[논문 리뷰]

Distribution-Free Multivariate Density Forecast

Jungyoon Song
Financial Risk Engineering Lab
Department of Industrial Engineering
Seoul National University

DEC 16, 2024

1. DAN-NFN [1]
2. JDAN-NFN [2]

DAN-NFN

Problem Description of Density Forecast

- A positive-weighted ANN, named distribution approximation network (DAN)
- The other ANN, named network forecast network (NFN), is built as the forecaster of DAN
- DAN-NFN is trained through maximum likelihood estimation (MLE)

$$f(y_{t+k}|X_t) = \begin{cases} w_0 \cdot \delta(y_{t+k}), & y_{t+k} = 0 \\ \bar{f}(y_{t+k}), & y_{t+k} \in (0,1) \\ w_1 \cdot \delta(y_{t+k} - 1), & y_{t+k} = 1 \end{cases} \quad (1)$$

$$\mathbf{\Lambda} = \left\{ \begin{array}{c} \mathbf{\Psi}|_{X_t} \text{ with} \\ \text{PDF } f(\cdot|X_t) \end{array} \middle| \begin{array}{l} w_0 \in [0,1], w_1 \in [0,1], \\ w_0 + w_1 \in [0,1], \\ \bar{f}(\cdot) \text{ is nonnegative on } (0,1), \\ \bar{f}(\cdot) \text{ is bounded on } (0,1), \\ \int_0^1 \bar{f}(y)\,dy = 1 - w_0 - w_1. \end{array} \right\} \quad (3)$$

$$\mathbf{y} = s\{\mathbf{w}_{H+1} \cdots s[\mathbf{w}_2 \cdot s(\mathbf{w}_1 \cdot \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2] \cdots + \mathbf{b}_{H+1}\} \quad (10)$$

$$F(y_{t+k}|X_t) = \begin{cases} w_0, & y_{t+k} = 0 \\ w_0 + \bar{F}(y_{t+k}), & y_{t+k} \in (0,1) \\ 1, & y_{t+k} = 1 \end{cases} \quad (2)$$

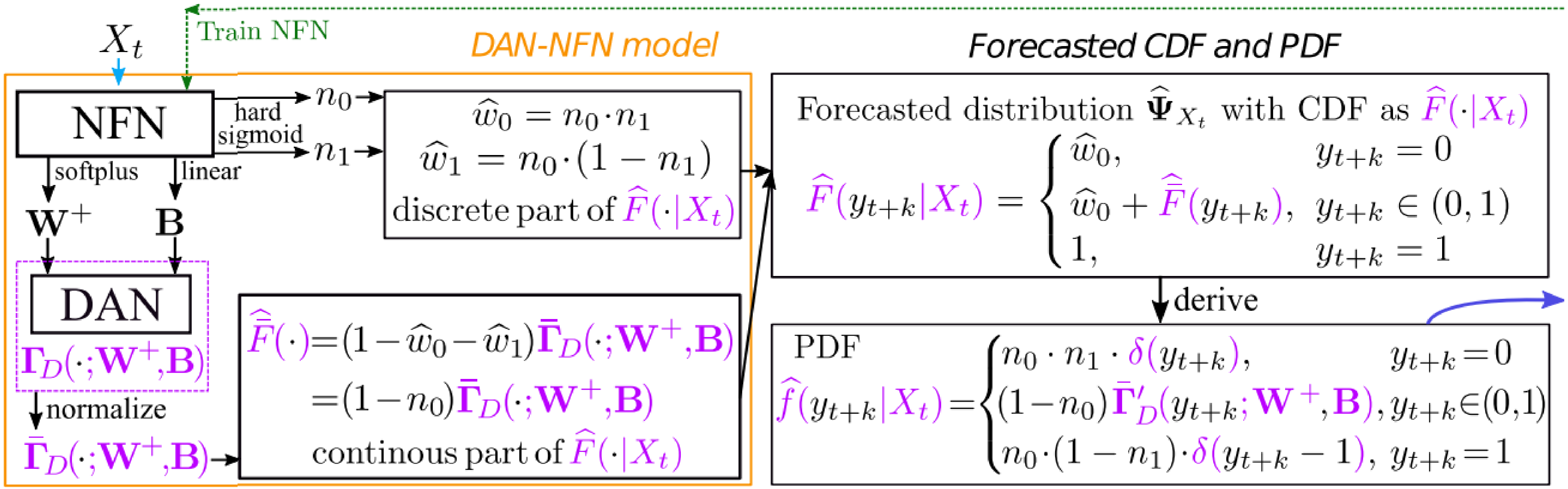
$$\mathbf{\Lambda} = \left\{ \begin{array}{c} \mathbf{\Psi}|_{X_t} \text{ with} \\ \text{CDF } F(\cdot|X_t) \end{array} \middle| \begin{array}{l} w_0 \in [0,1], w_1 \in [0,1], \\ w_0 + w_1 \in [0,1], \\ \lim_{y\to 0} \bar{F}(y) = 0, \\ \lim_{y\to 1} \bar{F}(y) = 1 - w_0 - w_1, \\ \bar{F}(\cdot) \text{ is monotone nondecreasing} \\ \quad \text{and continuous on } (0,1) \end{array} \right\}. \quad (4)$$

