

Machine Learning-Based Methods for Precipitation Phase Discrimination in China

[CRYOSPHERE & HYDROLOGY: 95462]

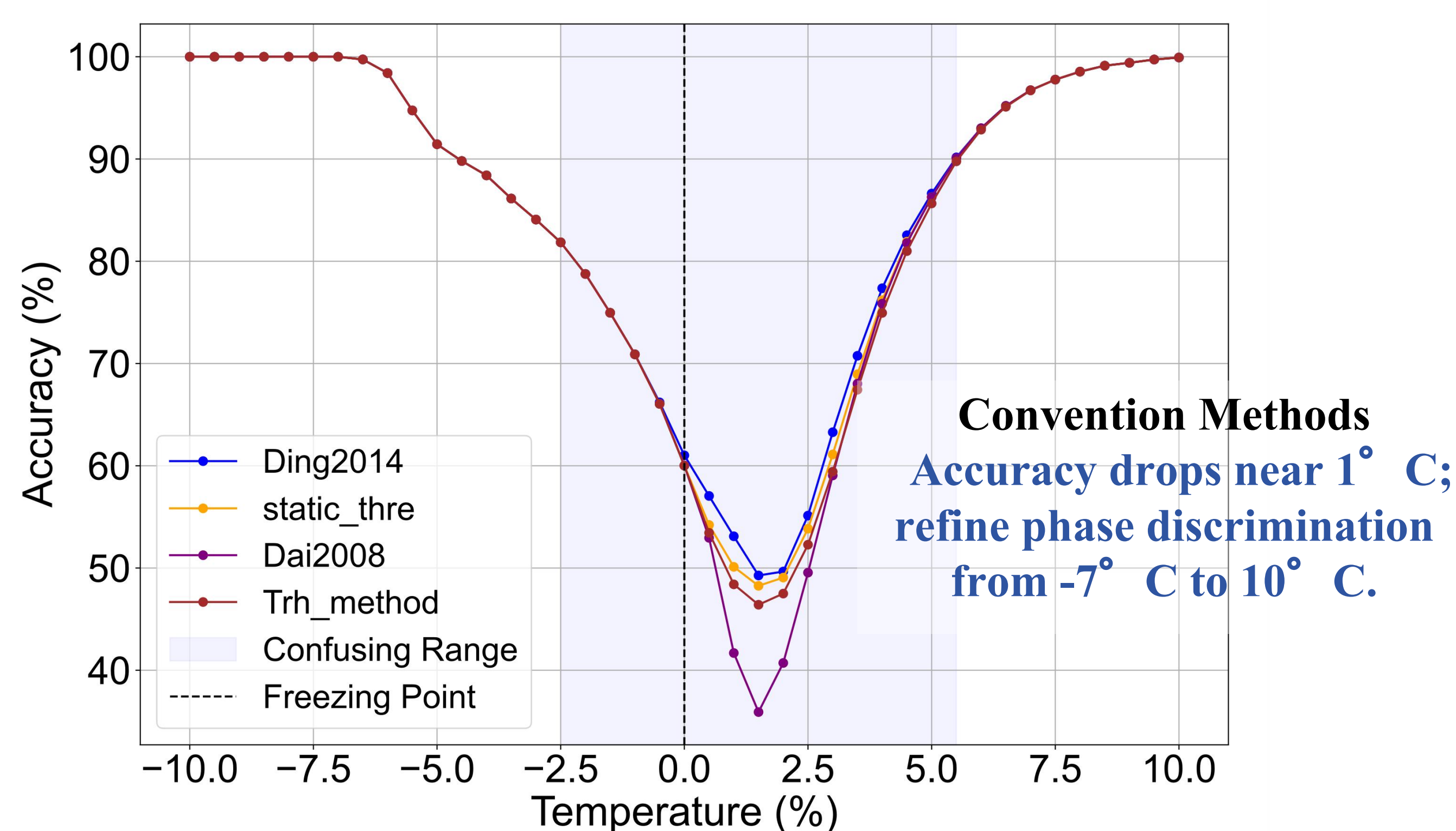


清华大学地球系统科学系
Department of Earth System Science, Tsinghua University

Department of Earth System Science, Ministry of Education Key Laboratory for Earth System Modeling,
Global Change Research Institute, Tsinghua University, Beijing 100084, China

Introduction

- Precipitation phase plays a crucial role in surface energy balance and hydrological processes.
- Existing empirical method based on surface meteorological conditions often struggle to accurately classify precipitation near the freezing point.



static_thre: fixed threshold = 0.6°C . Ding2014: from Ding et al., 2014.
Dai2008: from Dai, 2008. Trh_method: from Harder & Pomeroy, 2013.

