Supplementary Material for: Non-linear Causal Inference using Gaussianity Measures

Daniel Hernández-Lobato
daniel.hernandez@uam.es
Universidad Autónoma de Madrid
Calle Francisco Tomás y Valiente 11,
Madrid, 28049, Spain

Pablo Morales-Mombiela
pablo.morales@estudiante.uam.es
Quantitative Risk Research
Calle Faraday 7,
Madrid, 28049, Spain

David Lopez-Paz
dlp@fb.com
Facebook AI Research
6 rue Menars,
Paris 75002, France

Alberto Suárez
alberto.suarez@uam.es
Universidad Autónoma de Madrid
Calle Francisco Tomás y Valiente 11,
Madrid, 28049, Spain

1 An Illustrative Example

In this section we illustrate with a simple example that the proposed approach is different from the method of Hoyer et al. [1].

Assume that x is the cause and that y is the effect. Let ϵ be the residual of a (nonlinear) model in the causal direction and $\tilde{\epsilon}$ the residual of a (nonlinear) model in the anticausal direction. That is,

$$y = f(x) + \epsilon\,, \qquad \epsilon \perp x\,, \qquad x = g(x) + \tilde{\epsilon}\,. \tag{1}$$

In [1] it is used the fact that x and ϵ, the residual in the causal direction, are independent. Therefore, the mutual information between x and ϵ is zero ($I(x, \epsilon) = 0$). By contrast, y and $\tilde{\epsilon}$, the residual in the anticausal direction, are in general dependent. This means that the mutual information between y and $\tilde{\epsilon}$ is larger than or equal to zero ($I(y, \tilde{\epsilon}) \geq 0$). Lemma 1 of Kpotufe et al. [3] shows that this result is equivalent to:

$$H(x) + H(\epsilon) \leq H(y) + H(\tilde{\epsilon})\,, \tag{2}$$

where $H(\cdot)$ denotes differential entropy. This results is also used by Hyvärinen and Smith [2].

If x and y have the same distribution, $H(x) = H(y)$, and

$$H(\epsilon) \leq H(\tilde{\epsilon})\,. \tag{3}$$

See Eq. (16) in the main manuscript. That is, the entropy of the residuals in the anticausal direction is larger or equal to the entropy of the residuals in the causal direction. However, this does not necessarily mean that $\tilde{\epsilon}$ is more Gaussian than ϵ because the residuals of these nonlinear models have, in general, different variances.

The difference between our method and the one of Hoyer et al. [1] can be illustrated with the following example, inspired by a time series model analyzed in [4]. Consider a linear model with additive dicotomic noise in the causal direction. That is,

$$y_n = \frac{1}{2} x_n + \epsilon_n\,, \qquad \epsilon_n = \begin{cases} +\frac{1}{4} & \text{with prob.} \quad \frac{1}{2}\,. \\ -\frac{1}{4} & \text{with prob.} \quad \frac{1}{2}\,. \end{cases}\,, \qquad \epsilon_n \perp x_n\,. \tag{4}$$

Furthermore, assume that $x_n \sim U[-0.5, 0.5]$. This implies that $y_n \sim U[-0.5, 0.5]$.

In the anticausal direction the model is non-linear and deterministic (has zero noise). That is,

$$y_n = 2y_n \text{mod} 1 - \frac{1}{2}\,. \tag{5}$$

See Figure 1 for further details.

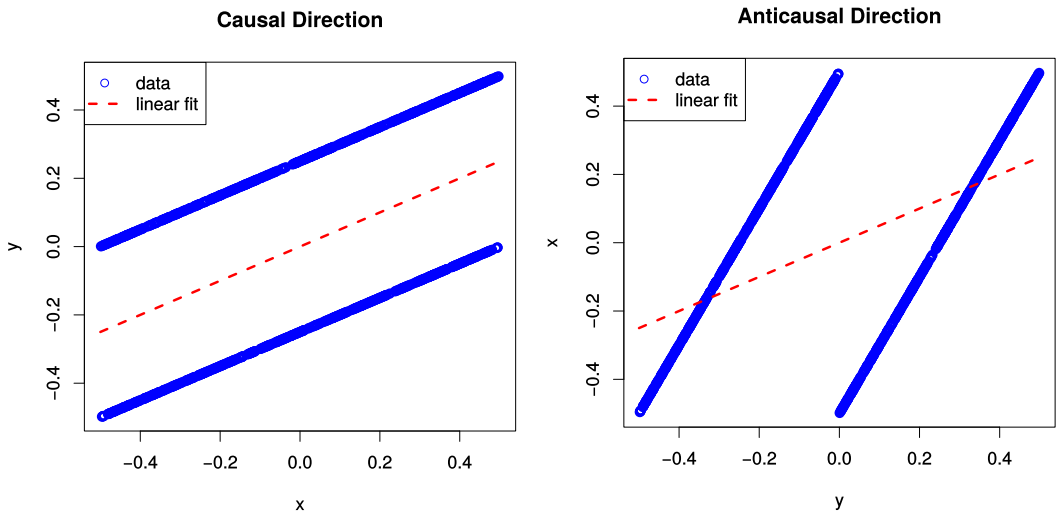

Figure 1: Plot, for the illustrative example, of the observed data and the corresponding fit of a linear model in the causal and the anticausal direction.

In the method proposed in this manuscript, x and y are assumed to have the same distribution. Then, we make a linear fit in feature space both for the causal and anticausal directions.

