Improving Seasonal Forecast Using Probabilistic Deep Learning

Pan et al., 2021, JAMES. @LLNL

Wen @ 2022-06-10

Course Presentation

Background1: Uncertainty in Machine Learning

Basic Questions

- How the observational data uncertainty affect/propagates in the ML model
- How confident the model is in its predictions

Uncertainty

Aleatoric

network weight uncertainty term

Epistemic

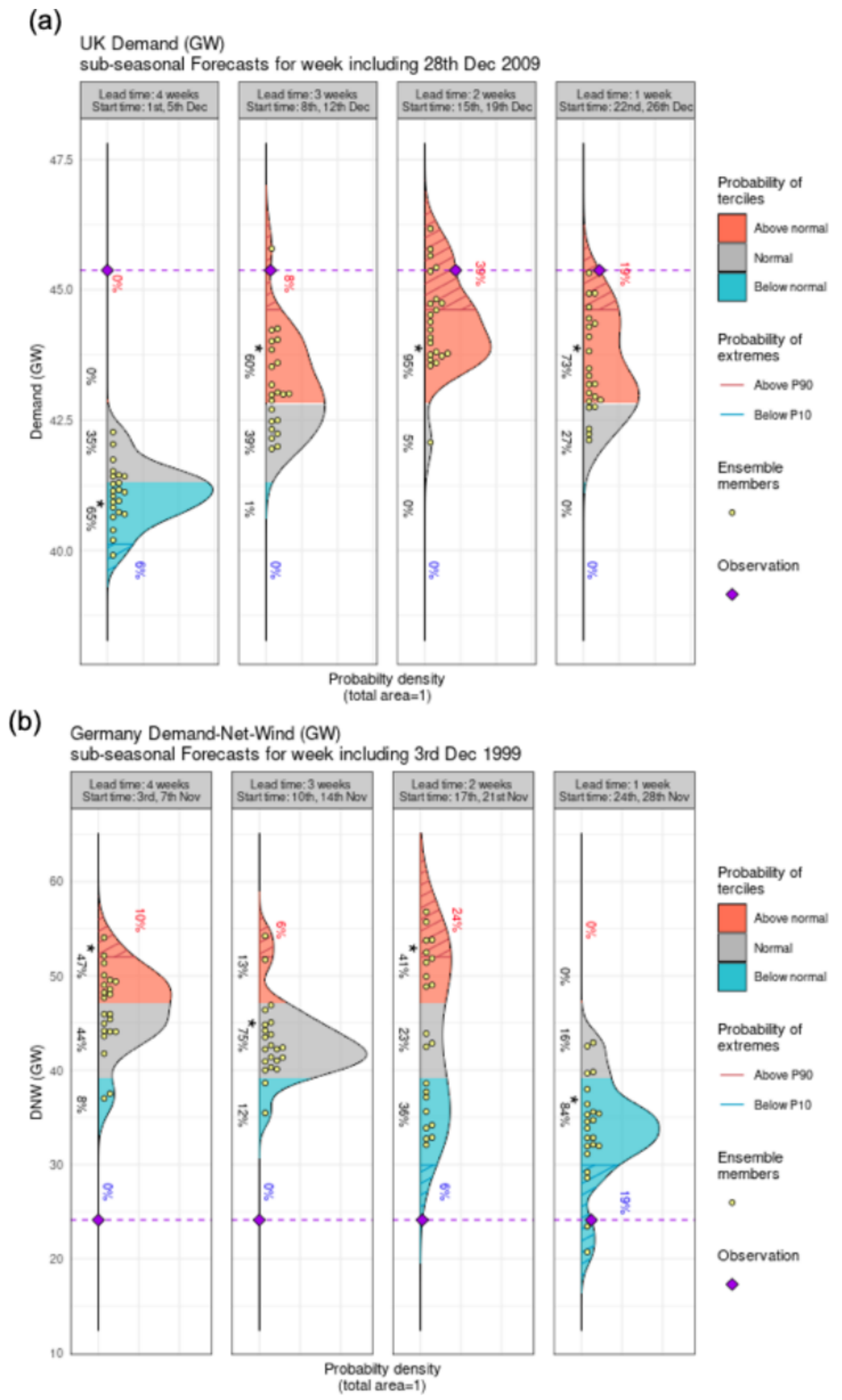
homoscedastic

heteroscedastic

input-dependent data noise term

Background2: Problistic Forecasts

• Dynamical Models



Ensemble members of:

- Seasonal models

机构	模式	成员数
ECMWF	SEAS5	25
MF	Météo-France System 6	25
UKMO	GloSea5-GC2-LI	28
CMCC	CMCC-SPS3	40
DWD	GCFS 2.0	30

- Subseasonal models

机构	模式	成员数
ECMWF	CY47R3	11
CMA	BCC-CPS-S2Sv2	4
Meteo-Franc	CNRM-CM 6.1	10
ECCC	GEPS 7	4
NCEP	CFSv2	4
IAP	CAS-FGOALS-f2-V1.3	4
UKMO	GloSea6	7

• Data Driven Models

From point estimation to: p.d.f. / interval

- traditional statistical methods:
  computationally expensive for deep models
- Ensemble learnig
- Bayesian deep learning (BDL):
  - Monte Carlo dropout
  - **Variational Bayesian methods**