- This study proposes a **machine learning-based approach** for precipitation phase discrimination within the ambiguous temperature range of -7°C to 10°C .

Data & Method

Daily Surface Climate Variables of China

Pre-processing & Quality Control

- Select data with correct quality control flags of T_a , RH, wind, prs, prec
- Remove data from abnormal year, abnormal stations and abnormal data
Abnormal Standard:
 $T_a(\text{rain}) < -5.4^{\circ}\text{C}$; $T_a(\text{snow}) > 8^{\circ}\text{C}$;
 $T_a(\text{sleet}) < -1.6^{\circ}\text{C}$ OR $T_a(\text{sleet}) > 9.6^{\circ}\text{C}$
Considering sub-freezing rainfall events

- Choose data records with T_a ranges from -7°C to 10°C
input: longitude, latitude, altitude; wind, T_a , Relative humidity, pressure; precipitation amount; T_w , T_d .

Wet-Bulb Temperature (T_w)

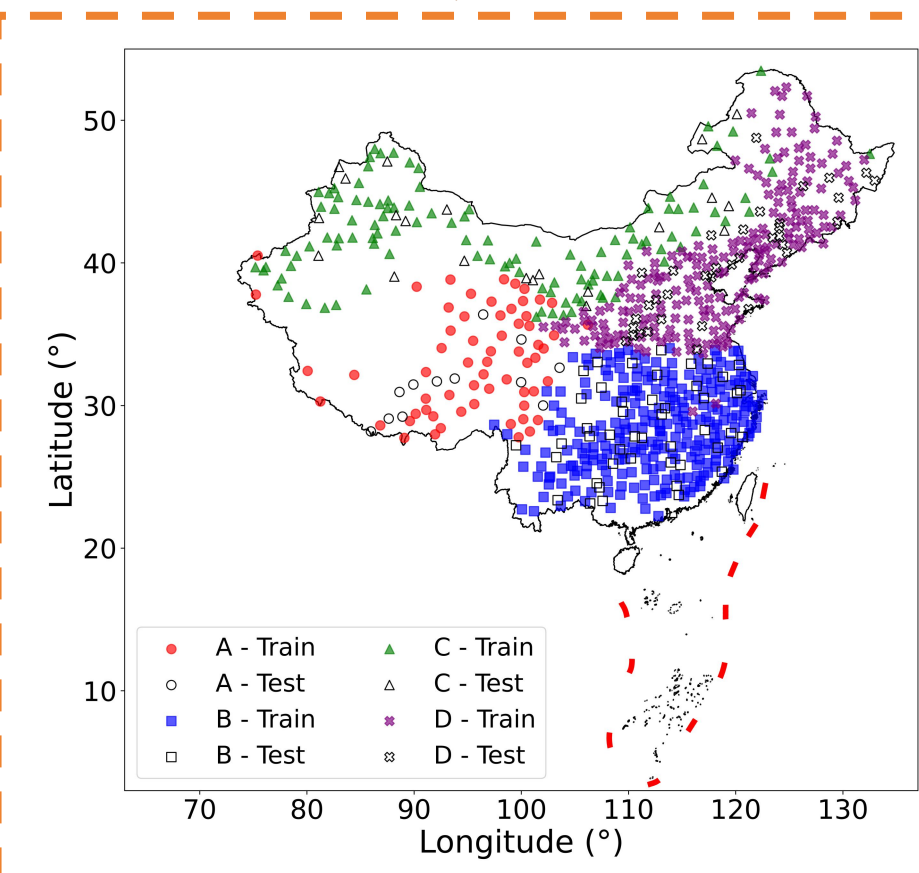
The lowest temperature air can reach through evaporation, reflecting temperature and humidity.

Dew Point Temperature (T_d)