In this particular case, x and y do have the same distribution. That is, they are uniform in the interval $[-0.5, 0.5]$. Furthermore, no embedding is necessary because the model is linear and additive in the causal direction. The residuals of a linear fit in the anticausal direction are

$$\tilde{\epsilon}_n = x_n - \frac{1}{2} y_n\,. \tag{6}$$

The distribution of these linear anticausal residuals is piecewise uniform with support in the interval [-0.5,0.5]. That is,

$$p(\tilde{\epsilon}_n) = \begin{cases} \frac{2}{3} & \text{if} \quad \tilde{\epsilon} \in [-\frac{1}{2}, -\frac{1}{4}]\,, \\ \frac{4}{3} & \text{if} \quad \tilde{\epsilon} \in [-\frac{1}{4}, \frac{1}{4}]\,, \\ \frac{2}{3} & \text{if} \quad \tilde{\epsilon} \in [\frac{1}{4}, \frac{1}{2}]\,, \end{cases} \quad . \tag{7}$$

Furthermore, the variances of both residuals is the same. That is,

$$\text{var}(\epsilon_n) = \text{var}(\tilde{\epsilon}_n) = \frac{1}{16}\,. \tag{8}$$

The Gaussianization effect can be observed by taking a look a the histograms of the residuals. Figure 2 shows a histogram for both ϵ_n and $\tilde{\epsilon}_n$.

If we applied the method of Hoyer et al. [1], the non-parametric fit in the causal direction necessarily leads to residuals with a finite variance: the distribution of y conditioned to a particular value of x is bimodal. Therefore, the residuals of a fit of y in terms of x are necessarily non-zero. By contrast, since the model is deterministic in the anticausal direction, a sufficiently flexible non-parametric fit in this direction would lead to residuals that are close to zero. Therefore, the residuals in the causal and anticausal direction have very different variances. This means that it is not possible to make a decision directly on the basis of a Gaussianization effect on the residuals of nonlinear models in input (x,y) space.

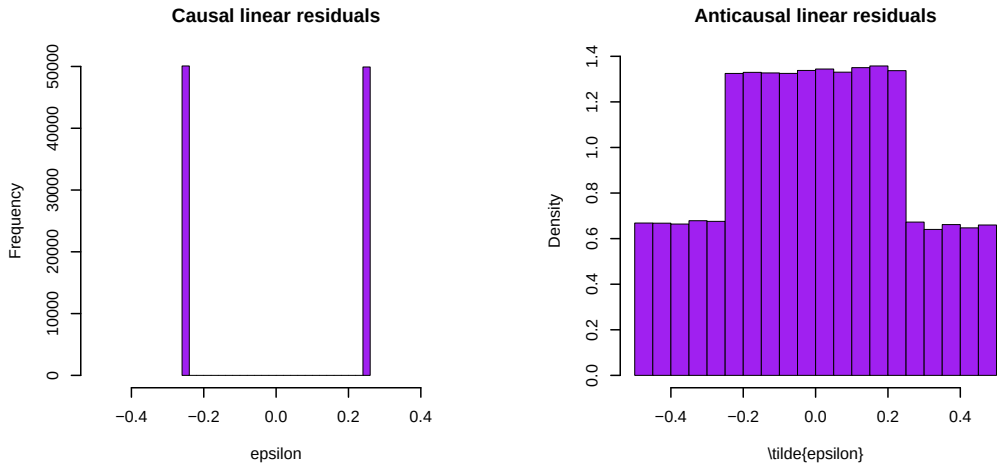

Figure 2: Histogram, for the illustrative example, of the residuals of a linear fit in the causal and the anticausal direction.

2 Additional Results for the Synthetic Experiments

This section contains more results for the synthetic experiments where the number of samples is respectively 100, 200, 300 and 1000. Furthermore, we also-show more results for the experiments involving the 184 cause-effect pairs selected from the ChaLearn Challenge when the number of samples is respectively 100, 200 and 300.

References

[1] P. O. Hoyer, D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf. Nonlinear causal discovery with additive noise models. In *Advances in Neural Information Processing Systems 21*, pages 689–696, 2009.

[2] A. Hyvärinen and S. M. Smith. Pairwise likelihood ratios for estimation of non-Gaussian structural equation models. *Journal of Machine Learning Research*, 14(1):111–152, 2013.

[3] S. Kpotufe, E. Sgouritsa, D. Janzing, and B. Schölkopf. Consistency of causal inference under the additive noise model. In *International Conference on Machine Learning*, pages 478–486, 2014.

[4] P. Srinivasa Rao, Don H. Johnson, and David D. Becker. Generation and analysis of non-Gaussian markov time series. *Signal Processing, IEEE Transactions on*, 40:845–856, 1992.

Table 1: Accuracy on the synthetic data for each method, causal mechanisms and type of noise. The number of samples employed is equal to 100

