Generative Time Series Forecasting with Diffusion, Denoise, and Disentanglement

Yan Li[*], Xinjiang Lu[‡], Yaqing Wang[‡], Dejing Dou[‡]

[‡]Baidu Research
[*] Zhejiang University

Challenges in Time Series Forecasting

- Uncertainty issues of deep neural networks
  - CNN-, RNN-, and Transformer-based models
  - Generative models can deliver more reliable predictions, e.g., VAR (but not satisfactory)

- Interpretable prediction
  - The interpretability of VAEs might be lowered due to the entangled latent variables
  - Disentangled representation learning is of importance for time series forecasting

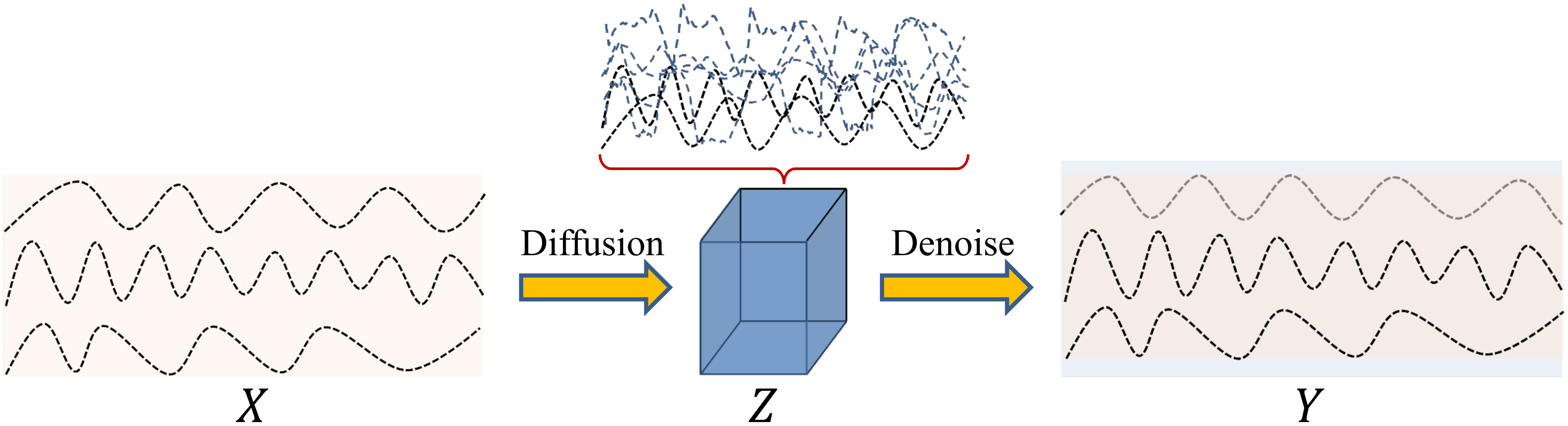
- Real-world time series data are often recorded in a short time period
  - Overfitting and generalization issues