Task: $x \to y$; Model: w

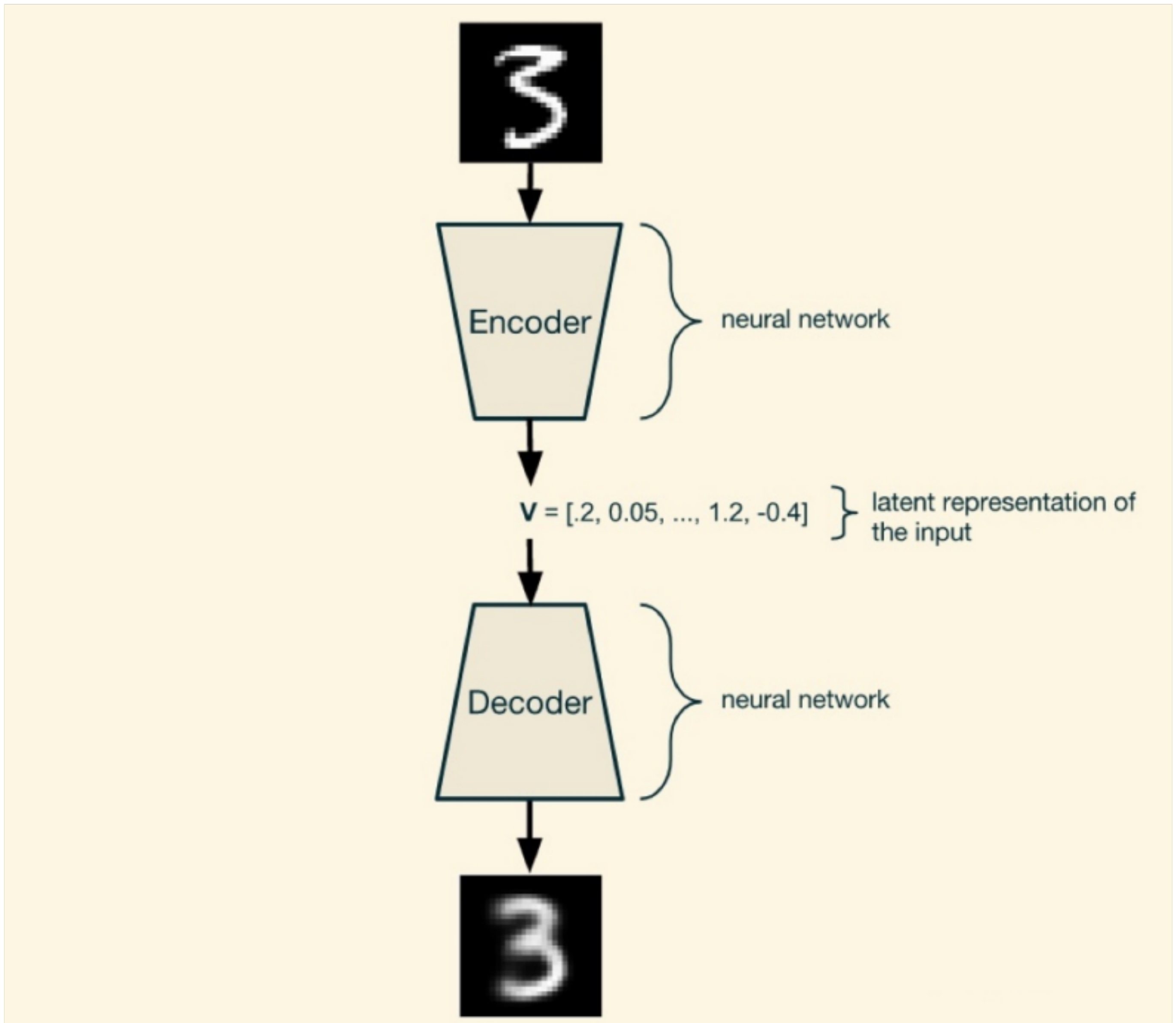
Question: $p(w \mid x, y)$

Construct a family: $q(w; \theta)$

Minimize the difference between p and q: $KL(p(w \mid x, y) \mid\mid q_\theta(w))$

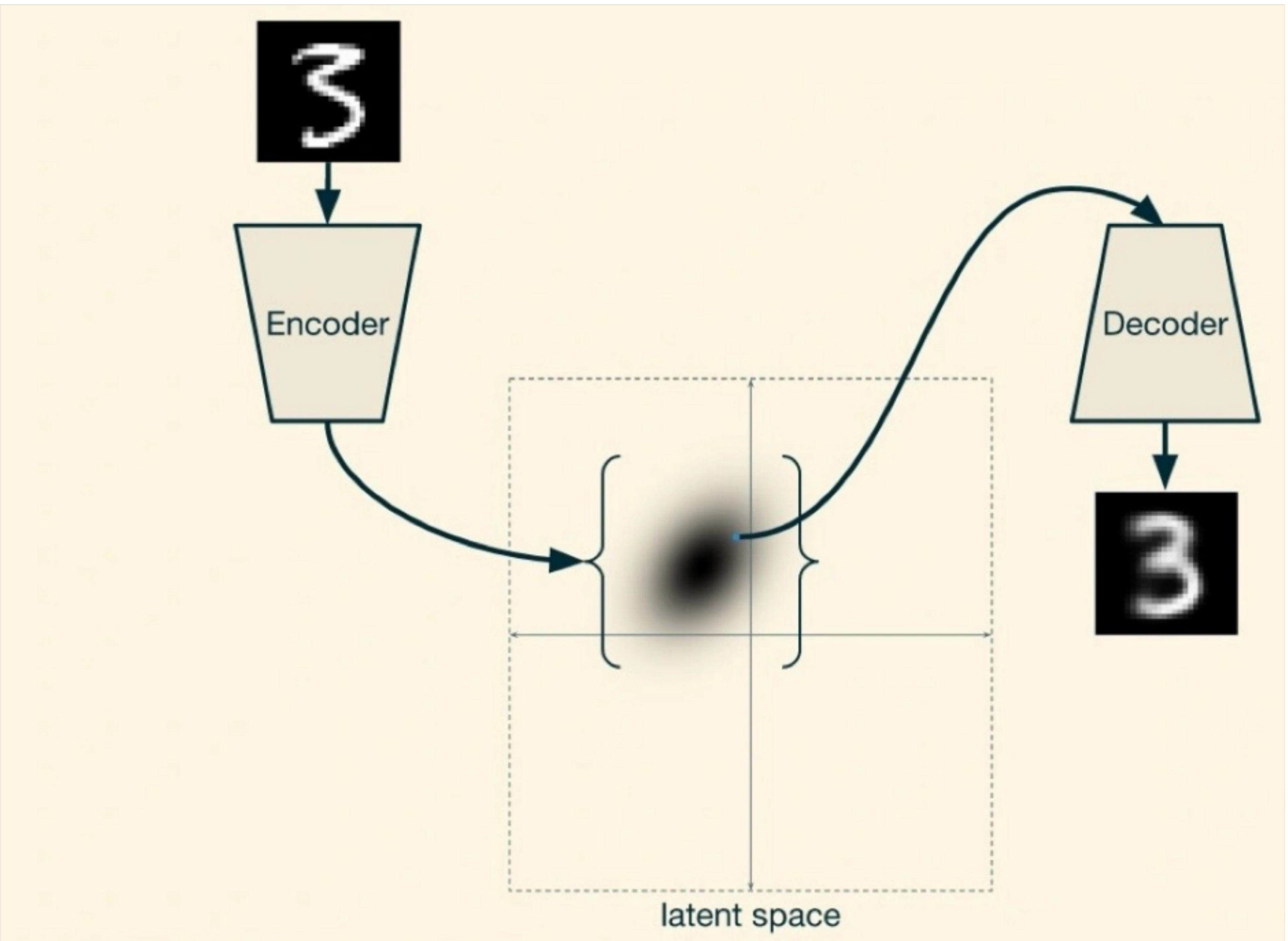
Background3A: Conditional Variational Autoencoder