Noise	**Generalized Gaussian**				**Laplacian**				**Gaussian**				**Bimodal**			
Algorithm	**Mechanism**				**Mechanism**				**Mechanism**				**Mechanism**			
	M1	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**
$p_1(x)$																
LINGAM	98%	5%	15%	71%	89%	0%	26%	64%	61%	2%	11%	34%	99%	0%	19%	98%
IGCI	44%	97%	99%	39%	68%	100%	100%	80%	50%	99%	100%	61%	42%	91%	99%	43%
EMD	52%	67%	80%	53%	42%	92%	84%	59%	75%	88%	73%	52%	47%	75%	56%	50%
IR-AN	88%	45%	87%	84%	72%	30%	80%	70%	52%	34%	74%	60%	99%	43%	82%	91%
NLME	84%	50%	48%	72%	86%	27%	42%	86%	44%	41%	33%	69%	97%	48%	45%	77%
MAD	11%	85%	96%	16%	90%	93%	100%	100%	42%	90%	100%	72%	3%	76%	87%	2%
GR-AN	87%	45%	75%	87%	92%	44%	61%	85%	52%	45%	37%	47%	100%	47%	88%	96%
GR-AN*	91%	4%	76%	93%	95%	27%	59%	80%	51%	17%	27%	42%	98%	11%	95%	97%
GR-K4	99%	66%	71%	87%	80%	51%	58%	82%	57%	47%	25%	34%	99%	53%	85%	95%
GR-ENT	75%	51%	69%	71%	82%	51%	45%	64%	40%	54%	44%	49%	99%	52%	75%	88%
$p_2(x)$																
LINGAM	100%	40%	71%	98%	98%	24%	87%	100%	100%	23%	72%	92%	100%	41%	88%	99%
IGCI	45%	88%	96%	57%	84%	100%	98%	88%	53%	99%	96%	77%	46%	74%	95%	48%
EMD	55%	84%	88%	76%	82%	96%	84%	84%	57%	98%	90%	79%	72%	80%	87%	59%
IR-AN	100%	48%	99%	100%	96%	45%	99%	100%	99%	46%	100%	100%	100%	57%	100%	100%
NLME	99%	49%	99%	99%	76%	38%	99%	98%	80%	38%	98%	95%	99%	44%	99%	99%
MAD	2%	91%	97%	36%	31%	96%	100%	87%	11%	93%	99%	76%	0%	95%	76%	7%
GR-AN	85%	41%	78%	95%	72%	47%	70%	82%	40%	42%	33%	47%	97%	45%	98%	100%
GR-AN*	96%	1%	68%	78%	87%	5%	47%	74%	43%	3%	13%	36%	98%	9%	88%	95%
GR-K4	99%	49%	88%	100%	66%	56%	77%	77%	46%	40%	29%	44%	97%	51%	96%	100%
GR-ENT	67%	42%	62%	74%	59%	46%	60%	69%	37%	49%	39%	54%	83%	41%	80%	93%
$p_3(x)$																
LINGAM	92%	0%	32%	81%	96%	1%	78%	84%	60%	0%	47%	46%	97%	15%	28%	90%
IGCI	65%	100%	99%	80%	82%	100%	99%	91%	63%	100%	100%	80%	69%	99%	100%	74%
EMD	62%	95%	92%	64%	81%	99%	96%	86%	64%	92%	96%	81%	40%	84%	97%	63%
IR-AN	96%	45%	99%	96%	90%	28%	98%	86%	64%	34%	98%	81%	99%	44%	98%	97%
NLME	93%	47%	97%	96%	77%	25%	96%	88%	48%	26%	90%	81%	97%	41%	92%	95%
MAD	12%	91%	96%	28%	72%	95%	100%	94%	48%	85%	100%	81%	7%	80%	88%	6%
GR-AN	90%	49%	70%	88%	85%	50%	78%	84%	47%	51%	40%	43%	99%	62%	77%	98%
GR-AN*	91%	3%	82%	84%	88%	30%	82%	89%	41%	9%	41%	37%	98%	15%	93%	99%
GR-K4	97%	57%	74%	96%	79%	38%	83%	76%	49%	51%	30%	41%	97%	59%	79%	95%
GR-ENT	73%	50%	53%	81%	67%	56%	59%	74%	50%	54%	40%	42%	92%	62%	66%	87%

Table 2: Accuracy on the synthetic data for each method, causal mechanisms and type of noise. The number of samples employed is equal to 200

Noise	**Generalized Gaussian**				**Laplacian**				**Gaussian**				**Bimodal**			
Algorithm	**Mechanism**				**Mechanism**				**Mechanism**				**Mechanism**			
	M1	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**
$p_1(x)$																
LINGAM	100%	5%	4%	89%	96%	0%	25%	74%	56%	0%	4%	28%	100%	0%	9%	100%
IGCI	31%	99%	100%	48%	65%	100%	100%	84%	51%	100%	100%	60%	46%	96%	100%	59%
EMD	46%	91%	92%	23%	64%	99%	99%	60%	36%	97%	97%	56%	35%	84%	96%	67%
IR-AN	100%	29%	97%	97%	91%	21%	86%	90%	43%	32%	90%	61%	100%	39%	97%	99%
NLME	96%	30%	33%	88%	92%	17%	29%	97%	46%	33%	28%	80%	100%	49%	33%	96%
MAD	0%	93%	100%	2%	96%	99%	100%	100%	53%	99%	100%	83%	0%	88%	97%	0%
GR-AN	100%	53%	88%	96%	97%	43%	71%	97%	50%	50%	9%	37%	100%	52%	100%	100%
GR-AN*	99%	0%	90%	98%	99%	7%	75%	97%	47%	2%	9%	34%	100%	0%	99%	100%
GR-K4	100%	67%	86%	100%	92%	51%	86%	92%	49%	50%	7%	36%	100%	64%	95%	100%
GR-ENT	98%	65%	67%	82%	84%	48%	59%	79%	43%	48%	35%	43%	100%	60%	76%	97%
$p_2(x)$																
LINGAM	100%	40%	66%	100%	100%	24%	88%	100%	100%	17%	73%	97%	100%	45%	84%	100%
IGCI	37%	94%	100%	51%	82%	100%	100%	96%	69%	100%	100%	78%	35%	90%	100%	62%
EMD	40%	85%	98%	89%	96%	98%	99%	98%	92%	99%	99%	89%	78%	91%	98%	82%
IR-AN	100%	44%	100%	100%	99%	40%	100%	100%	100%	43%	100%	100%	100%	47%	100%	100%
NLME	100%	37%	100%	100%	86%	33%	100%	100%	92%	38%	99%	100%	100%	47%	99%	100%
MAD	2%	99%	99%	22%	22%	100%	100%	94%	2%	100%	100%	91%	0%	97%	80%	1%
GR-AN	94%	45%	94%	100%	86%	38%	84%	96%	35%	57%	41%	45%	100%	36%	100%	100%
GR-AN*	98%	0%	74%	96%	94%	0%	45%	94%	51%	1%	8%	18%	100%	2%	99%	100%
GR-K4	100%	57%	99%	100%	81%	46%	92%	93%	38%	60%	33%	52%	100%	46%	99%	100%
GR-ENT	80%	51%	82%	92%	77%	37%	69%	85%	39%	56%	49%	47%	96%	43%	96%	99%
$p_3(x)$																
LINGAM	100%	0%	23%	88%	98%	0%	88%	96%	74%	0%	47%	61%	100%	8%	16%	100%
IGCI	69%	100%	100%	80%	92%	100%	100%	94%	83%	100%	100%	88%	67%	100%	100%	72%
EMD	72%	90%	99%	88%	89%	100%	100%	95%	83%	99%	100%	97%	67%	99%	99%	98%
IR-AN	100%	35%	100%	99%	97%	27%	99%	98%	74%	34%	99%	96%	100%	47%	100%	100%
NLME	100%	36%	95%	99%	95%	17%	95%	99%	52%	27%	94%	97%	99%	40%	96%	100%
MAD	2%	94%	100%	8%	93%	96%	100%	100%	47%	94%	100%	96%	0%	90%	94%	0%
GR-AN	99%	51%	80%	97%	97%	54%	95%	97%	46%	49%	43%	41%	100%	58%	85%	100%
GR-AN*	100%	2%	94%	99%	100%	1%	91%	97%	51%	0%	38%	41%	100%	1%	100%	100%
GR-K4	100%	60%	84%	99%	86%	44%	94%	93%	54%	50%	28%	28%	100%	51%	92%	100%
GR-ENT	95%	55%	56%	82%	85%	58%	68%	82%	41%	56%	45%	42%	99%	58%	45%	96%

