Table 1: Generalized Linear Models: Comparison of C2ST scores with a Random Forest (RF) and a Neural Network (NN). For the NN we follow the setup of Lueckmann et al., 2021. Evaluation across seven distinct scenarios on 50 synthetic and 17 real-world datasets. All results within two standard errors of the best average result in each scenario are marked in **bold**. Across all scenarios, both RF and NN classifiers yield quite consistent rankings of model performance with only insubstantial deviations in terms of the big picture. In particular, ICL is consistently among the top-performing approaches under both evaluation metrics. Out of the 14 total scenario–domain combinations (7 scenarios × 2 dataset types), the RF and NN metrics identify the same best-performing model in 12 cases.

Scenario	**Model**	**Synthetic Evaluation**		**Real-World Evaluation**	
		C2ST RF (↓)	C2ST NN (↓)	C2ST RF (↓)	C2ST NN (↓)
Scenario 1	Laplace Approximation	1.000 (± 0.000)	0.998 (± 0.000)	1.000 (± 0.000)	0.998 (± 0.000)
	VI: DiagonalNormal	0.904 (± 0.076)	0.857 (± 0.001)	0.797 (± 0.083)	0.803 (± 0.004)
	VI: MultivariateNormal	**0.750** (± 0.128)	0.780 (± 0.002)	**0.607** (± 0.070)	0.713 (± 0.004)
	VI: Structured Normal	**0.753** (± 0.126)	0.781 (± 0.002)	**0.600** (± 0.070)	0.705 (± 0.004)
	VI: IAF	**0.777** (± 0.122)	0.793 (± 0.002)	0.683 (± 0.132)	0.746 (± 0.006)
	HMC	**0.745** (± 0.130)	0.777 (± 0.002)	**0.595** (± 0.075)	**0.702** (± 0.004)
	ICL (ours)	**0.765** (± 0.123)	**0.712** (± 0.002)	**0.614** (± 0.074)	**0.701** (± 0.004)
Scenario 2	Laplace Approximation	1.000 (± 0.000)	0.998 (± 0.000)	1.000 (± 0.000)	0.998 (± 0.000)
	VI: DiagonalNormal	0.957 (± 0.091)	0.883 (± 0.002)	0.892 (± 0.044)	0.851 (± 0.003)
	VI: MultivariateNormal	0.910 (± 0.131)	0.860 (± 0.002)	**0.820** (± 0.031)	0.815 (± 0.003)
	VI: Structured Normal	0.908 (± 0.119)	0.859 (± 0.002)	**0.824** (± 0.023)	0.817 (± 0.003)
	VI: IAF	0.968 (± 0.063)	0.889 (± 0.001)	0.888 (± 0.067)	0.849 (± 0.004)
	ICL (ours)	**0.839** (± 0.072)	**0.824** (± 0.001)	**0.768** (± 0.033)	**0.789** (± 0.003)
Scenario 3	Laplace Approximation	1.000 (± 0.000)	0.998 (± 0.000)	1.000 (± 0.000)	0.998 (± 0.000)
	VI: DiagonalNormal	0.866 (± 0.101)	0.838 (± 0.002)	0.797 (± 0.083)	0.803 (± 0.004)
	VI: MultivariateNormal	0.656 (± 0.131)	0.733 (± 0.002)	**0.590** (± 0.035)	0.685 (± 0.003)
	VI: Structured Normal	0.653 (± 0.125)	0.731 (± 0.002)	**0.582** (± 0.028)	0.681 (± 0.003)
	VI: IAF	0.751 (± 0.148)	0.780 (± 0.002)	0.673 (± 0.141)	0.741 (± 0.006)
	ICL (ours)	**0.611** (± 0.070)	**0.710** (± 0.001)	**0.576** (± 0.027)	**0.693** (± 0.003)
Scenario 4	Laplace Approximation	1.000 (± 0.000)	0.998 (± 0.000)	1.000 (± 0.000)	0.998 (± 0.000)