Autoencoder



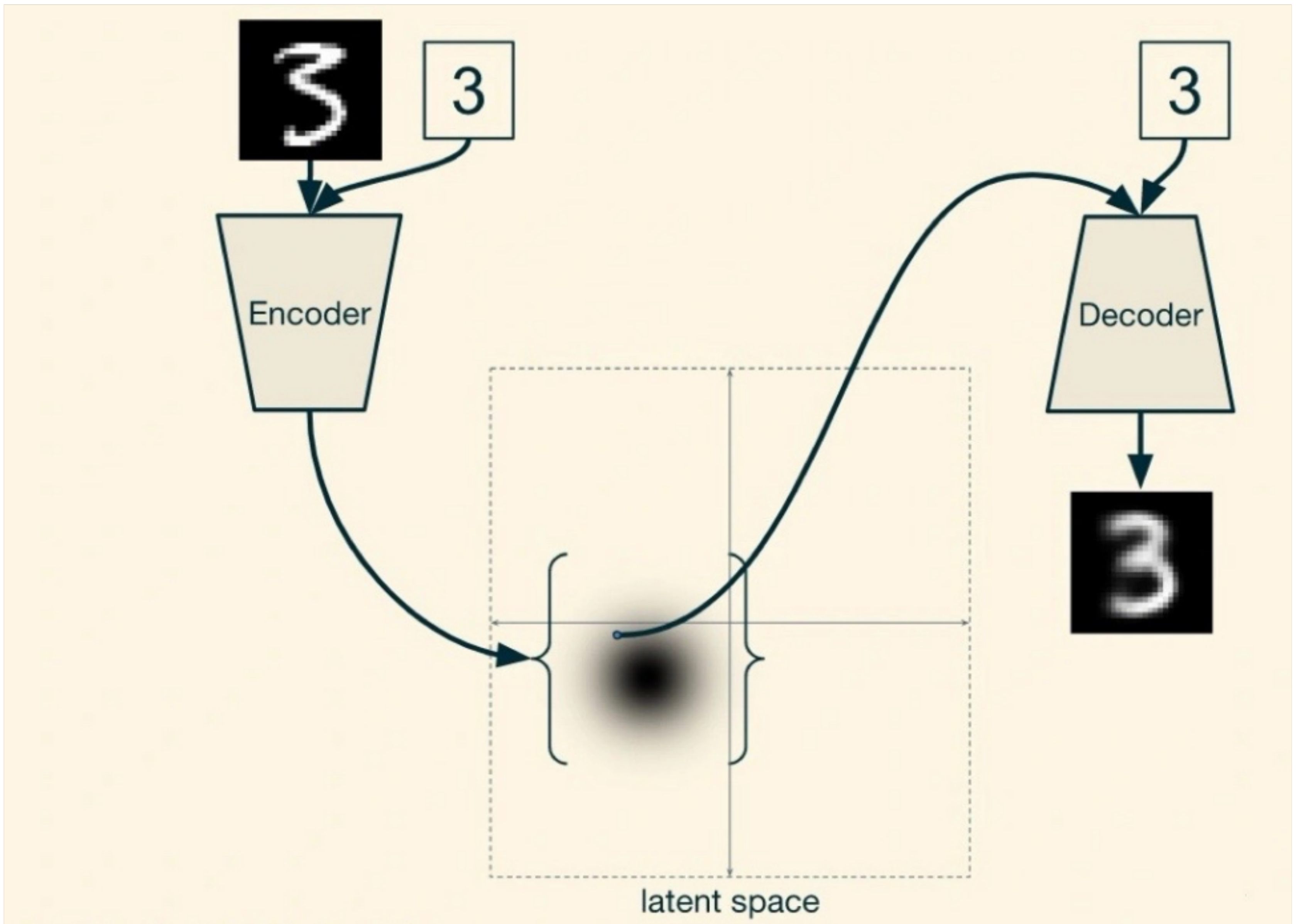
A standard autoencoder

VAE



A variational autoencoder

CVAE



A conditional variational autoencoder

Background3B: Conditional Variational Autoencoder

- **Variational Bayesian**

Task: $x \rightarrow y$; Model: w

Question: $p(w|x,y)$

Construct a family: $q(w;\theta)$

Minimize the difference between p and q: $KL(p(w|x,y)||q_\theta(w))$

Maximize the evidence lower bound (ELBO)

$$KL(p(w|x,y)||q_\theta(w)) \geq ELBO = L(q) - KL(q_\theta(\omega)\|p_0(\omega))$$

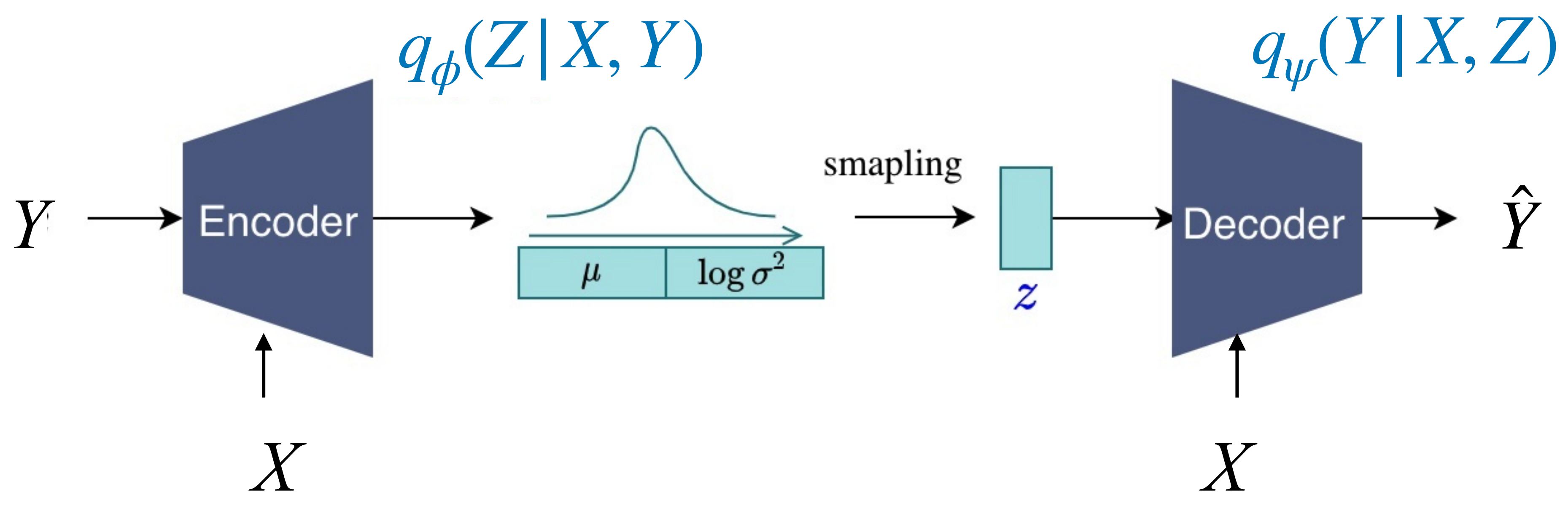
$$\approx \underbrace{-\frac{N}{M}\sum_{m=1}^{M} \|\mathrm{Y}^m - \mathrm{f}_{\mathrm{Y}}^{\hat{\omega}}(\mathrm{X}^m)\|^2}_{\text{Reconstruction}} - \underbrace{KL(q_\theta(\omega)\|p_0(\omega))}_{\text{Regularization}}$$