Table 3: Accuracy on the synthetic data for each method, causal mechanisms and type of noise. The number of samples employed is equal to 300

Noise	**Generalized Gaussian**				**Laplacian**				**Gaussian**				**Bimodal**			
Algorithm	**Mechanism**				**Mechanism**				**Mechanism**				**Mechanism**			
	M1	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**
$p_1(x)$																
LINGAM	100%	2%	6%	89%	100%	0%	30%	76%	58%	0%	13%	22%	100%	0%	11%	100%
IGCI	37%	100%	100%	36%	77%	100%	100%	84%	51%	100%	100%	60%	24%	100%	100%	61%
EMD	66%	86%	97%	57%	60%	98%	100%	62%	67%	100%	98%	73%	53%	84%	98%	76%
IR-AN	100%	27%	99%	100%	100%	19%	97%	94%	51%	25%	97%	65%	100%	42%	99%	100%
NLME	100%	30%	23%	95%	100%	16%	22%	100%	56%	26%	12%	88%	100%	45%	29%	96%
MAD	0%	93%	100%	1%	100%	99%	100%	100%	48%	98%	100%	97%	0%	96%	98%	0%
GR-AN	100%	53%	89%	100%	100%	45%	81%	100%	55%	52%	11%	26%	100%	46%	100%	100%
GR-AN*	100%	0%	97%	100%	100%	0%	86%	100%	59%	0%	5%	20%	100%	0%	100%	100%
GR-K4	100%	71%	89%	100%	99%	58%	90%	95%	62%	55%	7%	29%	100%	56%	100%	100%
GR-ENT	100%	64%	62%	91%	96%	50%	58%	90%	55%	57%	34%	42%	100%	56%	78%	99%
$p_2(x)$																
LINGAM	100%	40%	69%	100%	100%	29%	94%	100%	100%	22%	79%	100%	100%	53%	84%	100%
IGCI	40%	97%	100%	65%	90%	100%	100%	99%	62%	100%	100%	90%	42%	94%	100%	59%
EMD	96%	94%	97%	89%	99%	98%	100%	97%	95%	99%	100%	92%	86%	91%	99%	92%
IR-AN	100%	48%	100%	100%	100%	33%	100%	100%	100%	48%	100%	100%	100%	52%	100%	100%
NLME	100%	39%	99%	100%	92%	29%	99%	100%	97%	39%	100%	100%	100%	37%	100%	100%
MAD	1%	100%	100%	21%	10%	99%	100%	99%	1%	98%	100%	96%	0%	98%	80%	0%
GR-AN	99%	51%	96%	100%	91%	38%	90%	99%	37%	39%	28%	53%	99%	44%	100%	100%
GR-AN*	100%	0%	85%	99%	97%	0%	39%	99%	40%	0%	3%	19%	100%	2%	100%	100%
GR-K4	100%	63%	100%	100%	77%	50%	98%	91%	31%	54%	30%	43%	100%	47%	100%	100%
GR-ENT	89%	56%	88%	99%	76%	45%	67%	88%	43%	41%	36%	56%	96%	49%	96%	100%
$p_3(x)$																
LINGAM	100%	0%	19%	96%	100%	1%	97%	99%	85%	0%	52%	52%	100%	10%	17%	100%
IGCI	68%	100%	100%	77%	94%	100%	100%	99%	83%	100%	100%	95%	73%	100%	100%	84%
EMD	78%	96%	100%	88%	96%	100%	100%	94%	93%	99%	100%	92%	75%	91%	99%	83%
IR-AN	100%	43%	100%	99%	97%	34%	100%	99%	87%	21%	100%	96%	100%	36%	100%	100%
NLME	100%	33%	95%	100%	97%	18%	99%	100%	65%	27%	96%	98%	100%	33%	97%	100%
MAD	1%	98%	100%	4%	95%	97%	100%	100%	52%	100%	100%	88%	0%	93%	95%	0%
GR-AN	100%	57%	90%	99%	97%	53%	99%	98%	57%	52%	48%	49%	100%	53%	93%	100%
GR-AN*	100%	0%	99%	99%	100%	0%	100%	99%	61%	0%	23%	36%	100%	0%	100%	100%
GR-K4	100%	67%	90%	99%	91%	35%	99%	90%	45%	53%	27%	26%	100%	54%	98%	100%
GR-ENT	97%	60%	47%	89%	83%	56%	80%	85%	44%	51%	54%	41%	100%	50%	37%	98%