Generative Time Series Forecasting

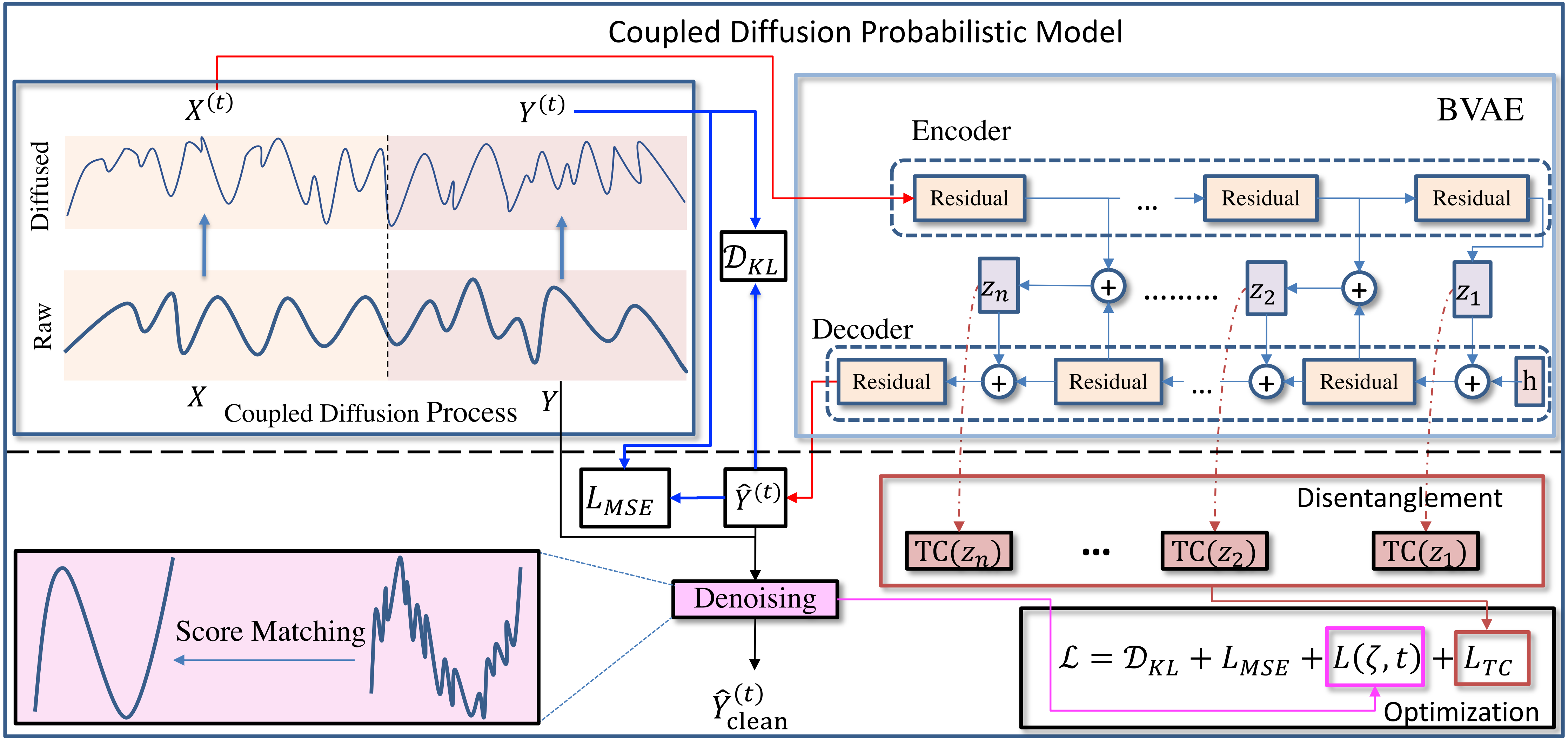
- Given an input time series $X = \{x_1, \cdots, x_n\}$ and the corresponding output series $Y = \{y_{n+1}, \cdots, x_{n+m}\}$, we assume that $Z \to Y$ $(Z \sim p(Z|X))$ and $p_\phi(Z|X) = g_\phi(X)$. Then, the data density of Y is given by:

$$p_\theta(Y) = \int_{\Omega_Z} p_\phi(Z|X)\big(Y - f_\theta(Z)\big)dZ$$

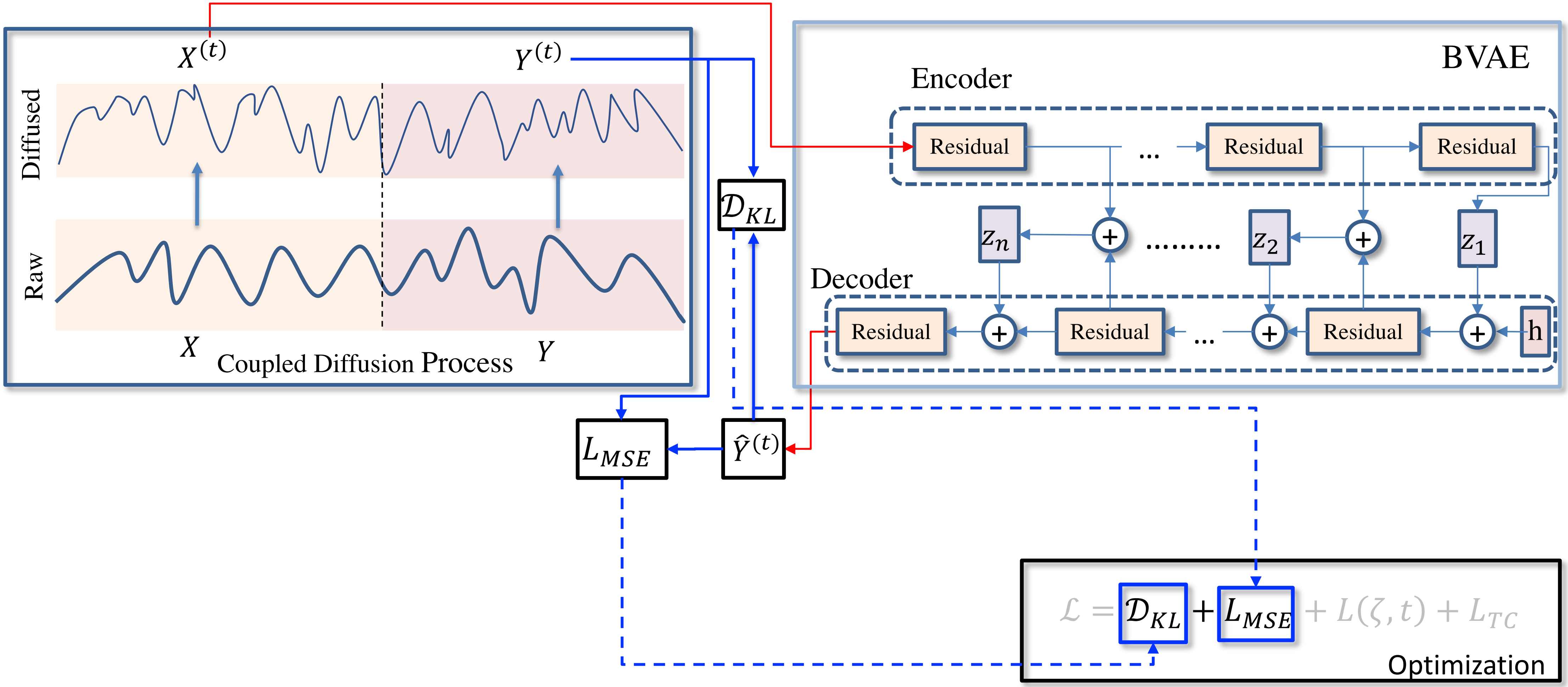
The target time series can be obtained by sampling from $p_\theta(Y)$ directly.



