % (27 / 300)
mul_b	19.667 % (59 / 300)
mul_c	35.000 % (105 / 300)
com_a	0.667 % (2 / 300)
com_b	99.333 % (298 / 300)
com_c	2.000 % (6 / 300)

Correct predictions averaged for noise type:

normal	42.000 % (504 / 1200)
uniform	53.833 % (646 / 1200)
exponential	48.583 % (583 / 1200)

CAM Results for Synthetic (100)

Correct estimates overall: 70.389 % (2534 / 3600, bootstrap err. 0.749 %)

Correct predictions for individual setups:

lin_a + normal	45.000 % (45 / 100)
lin_a + uniform	53.000 % (53 / 100)
lin_a + exponential	4.000 % (4 / 100)
lin_b + normal	30.000 % (30 / 100)
lin_b + uniform	15.000 % (15 / 100)
lin_b + exponential	52.000 % (52 / 100)
lin_c + normal	42.000 % (42 / 100)
lin_c + uniform	48.000 % (48 / 100)
lin_c + exponential	42.000 % (42 / 100)
add_a + normal	100.000 % (100 / 100)
add_a + uniform	100.000 % (100 / 100)
add_a + exponential	100.000 % (100 / 100)
add_b + normal	100.000 % (100 / 100)
add_b + uniform	100.000 % (100 / 100)
add_b + exponential	100.000 % (100 / 100)
add_c + normal	100.000 % (100 / 100)
add_c + uniform	100.000 % (100 / 100)
add_c + exponential	100.000 % (100 / 100)
mul_a + normal	6.000 % (6 / 100)
mul_a + uniform	6.000 % (6 / 100)
mul_a + exponential	1.000 % (1 / 100)
mul_b + normal	69.000 % (69 / 100)
mul_b + uniform	96.000 % (96 / 100)
mul_b + exponential	60.000 % (60 / 100)
mul_c + normal	100.000 % (100 / 100)
mul_c + uniform	100.000 % (100 / 100)
mul_c + exponential	87.000 % (87 / 100)
com_a + normal	69.000 % (69 / 100)
com_a + uniform	59.000 % (59 / 100)
com_a + exponential	64.000 % (64 / 100)
com_b + normal	97.000 % (97 / 100)
com_b + uniform	100.000 % (100 / 100)
com_b + exponential	89.000 % (89 / 100)
com_c + normal	100.000 % (100 / 100)
com_c + uniform	100.000 % (100 / 100)
com_c + exponential	100.000 % (100 / 100)

Correct predictions averaged for function type:

lin_a	34.000 % (102 / 300)
lin_b	32.333 % (97 / 300)
lin_c	44.000 % (132 / 300)
add_a	100.000 % (300 / 300)
add_b	100.000 % (300 / 300)
add_c	100.000 % (300 / 300)
mul_a	4.333 % (13 / 300)
mul_b	75.000 % (225 / 300)
mul_c	95.667 % (287 / 300)
com_a	64.000 % (192 / 300)
com_b	95.333 % (286 / 300)
com_c	100.000 % (300 / 300)

Correct predictions averaged for noise type:

normal: 71.500 % (858 / 1200)
uniform: 73.083 % (877 / 1200)
exponential: 66.583 % (799 / 1200)

CAM Results for Synthetic (1000)

Correct estimates overall: 69.306 % (2495 / 3600, bootstrap err. 0.769 %)

Correct predictions for individual setups:

lin_a + normal	53.000 % (53 / 100)
lin_a + uniform	43.000 % (43 / 100)
lin_a + exponential	0.000 % (0 / 100)
lin_b + normal	0.000 % (0 / 100)
lin_b + uniform	0.000 % (0 / 100)
lin_b + exponential	43.000 % (43 / 100)
lin_c + normal	40.000 % (40 / 100)
lin_c + uniform	6.000 % (6 / 100)
lin_c + exponential	40.000 % (40 / 100)
add_a + normal	100.000 % (100 / 100)
add_a + uniform	100.000 % (100 / 100)
add_a + exponential	100.000 % (100 / 100)
add_b + normal	100.000 % (100 / 100)
add_b + uniform	100.000 % (100 / 100)
add_b + exponential	100.000 % (100 / 100)
add_c + normal	100.000 % (100 / 100)
add_c + uniform	100.000 % (100 / 100)
add_c + exponential	100.000 % (100 / 100)
mul_a + normal	0.000 % (0 / 100)
mul_a + uniform	0.000 % (0 / 100)
mul_a + exponential	0.000 % (0 / 100)
mul_b + normal	85.000 % (85 / 100)
mul_b + uniform	100.000 % (100 / 100)
mul_b + exponential	41.000 % (41 / 100)
mul_c + normal	100.000 % (100 / 100)
mul_c + uniform	100.000 % (100 / 100)
mul_c + exponential	88.000 % (88 / 100)
com_a + normal	97.000 % (97 / 100)
com_a + uniform	95.000 % (95 / 100)
com_a + exponential	74.000 % (74 / 100)
com_b + normal	99.000 % (99 / 100)
com_b + uniform	100.000 % (100 / 100)
com_b + exponential	91.000 % (91 / 100)
com_c + normal	100.000 % (100 / 100)
com_c + uniform	100.000 % (100 / 100)
com_c + exponential	100.000 % (100 / 100)

Correct predictions averaged for function type:

lin_a	32.000 % (96 / 300)
lin_b	14.333 % (43 / 300)
lin_c	28.667 % (86 / 300)
add_a	100.000 % (300 / 300)
add_b	100.000 % (300 / 300)
add_c	100.000 % (300 / 300)
mul_a	0.000 % (0 / 300)
mul_b	75.333 % (226 / 300)
mul_c	96.000 % (288 / 300)
com_a	88.667 % (266 / 300)
com_b	96.667 % (290 / 300)
com_c	100.000 % (300 / 300)

Correct predictions averaged for noise type:

normal: 72.833 % (874 / 1200)
uniform: 70.333 % (844 / 1200)
exponential: 64.750 % (777 / 1200)

CCDr Results for Synthetic (100)

Correct estimates overall: 50.028 % (1801 / 3600, bootstrap err. 0.832 %)

Correct predictions for individual setups:

lin_a + normal	55.000 % (55 / 100)
lin_a + uniform	51.000 % (51 / 100)
lin_a + exponential	50.000 % (50 / 100)
lin_b + normal	46.000 % (46 / 100)
lin_b + uniform	50.000 % (50 / 100)
lin_b + exponential	52.000 % (52 / 100)
lin_c + normal	53.000 % (53 / 100)
lin_c + uniform	47.000 % (47 / 100)
lin_c + exponential	47.000 % (47 / 100)
add_a + normal	57.000 % (57 / 100)
add_a + uniform	56.000 % (56 / 100)
add_a + exponential	50.000 % (50 / 100)
add_b + normal	43.000 % (43 / 100)
add_b + uniform	48.000 % (48 / 100)
add_b + exponential	40.000 % (40 / 100)
add_c + normal	54.000 % (54 / 100)
add_c + uniform	48.000 % (48 / 100)
add_c + exponential	51.000 % (51 / 100)
mul_a + normal	50.000 % (50 / 100)
mul_a + uniform	53.000 % (53 / 100)
mul_a + exponential	46.000 % (46 / 100)
mul_b + normal	60.000 % (60 / 100)
mul_b + uniform	50.000 % (50 / 100)
mul_b + exponential	60.000 % (60 / 100)
mul_c + normal	50.000 % (50 / 100)
mul_c + uniform	43.000 % (43 / 100)
mul_c + exponential	49.000 % (49 / 100)
com_a + normal	52.000 % (52 / 100)
com_a + uniform	48.000 % (48 / 100)
com_a + exponential	50.000 % (50 / 100)
com_b + normal	50.000 % (50 / 100)
com_b + uniform	42.000 % (42 / 100)
com_b + exponential	52.000 % (52 / 100)
com_c + normal	50.000 % (50 / 100)
com_c + uniform	47.000 % (47 / 100)
com_c + exponential	51.000 % (51 / 100)