Table 4: Accuracy on the synthetic data for each method, causal mechanisms and type of noise. The number of samples employed is equal to 1000

Noise	**Generalized Gaussian**				**Laplacian**				**Gaussian**				**Bimodal**			
Algorithm	**Mechanism**				**Mechanism**				**Mechanism**				**Mechanism**			
	M1	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**	**M1**	**M2**	**M3**	**M4**
$p_1(x)$																
LINGAM	100%	0%	7%	100%	100%	0%	22%	99%	55%	0%	7%	14%	100%	0%	2%	100%
IGCI	26%	100%	100%	33%	95%	100%	100%	97%	55%	100%	100%	67%	21%	100%	100%	46%
EMD	45%	100%	100%	81%	70%	100%	100%	87%	44%	100%	100%	75%	69%	100%	100%	62%
IR-AN	100%	39%	100%	100%	100%	19%	100%	100%	54%	31%	100%	91%	100%	39%	100%	100%
NLME	100%	34%	6%	100%	100%	9%	6%	100%	55%	19%	1%	97%	100%	46%	15%	100%
MAD	0%	99%	100%	0%	100%	100%	100%	100%	60%	99%	100%	98%	0%	97%	100%	0%
GR-AN	100%	49%	100%	100%	100%	48%	78%	100%	52%	40%	1%	9%	100%	50%	100%	100%
GR-AN*	100%	0%	99%	100%	100%	0%	99%	100%	52%	0%	0%	6%	100%	0%	100%	100%
GR-K4	100%	63%	100%	100%	100%	52%	100%	100%	48%	62%	0%	12%	100%	59%	99%	100%
GR-ENT	100%	56%	71%	100%	100%	53%	56%	100%	43%	49%	22%	33%	100%	59%	87%	100%
$p_2(x)$																
LINGAM	100%	44%	61%	100%	100%	24%	100%	100%	100%	17%	84%	100%	100%	56%	96%	100%
IGCI	33%	98%	100%	59%	100%	100%	100%	100%	83%	100%	100%	98%	34%	98%	100%	65%
EMD	94%	100%	100%	99%	99%	100%	100%	100%	99%	100%	100%	99%	96%	100%	100%	98%
IR-AN	100%	62%	100%	100%	100%	36%	100%	100%	100%	48%	100%	100%	100%	56%	100%	100%
NLME	100%	56%	100%	100%	100%	38%	100%	100%	100%	42%	100%	100%	100%	50%	100%	100%
MAD	0%	100%	100%	0%	4%	100%	100%	100%	0%	100%	100%	100%	0%	100%	90%	0%
GR-AN	100%	40%	100%	100%	99%	41%	99%	100%	17%	30%	11%	52%	100%	31%	100%	100%
GR-AN*	100%	0%	99%	100%	100%	0%	73%	100%	24%	0%	0%	2%	100%	0%	100%	100%
GR-K4	100%	53%	100%	100%	92%	53%	100%	98%	18%	42%	13%	35%	100%	34%	100%	100%
GR-ENT	99%	49%	97%	100%	90%	41%	71%	96%	44%	34%	40%	51%	100%	35%	100%	100%
$p_3(x)$																
LINGAM	100%	0%	6%	100%	100%	0%	100%	100%	99%	0%	50%	73%	100%	8%	9%	100%
IGCI	76%	100%	100%	87%	100%	100%	100%	100%	96%	100%	100%	98%	69%	100%	100%	84%
EMD	98%	100%	100%	99%	100%	100%	100%	100%	97%	100%	100%	100%	92%	100%	100%	91%
IR-AN	100%	35%	100%	100%	100%	24%	100%	100%	99%	40%	100%	100%	100%	49%	100%	100%
NLME	100%	41%	100%	100%	100%	20%	100%	100%	73%	33%	99%	100%	100%	47%	100%	100%
MAD	0%	100%	100%	0%	100%	99%	100%	100%	43%	99%	100%	99%	0%	99%	99%	0%
GR-AN	100%	48%	100%	100%	100%	47%	100%	100%	52%	50%	25%	34%	100%	50%	100%	100%
GR-AN*	100%	0%	100%	100%	100%	0%	100%	100%	52%	0%	2%	16%	100%	0%	100%	100%
GR-K4	100%	64%	99%	100%	98%	49%	100%	100%	43%	60%	5%	16%	100%	48%	100%	100%
GR-ENT	100%	51%	42%	100%	99%	57%	93%	98%	54%	49%	41%	42%	100%	44%	32%	100%