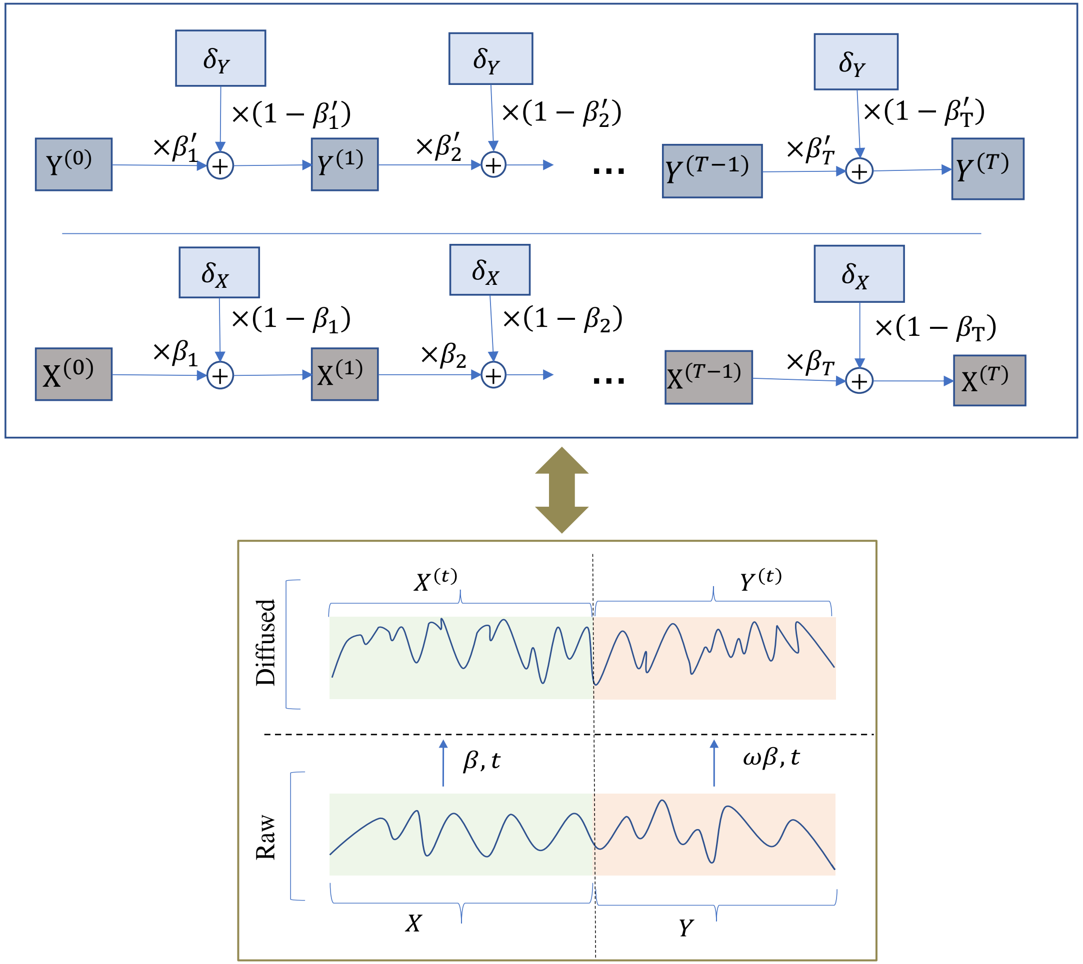
Framework Overview of Proposed *D³VAE*



Coupled Diffusion Probabilistic Model



Coupled Diffusion Process



- Key idea: Diffuse input X and output Y simultaneously

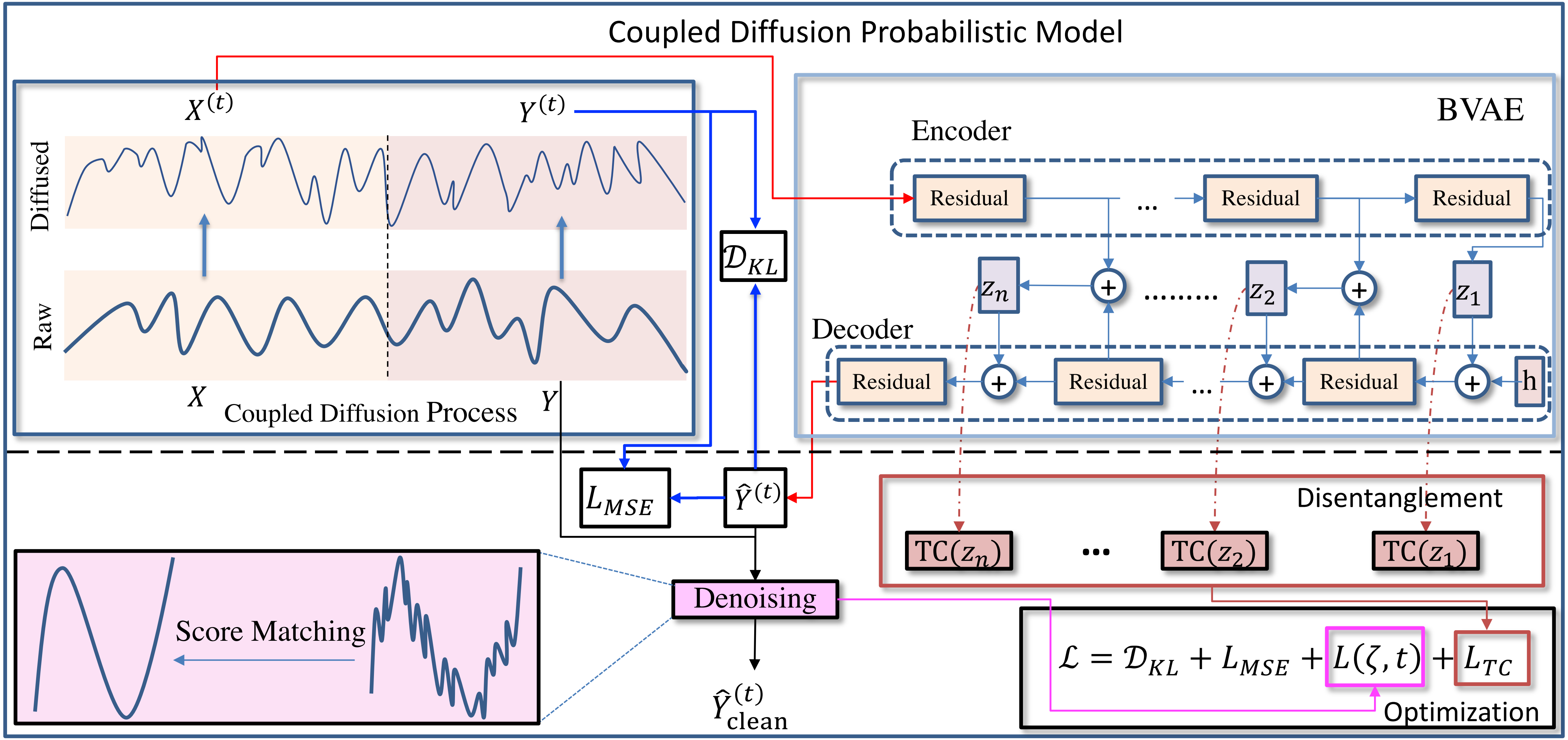
$$\alpha_t = 1 - \beta_t, \quad \bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$$