Correct predictions averaged for function type:

lin_a	52.000 % (156 / 300)
lin_b	49.333 % (148 / 300)
lin_c	49.000 % (147 / 300)
add_a	54.333 % (163 / 300)
add_b	43.667 % (131 / 300)
add_c	51.000 % (153 / 300)
mul_a	49.667 % (149 / 300)
mul_b	56.667 % (170 / 300)
mul_c	47.333 % (142 / 300)
com_a	50.000 % (150 / 300)
com_b	48.000 % (144 / 300)
com_c	49.333 % (148 / 300)

Correct predictions averaged for noise type:

normal: 51.667 % (620 / 1200)
uniform: 48.583 % (583 / 1200)
exponential: 49.833 % (598 / 1200)

CCDr Results for Synthetic (1000)

Correct estimates overall: 47.028 % (1693 / 3600, bootstrap err. 0.806 %)

Correct predictions for individual setups:

lin_a + normal	53.000 % (53 / 100)
lin_a + uniform	59.000 % (59 / 100)
lin_a + exponential	51.000 % (51 / 100)
lin_b + normal	36.000 % (36 / 100)
lin_b + uniform	44.000 % (44 / 100)
lin_b + exponential	47.000 % (47 / 100)
lin_c + normal	32.000 % (32 / 100)
lin_c + uniform	50.000 % (50 / 100)
lin_c + exponential	39.000 % (39 / 100)
add_a + normal	55.000 % (55 / 100)
add_a + uniform	49.000 % (49 / 100)
add_a + exponential	52.000 % (52 / 100)
add_b + normal	50.000 % (50 / 100)
add_b + uniform	44.000 % (44 / 100)
add_b + exponential	38.000 % (38 / 100)
add_c + normal	47.000 % (47 / 100)
add_c + uniform	47.000 % (47 / 100)
add_c + exponential	58.000 % (58 / 100)
mul_a + normal	48.000 % (48 / 100)
mul_a + uniform	51.000 % (51 / 100)
mul_a + exponential	46.000 % (46 / 100)
mul_b + normal	50.000 % (50 / 100)
mul_b + uniform	45.000 % (45 / 100)
mul_b + exponential	45.000 % (45 / 100)
mul_c + normal	55.000 % (55 / 100)
mul_c + uniform	44.000 % (44 / 100)
mul_c + exponential	42.000 % (42 / 100)
com_a + normal	39.000 % (39 / 100)
com_a + uniform	41.000 % (41 / 100)
com_a + exponential	51.000 % (51 / 100)
com_b + normal	35.000 % (35 / 100)
com_b + uniform	47.000 % (47 / 100)
com_b + exponential	58.000 % (58 / 100)
com_c + normal	54.000 % (54 / 100)
com_c + uniform	44.000 % (44 / 100)
com_c + exponential	47.000 % (47 / 100)

Correct predictions averaged for function type:

lin_a	54.333 % (163 / 300)
lin_b	42.333 % (127 / 300)
lin_c	40.333 % (121 / 300)
add_a	52.000 % (156 / 300)
add_b	44.000 % (132 / 300)
add_c	50.667 % (152 / 300)
mul_a	48.333 % (145 / 300)
mul_b	46.667 % (140 / 300)
mul_c	47.000 % (141 / 300)
com_a	43.667 % (131 / 300)
com_b	46.667 % (140 / 300)
com_c	48.333 % (145 / 300)

Correct predictions averaged for noise type:

normal: 46.167 % (554 / 1200)
uniform: 47.083 % (565 / 1200)
exponential: 47.833 % (574 / 1200)

CDS Results for Synthetic (100)

Correct estimates overall: 53.472 % (1925 / 3600, bootstrap err. 0.832 %)

Correct predictions for individual setups:

lin_a + normal	53.000 % (53 / 100)
lin_a + uniform	67.000 % (67 / 100)
lin_a + exponential	55.000 % (55 / 100)
lin_b + normal	46.000 % (46 / 100)
lin_b + uniform	51.000 % (51 / 100)
lin_b + exponential	22.000 % (22 / 100)
lin_c + normal	7.000 % (7 / 100)
lin_c + uniform	6.000 % (6 / 100)
lin_c + exponential	3.000 % (3 / 100)
add_a + normal	63.000 % (63 / 100)
add_a + uniform	78.000 % (78 / 100)
add_a + exponential	50.000 % (50 / 100)
add_b + normal	85.000 % (85 / 100)
add_b + uniform	76.000 % (76 / 100)
add_b + exponential	86.000 % (86 / 100)
add_c + normal	68.000 % (68 / 100)
add_c + uniform	68.000 % (68 / 100)
add_c + exponential	81.000 % (81 / 100)
mul_a + normal	33.000 % (33 / 100)
mul_a + uniform	43.000 % (43 / 100)
mul_a + exponential	37.000 % (37 / 100)
mul_b + normal	62.000 % (62 / 100)
mul_b + uniform	74.000 % (74 / 100)
mul_b + exponential	55.000 % (55 / 100)
mul_c + normal	91.000 % (91 / 100)
mul_c + uniform	90.000 % (90 / 100)
mul_c + exponential	65.000 % (65 / 100)
com_a + normal	29.000 % (29 / 100)
com_a + uniform	28.000 % (28 / 100)
com_a + exponential	41.000 % (41 / 100)
com_b + normal	98.000 % (98 / 100)
com_b + uniform	97.000 % (97 / 100)
com_b + exponential	90.000 % (90 / 100)
com_c + normal	7.000 % (7 / 100)
com_c + uniform	6.000 % (6 / 100)
com_c + exponential	14.000 % (14 / 100)

Correct predictions averaged for function type:

lin_a	58.333 % (175 / 300)
lin_b	39.667 % (119 / 300)
lin_c	5.333 % (16 / 300)
add_a	63.667 % (191 / 300)
add_b	82.333 % (247 / 300)
add_c	72.333 % (217 / 300)
mul_a	37.667 % (113 / 300)
mul_b	63.667 % (191 / 300)
mul_c	82.000 % (246 / 300)
com_a	32.667 % (98 / 300)
com_b	95.000 % (285 / 300)
com_c	9.000 % (27 / 300)

Correct predictions averaged for noise type:

normal: 53.500 % (642 / 1200)
uniform: 57.000 % (684 / 1200)
exponential: 49.917 % (599 / 1200)

DLR

CDS Results for Synthetic (1000)

Correct estimates overall: 51.139 % (1841 / 3600, bootstrap err. 0.817 %)

Correct predictions for individual setups:

lin_a + normal	47.000 % (47 / 100)
lin_a + uniform	99.000 % (99 / 100)
lin_a + exponential	99.000 % (99 / 100)
lin_b + normal	43.000 % (43 / 100)
lin_b + uniform	25.000 % (25 / 100)
lin_b + exponential	6.000 % (6 / 100)
lin_c + normal	12.000 % (12 / 100)
lin_c + uniform	1.000 % (1 / 100)
lin_c + exponential	10.000 % (10 / 100)
add_a + normal	100.000 % (100 / 100)
add_a + uniform	60.000 % (60 / 100)
add_a + exponential	95.000 % (95 / 100)
add_b + normal	100.000 % (100 / 100)
add_b + uniform	100.000 % (100 / 100)
add_b + exponential	98.000 % (98 / 100)
add_c + normal	91.000 % (91 / 100)
add_c + uniform	97.000 % (97 / 100)
add_c + exponential	91.000 % (91 / 100)
mul_a + normal	0.000 % (0 / 100)
mul_a + uniform	0.000 % (0 / 100)
mul_a + exponential	5.000 % (5 / 100)
mul_b + normal	31.000 % (31 / 100)
mul_b + uniform	58.000 % (58 / 100)
mul_b + exponential	21.000 % (21 / 100)
mul_c + normal	33.000 % (33 / 100)
mul_c + uniform	77.000 % (77 / 100)
mul_c + exponential	29.000 % (29 / 100)
com_a + normal	3.000 % (3 / 100)
com_a + uniform	69.000 % (69 / 100)
com_a + exponential	9.000 % (9 / 100)
com_b + normal	34.000 % (34 / 100)
com_b + uniform	100.000 % (100 / 100)
com_b + exponential	14.000 % (14 / 100)
com_c + normal	64.000 % (64 / 100)
com_c + uniform	58.000 % (58 / 100)
com_c + exponential	62.000 % (62 / 100)