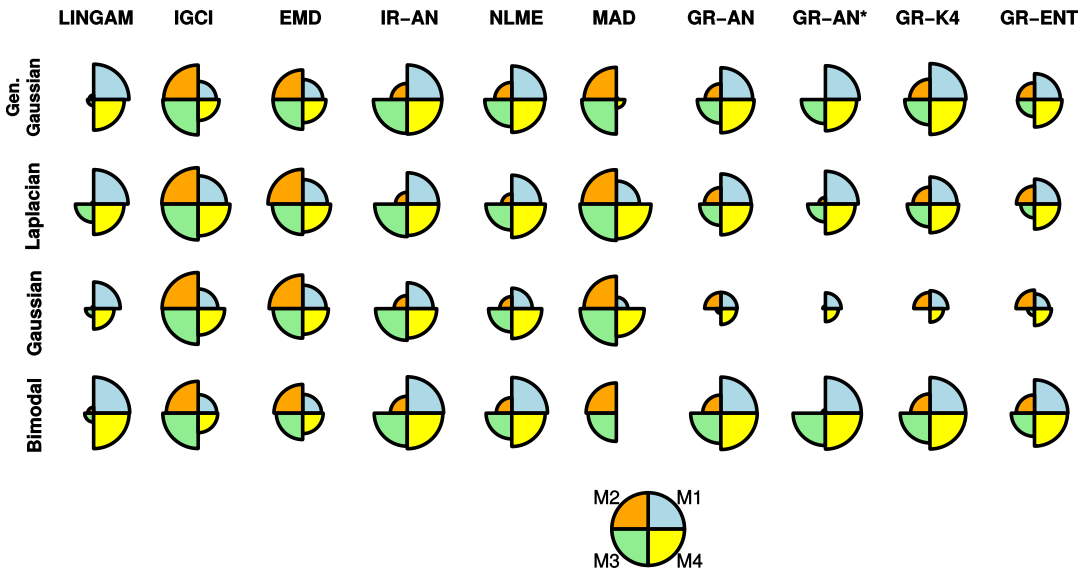

Figure 3: Radar charts showing the average accuracy of each method for the different types of noise considered and for each mechanism M1, M2, M3 and M4. For a particular method and type of noise, the radius of each portion of the pie is proportional to the corresponding average accuracy of the method across the distributions p_1, p_2 and p_3 for the cause. The pie at the bottom corresponds to 100% accuracy for each mechanism. The number of samples is equal to 100.

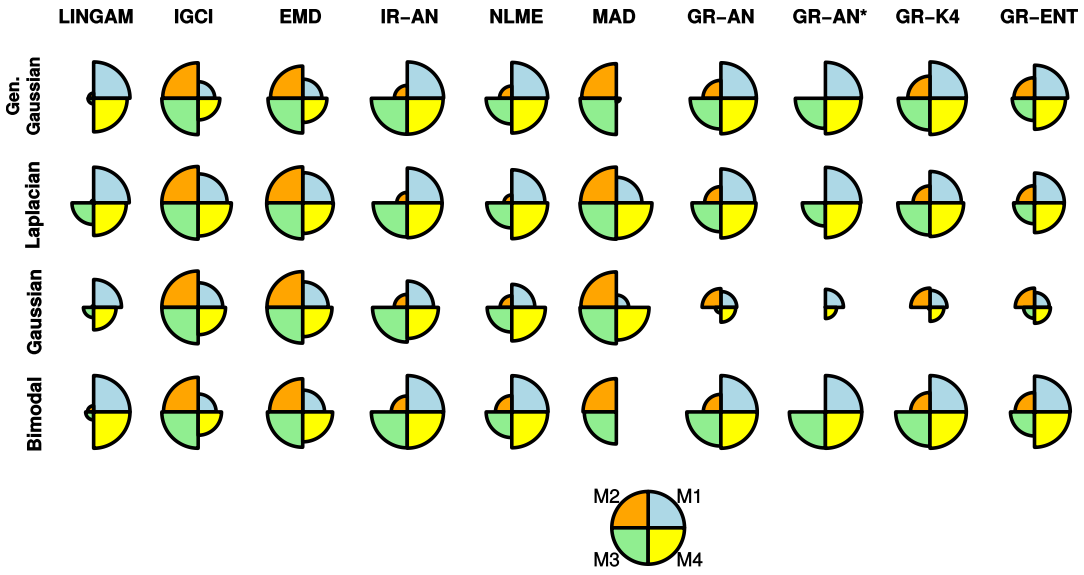

Figure 4: Radar charts showing the average accuracy of each method for the different types of noise considered and for each mechanism M1, M2, M3 and M4. For a particular method and type of noise, the radius of each portion of the pie is proportional to the corresponding average accuracy of the method across the distributions p_1, p_2 and p_3 for the cause. The pie at the bottom corresponds to 100% accuracy for each mechanism. The number of samples is equal to 200.