$$\mathbf{y} = \mathbf{\Gamma}(\mathbf{x}; [\mathbf{w}_*], [\mathbf{b}_*]) \quad (11)$$

DAN-NFN

Framework of DAN-NFN

- NFN 에서 DAN의 W^+, B, n_0, n_1 출력, DAN에서 입력 받은 weight를 이용하여 monotone non-decreasing CDF로 활용



$$\bar{\mathbf{\Gamma}}_D(y; \mathbf{W}^+, \mathbf{B}) = \frac{\mathbf{\Gamma}_D(y; \mathbf{W}^+, \mathbf{B}) - \mathbf{\Gamma}_D(L_y; \mathbf{W}^+, \mathbf{B})}{\mathbf{\Gamma}_D(U_y; \mathbf{W}^+, \mathbf{B}) - \mathbf{\Gamma}_D(L_y; \mathbf{W}^+, \mathbf{B})} \quad (16)$$

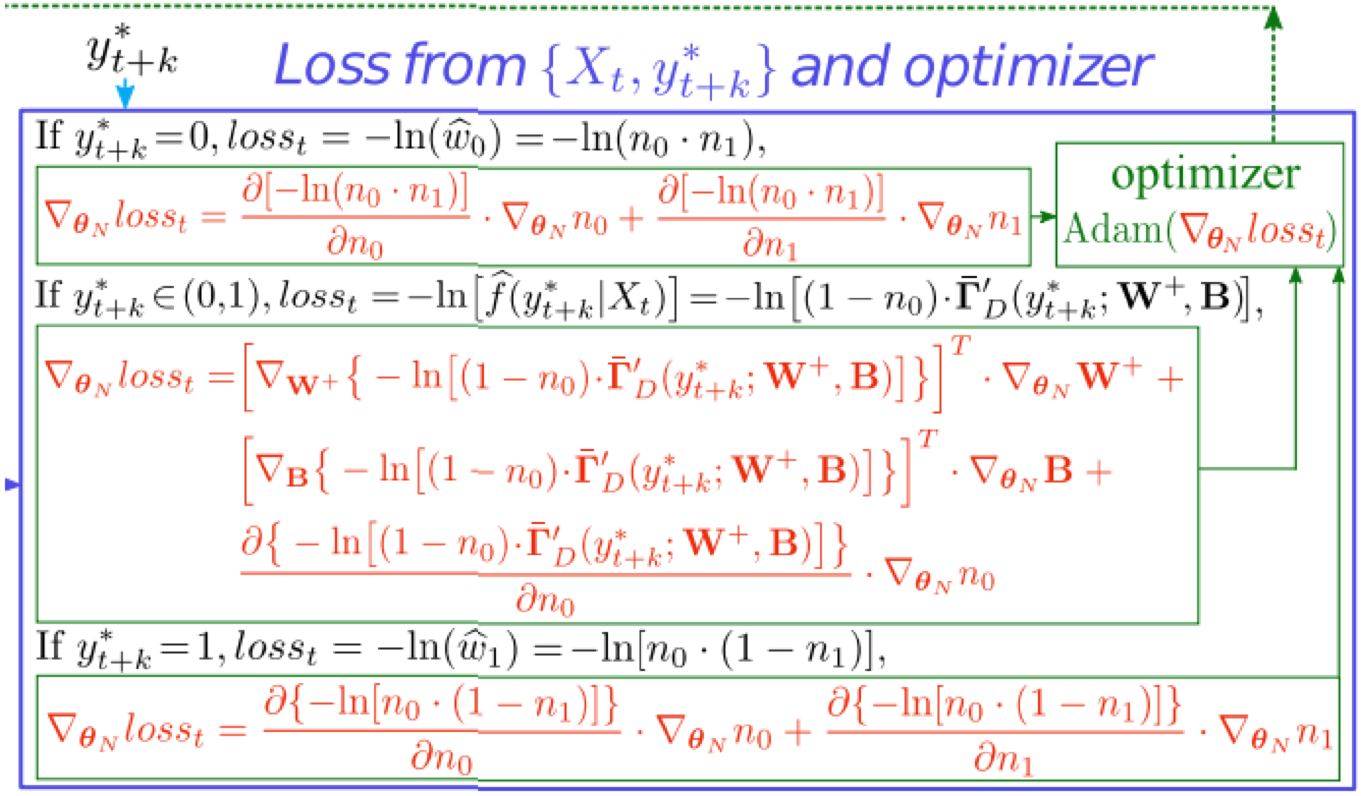
DAN-NFN

Framework of DAN-NFN

- CDF를 PDF로 변환 가능하고, 이를 MLE기반 NLL을 통해 최적화