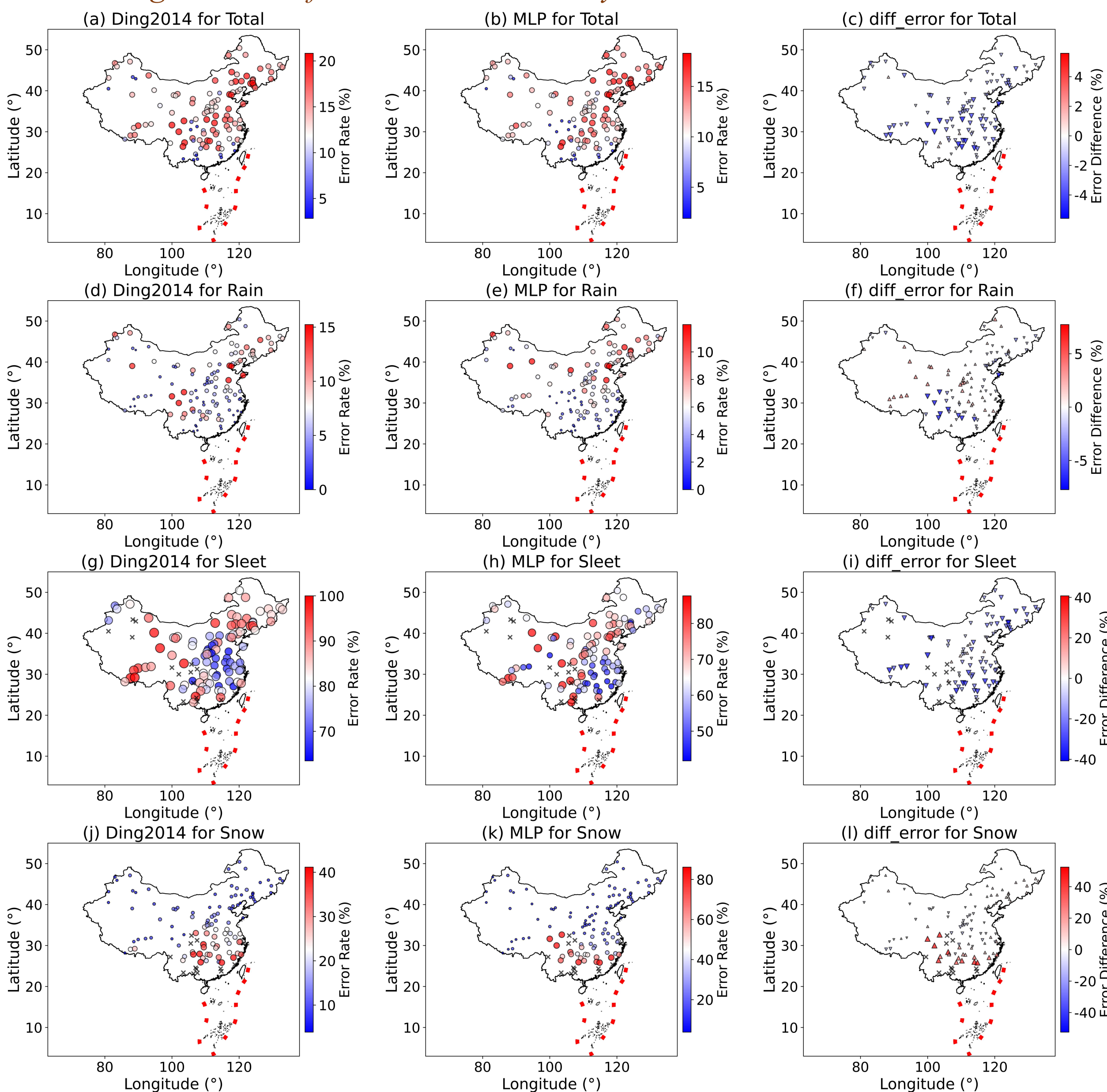
The temperature at which air becomes saturated and water vapor begins to condense.

Five machine learning methods are explored:
SVM (Support Vector Machines); RF (Random Forest);
CatBoost (a gradient boosting algorithm);
MLP (Multilayer Perceptron); CNN (Convolutional Neural Network)

Divided into 4 regions based on elevation, January T_a , annual prec



In each region, 15% of stations are randomly used as test set



Result: ML & Ding2014 Comparison

Optimal INPUT parameter combinations used by each ML method

Method	Optimal input parameter combinations
SVM	$lon, lat, p_s, RH, wind, amount, T_w, T_d$
RF	$lon, lat, p_s, RH, amount, T_w, T_d$
CatBoost	$lon, lat, p_s, RH, amount, T_w, T_d$
CNN	$lon, lat, T_a, p_s, wind, RH, amount, T_w$
MLP	$lon, lat, T_a, p_s, wind, RH, amount, T_w$