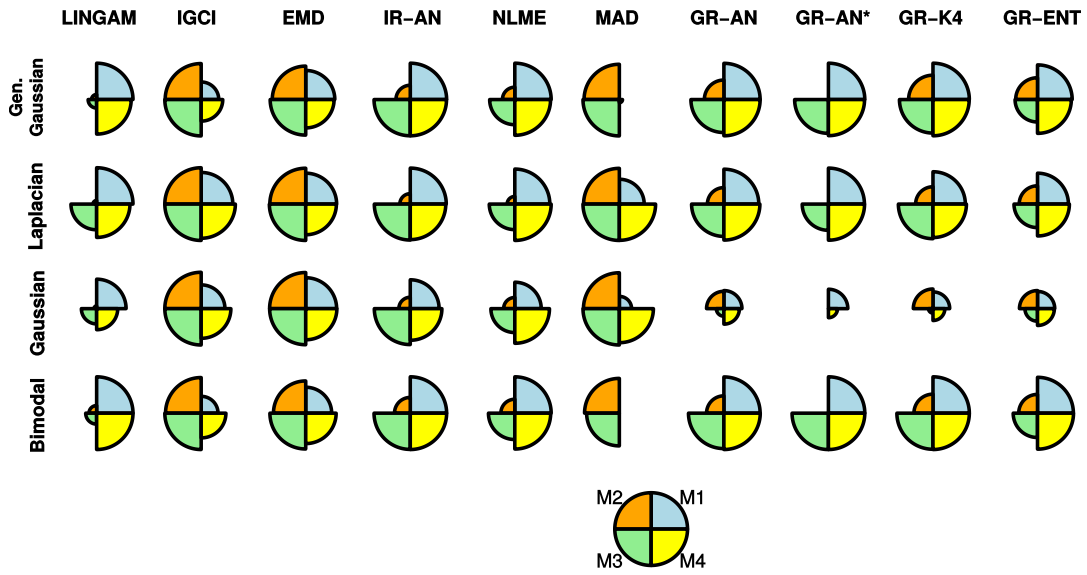

Figure 5: Radar charts showing the average accuracy of each method for the different types of noise considered and for each mechanism M1, M2, M3 and M4. For a particular method and type of noise, the radius of each portion of the pie is proportional to the corresponding average accuracy of the method across the distributions p_1, p_2 and p_3 for the cause. The pie at the bottom corresponds to 100% accuracy for each mechanism. The number of samples is equal to 300.

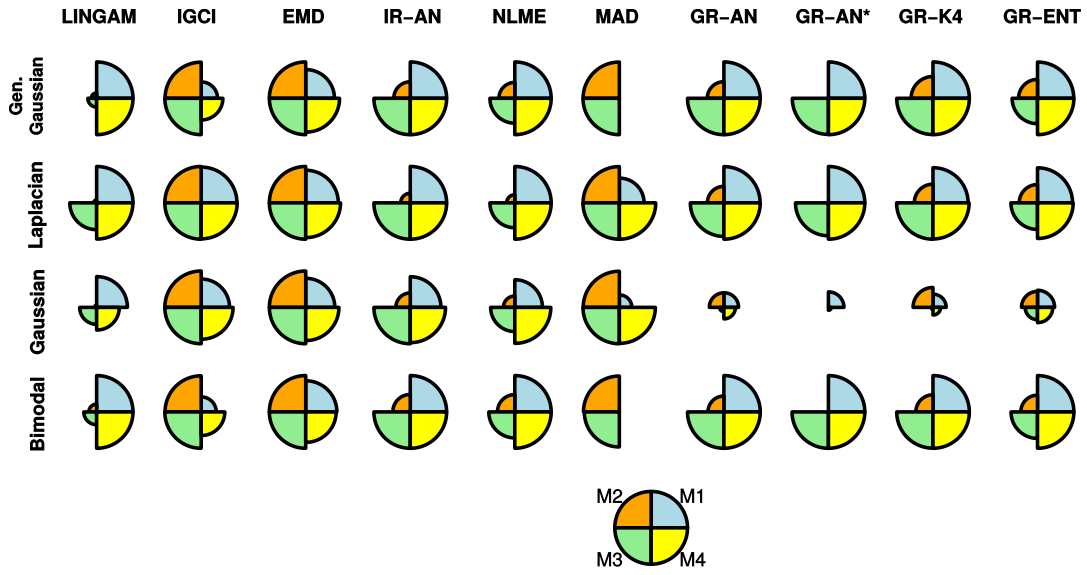

Figure 6: Radar charts showing the average accuracy of each method for the different types of noise considered and for each mechanism M1, M2, M3 and M4. For a particular method and type of noise, the radius of each portion of the pie is proportional to the corresponding average accuracy of the method across the distributions p_1, p_2 and p_3 for the cause. The pie at the bottom corresponds to 100% accuracy for each mechanism. The number of samples is equal to 1000.

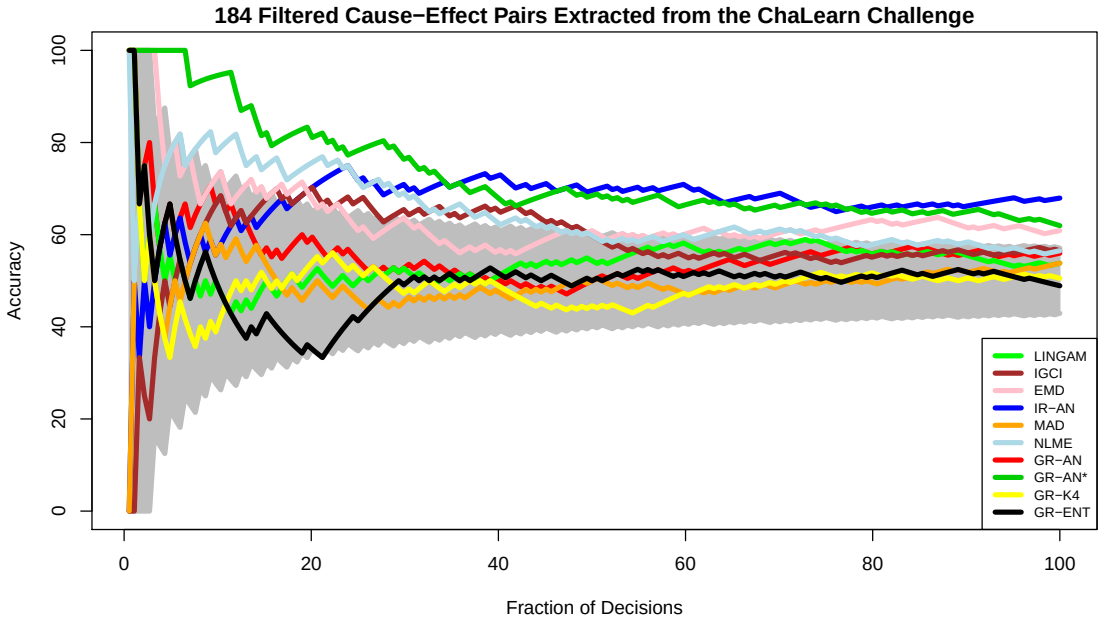

Figure 7: Accuracy of each method, as a fraction of the decisions made, on the 184 filtered cause-effect pairs extracted from the ChaLearn challenge. The number of samples of each pair is equal to 100. Best seen in color.