Reconstruction Regularization

The output of decoder gives a right answer

The distribution will not collapse into point extimation

- **CVAE**



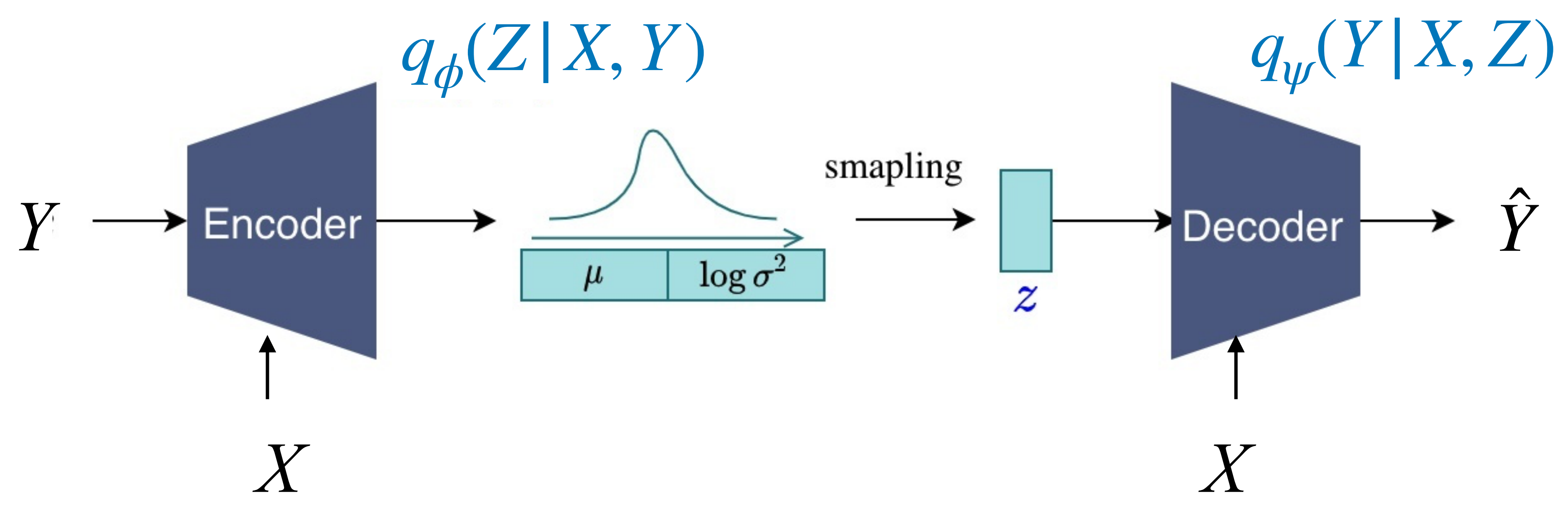
- Maximize the likelihood:

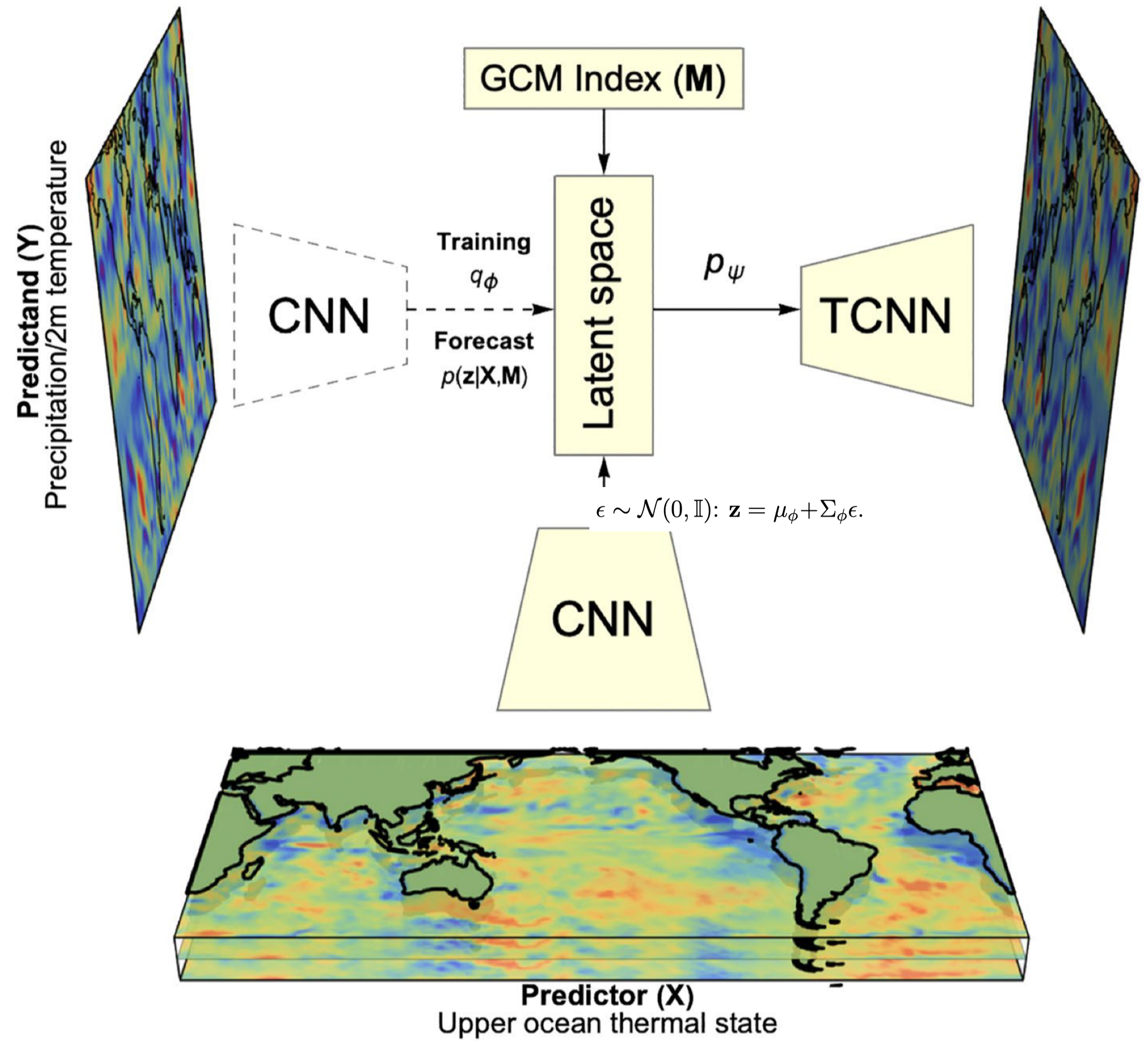
$$\log p_\theta(Y|X) = \overbrace{\underbrace{E_{Z\sim q_\phi}[\log p_\psi(Y|X,Z)]}_{\text{Reconstruction term}} - \underbrace{KL\big(q_\phi(Z|X,Y)\|p(Z|X)\big)}_{\text{Regularization term}}}^{\text{Evidence lower bound (ELBO)}} + \underbrace{KL\big(q_\phi(Z|X,Y)\|P(Z|X,Y)\big)}_{\text{Residual term}}$$