- $NLL = \sum^{T} -ln\hat{f}(y_{t+k}|X_t)$

PDF

$$\hat{f}(y_{t+k}|X_t) = \begin{cases} n_0 \cdot n_1 \cdot \delta(y_{t+k}), & y_{t+k}=0 \\ (1-n_0)\bar{\Gamma}'_D(y_{t+k};\mathbf{W}^+,\mathbf{B}), & y_{t+k}\in(0,1) \\ n_0\cdot(1-n_1)\cdot\delta(y_{t+k}-1), & y_{t+k}=1 \end{cases}$$



DAN-NFN

Architecture of DAN-NFN

- Time series를 다룰 수 있도록 LSTM Layer 사용
- Gradient diffusion을 방지하기 위한 residual architecture

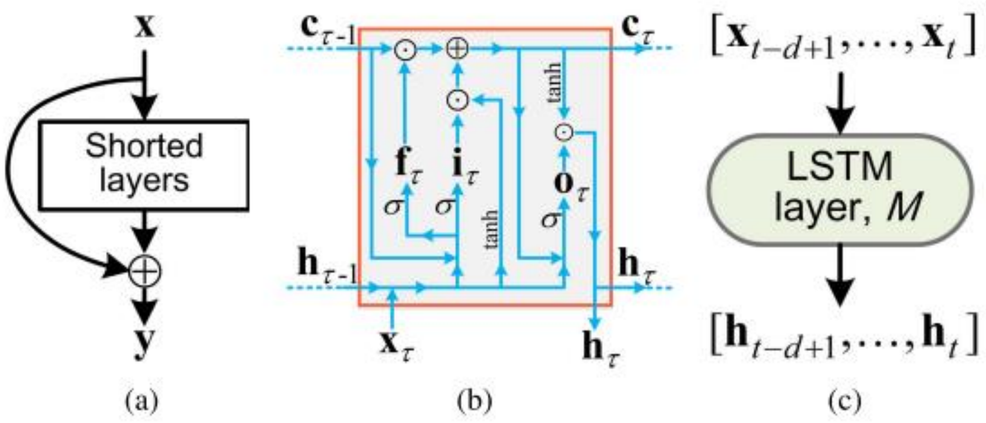


Fig. 2. Residual architecture and LSTM. (a) Structure of the residual network. (b) Information flow in LSTM. (c) LSTM layer.