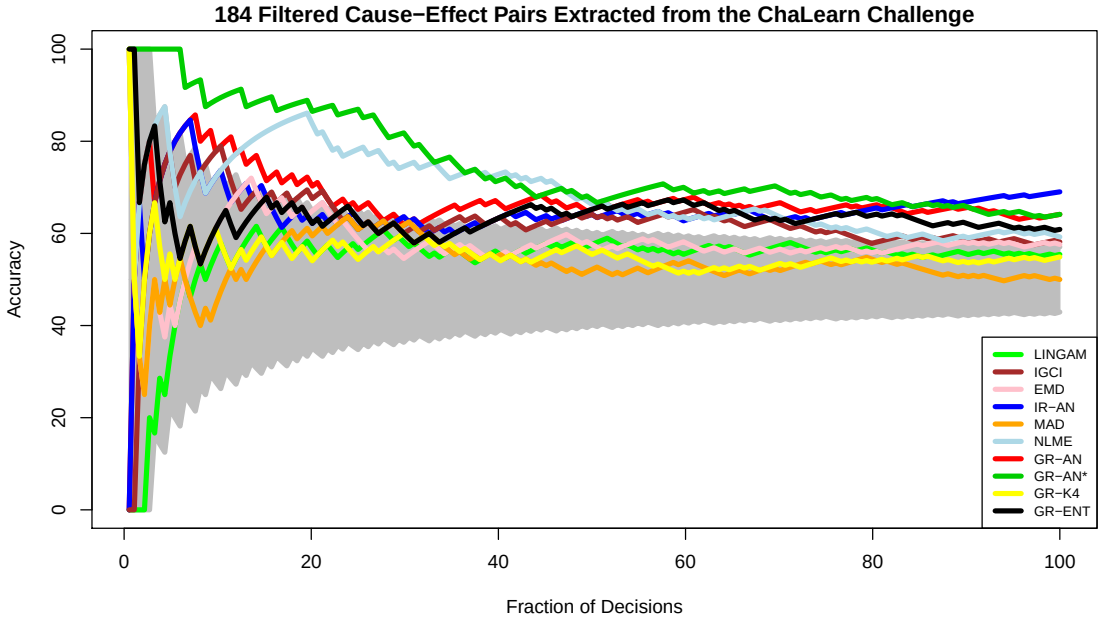

Figure 8: Accuracy of each method, as a fraction of the decisions made, on the 184 filtered cause-effect pairs extracted from the ChaLearn challenge. The number of samples of each pair is equal to 200. Best seen in color.

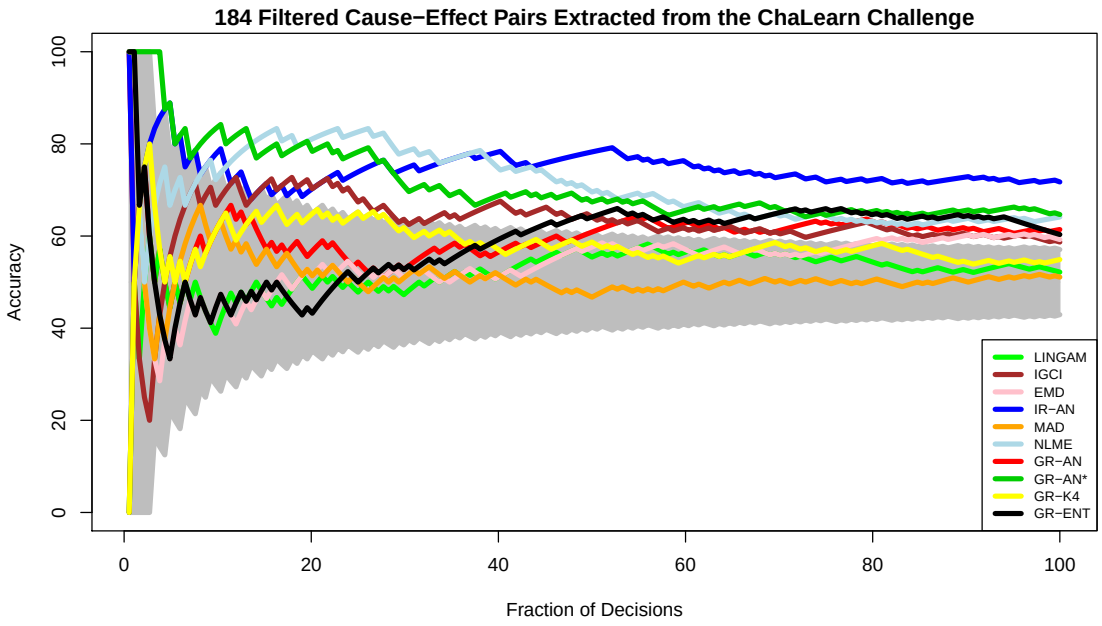

Figure 9: Accuracy of each method, as a fraction of the decisions made, on the 184 filtered cause-effect pairs extracted from the ChaLearn challenge. The number of samples of each pair is equal to 300. Best seen in color.