	VI: DiagonalNormal	0.968 (± 0.036)	0.889 (± 0.001)	0.916 (± 0.040)	0.863 (± 0.003)
	VI: MultivariateNormal	0.855 (± 0.123)	0.832 (± 0.002)	0.771 (± 0.017)	0.790 (± 0.002)
	VI: Structured Normal	0.847 (± 0.116)	0.828 (± 0.002)	0.769 (± 0.012)	0.789 (± 0.002)
	VI: IAF	0.942 (± 0.077)	0.876 (± 0.001)	0.833 (± 0.069)	0.821 (± 0.004)
	ICL (ours)	**0.753** (± 0.049)	**0.781** (± 0.001)	**0.762** (± 0.015)	**0.786** (± 0.002)
Scenario 5	Laplace Approximation	1.000 (± 0.000)	0.998 (± 0.000)	1.000 (± 0.000)	0.998 (± 0.000)
	VI: DiagonalNormal	0.866 (± 0.085)	0.838 (± 0.002)	0.810 (± 0.036)	0.810 (± 0.003)
	VI: MultivariateNormal	0.765 (± 0.100)	0.787 (± 0.002)	0.711 (± 0.038)	0.760 (± 0.003)
	VI: Structured Normal	0.758 (± 0.098)	0.784 (± 0.002)	0.705 (± 0.032)	0.757 (± 0.003)
	VI: IAF	0.814 (± 0.105)	0.812 (± 0.002)	0.777 (± 0.106)	0.793 (± 0.005)
	ICL (ours)	**0.621** (± 0.063)	**0.715** (± 0.001)	**0.610** (± 0.045)	**0.710** (± 0.003)
Scenario 6	Laplace Approximation	1.000 (± 0.000)	0.998 (± 0.000)	1.000 (± 0.000)	0.998 (± 0.000)
	VI: DiagonalNormal	0.724 (± 0.060)	0.767 (± 0.001)	0.703 (± 0.039)	0.756 (± 0.003)
	VI: MultivariateNormal	**0.534** (± 0.018)	**0.672** (± 0.001)	**0.538** (± 0.019)	0.674 (± 0.002)
	VI: Structured Normal	**0.536** (± 0.016)	**0.673** (± 0.001)	**0.536** (± 0.019)	0.673 (± 0.002)
	VI: IAF	0.542 (± 0.026)	0.676 (± 0.001)	**0.535** (± 0.015)	0.672 (± 0.002)
	ICL (ours)	**0.532** (± 0.019)	**0.671** (± 0.001)	0.556 (± 0.017)	**0.653** (± 0.002)
Scenario 7	Laplace Approximation	1.000 (± 0.000)	0.998 (± 0.000)	1.000 (± 0.000)	0.998 (± 0.000)
	VI: DiagonalNormal	0.938 (± 0.074)	0.874 (± 0.001)	0.936 (± 0.024)	0.873 (± 0.003)
	VI: MultivariateNormal	0.814 (± 0.181)	0.812 (± 0.002)	**0.741** (± 0.020)	0.775 (± 0.003)
	VI: Structured Normal	0.824 (± 0.177)	0.817 (± 0.002)	**0.734** (± 0.025)	0.772 (± 0.003)
	VI: IAF	0.939 (± 0.091)	0.874 (± 0.002)	0.864 (± 0.093)	0.837 (± 0.005)
	ICL (ours)	**0.700** (± 0.116)	**0.721** (± 0.002)	0.773 (± 0.048)	**0.751** (± 0.003)

Table 2: Factor Analysis: Comparison of C2ST scores using a Random Forest (RF) and a Neural Network (NN) classifier across six different scenarios on 50 synthetic and 17 real-world datasets. For the NN we follow the setup of Lueckmann et al., 2021. All results within two standard errors of the best average result in each scenario are marked in **bold**. Across both classifiers and dataset types, the ICL method consistently achieves top-tier performance, often outperforming standard variational approximations and the Laplace baseline. The RF and NN classifiers yield largely consistent rankings, identifying ICL as the best-performing model in the majority of cases.