$$X^{(t)} = \sqrt{\bar{\alpha}_t} X^{(0)} + (1 - \bar{\alpha}_t)\delta_X$$

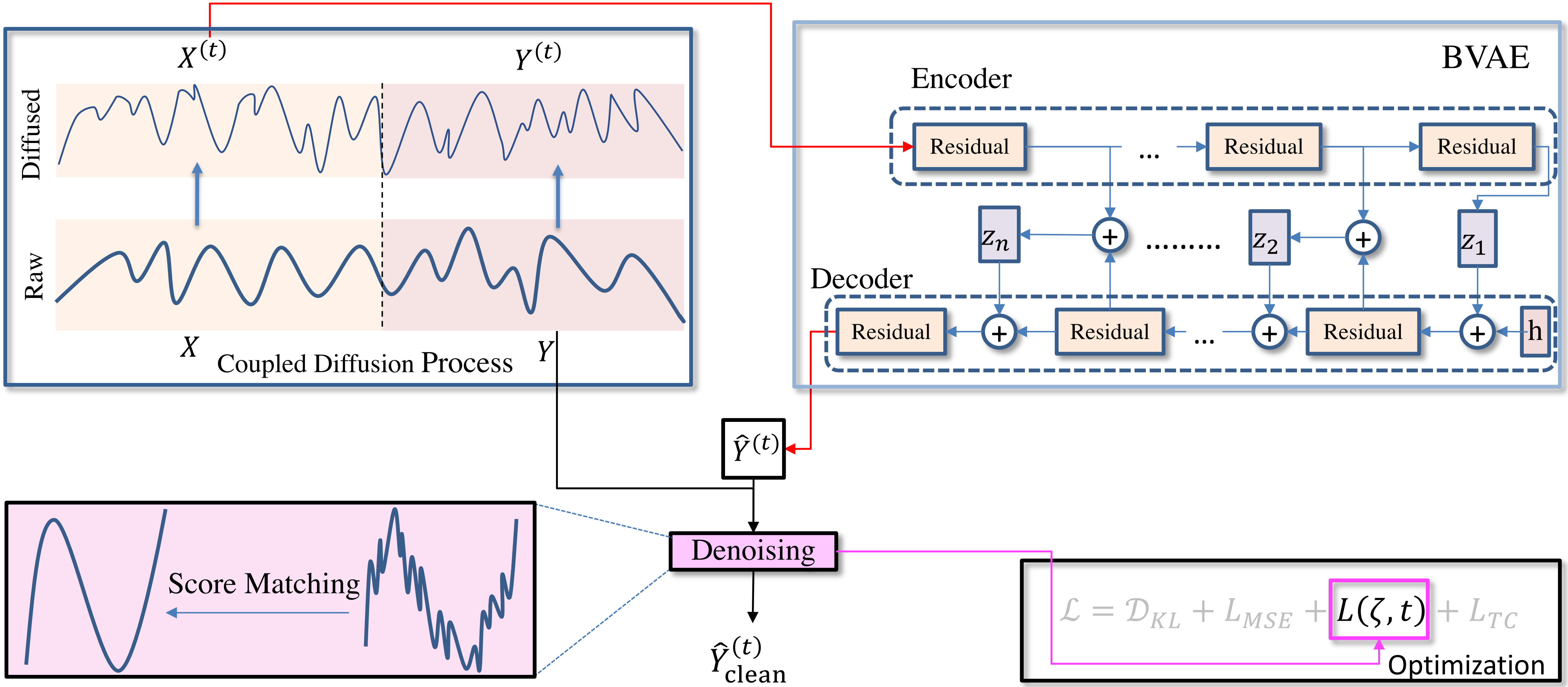
$$\alpha'_t = 1 - \omega\beta_t, \quad \bar{\alpha}'_t = \prod_{s=1}^{t} \alpha'_s$$

$$Y^{(t)} = \sqrt{\bar{\alpha}_t{}'} Y^{(0)} + (1 - \bar{\alpha}'_t)\delta_Y$$

Scaled Denoising Score Matching



Scaled Denoising Score Matching



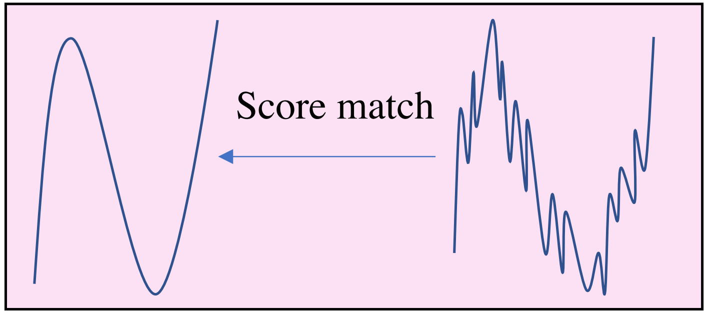
Scaled Denoising Score Matching

Training:

In the case of gaussian noise, we denoise the generated target series with the score matching:

$$L(\zeta, t) = \mathbb{E}_{q_\sigma(\hat{Y}^{(t)}|Y)p(Y)}\,\ell(\sigma_t)\left\|Y - \hat{Y}^{(t)} + \sigma_0^2 \nabla_{\hat{Y}^{(t)}} E\left(\hat{Y}^{(t)};\ \zeta\right)\right\|^2$$

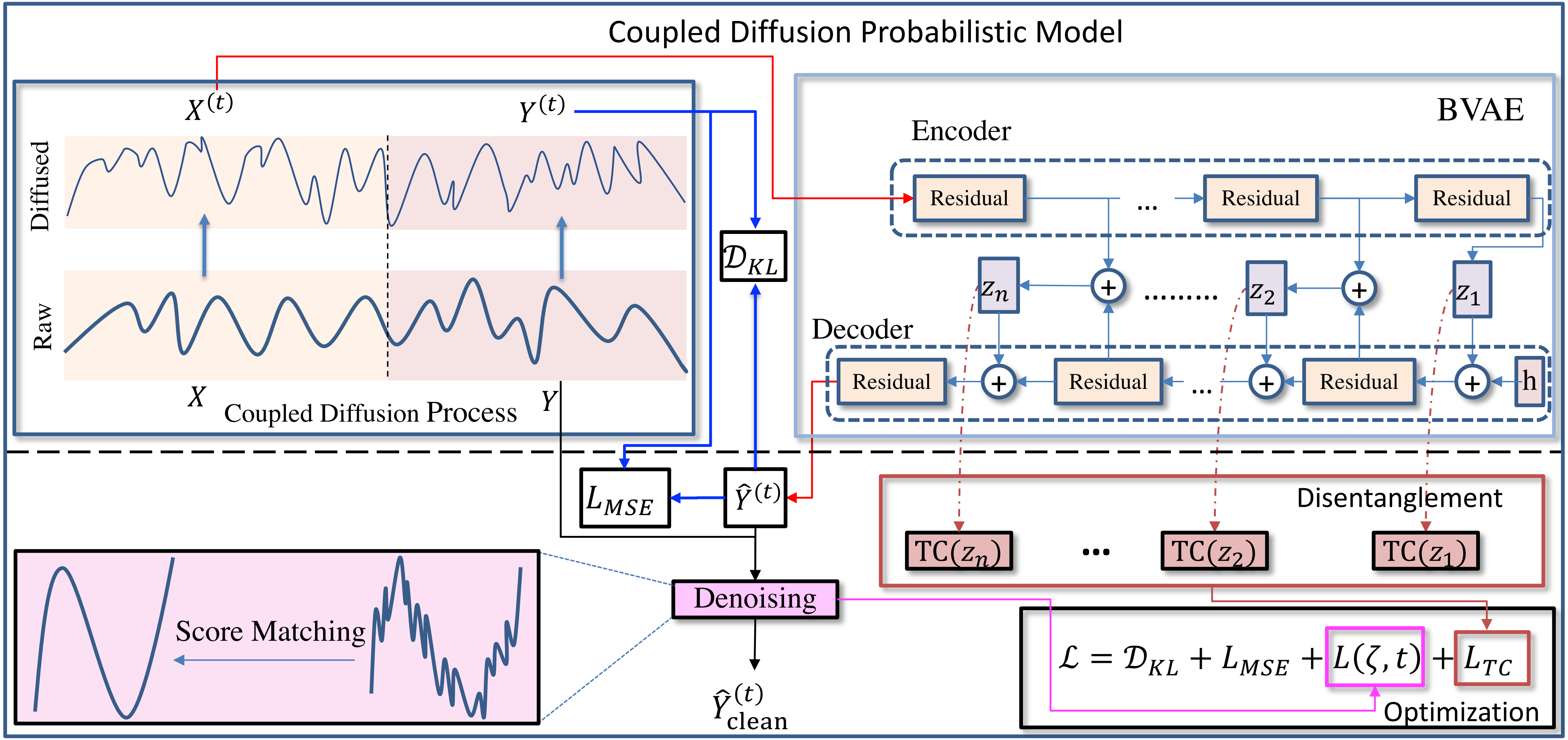
Testing:



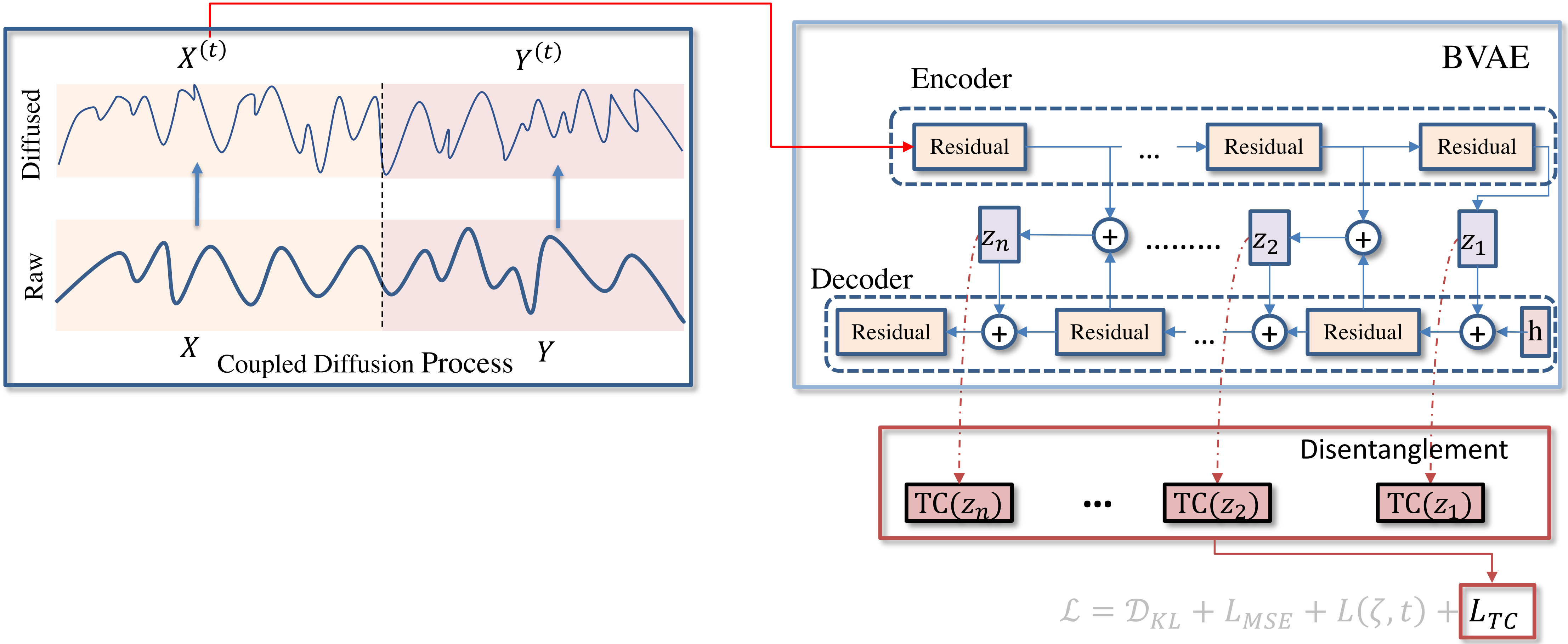
Then the cleaned estimated time series can be obtained:

$$\hat{Y}_{clean} = \hat{Y} - \sigma_0^2 \nabla_{\hat{Y}} E\left(\hat{Y};\ \zeta\right)$$

Disentanglement

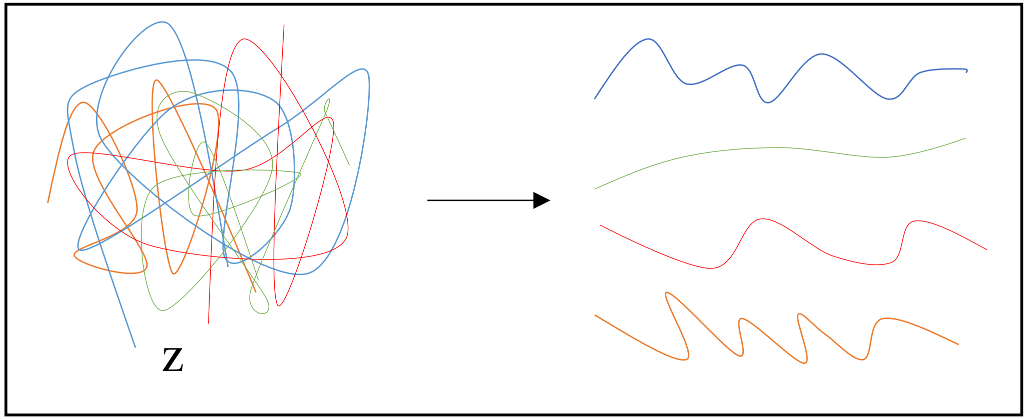


Disentanglement



Disentanglement

For better interpretability and reliability, the latent variables need to be disentangled.