Encourage accurate estimation

Force the posterior of z to fully exploit the prior distribution space

Conditional Generative Forecasting (CGF) model





- Images → **predictand**: (boreal winter) precip, t2m
- Label/condition → **predictor**: upper ocean θ_e profile (previous July)
- Fitting the model: maximinze ELBO
- Seasonal forecast: $q_\psi(Y|X,Z_0)$
- Probalistic forecasts: $\int q_\psi(Y|X,Z)p(Z|X)dZ$

Model details

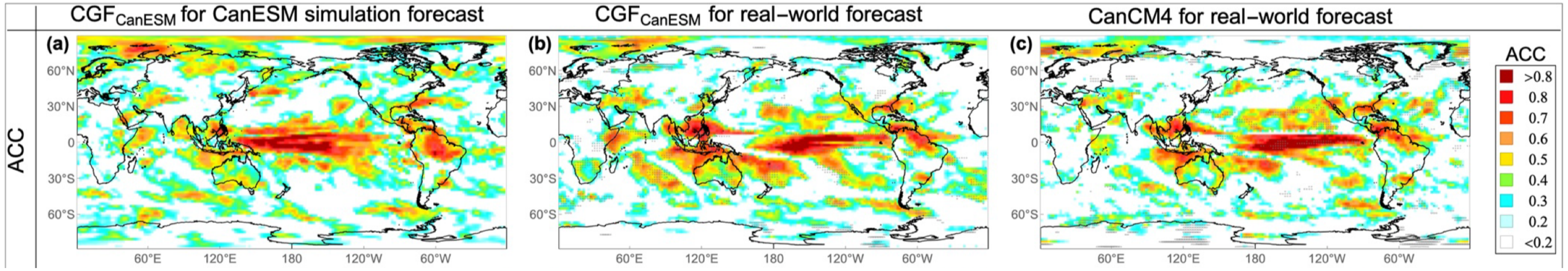
- CNN & TCNN
  - Learn hierarchical feature representations of data
  - Cross-correlate the input feature maps
- Entity embedding
  - Encode the Climate Model Information
  - M: a vactor to index GCMs

Results1: Individual GCM Analog

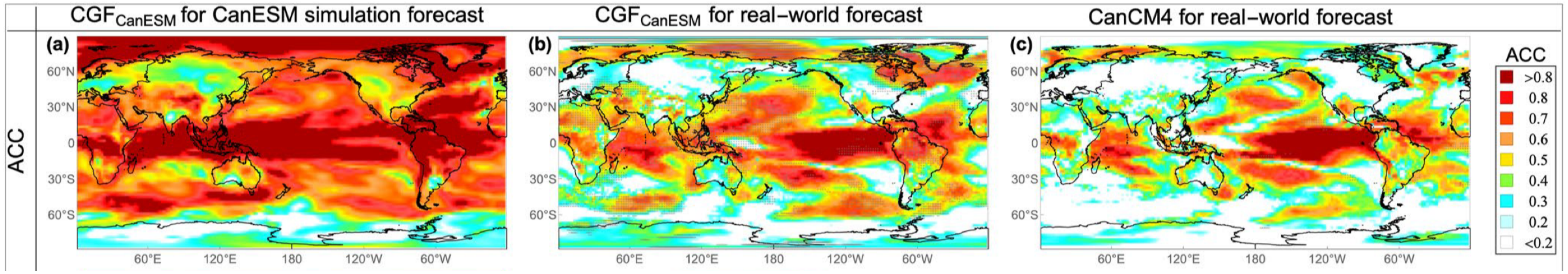
Datasets

- GCM samples
  CMIP5/6 DECK, senarios, … **CanESM**
  to learn the probabilistic dependency **(analog)**
- Dynamical seasonal forecast systems:
  NMME hindcasts **CanCM4**
  for evaluation
- Observational references
  ECMWF atmos/ocean reanalysis, GPCP
  for comparasion and provide initial state

precip



t2m



CGF model is

in general **comparable to** CanCM4, and significantly **outperforms** CanCM4 for a broad range of regions

- CGF skill for the 'model-world'
- CGF skill for the 'real-world'
- Dynamical benchmark

- CGF model captures the seasonal co-variability relationship as revealed by the climate simulation training data.
- Ocean reanalysis as input

