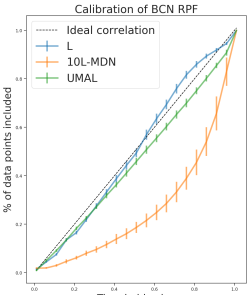
First of all, we would like to thank all reviewers for their suggestions to improve our paper submission.

Reviewers **#1** and **#2** suggest experiments to measure if UMAL yields calibrated outputs. We have performed an additional empirical study to assess this point. We highlight that our system predicts an output distribution $p(y|x, \boldsymbol{w})$ (not a confidence value). In particular, we have computed the % of actual test data that falls into different thresholds of predicted probability. Ideally, given a certain threshold $\theta \in [0, 1]$, the amount of data points with a predicted probability above or equal to $1-\theta$ should be similar to θ. In the left part of Figure 1 we plot these measures for different methods (in green, our model) when considering the BCN RPF dataset. Furthermore, (following reviewers **#1** and **#2** suggestions) we have evaluated calibration quality on the UCI datasets with the same architectures as [2,3]. In the right part of Figure 1 we report the mean absolute error between the empirical measures and the ideal ones for all datasets. As it is shown, the conditional distribution predicted by UMAL has low error values. **Therefore, we can state that UMAL produces proper and calibrated conditional distributions that are especially suitable for heterogeneous problems**.

Figure 1: *Plot with the performance of three different models in terms of calibration. The mean and standard deviation for all folds of the mean absolute error between the predicted calibration and the perfect ideal calibration is represented in the table.*



	BCN	YVC	Housing	Concrete	Energy	Kin8nm	Naval	Power	Protein	Wine	Yacht
N	$.12 \pm .04$	$\mathbf{.04 \pm .01}$	$\mathbf{.08 \pm .04}$	$\mathbf{.03 \pm .01}$	$\mathbf{.03 \pm .01}$	$\mathbf{.02 \pm .01}$	$.39 \pm .02$	$\mathbf{.02 \pm .01}$	$.03 \pm .00$	$\mathbf{.03 \pm .01}$	$\mathbf{.06 \pm .02}$
L	$\mathbf{.03 \pm .00}$	$.06 \pm .01$	$\mathbf{.07 \pm .04}$	$\mathbf{.05 \pm .02}$	$\mathbf{.04 \pm .01}$	$\mathbf{.04 \pm .01}$	$.35 \pm .03$	$.05 \pm .01$	$\mathbf{.02 \pm .00}$	$\mathbf{.06 \pm .02}$	$\mathbf{.07 \pm .02}$
I-QR	$.10 \pm .02$	$.12 \pm .02$	$.20 \pm .05$	$.18 \pm .02$	$.15 \pm .04$	$.17 \pm .01$	$\mathbf{.12 \pm .05}$	$.20 \pm .02$	$.06 \pm .01$	$.19 \pm .03$	$.14 \pm .06$
2N-MDN	$\mathbf{.05 \pm .02}$	$.12 \pm .05$	$\mathbf{.07 \pm .05}$	$\mathbf{.04 \pm .02}$	$\mathbf{.05 \pm .02}$	$\mathbf{.01 \pm .01}$	$.36 \pm .04$	$\mathbf{.03 \pm .01}$	$.06 \pm .01$	$.08 \pm .03$	$.06 \pm .02$
3N-MDN	$.07 \pm .02$	$.14 \pm .04$	$\mathbf{.07 \pm .05}$	$\mathbf{.07 \pm .03}$	$.04 \pm .02$	$\mathbf{.02 \pm .01}$	$.37 \pm .04$	$\mathbf{.03 \pm .01}$	$.11 \pm .01$	$.15 \pm .04$	$\mathbf{.07 \pm .02}$
4N-MDN	$.10 \pm .03$	$.17 \pm .06$	$\mathbf{.09 \pm .05}$	$.10 \pm .03$	$\mathbf{.05 \pm .02}$	$\mathbf{.03 \pm .01}$	$.36 \pm .05$	$\mathbf{.03 \pm .01}$	$.15 \pm .01$	$.18 \pm .03$	$\mathbf{.07 \pm .04}$
10N-MDN	$.19 \pm .04$	$.19 \pm .06$	$.12 \pm .06$	$.22 \pm .06$	$.09 \pm .04$	$.08 \pm .01$	$.33 \pm .05$	$\mathbf{.03 \pm .01}$	$.22 \pm .01$	$.18 \pm .02$	$\mathbf{.09 \pm .05}$
2L-MDN	$.05 \pm .01$	$.09 \pm .01$	$\mathbf{.09 \pm .05}$	$\mathbf{.06 \pm .02}$	$\mathbf{.05 \pm .02}$	$\mathbf{.04 \pm .01}$	$.32 \pm .04$	$.07 \pm .02$	$.07 \pm .00$	$\mathbf{.06 \pm .02}$	$\mathbf{.06 \pm .03}$
3L-MDN	$.08 \pm .02$	$.11 \pm .02$	$\mathbf{.11 \pm .05}$	$.10 \pm .03$	$.08 \pm .03$	$.05 \pm .01$	$.29 \pm .04$	$.08 \pm .02$	$.12 \pm .01$	$.16 \pm .03$	$\mathbf{.06 \pm .02}$
4L-MDN	$.13 \pm .05$	$.12 \pm .03$	$\mathbf{.14 \pm .06}$	$.12 \pm .03$	$\mathbf{.08 \pm .04}$	$.06 \pm .01$	$.31 \pm .04$	$.07 \pm .02$	$.15 \pm .01$	$.15 \pm .02$	$\mathbf{.05 \pm .02}$
10L-MDN	$.24 \pm .03$	$.18 \pm .05$	$.21 \pm .05$	$.18 \pm .04$	$.16 \pm .05$	$.11 \pm .01$	$.27 \pm .06$	$.08 \pm .02$	$.22 \pm .01$	$.17 \pm .01$	$\mathbf{.10 \pm .04}$
I-$\mathcal{A}$LD	$.06 \pm .01$	$\mathbf{.02 \pm .01}$	$\mathbf{.07 \pm .06}$	$\mathbf{.04 \pm .01}$	$\mathbf{.05 \pm .02}$	$\mathbf{.04 \pm .01}$	$.44 \pm .01$	$\mathbf{.04 \pm .01}$	$.07 \pm .00$	$\mathbf{.03 \pm .01}$	$\mathbf{.09 \pm .04}$
UMAL	$\mathbf{.04 \pm .01}$	$.07 \pm .01$	$\mathbf{.10 \pm .05}$	$\mathbf{.07 \pm .04}$	$\mathbf{.06 \pm .02}$	$\mathbf{.02 \pm .01}$	$.43 \pm .01$	$\mathbf{.02 \pm .01}$	$\mathbf{.02 \pm .00}$	$.13 \pm .06$	$\mathbf{.06 \pm .03}$