Scenario	**Model**	**Synthetic Evaluation**		**Real-World Evaluation**	
		C2ST RF (↓)	C2ST NN (↓)	C2ST RF (↓)	C2ST NN (↓)
Scenario 1	Laplace Approximation	1.000 (± 0.000)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: DiagonalNormal	1.000 (± 0.001)	0.997 (± 0.000)	0.979 (± 0.008)	0.950 (± 0.001)
	VI: MultivariateNormal	0.998 (± 0.003)	0.960 (± 0.000)	0.966 (± 0.010)	0.944 (± 0.001)
	VI: Structured Normal	0.997 (± 0.004)	0.959 (± 0.000)	0.979 (± 0.010)	0.950 (± 0.001)
	VI: IAF	0.953 (± 0.104)	0.937 (± 0.001)	0.849 (± 0.075)	0.885 (± 0.003)
	ICL (ours)	**0.552** (± 0.028)	**0.737** (± 0.000)	**0.606** (± 0.038)	**0.764** (± 0.001)
Scenario 2	Laplace Approximation	1.000 (± 0.000)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: DiagonalNormal	0.998 (± 0.002)	0.960 (± 0.000)	0.975 (± 0.010)	0.948 (± 0.001)
	VI: MultivariateNormal	0.989 (± 0.009)	0.955 (± 0.000)	0.951 (± 0.025)	0.936 (± 0.001)
	VI: Structured Normal	0.984 (± 0.031)	0.953 (± 0.000)	0.958 (± 0.025)	0.940 (± 0.001)
	VI: IAF	0.966 (± 0.066)	0.944 (± 0.001)	0.799 (± 0.058)	0.860 (± 0.002)
	ICL (ours)	**0.542** (± 0.006)	**0.732** (± 0.000)	**0.622** (± 0.032)	**0.772** (± 0.001)
Scenario 3	Laplace Approximation	1.000 (± 0.000)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: DiagonalNormal	0.999 (± 0.002)	0.960 (± 0.000)	0.951 (± 0.007)	0.936 (± 0.001)
	VI: MultivariateNormal	0.994 (± 0.007)	0.958 (± 0.000)	0.945 (± 0.007)	0.933 (± 0.001)
	VI: Structured Normal	0.997 (± 0.003)	0.959 (± 0.000)	0.942 (± 0.009)	0.932 (± 0.001)
	VI: IAF	0.990 (± 0.011)	0.987 (± 0.000)	0.928 (± 0.015)	0.925 (± 0.001)
	ICL (ours)	**0.537** (± 0.023)	**0.729** (± 0.000)	**0.609** (± 0.019)	**0.765** (± 0.001)
Scenario 4	Laplace Approximation	1.000 (± 0.000)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: DiagonalNormal	1.000 (± 0.000)	0.997 (± 0.000)	0.977 (± 0.003)	0.949 (± 0.000)
	VI: MultivariateNormal	0.999 (± 0.001)	0.960 (± 0.000)	0.973 (± 0.008)	0.947 (± 0.001)
	VI: Structured Normal	1.000 (± 0.000)	0.997 (± 0.000)	0.973 (± 0.007)	0.947 (± 0.001)
	VI: IAF	0.999 (± 0.001)	0.960 (± 0.000)	**0.961** (± 0.018)	0.941 (± 0.001)
	ICL (ours)	**0.684** (± 0.060)	**0.803** (± 0.001)	0.988 (± 0.003)	**0.955** (± 0.000)
Scenario 5	Laplace Approximation	1.000 (± 0.000)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: DiagonalNormal	0.999 (± 0.002)	0.960 (± 0.000)	0.944 (± 0.010)	0.933 (± 0.001)
	VI: MultivariateNormal	0.995 (± 0.007)	0.958 (± 0.000)	0.930 (± 0.017)	0.926 (± 0.001)
	VI: Structured Normal	0.998 (± 0.005)	0.960 (± 0.000)	0.934 (± 0.011)	0.928 (± 0.001)
	VI: IAF	0.992 (± 0.012)	0.957 (± 0.000)	0.910 (± 0.011)	0.916 (± 0.001)
	ICL (ours)	**0.535** (± 0.016)	**0.728** (± 0.000)	**0.886** (± 0.017)	**0.904** (± 0.001)
Scenario 6	Laplace Approximation	1.000 (± 0.000)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: DiagonalNormal	0.998 (± 0.002)	0.960 (± 0.000)	0.949 (± 0.008)	0.935 (± 0.001)
	VI: MultivariateNormal	0.991 (± 0.013)	0.956 (± 0.000)	0.938 (± 0.009)	0.930 (± 0.001)
	VI: Structured Normal	0.997 (± 0.005)	0.959 (± 0.000)	0.944 (± 0.006)	0.933 (± 0.001)
	VI: IAF	0.989 (± 0.029)	0.955 (± 0.000)	0.865 (± 0.027)	0.893 (± 0.001)
	ICL (ours)	**0.543** (± 0.021)	**0.732** (± 0.000)	**0.666** (± 0.020)	**0.794** (± 0.001)

Table 3: Gaussian Mixture Models: Comparison of C2ST scores using a Random Forest (RF) and a Neural Network (NN) classifier across six distinct scenarios on 50 synthetic and 17 real-world datasets. All results within two standard errors of the best average result in each scenario are marked in **bold**. For the NN we follow the setup of Lueckmann et al., 2021. Both RF and NN classifiers yield consistent rankings, with ICL emerging as the top method in scenarios with more pronounced model mismatch.