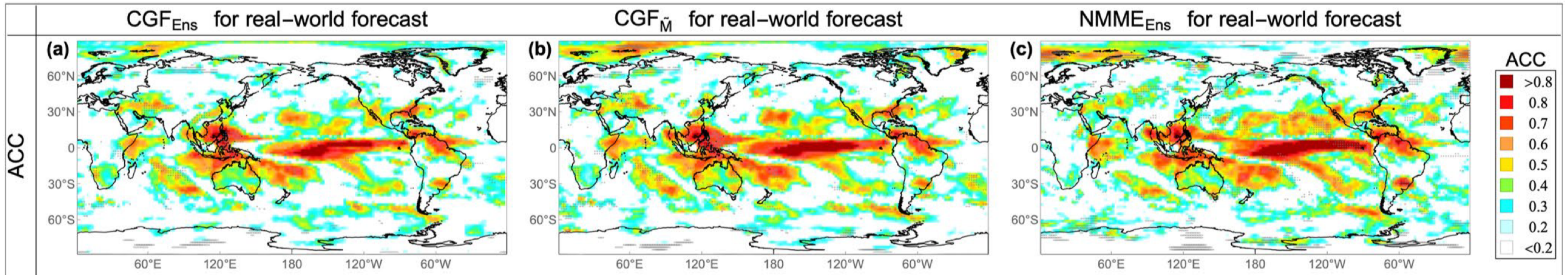
Initial shock

Results2: Multi-Model Ensemble Analog

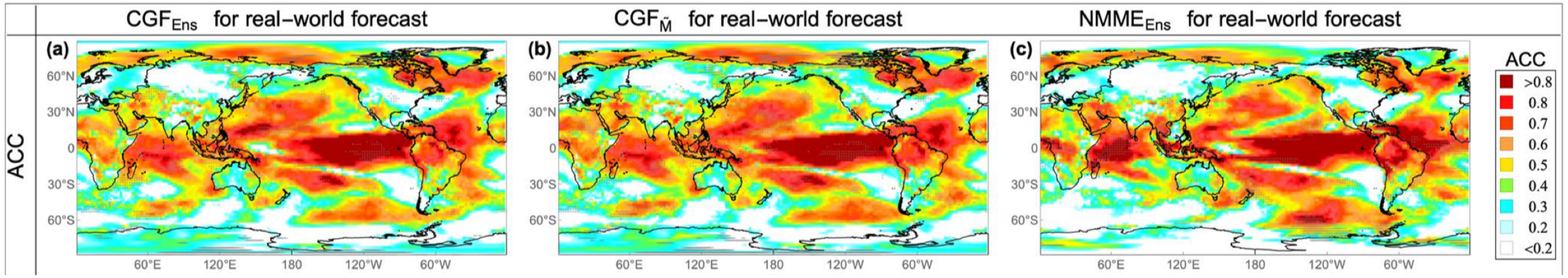
Datasets

- GCM samples
  CMIP5/6 DECK, senarios, …
  with encoding vector *M* for each GCM
- Dynamical seasonal forecast systems:
  NMME hindcasts
  for evaluation
- Observational references
  ECMWF atmos/ocean reanalysis, GPCP
  for transfer learning to get an opitmimal $\tilde{M}$
  (fine-tuning the embedding module)

precip



t2m



CGF model is

able to offer an optimal strategy for composing GCM-ensemble forecast

- CGF ensemble
- CGF optimal ensemble
- Dynamical benchmark

- Does not outperform the single model analog

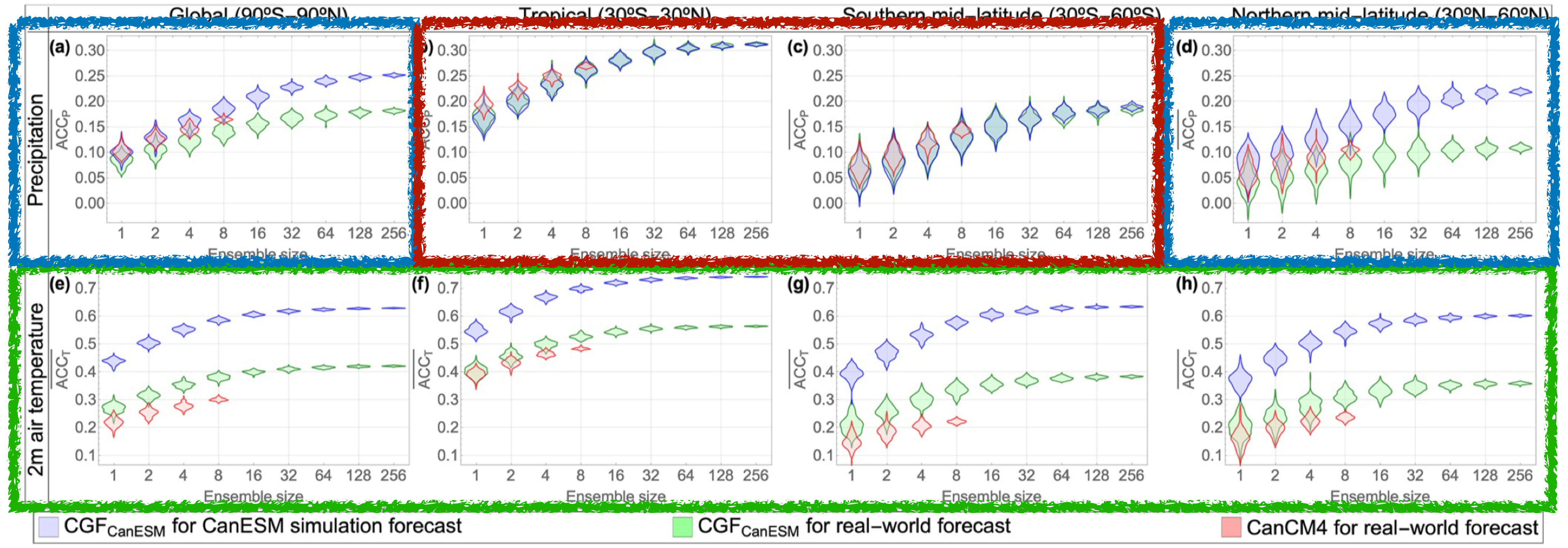
Alleviates the GCM formulation deficiencies