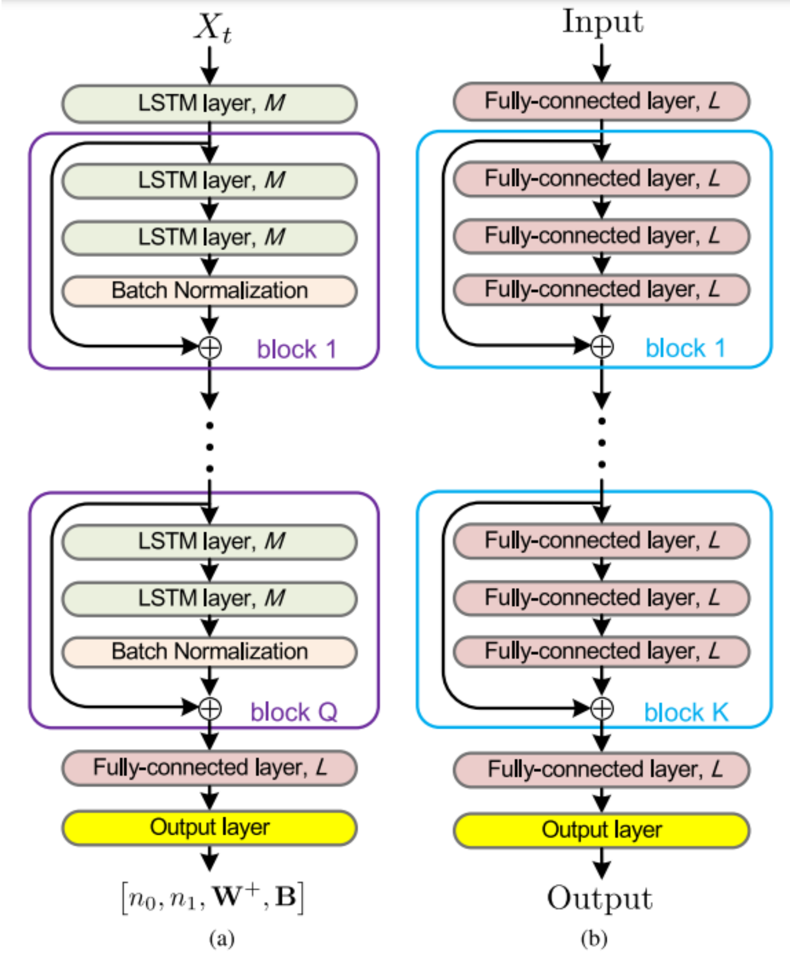


Fig. 3. Structures of NFN and DAN. (a) NFN. (b) DAN.

JDAN-NFN

Problem Description of Multivariate Density Forecast

- A joint distribution approximation network (JDAN) is a generic framework for approximating real continuous joint CDFs through a deep NN
- NFN outputs all parameters of JDAN
- JDAN-NFN is trained through maximum likelihood estimation (MLE)

Let $\boldsymbol{y}_{t+\tau}$ be a D-dimensional random vector, then $f(\cdot)$ should satisfy the following conditions:

$$\begin{cases} f(\cdot) \quad \text{is nonnegative and bounded,} \\ \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(\boldsymbol{y}) dy^1 \cdots dy^D = 1. \end{cases} \qquad (1)$$

For $F(\cdot)$, it can be written as

$$F(\boldsymbol{y}_{t+\tau}) = \int_{-\infty}^{y^1_{t+\tau}} \cdots \int_{-\infty}^{y^D_{t+\tau}} f(\boldsymbol{y}) dy^1 \cdots dy^D. \qquad (2)$$

$$\begin{cases} \text{(i) } F(\cdot) \text{ is continuously differentiable,} \\ \text{(ii) } \frac{\partial^D (F(\boldsymbol{y}))}{\partial y^1 \cdots \partial y^D} \Big|_{\boldsymbol{y}=\boldsymbol{y}_{t+\tau}} \geq 0, \\ \text{(iii) } \lim\limits_{\boldsymbol{y}_{t+\tau} \to +\infty} F(\boldsymbol{y}_{t+\tau}) = 1, \\ \text{(iv) } \lim\limits_{y^d_{t+\tau} \to -\infty} F(\boldsymbol{y}_{t+\tau}) = 0, \forall d \in [1, D]. \end{cases} \qquad (3)$$