Correct predictions averaged for function type:

lin_a	81.667 % (245 / 300)
lin_b	24.667 % (74 / 300)
lin_c	7.667 % (23 / 300)
add_a	85.000 % (255 / 300)
add_b	99.333 % (298 / 300)
add_c	93.000 % (279 / 300)
mul_a	1.667 % (5 / 300)
mul_b	36.667 % (110 / 300)
mul_c	46.333 % (139 / 300)
com_a	27.000 % (81 / 300)
com_b	49.333 % (148 / 300)
com_c	61.333 % (184 / 300)

Correct predictions averaged for noise type:

normal: 46.500 % (558 / 1200)
uniform: 62.000 % (744 / 1200)
exponential: 44.917 % (539 / 1200)

IGCI Results for Synthetic (100)

Correct estimates overall: 93.389 % (3362 / 3600, bootstrap err. 0.415 %)

Correct predictions for individual setups:

lin_a + normal	42.000 % (42 / 100)
lin_a + uniform	54.000 % (54 / 100)
lin_a + exponential	61.000 % (61 / 100)
lin_b + normal	40.000 % (40 / 100)
lin_b + uniform	79.000 % (79 / 100)
lin_b + exponential	100.000 % (100 / 100)
lin_c + normal	100.000 % (100 / 100)
lin_c + uniform	100.000 % (100 / 100)
lin_c + exponential	100.000 % (100 / 100)
add_a + normal	96.000 % (96 / 100)
add_a + uniform	98.000 % (98 / 100)
add_a + exponential	99.000 % (99 / 100)
add_b + normal	100.000 % (100 / 100)
add_b + uniform	100.000 % (100 / 100)
add_b + exponential	100.000 % (100 / 100)
add_c + normal	100.000 % (100 / 100)
add_c + uniform	100.000 % (100 / 100)
add_c + exponential	100.000 % (100 / 100)
mul_a + normal	100.000 % (100 / 100)
mul_a + uniform	99.000 % (99 / 100)
mul_a + exponential	100.000 % (100 / 100)
mul_b + normal	100.000 % (100 / 100)
mul_b + uniform	100.000 % (100 / 100)
mul_b + exponential	100.000 % (100 / 100)
mul_c + normal	100.000 % (100 / 100)
mul_c + uniform	100.000 % (100 / 100)
mul_c + exponential	100.000 % (100 / 100)
com_a + normal	100.000 % (100 / 100)
com_a + uniform	99.000 % (99 / 100)
com_a + exponential	100.000 % (100 / 100)
com_b + normal	100.000 % (100 / 100)
com_b + uniform	95.000 % (95 / 100)
com_b + exponential	100.000 % (100 / 100)
com_c + normal	100.000 % (100 / 100)
com_c + uniform	100.000 % (100 / 100)
com_c + exponential	100.000 % (100 / 100)

Correct predictions averaged for function type:

lin_a	52.333 % (157 / 300)
lin_b	73.000 % (219 / 300)
lin_c	100.000 % (300 / 300)
add_a	97.667 % (293 / 300)
add_b	100.000 % (300 / 300)
add_c	100.000 % (300 / 300)
mul_a	99.667 % (299 / 300)
mul_b	100.000 % (300 / 300)
mul_c	100.000 % (300 / 300)
com_a	99.667 % (299 / 300)
com_b	98.333 % (295 / 300)
com_c	100.000 % (300 / 300)

Correct predictions averaged for noise type:

normal: 89.833 % (1078 / 1200)
uniform: 93.667 % (1124 / 1200)
exponential: 96.667 % (1160 / 1200)

DLR

IGCI Results for Synthetic (1000)

Correct estimates overall: 95.306 % (3431 / 3600, bootstrap err. 0.352 %)

Correct predictions for individual setups:

lin_a + normal	61.000 % (61 / 100)
lin_a + uniform	50.000 % (50 / 100)
lin_a + exponential	74.000 % (74 / 100)
lin_b + normal	46.000 % (46 / 100)
lin_b + uniform	100.000 % (100 / 100)
lin_b + exponential	100.000 % (100 / 100)
lin_c + normal	100.000 % (100 / 100)
lin_c + uniform	100.000 % (100 / 100)
lin_c + exponential	100.000 % (100 / 100)
add_a + normal	100.000 % (100 / 100)
add_a + uniform	100.000 % (100 / 100)
add_a + exponential	100.000 % (100 / 100)
add_b + normal	100.000 % (100 / 100)
add_b + uniform	100.000 % (100 / 100)
add_b + exponential	100.000 % (100 / 100)
add_c + normal	100.000 % (100 / 100)
add_c + uniform	100.000 % (100 / 100)
add_c + exponential	100.000 % (100 / 100)
mul_a + normal	100.000 % (100 / 100)
mul_a + uniform	100.000 % (100 / 100)
mul_a + exponential	100.000 % (100 / 100)
mul_b + normal	100.000 % (100 / 100)
mul_b + uniform	100.000 % (100 / 100)
mul_b + exponential	100.000 % (100 / 100)
mul_c + normal	100.000 % (100 / 100)
mul_c + uniform	100.000 % (100 / 100)
mul_c + exponential	100.000 % (100 / 100)
com_a + normal	100.000 % (100 / 100)
com_a + uniform	100.000 % (100 / 100)
com_a + exponential	100.000 % (100 / 100)
com_b + normal	100.000 % (100 / 100)
com_b + uniform	100.000 % (100 / 100)
com_b + exponential	100.000 % (100 / 100)
com_c + normal	100.000 % (100 / 100)
com_c + uniform	100.000 % (100 / 100)
com_c + exponential	100.000 % (100 / 100)

Correct predictions averaged for function type:

lin_a	61.667 % (185 / 300)
lin_b	82.000 % (246 / 300)
lin_c	100.000 % (300 / 300)
add_a	100.000 % (300 / 300)
add_b	100.000 % (300 / 300)
add_c	100.000 % (300 / 300)
mul_a	100.000 % (300 / 300)
mul_b	100.000 % (300 / 300)
mul_c	100.000 % (300 / 300)
com_a	100.000 % (300 / 300)
com_b	100.000 % (300 / 300)
com_c	100.000 % (300 / 300)

Correct predictions averaged for noise type:

normal: 92.250 % (1107 / 1200)
uniform: 95.833 % (1150 / 1200)
exponential: 97.833 % (1174 / 1200)

KCDC Results for Synthetic (100)

Correct estimates overall: overall: 9.278 % (334 / 3600, bootstrap err. 0.478 %)

Correct predictions for individual setups:

lin_a + normal	0.000 % (0 / 100)
lin_a + uniform	1.000 % (1 / 100)
lin_a + exponential	0.000 % (0 / 100)
lin_b + normal	27.000 % (27 / 100)
lin_b + uniform	0.000 % (0 / 100)
lin_b + exponential	1.000 % (1 / 100)
lin_c + normal	0.000 % (0 / 100)
lin_c + uniform	0.000 % (0 / 100)
lin_c + exponential	0.000 % (0 / 100)
add_a + normal	0.000 % (0 / 100)
add_a + uniform	0.000 % (0 / 100)
add_a + exponential	0.000 % (0 / 100)
add_b + normal	4.000 % (4 / 100)
add_b + uniform	1.000 % (1 / 100)
add_b + exponential	3.000 % (3 / 100)
add_c + normal	0.000 % (0 / 100)
add_c + uniform	0.000 % (0 / 100)
add_c + exponential	0.000 % (0 / 100)
mul_a + normal	0.000 % (0 / 100)
mul_a + uniform	0.000 % (0 / 100)
mul_a + exponential	35.000 % (35 / 100)
mul_b + normal	0.000 % (0 / 100)
mul_b + uniform	0.000 % (0 / 100)
mul_b + exponential	0.000 % (0 / 100)
mul_c + normal	6.000 % (6 / 100)
mul_c + uniform	4.000 % (4 / 100)
mul_c + exponential	28.000 % (28 / 100)
com_a + normal	0.000 % (0 / 100)
com_a + uniform	95.000 % (95 / 100)
com_a + exponential	0.000 % (0 / 100)
com_b + normal	26.000 % (26 / 100)
com_b + uniform	100.000 % (100 / 100)
com_b + exponential	3.000 % (3 / 100)
com_c + normal	0.000 % (0 / 100)
com_c + uniform	0.000 % (0 / 100)
com_c + exponential	0.000 % (0 / 100)