$Z = [z_1, z_2, \ldots, z_n]$

The total correlation (TC) of z_i:

$$TC(z_i) = \mathcal{D}_{KL}(p_\phi(z_i) || \bar{p}_\phi(z_i))$$

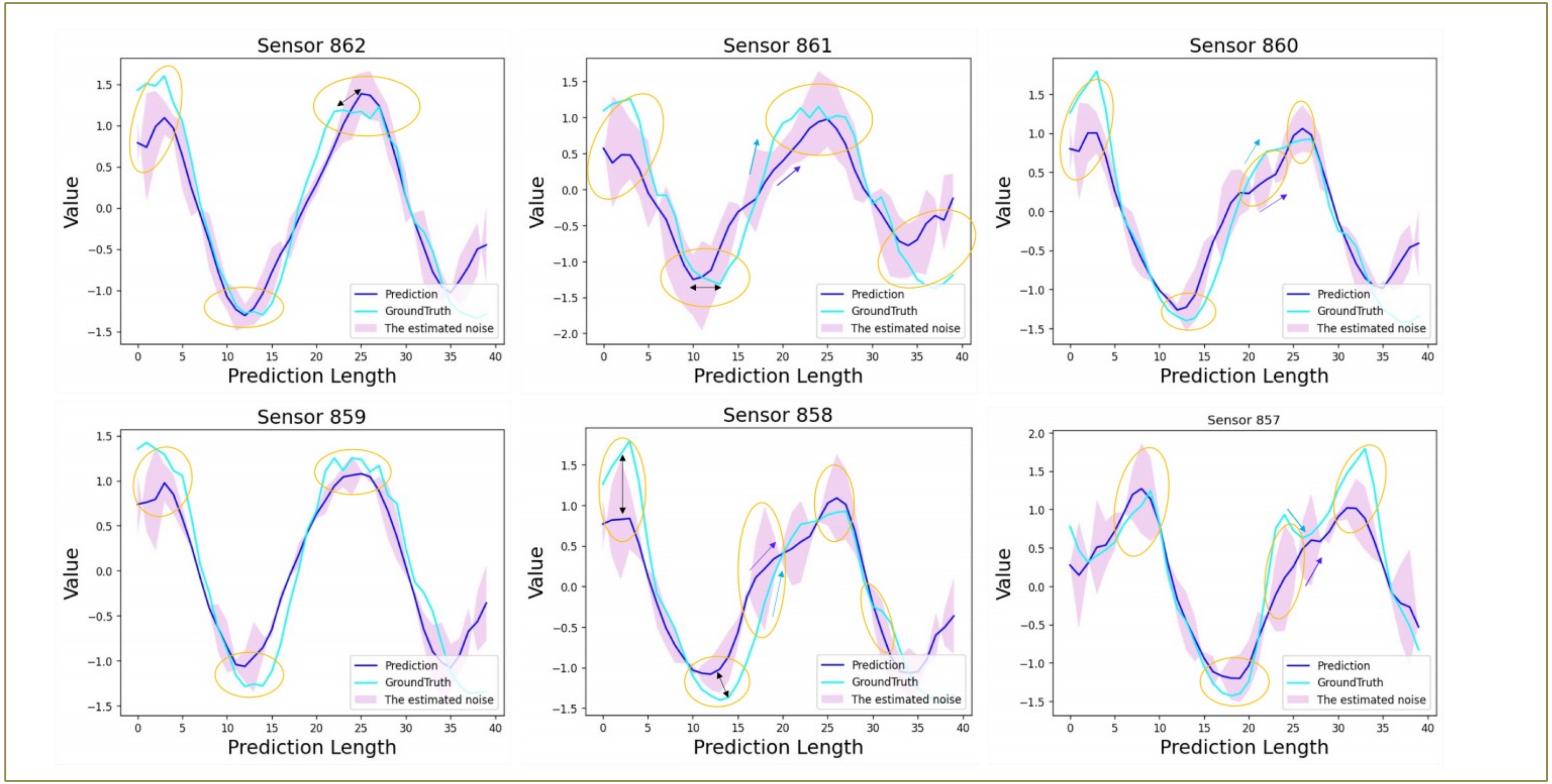
$$\bar{p}_\phi(z_i) = \prod_{j=1}^{m} p_\phi(z_{i,j})$$

Minimize the average total correlation of latent variables.

$$L_{TC} = \frac{1}{n}\sum_{i=1}^{n} TC(z_i)$$

Show Cases of Time Series Forecasting

Time series forecasting with uncertainty estimation on the last six dimensions of the Traffic Dataset.



Thanks!