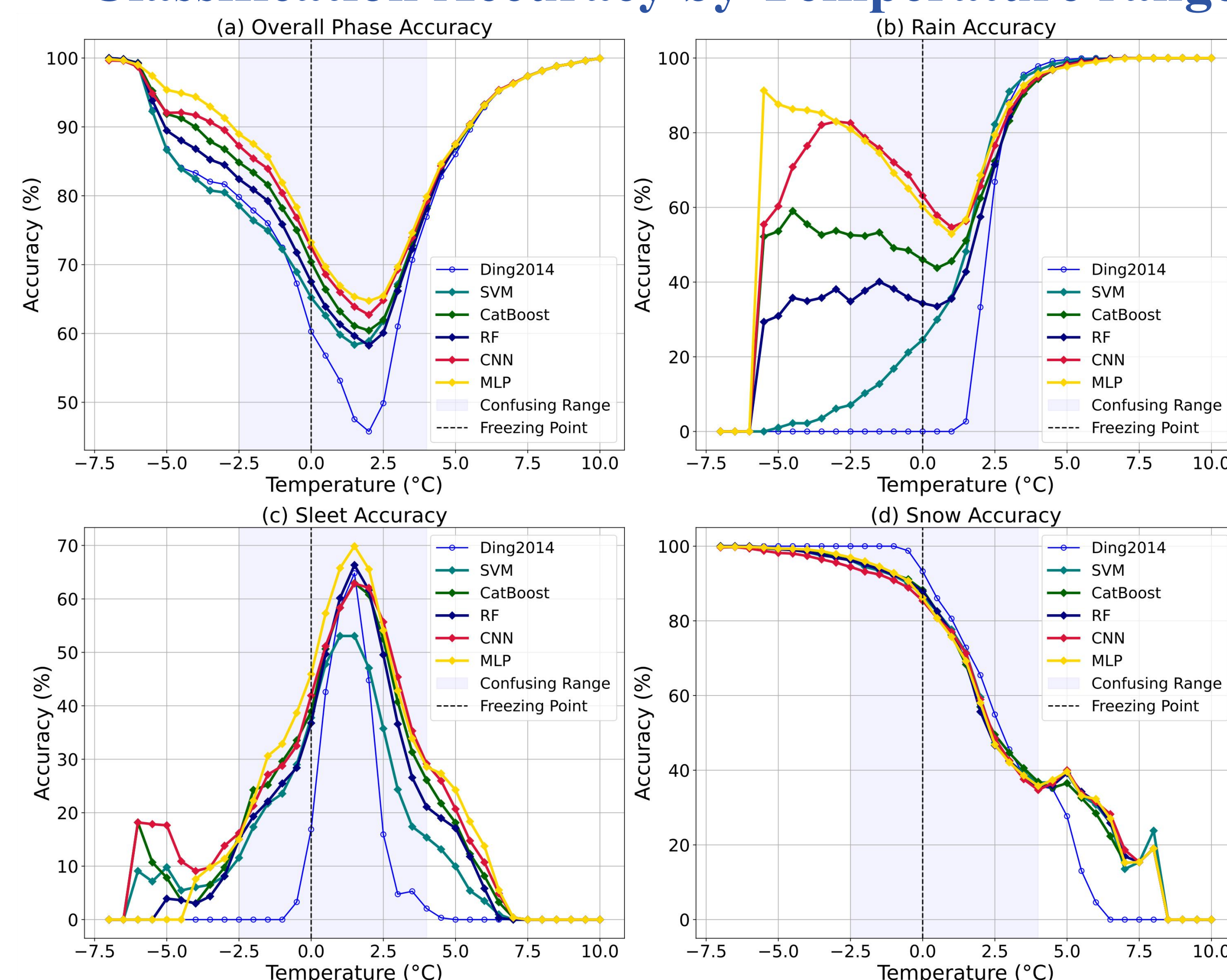
Phase recognition ACCURACY of ML and Ding2014 for test set

Method	Overall	A accuracy	B accuracy	C accuracy	D accuracy
Ding2014	83.7%	78.6%	86.0%	84.7%	83.8%
SVM	85.4%	81.8%	87.2%	85.9%	85.3%
RF	86.0%	84.2%	87.3%	85.9%	85.3%
CatBoost	86.7%	87.0%	87.6%	85.7%	85.4%
CNN	87.6%	91.1%	87.3%	85.8%	85.4%
MLP	88.1%	91.4%	88.3%	86.3%	85.3%

A: Tibetan Plateau; B: South; C: Northwest; D: Northeast & North

ALL ML BETTER AND **MLP method achieved an overall test set accuracy of 88.1%, improving by 4.4% over Ding2014.**
In Tibetan Plateau, the accuracy reached 91.4%, an increasement of 12.8% compared to Ding2014.