Correct predictions averaged for function type:

lin_a	0.333 % (1 / 300)
lin_b	9.333 % (28 / 300)
lin_c	0.000 % (0 / 300)
add_a	0.000 % (0 / 300)
add_b	2.667 % (8 / 300)
add_c	0.000 % (0 / 300)
mul_a	11.667 % (35 / 300)
mul_b	0.000 % (0 / 300)
mul_c	12.667 % (38 / 300)
com_a	31.667 % (95 / 300)
com_b	43.000 % (129 / 300)
com_c	0.000 % (0 / 300)

Correct predictions averaged for noise type:

normal: 5.250 % (63 / 1200)
uniform: 16.750 % (201 / 1200)
exponential: 5.833 % (70 / 1200)

Note: Strongly affected by hyperparameter!

KCDC Results for Synthetic (1000)

Correct estimates overall: overall: 9.917 % (357 / 3600, bootstrap err. 0.491 %)

Correct predictions for individual setups:

lin_a + normal	0.000 % (0 / 100)
lin_a + uniform	0.000 % (0 / 100)
lin_a + exponential	0.000 % (0 / 100)
lin_b + normal	52.000 % (52 / 100)
lin_b + uniform	0.000 % (0 / 100)
lin_b + exponential	0.000 % (0 / 100)
lin_c + normal	0.000 % (0 / 100)
lin_c + uniform	0.000 % (0 / 100)
lin_c + exponential	0.000 % (0 / 100)
add_a + normal	0.000 % (0 / 100)
add_a + uniform	0.000 % (0 / 100)
add_a + exponential	0.000 % (0 / 100)
add_b + normal	0.000 % (0 / 100)
add_b + uniform	0.000 % (0 / 100)
add_b + exponential	0.000 % (0 / 100)
add_c + normal	0.000 % (0 / 100)
add_c + uniform	0.000 % (0 / 100)
add_c + exponential	0.000 % (0 / 100)
mul_a + normal	0.000 % (0 / 100)
mul_a + uniform	0.000 % (0 / 100)
mul_a + exponential	66.000 % (66 / 100)
mul_b + normal	0.000 % (0 / 100)
mul_b + uniform	0.000 % (0 / 100)
mul_b + exponential	0.000 % (0 / 100)
mul_c + normal	0.000 % (0 / 100)
mul_c + uniform	0.000 % (0 / 100)
mul_c + exponential	38.000 % (38 / 100)
com_a + normal	0.000 % (0 / 100)
com_a + uniform	100.000 % (100 / 100)
com_a + exponential	0.000 % (0 / 100)
com_b + normal	1.000 % (1 / 100)
com_b + uniform	100.000 % (100 / 100)
com_b + exponential	0.000 % (0 / 100)
com_c + normal	0.000 % (0 / 100)
com_c + uniform	0.000 % (0 / 100)
com_c + exponential	0.000 % (0 / 100)

Correct predictions averaged for function type:

lin_a	0.000 % (0 / 300)
lin_b	17.333 % (52 / 300)
lin_c	0.000 % (0 / 300)
add_a	0.000 % (0 / 300)
add_b	0.000 % (0 / 300)
add_c	0.000 % (0 / 300)
mul_a	22.000 % (66 / 300)
mul_b	0.000 % (0 / 300)
mul_c	12.667 % (38 / 300)
com_a	33.333 % (100 / 300)
com_b	33.667 % (101 / 300)
com_c	0.000 % (0 / 300)

Correct predictions averaged for noise type:

normal: 4.417 % (53 / 1200)
uniform: 16.667 % (200 / 1200)
exponential: 8.667 % (104 / 1200)

Note: Strongly affected by hyperparameter!

LiNGAM Results for Synthetic (100)

Correct estimates overall: overall: 1.694 % (61 / 3600, bootstrap err. 0.215 %)

Correct predictions for individual setups:

lin_a + normal	34.000 % (34 / 100)
lin_a + uniform	7.000 % (7 / 100)
lin_a + exponential	10.000 % (10 / 100)
lin_b + normal	0.000 % (0 / 100)
lin_b + uniform	0.000 % (0 / 100)
lin_b + exponential	0.000 % (0 / 100)
lin_c + normal	0.000 % (0 / 100)
lin_c + uniform	0.000 % (0 / 100)
lin_c + exponential	0.000 % (0 / 100)
add_a + normal	0.000 % (0 / 100)
add_a + uniform	0.000 % (0 / 100)
add_a + exponential	0.000 % (0 / 100)
add_b + normal	0.000 % (0 / 100)
add_b + uniform	0.000 % (0 / 100)
add_b + exponential	0.000 % (0 / 100)
add_c + normal	0.000 % (0 / 100)
add_c + uniform	0.000 % (0 / 100)
add_c + exponential	0.000 % (0 / 100)
mul_a + normal	0.000 % (0 / 100)
mul_a + uniform	0.000 % (0 / 100)
mul_a + exponential	0.000 % (0 / 100)
mul_b + normal	0.000 % (0 / 100)
mul_b + uniform	0.000 % (0 / 100)
mul_b + exponential	0.000 % (0 / 100)
mul_c + normal	0.000 % (0 / 100)
mul_c + uniform	0.000 % (0 / 100)
mul_c + exponential	0.000 % (0 / 100)
com_a + normal	0.000 % (0 / 100)
com_a + uniform	6.000 % (6 / 100)
com_a + exponential	0.000 % (0 / 100)
com_b + normal	1.000 % (1 / 100)
com_b + uniform	3.000 % (3 / 100)
com_b + exponential	0.000 % (0 / 100)
com_c + normal	0.000 % (0 / 100)
com_c + uniform	0.000 % (0 / 100)
com_c + exponential	0.000 % (0 / 100)

Correct predictions averaged for function type:

lin_a	17.000 % (51 / 300)
lin_b	0.000 % (0 / 300)
lin_c	0.000 % (0 / 300)
add_a	0.000 % (0 / 300)
add_b	0.000 % (0 / 300)
add_c	0.000 % (0 / 300)
mul_a	0.000 % (0 / 300)
mul_b	0.000 % (0 / 300)
mul_c	0.000 % (0 / 300)
com_a	2.000 % (6 / 300)
com_b	1.333 % (4 / 300)
com_c	0.000 % (0 / 300)

Correct predictions averaged for noise type:

normal: 2.917 % (35 / 1200)
uniform: 1.333 % (16 / 1200)
exponential: 0.833 % (10 / 1200)

Note: No data normalization!