- better t2m skill
- lower precip skill

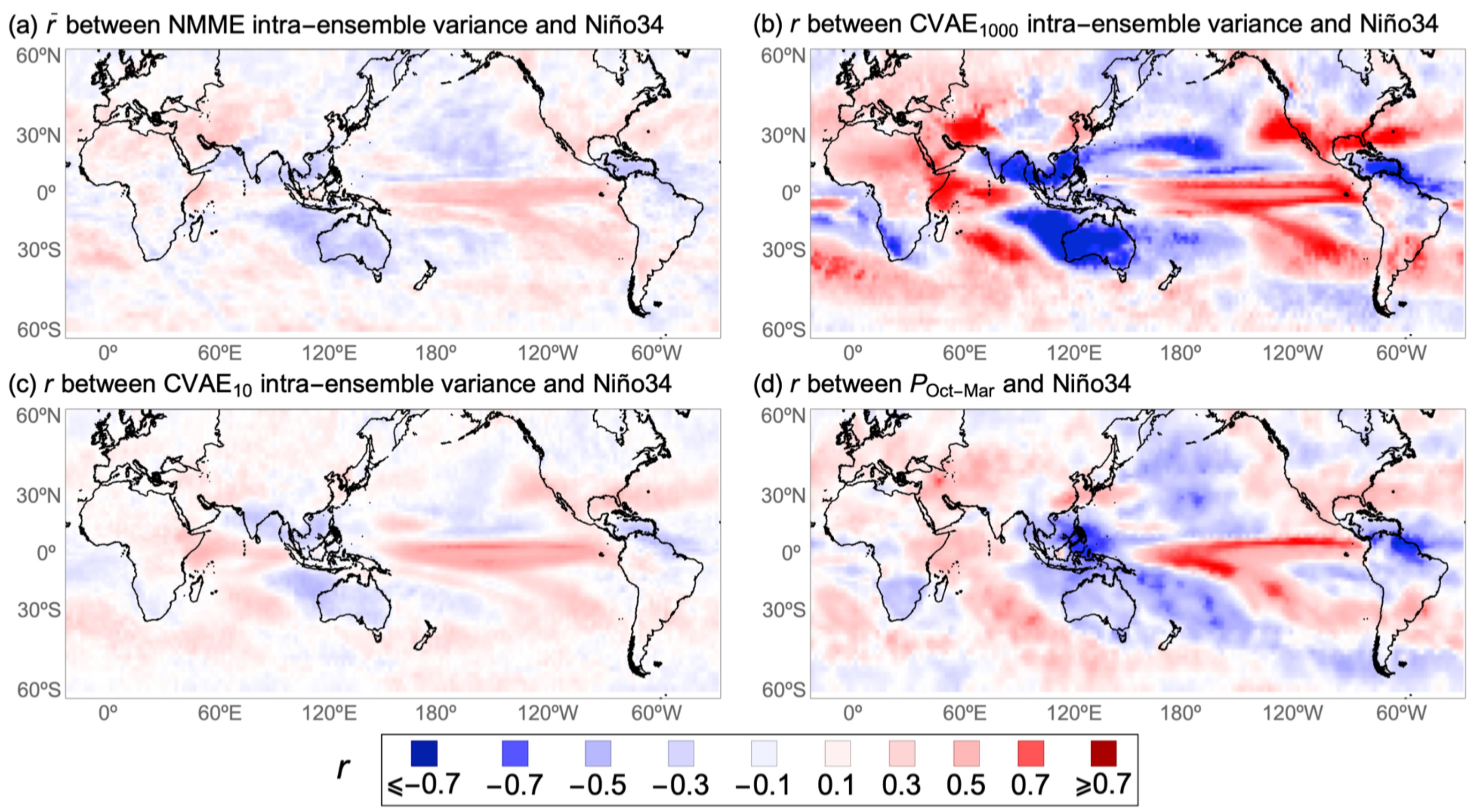
Contributions & Exploring the Interpretability