JDAN-NFN

Framework of JDAN-NFN

- NFN은 JDAN의 parameters인 W^+, B, C 출력
- JDAN은 Tensor W^+, B 으로 D차원의 CDFs 생성
- 각각 CDFs에 대한 correlation term을 C $(\binom{D}{2}\ dim)$로 표현 (time-variant correlation)
- JDAN-NFN 의 Ψ 는 DAN-NFN의 Γ 와 동일한 의미

$$\Psi^{\mathcal{C}} = \frac{1}{\binom{D}{2}} \cdot \sum_{i>d}^{D} \sum_{d=1}^{D-1} [\mathbf{C}_{di} \cdot (1 - \overline{\Psi}^{d}) \cdot (1 - \overline{\Psi}^{i}) + 1], \quad (10)$$

$$\Psi_{\mathcal{J}}(\boldsymbol{y}_{t+\tau}; \mathbf{W}^{+}, \mathbf{B}, \mathbf{C}) = \Psi^{\mathcal{C}} \cdot \prod_{d=1}^{D} \overline{\Psi}^{d}. \quad (11)$$

- Loss function $L(\boldsymbol{\theta}_N, \boldsymbol{X}_t, \boldsymbol{y}^*_{t+\tau}) = -\ln[\widehat{f}(\boldsymbol{y}^*_{t+\tau}|\boldsymbol{X}_t)],$

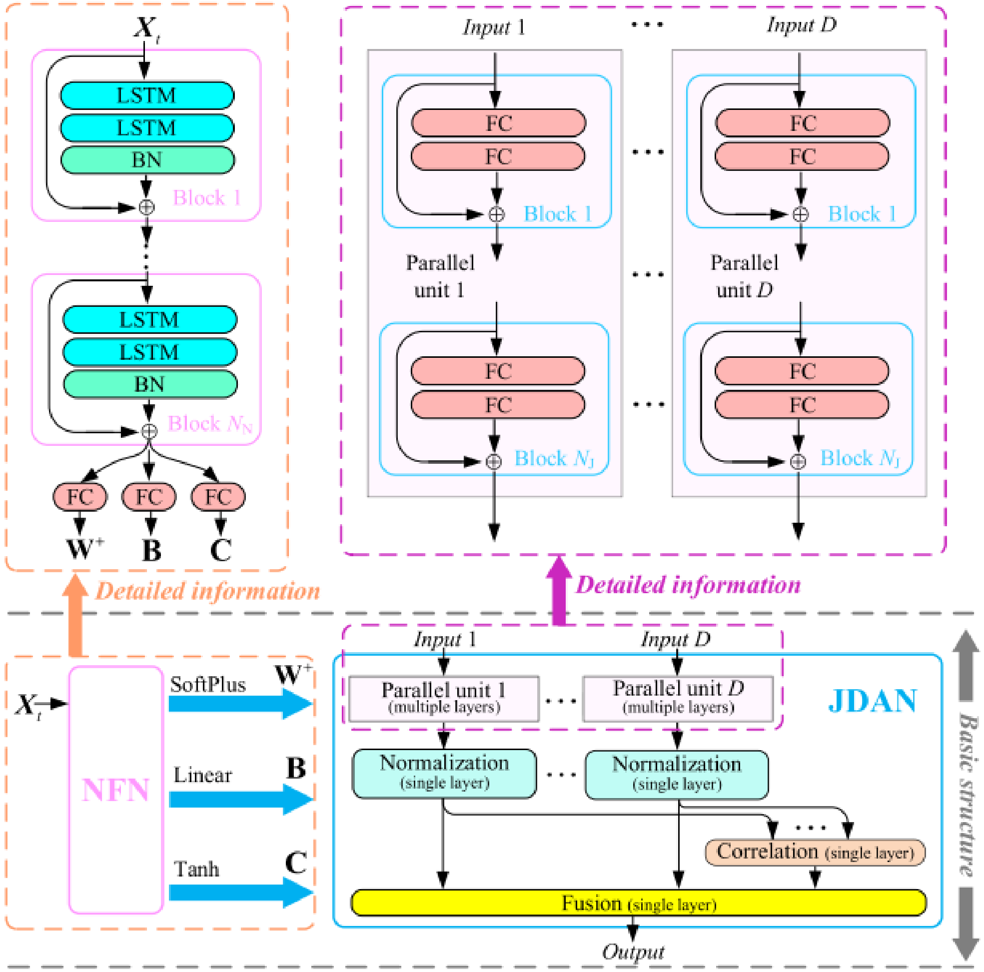


Fig. 2. JDAN-NFN model.