Classification Accuracy by Temperature range



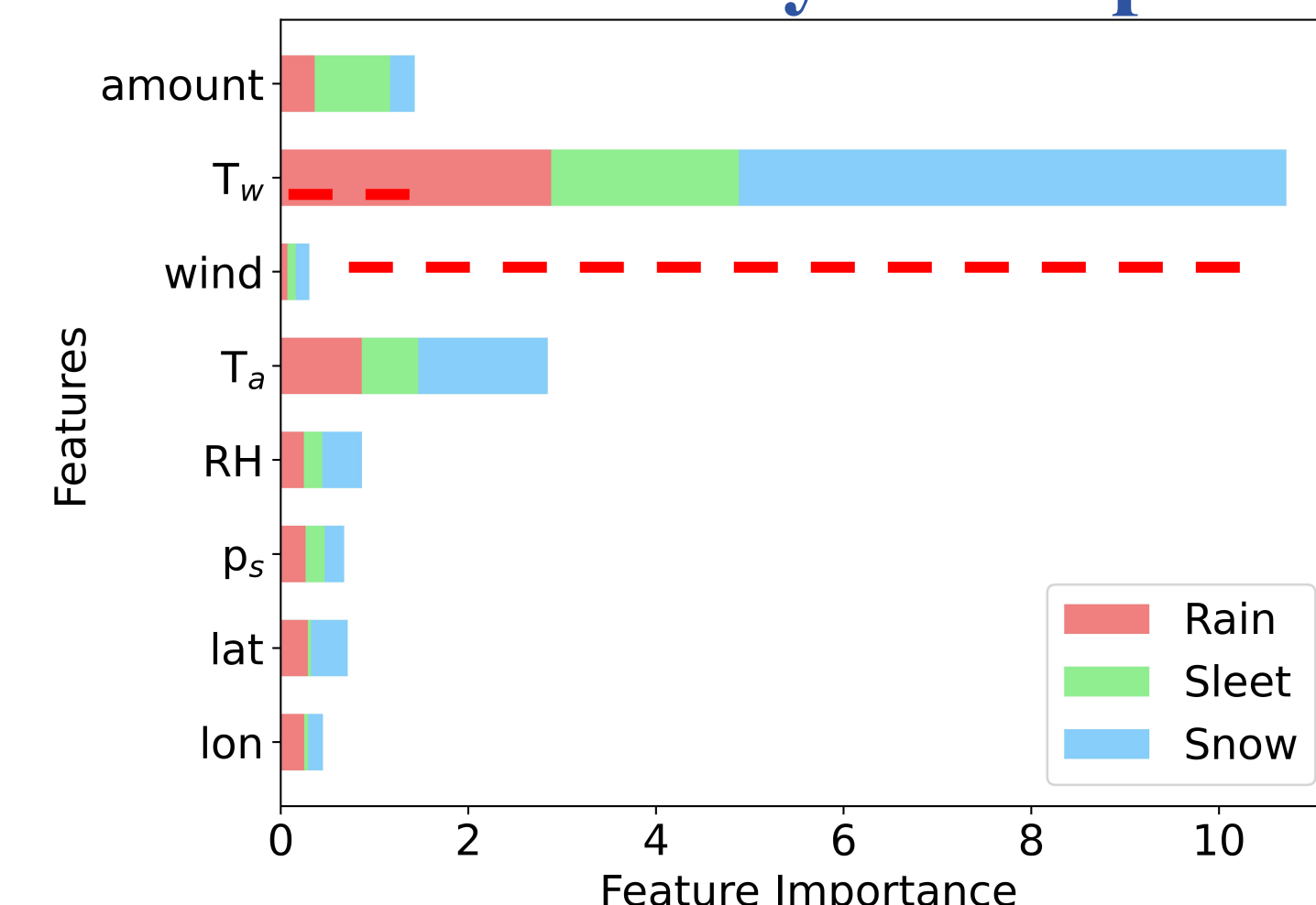
MLP performed well in rainfall (especially freezing rain) and showed high sensitivity to sleet despite limited sample size.

Station error map of MLP and Ding2014, and their difference

Accuracy varies with the **phase sample distribution** and shows a **pronounced north-south contrast**, reflecting ability to learn region-specific characteristics.

The **improvement** is mainly concentrated in the **eastern Tibetan Plateau** and several provinces in **southern China**, where **rainfall and sleet occur frequently**.

SHAP analysis: Importance of Input Features in Rain, Sleet, and Snow Classification Probabilities



T_w incorporates information on temperature, humidity and pressure, has the greatest impact on MLP model's performance.

Precipitation amount also significantly affects the probability of sleet classification.

MLP is OK for phase discrimination AND

Future work may focus on identifying **the hidden physical factors underlying the effectiveness of geographic coordinates**, and **better handle sleet** with limited and ambiguous samples.