• Ensemble size

- Skill against ensemble size



- Impact of ENSO on predictability

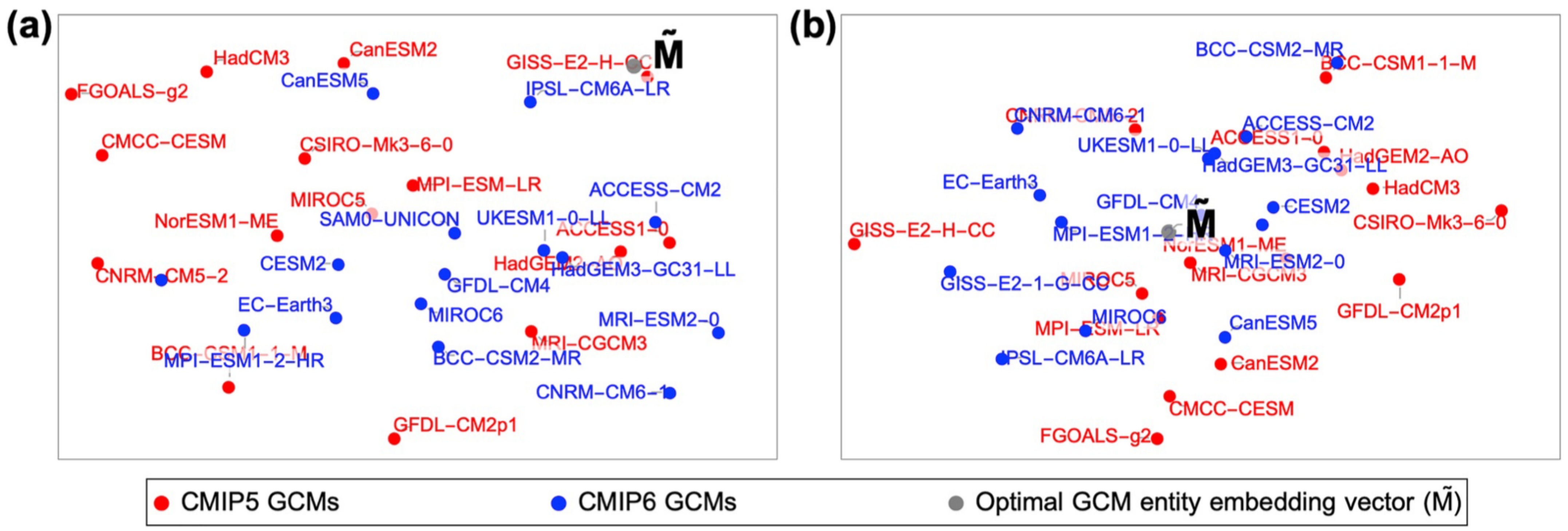


• Optimal embedding

- Reduce the dimension of embedding vector to visualize

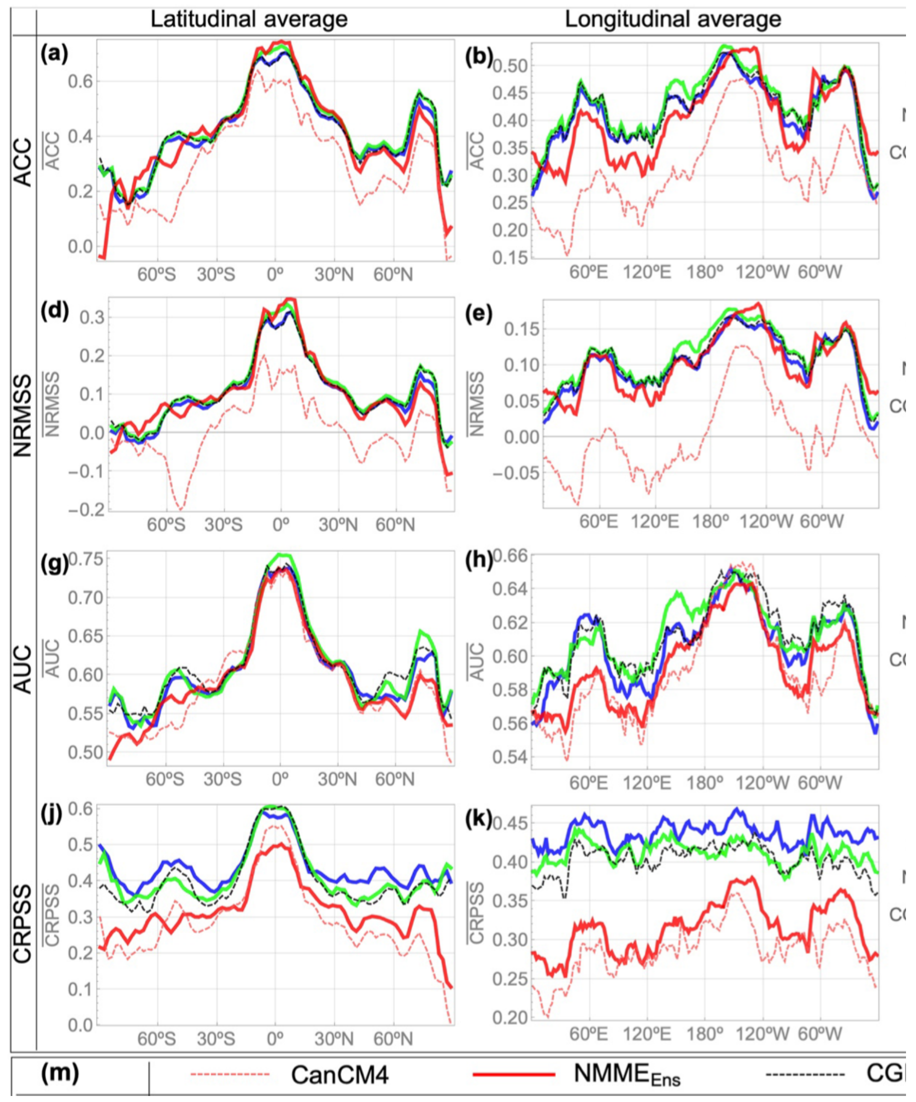
precip

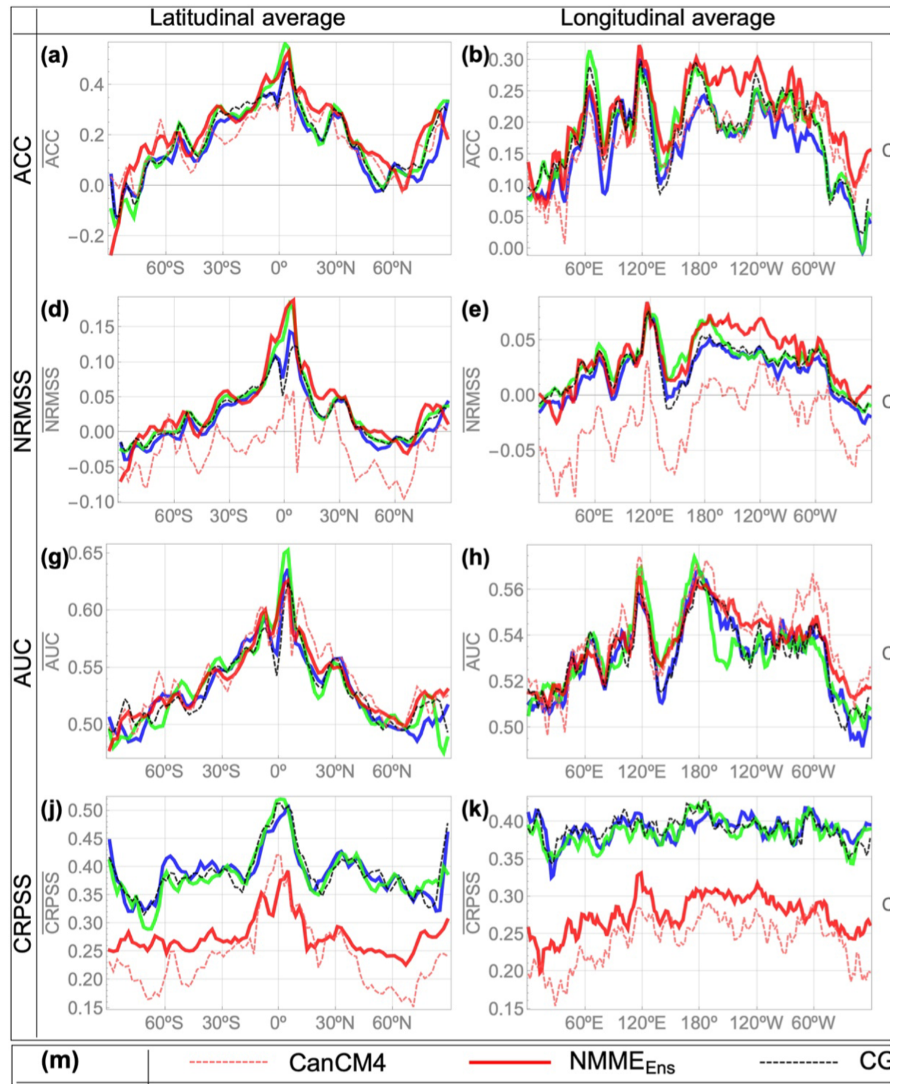
t2m



- CGF method offers an efficient way to extensively sample the plausible forecasting outcomes
- CGF method visualize the bias of seasonal co-variability relationship in each GCM

Summary





Uncertainty Quantification in Machine Learning

Problistic Forecasts

Variation

Autoencoder

Predictor

CGF model based on CVAE

Deterministic skill

- Comparable and outperform in some region
- Improvement t2m>precip

Problistic forecasts

- Reproduce the predictabiity and its dependency
- Efficiently generate large ensembles
- Addressing heteroscedastic aleatoric uncertainty (Anderson et al, 2021)

Interpretability

- Represent the initial shock
- Visualize GCM formulation deficiencies

Thank you!