For the sake of completeness, we also have computed UMAL negative log-likelihood for UCI datasets (see Table 1) following [2]. These results restate that UMAL is always in the best positions. However, it should be noted that most of these databases have a **small sample size** and that in this regime, aleatoric uncertainty cannot be reliably estimated. We hypothesize that a better solution would be to simultaneous estimate epistemic (as [1,2,3]) and aleatoric uncertainty.

Minor comments:

- Quantile Regression allows us to approximate a desired quantile, unlike the classical regression that only estimates the mean or the median. This is useful since we could capture confidence intervals without strong assumptions about the distribution function to approximate (**Rev. #3. Q5.1**).

- IQR is a function, $\phi(\mathbf{x}, \tau)$, that can be evaluated for any real value of $\tau \in (0.1)$ to give us a point-wise estimation of an infinite number of quantiles $(\mu_\tau)_{\tau \in (0,1)}$ (**Rev. #3. Q5.3**). On the other hand, I-$\mathcal{A}$LD is a function, $\phi(\mathbf{x}, \tau)$, that predicts the $(\mu_\tau, b_\tau)_{\tau \in (0,1)}$ parameters of each ALD that is identified with a real asymmetry value, τ. Thus, now each ALD tries to estimate, in a non-point-wise manner, their corresponding quantile (**Rev. #3.Q5.2**).

- If we consider each ALD as a component of a single mixture model we arrive to UMAL and, in turn, we solve the crossing quantiles problem because all ALDs are optimized jointly to produce the output distribution (**Rev. #3. Q5.4**).

- To obtain the distribution predicted by UMAL we will evaluate the learnt function $\psi(\mathbf{x}, \tau)$ for each τ in any discretitzation of its interval of definition, $(0, 1)$. For instance, by considering the partition $sel_\tau = [0.01, 0.02, \ldots, 0.99]$ as parameter of the function defined in Algorithm 3 of the paper (**Rev. #1. Q2.2**).

We are seriously considering the suggestions made by reviewers **#1** and **#2** regarding typos and polishing. Thus, the paper is being revised by several native English speakers to improve its flow.

Table 1: *Comparison of the Negative Log-Likelihood of the test set over different train-test folds proposed in* [3].