JDAN-NFN

Constraints of JDAN-NFN

- CDF of JDAN-NQF

$$\Psi^{\mathcal{C}} = \frac{1}{\binom{D}{2}} \cdot \sum_{i>d}^{D} \sum_{d=1}^{D-1} [\mathbf{C}_{di} \cdot (1 - \overline{\Psi}^{d}) \cdot (1 - \overline{\Psi}^{i}) + 1], \quad (10)$$

$$\Psi_{\mathcal{J}}(\boldsymbol{y}_{t+\tau}; \mathbf{W}^{+}, \mathbf{B}, \mathbf{C}) = \Psi^{\mathcal{C}} \cdot \prod_{d=1}^{D} \overline{\Psi}^{d}. \quad (11)$$

- Nonnegativity of JDAN-NQF (d=2), d>2는 귀납적 증명

$$\Psi_{\mathcal{J}}^{1,2} = \overline{\Psi}^{1} \cdot \overline{\Psi}^{2} \cdot [\mathbf{C}_{12} \cdot (1 - \overline{\Psi}^{1}) \cdot (1 - \overline{\Psi}^{2}) + 1]. \quad (40)$$

$$\frac{\partial^2 (\Psi_{\mathcal{J}}^{1,2})}{\partial y^1 \partial y^2} = \frac{\partial \overline{\Psi}^{1}}{\partial y^1} \cdot \frac{\partial \overline{\Psi}^{2}}{\partial y^2} \cdot \left[\mathbf{C}_{12} \cdot (1 - 2\overline{\Psi}^{1}) \cdot (1 - 2\overline{\Psi}^{2}) + 1\right]. \quad (41)$$

- Limits of JDAN-NQF

$$\lim_{\boldsymbol{y}_{t+\tau} \to +\infty} \Psi_{\mathcal{J}}(\boldsymbol{y}_{t+\tau}; \mathbf{W}^{+}, \mathbf{B}, \mathbf{C})$$
$$= \lim_{\boldsymbol{y}_{t+\tau} \to +\infty} \Psi^{\mathcal{C}} \cdot \left(\prod_{d=1}^{D} \lim_{y_{t+\tau}^{i} \to +\infty} \overline{\Psi}^{d} \right) = 1, \quad (13)$$

$$\lim_{y_{t+\tau}^{d} \to -\infty} \Psi_{\mathcal{J}}(\boldsymbol{y}_{t+\tau}; \mathbf{W}^{+}, \mathbf{B}, \mathbf{C}) = 0, \forall d \in [1, D]. \quad (14)$$

JDAN-NFN

Evaluation measure

- Reliability

$$b_{\tau}^{\alpha_{j,d}} = \alpha_{j,d} - \frac{1}{N}\sum_{i=1}^{N} H(\hat{q}_{i+\tau|i}^{\alpha_{j,d}} - y_{i+\tau}^{d*}),$$

- Sharpness

$$\delta_{\tau}^{\alpha_{j,d}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{q}_{i+\tau|i}^{1-\alpha_{j,d}/2} - \hat{q}_{i+\tau|i}^{\alpha_{j,d}/2}),$$

- Skill score

$$S_{t+\tau|t}^{d} = \sum_{j=1}^{J}\{[H(\hat{q}_{t+\tau|t}^{\alpha_{j,d}} - y_{t+\tau}^{d*}) - \alpha_{j,d}](y_{t+\tau}^{d*} - \hat{q}_{t+\tau|t}^{\alpha_{j,d}})\},$$