LiNGAM Results for Synthetic (1000)

Correct estimates overall: overall: 1.667 % (60 / 3600, bootstrap err. 0.213 %)

Correct predictions for individual setups:

lin_a + normal	33.000 % (33 / 100)
lin_a + uniform	0.000 % (0 / 100)
lin_a + exponential	0.000 % (0 / 100)
lin_b + normal	0.000 % (0 / 100)
lin_b + uniform	0.000 % (0 / 100)
lin_b + exponential	0.000 % (0 / 100)
lin_c + normal	0.000 % (0 / 100)
lin_c + uniform	0.000 % (0 / 100)
lin_c + exponential	0.000 % (0 / 100)
add_a + normal	0.000 % (0 / 100)
add_a + uniform	0.000 % (0 / 100)
add_a + exponential	0.000 % (0 / 100)
add_b + normal	0.000 % (0 / 100)
add_b + uniform	0.000 % (0 / 100)
add_b + exponential	0.000 % (0 / 100)
add_c + normal	0.000 % (0 / 100)
add_c + uniform	0.000 % (0 / 100)
add_c + exponential	0.000 % (0 / 100)
mul_a + normal	0.000 % (0 / 100)
mul_a + uniform	0.000 % (0 / 100)
mul_a + exponential	0.000 % (0 / 100)
mul_b + normal	0.000 % (0 / 100)
mul_b + uniform	0.000 % (0 / 100)
mul_b + exponential	0.000 % (0 / 100)
mul_c + normal	0.000 % (0 / 100)
mul_c + uniform	0.000 % (0 / 100)
mul_c + exponential	0.000 % (0 / 100)
com_a + normal	0.000 % (0 / 100)
com_a + uniform	4.000 % (4 / 100)
com_a + exponential	0.000 % (0 / 100)
com_b + normal	0.000 % (0 / 100)
com_b + uniform	23.000 % (23 / 100)
com_b + exponential	0.000 % (0 / 100)
com_c + normal	0.000 % (0 / 100)
com_c + uniform	0.000 % (0 / 100)
com_c + exponential	0.000 % (0 / 100)

Correct predictions averaged for function type:

lin_a	11.000 % (33 / 300)
lin_b	0.000 % (0 / 300)
lin_c	0.000 % (0 / 300)
add_a	0.000 % (0 / 300)
add_b	0.000 % (0 / 300)
add_c	0.000 % (0 / 300)
mul_a	0.000 % (0 / 300)
mul_b	0.000 % (0 / 300)
mul_c	0.000 % (0 / 300)
com_a	1.333 % (4 / 300)
com_b	7.667 % (23 / 300)
com_c	0.000 % (0 / 300)

Correct predictions averaged for noise type:

normal: 2.750 % (33 / 1200)
uniform: 2.250 % (27 / 1200)
exponential: 0.000 % (0 / 1200)

Note: No data normalization!

noTEARS Results for Synthetic (100)

Correct estimates overall: overall: 86.278 % (3106 / 3600, bootstrap err. 0.578 %)

Correct predictions for individual setups:

lin_a + normal	88.000 % (88 / 100)
lin_a + uniform	94.000 % (94 / 100)
lin_a + exponential	87.000 % (87 / 100)
lin_b + normal	78.000 % (78 / 100)
lin_b + uniform	58.000 % (58 / 100)
lin_b + exponential	86.000 % (86 / 100)
lin_c + normal	92.000 % (92 / 100)
lin_c + uniform	40.000 % (40 / 100)
lin_c + exponential	27.000 % (27 / 100)
add_a + normal	100.000 % (100 / 100)
add_a + uniform	100.000 % (100 / 100)
add_a + exponential	100.000 % (100 / 100)
add_b + normal	97.000 % (97 / 100)
add_b + uniform	98.000 % (98 / 100)
add_b + exponential	97.000 % (97 / 100)
add_c + normal	99.000 % (99 / 100)
add_c + uniform	100.000 % (100 / 100)
add_c + exponential	100.000 % (100 / 100)
mul_a + normal	100.000 % (100 / 100)
mul_a + uniform	100.000 % (100 / 100)
mul_a + exponential	100.000 % (100 / 100)
mul_b + normal	100.000 % (100 / 100)
mul_b + uniform	100.000 % (100 / 100)
mul_b + exponential	94.000 % (94 / 100)
mul_c + normal	99.000 % (99 / 100)
mul_c + uniform	99.000 % (99 / 100)
mul_c + exponential	93.000 % (93 / 100)
com_a + normal	64.000 % (64 / 100)
com_a + uniform	12.000 % (12 / 100)
com_a + exponential	97.000 % (97 / 100)
com_b + normal	73.000 % (73 / 100)
com_b + uniform	50.000 % (50 / 100)
com_b + exponential	84.000 % (84 / 100)
com_c + normal	100.000 % (100 / 100)
com_c + uniform	100.000 % (100 / 100)
com_c + exponential	100.000 % (100 / 100)

Correct predictions averaged for function type:

lin_a	89.667 % (269 / 300)
lin_b	74.000 % (222 / 300)
lin_c	53.000 % (159 / 300)
add_a	100.000 % (300 / 300)
add_b	97.333 % (292 / 300)
add_c	99.667 % (299 / 300)
mul_a	100.000 % (300 / 300)
mul_b	98.000 % (294 / 300)
mul_c	97.000 % (291 / 300)
com_a	57.667 % (173 / 300)
com_b	69.000 % (207 / 300)
com_c	100.000 % (300 / 300)

Correct predictions averaged for noise type:

normal: 90.833 % (1090 / 1200)
uniform: 79.250 % (951 / 1200)
exponential: 88.750 % (1065 / 1200)

noTEARS Results for Synthetic (1000)

Correct estimates overall: overall: 77.139 % (2777 / 3600, bootstrap err. 0.692 %)

Correct predictions for individual setups:

lin_a + normal	99.000 % (99 / 100)
lin_a + uniform	99.000 % (99 / 100)
lin_a + exponential	99.000 % (99 / 100)
lin_b + normal	39.000 % (39 / 100)
lin_b + uniform	9.000 % (9 / 100)
lin_b + exponential	95.000 % (95 / 100)
lin_c + normal	91.000 % (91 / 100)
lin_c + uniform	7.000 % (7 / 100)
lin_c + exponential	0.000 % (0 / 100)
add_a + normal	100.000 % (100 / 100)
add_a + uniform	100.000 % (100 / 100)
add_a + exponential	100.000 % (100 / 100)
add_b + normal	97.000 % (97 / 100)
add_b + uniform	92.000 % (92 / 100)
add_b + exponential	95.000 % (95 / 100)
add_c + normal	98.000 % (98 / 100)
add_c + uniform	99.000 % (99 / 100)
add_c + exponential	98.000 % (98 / 100)
mul_a + normal	100.000 % (100 / 100)
mul_a + uniform	100.000 % (100 / 100)
mul_a + exponential	98.000 % (98 / 100)
mul_b + normal	92.000 % (92 / 100)
mul_b + uniform	97.000 % (97 / 100)
mul_b + exponential	65.000 % (65 / 100)
mul_c + normal	97.000 % (97 / 100)
mul_c + uniform	94.000 % (94 / 100)
mul_c + exponential	70.000 % (70 / 100)
com_a + normal	44.000 % (44 / 100)
com_a + uniform	0.000 % (0 / 100)
com_a + exponential	67.000 % (67 / 100)
com_b + normal	45.000 % (45 / 100)
com_b + uniform	6.000 % (6 / 100)
com_b + exponential	85.000 % (85 / 100)
com_c + normal	100.000 % (100 / 100)
com_c + uniform	100.000 % (100 / 100)
com_c + exponential	100.000 % (100 / 100)

Correct predictions averaged for function type:

lin_a	99.000 % (297 / 300)
lin_b	47.667 % (143 / 300)
lin_c	32.667 % (98 / 300)
add_a	100.000 % (300 / 300)
add_b	94.667 % (284 / 300)
add_c	98.333 % (295 / 300)
mul_a	99.333 % (298 / 300)
mul_b	84.667 % (254 / 300)
mul_c	87.000 % (261 / 300)
com_a	37.000 % (111 / 300)
com_b	45.333 % (136 / 300)
com_c	100.000 % (300 / 300)

Correct predictions averaged for noise type:

normal: 83.500 % (1002 / 1200)
uniform: 66.917 % (803 / 1200)
exponential: 81.000 % (972 / 1200)

RCC Results for Synthetic (100)

Correct estimates overall: overall: 47.417 % (1707 / 3600, bootstrap err. 0.826 %)