Scenario	**Model**	**Synthetic Evaluation**		**Real-World Evaluation**	
		C2ST RF (↓)	C2ST NN (↓)	C2ST RF (↓)	C2ST NN (↓)
Scenario 1	Laplace Approximation	1.000 (± 0.000)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: DiagonalNormal	0.988 (± 0.013)	1.012 (± 0.000)	0.995 (± 0.006)	0.996 (± 0.001)
	VI: MultivariateNormal	0.988 (± 0.013)	1.012 (± 0.000)	0.994 (± 0.007)	0.993 (± 0.001)
	VI: Structured Normal	0.987 (± 0.015)	0.982 (± 0.000)	0.993 (± 0.009)	0.992 (± 0.001)
	VI: IAF	0.989 (± 0.013)	0.983 (± 0.000)	0.995 (± 0.010)	0.996 (± 0.001)
	ICL (ours)	**0.760** (± 0.092)	**0.825** (± 0.001)	**0.847** (± 0.082)	**0.869** (± 0.003)
Scenario 2	Laplace Approximation	1.000 (± 0.000)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: DiagonalNormal	0.989 (± 0.024)	0.983 (± 0.000)	0.998 (± 0.003)	0.997 (± 0.001)
	VI: MultivariateNormal	0.991 (± 0.021)	0.991 (± 0.000)	0.999 (± 0.002)	1.002 (± 0.001)
	VI: Structured Normal	0.992 (± 0.017)	0.988 (± 0.000)	0.999 (± 0.002)	1.002 (± 0.001)
	VI: IAF	0.992 (± 0.021)	0.988 (± 0.000)	0.998 (± 0.004)	0.997 (± 0.001)
	ICL (ours)	**0.812** (± 0.061)	**0.851** (± 0.001)	**0.937** (± 0.041)	**0.915** (± 0.002)
Scenario 3	Laplace Approximation	1.000 (± 0.000)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: DiagonalNormal	**0.996** (± 0.011)	1.004 (± 0.000)	**0.992** (± 0.018)	0.988 (± 0.001)
	VI: MultivariateNormal	0.997 (± 0.009)	1.007 (± 0.000)	**0.993** (± 0.016)	0.992 (± 0.001)
	VI: Structured Normal	**0.995** (± 0.017)	0.996 (± 0.000)	**0.993** (± 0.016)	0.992 (± 0.001)
	VI: IAF	**0.994** (± 0.018)	0.993 (± 0.000)	**0.993** (± 0.017)	0.992 (± 0.001)
	ICL (ours)	1.000 (± 0.000)	**0.997** (± 0.000)	1.000 (± 0.000)	**0.997** (± 0.000)
Scenario 4	Laplace Approximation	1.000 (± 0.000)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: DiagonalNormal	**1.000** (± 0.002)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: MultivariateNormal	**1.000** (± 0.002)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	VI: Structured Normal	**1.000** (± 0.001)	0.997 (± 0.000)	**0.996** (± 0.016)	1.004 (± 0.001)
	VI: IAF	**1.000** (± 0.002)	0.997 (± 0.000)	1.000 (± 0.000)	0.997 (± 0.000)
	ICL (ours)	1.000 (± 0.000)	**0.997** (± 0.000)	**1.000** (± 0.000)	**0.997** (± 0.000)