	Housing	Concrete	Energy	Kin8nm	Naval	Power	Protein	Wine	Yacht
Normal distribution	$2.76 \pm .34$	$\mathbf{3.20 \pm .16}$	$\mathbf{2.13 \pm .24}$	$\mathbf{-1.15 \pm .03}$	$-3.67 \pm .01$	$\mathbf{2.83 \pm .03}$	$2.84 \pm .03$	$1.05 \pm .14$	$\mathbf{1.86 \pm .31}$
Laplace distribution	$\mathbf{2.59 \pm .20}$	$\mathbf{3.21 \pm .13}$	$\mathbf{2.06 \pm .20}$	$\mathbf{-1.08 \pm .04}$	$\mathbf{-3.73 \pm .04}$	$\mathbf{2.87 \pm .03}$	$2.74 \pm .01$	$1.00 \pm .08$	$\mathbf{1.54 \pm .37}$
Independent QR	10.96 ± 2.4	$10.19 \pm .95$	9.45 ± 1.3	$9.22 \pm .66$	$5.14 \pm .89$	$8.39 \pm .45$	$8.14 \pm .52$	$12.30 \pm .91$	10.32 ± 2.9
2 comp. Normal MDN	$2.74 \pm .30$	$3.25 \pm .21$	$\mathbf{2.02 \pm .30}$	$\mathbf{-1.15 \pm .05}$	$-3.66 \pm .02$	$\mathbf{2.85 \pm .05}$	$2.56 \pm .03$	$1.33 \pm .61$	$\mathbf{1.55 \pm .32}$
3 comp. Normal MDN	$2.68 \pm .28$	$3.64 \pm .28$	$\mathbf{2.30 \pm .43}$	$-1.15 \pm .05$	$-3.66 \pm .01$	$\mathbf{2.85 \pm .04}$	$2.90 \pm .15$	0.69 ± 1.0	$\mathbf{1.54 \pm .52}$
4 comp. Normal MDN	$2.87 \pm .46$	$3.74 \pm .28$	$2.46 \pm .39$	$\mathbf{-1.12 \pm .04}$	$\mathbf{-3.66 \pm .03}$	$\mathbf{2.86 \pm .05}$	$3.32 \pm .11$	$\mathbf{0.52 \pm .90}$	$\mathbf{1.43 \pm .36}$
10 comp. Normal MDN	$3.10 \pm .46$	5.64 ± 1.1	$3.03 \pm .71$	$-0.99 \pm .06$	$-3.64 \pm .03$	$\mathbf{2.86 \pm .04}$	$4.94 \pm .75$	$0.75 \pm .95$	$\mathbf{1.75 \pm .49}$
2 comp. Laplace MDN	$2.61 \pm .23$	$\mathbf{3.28 \pm .14}$	$\mathbf{2.06 \pm .30}$	$\mathbf{-1.10 \pm .04}$	$-3.70 \pm .06$	$2.91 \pm .05$	$2.50 \pm .03$	$0.59 \pm .63$	$\mathbf{1.37 \pm .42}$
3 comp. Laplace MDN	$2.65 \pm .25$	$3.45 \pm .16$	$\mathbf{2.30 \pm .21}$	$\mathbf{-1.09 \pm .03}$	$\mathbf{-3.66 \pm .06}$	$2.95 \pm .04$	$2.65 \pm .06$	$\mathbf{-0.81 \pm .70}$	$\mathbf{1.39 \pm .35}$
4 comp. Laplace MDN	$2.76 \pm .42$	$3.57 \pm .14$	$\mathbf{2.31 \pm .35}$	$\mathbf{-1.10 \pm .05}$	$\mathbf{-3.68 \pm .06}$	$2.93 \pm .04$	$2.79 \pm .08$	$\mathbf{-0.65 \pm .96}$	$\mathbf{1.45 \pm .35}$
10 comp. Laplace MDN	$3.17 \pm .46$	$3.95 \pm .34$	$2.80 \pm .49$	$-0.98 \pm .07$	$-3.62 \pm .10$	$2.96 \pm .05$	$3.46 \pm .12$	$0.52 \pm .74$	$\mathbf{1.63 \pm .34}$
Independent $\mathcal{ALD}$	$2.79 \pm .56$	$3.87 \pm .12$	$\mathbf{2.28 \pm .11}$	$-1.00 \pm .05$	$-2.82 \pm .01$	$\mathbf{2.89 \pm .02}$	$2.68 \pm .01$	$1.01 \pm .07$	$\mathbf{1.78 \pm .41}$
UMAL model	$\mathbf{2.59 \pm .26}$	$3.74 \pm .15$	$\mathbf{2.13 \pm .14}$	$\mathbf{-1.09 \pm .03}$	$-2.81 \pm .01$	$\mathbf{2.85 \pm .03}$	$\mathbf{2.40 \pm .01}$	$\mathbf{0.14 \pm .70}$	$\mathbf{1.41 \pm .38}$

[1] Gal, Y. and Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. ICML, 2016.

[2] Lakshminarayanan, Balaji, et al. Simple and scalable predictive uncertainty estimation using deep ensembles. NIPS, 2017.

[3] J. M. Hernández-Lobato, et al. Probabilistic backpropagation for scalable learning of Bayesian neural networks. ICML, 2015.