Correct predictions for individual setups:

lin_a + normal	33.000 % (33 / 100)
lin_a + uniform	35.000 % (35 / 100)
lin_a + exponential	25.000 % (25 / 100)
lin_b + normal	29.000 % (29 / 100)
lin_b + uniform	29.000 % (29 / 100)
lin_b + exponential	43.000 % (43 / 100)
lin_c + normal	91.000 % (91 / 100)
lin_c + uniform	100.000 % (100 / 100)
lin_c + exponential	6.000 % (6 / 100)
add_a + normal	42.000 % (42 / 100)
add_a + uniform	44.000 % (44 / 100)
add_a + exponential	57.000 % (57 / 100)
add_b + normal	35.000 % (35 / 100)
add_b + uniform	63.000 % (63 / 100)
add_b + exponential	61.000 % (61 / 100)
add_c + normal	29.000 % (29 / 100)
add_c + uniform	39.000 % (39 / 100)
add_c + exponential	36.000 % (36 / 100)
mul_a + normal	74.000 % (74 / 100)
mul_a + uniform	86.000 % (86 / 100)
mul_a + exponential	48.000 % (48 / 100)
mul_b + normal	69.000 % (69 / 100)
mul_b + uniform	48.000 % (48 / 100)
mul_b + exponential	15.000 % (15 / 100)
mul_c + normal	36.000 % (36 / 100)
mul_c + uniform	10.000 % (10 / 100)
mul_c + exponential	23.000 % (23 / 100)
com_a + normal	89.000 % (89 / 100)
com_a + uniform	90.000 % (90 / 100)
com_a + exponential	4.000 % (4 / 100)
com_b + normal	59.000 % (59 / 100)
com_b + uniform	55.000 % (55 / 100)
com_b + exponential	56.000 % (56 / 100)
com_c + normal	48.000 % (48 / 100)
com_c + uniform	56.000 % (56 / 100)
com_c + exponential	44.000 % (44 / 100)

Correct predictions averaged for function type:

lin_a	31.000 % (93 / 300)
lin_b	33.667 % (101 / 300)
lin_c	65.667 % (197 / 300)
add_a	47.667 % (143 / 300)
add_b	53.000 % (159 / 300)
add_c	34.667 % (104 / 300)
mul_a	69.333 % (208 / 300)
mul_b	44.000 % (132 / 300)
mul_c	23.000 % (69 / 300)
com_a	61.000 % (183 / 300)
com_b	56.667 % (170 / 300)
com_c	49.333 % (148 / 300)

Correct predictions averaged for noise type:

normal: 52.833 % (634 / 1200)
uniform: 54.583 % (655 / 1200)
exponential: 34.833 % (418 / 1200)

RCC Results for Synthetic (1000)

Correct estimates overall: overall: 46.361 % (1669 / 3600, bootstrap err. 0.841 %)

Correct predictions for individual setups:

lin_a + normal	39.000 % (39 / 100)
lin_a + uniform	29.000 % (29 / 100)
lin_a + exponential	6.000 % (6 / 100)
lin_b + normal	0.000 % (0 / 100)
lin_b + uniform	0.000 % (0 / 100)
lin_b + exponential	1.000 % (1 / 100)
lin_c + normal	50.000 % (50 / 100)
lin_c + uniform	92.000 % (92 / 100)
lin_c + exponential	0.000 % (0 / 100)
add_a + normal	98.000 % (98 / 100)
add_a + uniform	100.000 % (100 / 100)
add_a + exponential	100.000 % (100 / 100)
add_b + normal	88.000 % (88 / 100)
add_b + uniform	46.000 % (46 / 100)
add_b + exponential	0.000 % (0 / 100)
add_c + normal	60.000 % (60 / 100)
add_c + uniform	67.000 % (67 / 100)
add_c + exponential	63.000 % (63 / 100)
mul_a + normal	84.000 % (84 / 100)
mul_a + uniform	100.000 % (100 / 100)
mul_a + exponential	100.000 % (100 / 100)
mul_b + normal	96.000 % (96 / 100)
mul_b + uniform	72.000 % (72 / 100)
mul_b + exponential	0.000 % (0 / 100)
mul_c + normal	45.000 % (45 / 100)
mul_c + uniform	0.000 % (0 / 100)
mul_c + exponential	0.000 % (0 / 100)
com_a + normal	68.000 % (68 / 100)
com_a + uniform	4.000 % (4 / 100)
com_a + exponential	0.000 % (0 / 100)
com_b + normal	65.000 % (65 / 100)
com_b + uniform	79.000 % (79 / 100)
com_b + exponential	50.000 % (50 / 100)
com_c + normal	12.000 % (12 / 100)
com_c + uniform	35.000 % (35 / 100)
com_c + exponential	20.000 % (20 / 100)

Correct predictions averaged for function type:

lin_a	24.667 % (74 / 300)
lin_b	0.333 % (1 / 300)
lin_c	47.333 % (142 / 300)
add_a	99.333 % (298 / 300)
add_b	44.667 % (134 / 300)
add_c	63.333 % (190 / 300)
mul_a	94.667 % (284 / 300)
mul_b	56.000 % (168 / 300)
mul_c	15.000 % (45 / 300)
com_a	24.000 % (72 / 300)
com_b	64.667 % (194 / 300)
com_c	22.333 % (67 / 300)

Correct predictions averaged for noise type:

normal: 58.750 % (705 / 1200)
uniform: 52.000 % (624 / 1200)
exponential: 28.333 % (340 / 1200)

RECI Results for Synthetic (100)

Correct estimates overall: 86.944 % (3130 / 3600, bootstrap err. 0.561 %)

Correct predictions for individual setups:

lin_a + normal	52.000 % (52 / 100)
lin_a + uniform	41.000 % (41 / 100)
lin_a + exponential	92.000 % (92 / 100)
lin_b + normal	60.000 % (60 / 100)
lin_b + uniform	0.000 % (0 / 100)
lin_b + exponential	57.000 % (57 / 100)
lin_c + normal	99.000 % (99 / 100)
lin_c + uniform	6.000 % (6 / 100)
lin_c + exponential	100.000 % (100 / 100)
add_a + normal	95.000 % (95 / 100)
add_a + uniform	90.000 % (90 / 100)
add_a + exponential	96.000 % (96 / 100)
add_b + normal	100.000 % (100 / 100)
add_b + uniform	100.000 % (100 / 100)
add_b + exponential	100.000 % (100 / 100)
add_c + normal	100.000 % (100 / 100)
add_c + uniform	100.000 % (100 / 100)
add_c + exponential	100.000 % (100 / 100)
mul_a + normal	100.000 % (100 / 100)
mul_a + uniform	96.000 % (96 / 100)
mul_a + exponential	100.000 % (100 / 100)
mul_b + normal	100.000 % (100 / 100)
mul_b + uniform	100.000 % (100 / 100)
mul_b + exponential	100.000 % (100 / 100)
mul_c + normal	99.000 % (99 / 100)
mul_c + uniform	99.000 % (99 / 100)
mul_c + exponential	100.000 % (100 / 100)
com_a + normal	93.000 % (93 / 100)
com_a + uniform	67.000 % (67 / 100)
com_a + exponential	100.000 % (100 / 100)
com_b + normal	97.000 % (97 / 100)
com_b + uniform	91.000 % (91 / 100)
com_b + exponential	100.000 % (100 / 100)
com_c + normal	100.000 % (100 / 100)
com_c + uniform	100.000 % (100 / 100)
com_c + exponential	100.000 % (100 / 100)

Correct predictions averaged for function type:

lin_a	61.667 % (185 / 300)
lin_b	39.000 % (117 / 300)
lin_c	68.333 % (205 / 300)
add_a	93.667 % (281 / 300)
add_b	100.000 % (300 / 300)
add_c	100.000 % (300 / 300)
mul_a	98.667 % (296 / 300)
mul_b	100.000 % (300 / 300)
mul_c	99.333 % (298 / 300)
com_a	86.667 % (260 / 300)
com_b	96.000 % (288 / 300)
com_c	100.000 % (300 / 300)

Correct predictions averaged for noise type:

normal: 91.250 % (1095 / 1200)
uniform: 74.167 % (890 / 1200)
exponential: 95.417 % (1145 / 1200)

RECI Results for Synthetic (1000)

Correct estimates overall: 89.639 % (3227 / 3600, bootstrap err. 0.499 %)