- Variogram score

$$VS_{t+\tau|t} = \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}(|y_{t+\tau}^{i*} - y_{t+\tau}^{j*}|^{p} - E_{\hat{\Phi}}|Y_i - Y_j|^{p})^2$$

$$\approx \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}(|y_{t+\tau}^{i*} - y_{t+\tau}^{j*}|^{p} - \frac{1}{m}\sum_{k=1}^{m}|Y_i^{(k)} - Y_j^{(k)}|^{p})^2,$$

JDAN-NFN

Dataset

TABLE I
DETAILS OF THE THREE DATA SETS

	Case $\mathcal{A}$	Case $\mathcal{B}$	Case $\mathcal{C}$
Scenario	wind speed	wind power	electricity load
Location	$34°12'23.04''$N, $102°44'32.64''$W Texas, USA	Brown Hill Wind Farm, $P_c = 94.5$MW Hallett Hill Wind Farm, $P_c = 71.4$MW North Brown Hill Wind Farm, $P_c = 132.3$MW Australia	Queensland, Australia
Covering period	2019/12/31 to 2020/12/31	2019/12/31 to 2020/12/31	2009/12/31 to 2020/12/31
Resolution	5 minutes	5 minutes	2 hours
$\boldsymbol{X}_t$	historical wind speed series in different directions, historical temperature series, historical relative humidity series, time of the day	historical wind power series of different wind farms, time of the day	historical load series, historical electricity price series, day of the week, time of the day
$\boldsymbol{y}_{t+\tau}$	speed vector of next 5 minutes	power vector of next 5 minutes	load vector of the next day
D	2	3	12

JDAN-NFN

Results

TABLE III
PERFORMANCE DEMONSTRATION

	Case $\mathcal{A}$				Case $\mathcal{B}$				Case $\mathcal{C}$			
	$\overline{b_\tau}\%$	$\overline{\delta_\tau}$	$\overline{S_\tau}$	$\overline{VS_\tau}$	$\overline{b_\tau}\%$	$\overline{\delta_\tau}$	$\overline{S_\tau}$	$\overline{VS_\tau}$	$\overline{b_\tau}\%$	$\overline{\delta_\tau}$	$\overline{S_\tau}$	$\overline{VS_\tau}$
Persistence	4.64	0.0592	−1.034	0.0529	4.78	0.0676	−1.556	0.1228	3.92	0.0885	−1.938	0.9732
VARX	2.40	0.0468	−0.860	0.0395	4.32	0.0772	−1.390	0.0827	3.67	0.1105	−1.887	0.7650
VCG-t-copula	2.89	0.0463	−0.846	0.0227	2.39	0.0683	−1.387	0.0789	3.06	0.0911	−1.884	0.6473
WKDE-R-copula	2.04	0.0483	−0.822	0.0121	2.61	0.0905	−1.370	0.0469	2.26	0.1087	−1.746	0.3407
Gaussian-E-copula	1.95	0.0350	−0.817	0.0061	2.15	0.0642	−1.338	0.0356	2.92	0.0731	−1.458	0.4145
AMISE-MKDE	1.99	0.0559	−0.833	0.0084	2.60	0.0915	−1.383	0.0454	3.26	0.1073	−1.826	0.5405
MQR	2.59	0.0413	−0.810	/	2.55	0.0627	−1.212	/	3.48	**0.0705**	−1.453	/
QR-E-copula	2.51	0.0339	−0.643	/	2.72	0.0596	−1.336	/	3.12	0.0706	−1.448	/
JDAN-NFN	**0.79**	**0.0280**	**−0.511**	**0.0050**	**1.40**	**0.0521**	**−0.8979**	**0.0269**	**2.12**	0.0787	**−1.365**	**0.3167**

denotes parametric approaches; denotes nonparametric approaches

REFERENCE

[1] Hu, T., Guo, Q., Li, Z., Shen, X., & Sun, H. (2019). Distribution-free probability density forecast through deep neural networks. IEEE transactions on neural networks and learning systems, 31(2), 612-625.

[2] Meng, Z., Guo, Y., Tang, W., & Sun, H. (2022). Nonparametric multivariate probability density forecast in smart grids with deep learning. IEEE Transactions on Power Systems, 38(5), 4900-4915.