Correct predictions for individual setups:

lin_a + normal	43.000 % (43 / 100)
lin_a + uniform	43.000 % (43 / 100)
lin_a + exponential	99.000 % (99 / 100)
lin_b + normal	60.000 % (60 / 100)
lin_b + uniform	0.000 % (0 / 100)
lin_b + exponential	86.000 % (86 / 100)
lin_c + normal	100.000 % (100 / 100)
lin_c + uniform	0.000 % (0 / 100)
lin_c + exponential	100.000 % (100 / 100)
add_a + normal	100.000 % (100 / 100)
add_a + uniform	100.000 % (100 / 100)
add_a + exponential	100.000 % (100 / 100)
add_b + normal	100.000 % (100 / 100)
add_b + uniform	100.000 % (100 / 100)
add_b + exponential	100.000 % (100 / 100)
add_c + normal	100.000 % (100 / 100)
add_c + uniform	100.000 % (100 / 100)
add_c + exponential	100.000 % (100 / 100)
mul_a + normal	100.000 % (100 / 100)
mul_a + uniform	100.000 % (100 / 100)
mul_a + exponential	100.000 % (100 / 100)
mul_b + normal	100.000 % (100 / 100)
mul_b + uniform	100.000 % (100 / 100)
mul_b + exponential	100.000 % (100 / 100)
mul_c + normal	100.000 % (100 / 100)
mul_c + uniform	100.000 % (100 / 100)
mul_c + exponential	100.000 % (100 / 100)
com_a + normal	100.000 % (100 / 100)
com_a + uniform	96.000 % (96 / 100)
com_a + exponential	100.000 % (100 / 100)
com_b + normal	100.000 % (100 / 100)
com_b + uniform	100.000 % (100 / 100)
com_b + exponential	100.000 % (100 / 100)
com_c + normal	100.000 % (100 / 100)
com_c + uniform	100.000 % (100 / 100)
com_c + exponential	100.000 % (100 / 100)

Correct predictions averaged for function type:

lin_a	61.667 % (185 / 300)
lin_b	48.667 % (146 / 300)
lin_c	66.667 % (200 / 300)
add_a	100.000 % (300 / 300)
add_b	100.000 % (300 / 300)
add_c	100.000 % (300 / 300)
mul_a	100.000 % (300 / 300)
mul_b	100.000 % (300 / 300)
mul_c	100.000 % (300 / 300)
com_a	98.667 % (296 / 300)
com_b	100.000 % (300 / 300)
com_c	100.000 % (300 / 300)

Correct predictions averaged for noise type:

normal: 91.917 % (1103 / 1200)
uniform: 78.250 % (939 / 1200)
exponential: 98.750 % (1185 / 1200)

Synthetic Results

	Synthetic (100)	Synthetic (1000)
ANM	48.361 % (1741 / 3600) bootstrap err. 0.812 %	48.139 % (1733 / 3600) bootstrap err. 0.844 %
CAM	70.389 % (2534 / 3600) bootstrap err. 0.748 %	69.306 % (2495 / 3600) bootstrap err. 0.769 %
CCDr	50.028 % (1801 / 3600) bootstrap err. 0.832 %	47.028 % (1693 / 3600) bootstrap err. 0.806 %
CDS	53.472 % (1925 / 3600) bootstrap err. 0.832 %	51.139 % (1841 / 3600) bootstrap err. 0.817 %
IGCI	**93.389 % (3362 / 3600)** bootstrap err. 0.415 %	**95.306 % (3431 / 3600)** bootstrap err. 0.352 %
KCDC	9.278 % (334 / 3600) bootstrap err. 0.478 %	9.917 % (357 / 3600) bootstrap err. 0.491 %
LiNGAM	1.694 % (61 / 3600) bootstrap err. 0.215 %	1.667 % (60 / 3600) bootstrap err. 0.213 %
noTEARS	86.278 % (3106 / 3600) bootstrap err. 0.578 %	77.139 % (2777 / 3600) bootstrap err. 0.692 %
RCC	47.417 % (1707 / 3600) bootstrap err. 0.826 %	46.361 % (1669 / 3600) bootstrap err. 0.841 %
RECI	86.944 % (3130 / 3600) bootstrap err. 0.561 %	89.639 % (3227 / 3600) bootstrap err. 0.499 %

Application Data

Application data allows for obtaining further performance indicators close to real-world problems.

We use data proposed by [Goudet et al., 2018]:

- **CE-Cha**: restricted set of continuous variable pairs from the cause effect pair challenge by [Guyon, 2013]
- **CE-Gauss**: pairs generated by random Gaussian processes with Gaussian mixtures for the cause
- **CE-Multi**: pairs generated with linear and polynomial mechanisms
- **CE-Net**: pairs generated with a randomly initialized neural network with various distributions of the cause

We use further the well-known Tübingen cause-effect pairs (**TCEP**) dataset by [Mooij et al., 2016].

Application Results I

	CE-Cha	CE-Gauss	CE-Multi	CE-Net
ANM	67.333 % (202 / 300) bootstrap err. 2.626 %	80.000 % (240 / 300) bootstrap err. 2.302 %	35.333 % (106 / 300) bootstrap err. 2.686 %	76.333 % (229 / 300) bootstrap err. 2.344 %
CAM	46.667 % (140 / 300) bootstrap err. 2.869 %	24.000 % (72 / 300) bootstrap err. 2.523 %	34.667 % (104 / 300) bootstrap err. 2.664 %	**78.333 % (235 / 300)** bootstrap err. 2.403 %
CCDr	47.333 % (142 / 300) bootstrap err. 2.778 %	48.667 % (146 / 300) bootstrap err. 2.949 %	49.667 % (149 / 300) bootstrap err. 2.888 %	53.667 % (161 / 300) bootstrap err. 2.747 %
CDS	**71.000 % (213 / 300)** bootstrap err. 2.570 %	**84.000 % (252 / 300)** bootstrap err. 2.138 %	43.667 % (131 / 300) bootstrap err. 2.850 %	**78.333 % (235 / 300)** bootstrap err. 2.442 %
IGCI	57.333 % (172 / 300) bootstrap err. 2.840 %	21.333 % (64 / 300) bootstrap err. 2.376 %	68.000 % (204 / 300) bootstrap err. 2.743 %	56.333 % (169 / 300) bootstrap err. 2.926 %
KCDC	49.000 % (147 / 300) bootstrap err. 2.929 %	61.667 % (185 / 300) bootstrap err. 2.847 %	**89.000 % (267 / 300)** bootstrap err. 1.780 %	60.667 % (182 / 300) bootstrap err. 2.730 %
LiNGAM	41.667 % (125 / 300) bootstrap err. 2.834 %	22.000 % (66 / 300) bootstrap err. 2.398 %	30.667 % (92 / 300) bootstrap err. 2.601 %	60.667 % (182 / 300) bootstrap err. 2.850 %
noTEARS	41.000 % (123 / 300) bootstrap err. 2.804 %	49.333 % (148 / 300) bootstrap err. 2.999 %	39.333 % (118 / 300) bootstrap err. 2.890 %	24.333 % (73 / 300) bootstrap err. 2.530 %
RCC	46.667 % (140 / 300) bootstrap err. 2.843 %	67.667 % (203 / 300) bootstrap err. 2.687 %	44.667 % (134 / 300) bootstrap err. 2.885 %	73.333 % (220 / 300) bootstrap err. 2.578 %
RECI	56.000 % (168 / 300) bootstrap err. 2.806 %	64.333 % (193 / 300) bootstrap err. 2.771 %	85.333 % (256 / 300) bootstrap err. 2.052 %	60.333 % (181 / 300) bootstrap err. 2.812 %

Application Results II

	TCEP v1.0 (plain)	TCEP v1.0 (weighted)
ANM	52.632 % (50 / 95) bootstrap err. 5.079 %	52.936 % bootstrap err. 6.909 %
CAM	60.000 % (57 / 95) bootstrap err. 4.911 %	59.238 % bootstrap err. 6.597 %
CCDr	73.684 % (70 / 95) bootstrap err. 4.549 %	72.142 % bootstrap err. 6.152 %
CDS	68.421 % (65 / 95) bootstrap err. 4.889 %	66.222 % bootstrap err. 6.595 %
IGCI	34.737 % (33 / 95) bootstrap err. 5.101 %	33.569 % bootstrap err. 6.515 %
KCDC	51.579 % (49 / 95) bootstrap err. 5.197 %	54.816 % bootstrap err. 6.756 %
LiNGAM	50.526 % (48 / 95) bootstrap err. 5.057 %	37.732 % bootstrap err. 6.244 %
noTEARS	48.421 % (46 / 95) bootstrap err. 5.104 %	50.831 % bootstrap err. 6.545 %
RCC	38.947 % (37 / 95) bootstrap err. 4.978 %	42.413 % bootstrap err. 6.633 %
RECI	**62.105 % (59 / 95)** bootstrap err. 4.958 %	**69.223 %** bootstrap err. 5.946 %

Summary

- This is an **intermediate** state of an ongoing investigation.
- Methods are realized by the code provided by authors, the CausalDiscoveryToolbox by [Kalainathan and Goudet, 2019], or are own implementations.
- There is **no** hyperparameter optimization, dataset normalization, or fine-tuning of the methods involved (we always use default settings).
- We plan to utilize further datasets such as more cause effect challenge pairs by [Guyon, 2013]
- We also have many more methods to evaluate (results already partially available): [Zhang and Hyvärinen, 2008, Mooij et al., 2010, de Almeida, 2013, Sgouritsa et al., 2015, Bontempi and Flauder, 2015, Hernandez-Lobato et al., 2016, Fonollosa, 2016, Budhathoki and Vreeken, 2017, Lopez-Paz et al., 2017, Ramsey et al., 2017, Goudet et al., 2018, Kalainathan et al., 2018, Chen et al., 2019, Zhu and Chen, 2019, Gnecco et al., 2019, Cai et al., 2019, Goldfarb and Evans, 2019, Monti et al., 2019, de Almeida, 2019, Samothrakis et al., 2019, Minnaert, 2019, Strobl and Visweswaran, 2019]
- Please tell us if we missed some interesting methods / datasets: christoph.kaeding@dlr.de

References I

Aragam, B. and Zhou, Q. (2015).
Concave penalized estimation of sparse gaussian bayesian networks.
Journal of Machine Learning Research.

Blöbaum, P., Janzing, D., Washio, T., Shimizu, S., and Schölkopf, B. (2018).
Cause-effect inference by comparing regression errors.
In *International Conference on Artificial Intelligence and Statistics*.

Bontempi, G. and Flauder, M. (2015).
From dependency to causality: a machine learning approach.
The Journal of Machine Learning Research (JMLR).

Budhathoki, K. and Vreeken, J. (2017).
Causal inference by stochastic complexity.
arXiv preprint arXiv:1702.06776.

Bühlmann, P., Peters, J., Ernest, J., et al. (2014).
Cam: Causal additive models, high-dimensional order search and penalized regression.
The Annals of Statistics.

Cai, R., Qiao, J., Zhang, K., Zhang, Z., and Hao, Z. (2019).
Causal discovery with cascade nonlinear additive noise models.
arXiv preprint arXiv:1905.09442.

Chen, Z., Zhu, S., Liu, Y., and Tse, T. (2019).
Causal discovery by kernel intrinsic invariance measure.
arXiv preprint arXiv:1909.00513.

References II

Daniusis, P., Janzing, D., Mooij, J., Zscheischler, J., Steudel, B., Zhang, K., and Schölkopf, B. (2012).
Inferring deterministic causal relations.
arXiv preprint arXiv:1203.3475.

de Almeida, D. M. (2013).
Automated feature engineering applied to causality.
In *NIPS Workshop on Causality*.

de Almeida, D. M. (2019).
Pattern-based causal feature extraction.
In *Cause Effect Pairs in Machine Learning*.

Fonollosa, J. A. (2016).
Conditional distribution variability measures for causality detection.
arXiv preprint arXiv:1601.06680.

Gnecco, N., Meinshausen, N., Peters, J., and Engelke, S. (2019).
Causal discovery in heavy-tailed models.
arXiv preprint arXiv:1908.05097.

Goldfarb, D. and Evans, S. (2019).
Causal inference via conditional kolmogorov complexity using mdl binning.
arXiv preprint arXiv:1911.00332.

Goudet, O., Kalainathan, D., Caillou, P., Guyon, I., Lopez-Paz, D., and Sebag, M. (2018).
Learning functional causal models with generative neural networks.
In *Explainable and Interpretable Models in Computer Vision and Machine Learning*.

References III

Guyon, I. (2013).
Chalearn cause effect pairs challenge.
http://www.causality.inf.ethz.ch/cause-effect.php.

Hernandez-Lobato, D., Morales-Mombiela, P., Lopez-Paz, D., and Suarez, A. (2016).
Non-linear causal inference using gaussianity measures.
The Journal of Machine Learning Research (JMLR).

Hoyer, P. O., Janzing, D., Mooij, J. M., Peters, J., and Schölkopf, B. (2009).
Nonlinear causal discovery with additive noise models.
In *Advances in Neural Information Processing Systems*.

Kalainathan, D. and Goudet, O. (2019).
Causal discovery toolbox: Uncover causal relationships in python.
arXiv preprint arXiv:1903.02278.

Kalainathan, D., Goudet, O., Guyon, I., Lopez-Paz, D., and Sebag, M. (2018).
Sam: Structural agnostic model, causal discovery and penalized adversarial learning.
arXiv preprint arXiv:1803.04929.

Lopez-Paz, D., Muandet, K., Schölkopf, B., and Tolstikhin, I. (2015).
Towards a learning theory of cause-effect inference.
In *International Conference on Machine Learning*.

Lopez-Paz, D., Nishihara, R., Chintala, S., Scholkopf, B., and Bottou, L. (2017).
Discovering causal signals in images.
In *IEEE Conference on Computer Vision and Pattern Recognition*.

References IV

Minnaert, B. (2019).
Feature importance in causal inference for numerical and categorical variables.
In *Cause Effect Pairs in Machine Learning*.

Mitrovic, J., Sejdinovic, D., and Teh, Y. W. (2018).
Causal inference via kernel deviance measures.
In *Advances in Neural Information Processing Systems*.

Monti, R. P., Zhang, K., and Hyvarinen, A. (2019).
Causal discovery with general non-linear relationships using non-linear ica.
arXiv preprint arXiv:1904.09096.

Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., and Schölkopf, B. (2016).
Distinguishing cause from effect using observational data: methods and benchmarks.
The Journal of Machine Learning Research.

Mooij, J. M., Stegle, O., Janzing, D., Zhang, K., and Schölkopf, B. (2010).
Probabilistic latent variable models for distinguishing between cause and effect.
In *Advances in Neural Information Processing Systems*.

Ramsey, J., Glymour, M., Sanchez-Romero, R., and Glymour, C. (2017).
A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images.
International journal of data science and analytics.

References V

Samothrakis, S., Perez, D., and Lucas, S. (2019).
Training gradient boosting machines using curve-fitting and information-theoretic features for causal direction detection.
In *Cause Effect Pairs in Machine Learning*.

Sgouritsa, E., Janzing, D., Hennig, P., and Schölkopf, B. (2015).
Inference of cause and effect with unsupervised inverse regression.
In *Artificial intelligence and statistics*.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A. (2006).
A linear non-gaussian acyclic model for causal discovery.
Journal of Machine Learning Research.

Strobl, E. V. and Visweswaran, S. (2019).
Markov blanket ranking using kernel-based conditional dependence measures.
In *Cause Effect Pairs in Machine Learning*.

Zhang, K. and Hyvärinen, A. (2008).
Distinguishing causes from effects using nonlinear acyclic causal models.
In *International Conference on Causality*.

Zheng, X., Aragam, B., Ravikumar, P. K., and Xing, E. P. (2018).
Dags with no tears: Continuous optimization for structure learning.
In *Advances in Neural Information Processing Systems*.

Zhu, S. and Chen, Z. (2019).
Causal discovery with reinforcement learning.
arXiv preprint arXiv:1906.04477.