APEC Workshop on Achieving Sustainability and Resilience in Water Management of APEC Developing Economies Using Open Environmental Data

APEC Policy Partnership on Science, Technology and Innovation

June 2025

Asia-Pacific
Economic Cooperation

APEC Workshop on Achieving Sustainability and Resilience in Water Management of APEC Developing Economies Using Open Environmental Data

APEC Policy Partnership on Science, Technology and Innovation

June 2025

APEC Project: PPSTI 02 2023

Produced by
Dr. Sirod Sirisup
National Electronics and Computer Technology Center (NECTEC)

For
Asia-Pacific Economic Cooperation Secretariat
35 Heng Mui Keng Terrace
Singapore 119616
Tel: (65) 68919 600
Fax: (65) 68919 690
Email: info@apec.org
Website: www.apec.org

APEC#225-PP-01.8

Overview

This report presents a framework for seasonal rainfall prediction to support sustainable and resilient water management across APEC economies. The framework and its implementation guidelines were developed by the organizers and discussed during a two-day workshop, where expert speakers shared insights on feature engineering and observational data sources, while participants contributed through presentations and discussions.

The workshop explored differences between tropical and midlatitude meteorology, emphasizing the need for localized feature engineering techniques to improve rainfall prediction models. Expert speakers provided key knowledge on meteorological variables, observational data sources, and high-resolution model outputs critical for accurate predictions. Discussions among participants led to the identification of use cases from different economies, illustrating how different climatic and geographical conditions shape feature selection for predictive modeling.

Based on the insights gathered from the workshop, this report also provides suggestions for refining the framework, focusing on region-specific adaptations, data integration strategies, and improvements in observational coverage. By synthesizing the organizers' framework, expert insights, and collaborative discussions, this report serves as a foundation for advancing data-driven water resource management, disaster mitigation, and climate adaptation strategies in APEC economies.

Table of Contents

Framework for Enhancing Sustainable and Resilient Water Management in APEC Economies

Water management is a critical challenge for APEC developing economies, where climate change, rapid urbanization, and increasing water demand pose significant threats to sustainability and resilience. This project aims to enhance the capacity of APEC developing economies in long-term water management through collaborations with developed economies, fostering a collective effort to address these challenges.

To achieve this objective, we have developed a Framework for Water Management in APEC Economies (Figure 1). The Framework offers a scalable and adaptable solution leveraging open environmental data. The framework is designed to provide actionable insights, promote data-driven decision-making, and foster stakeholder engagement to ensure effective and sustainable water resource management across APEC economies.

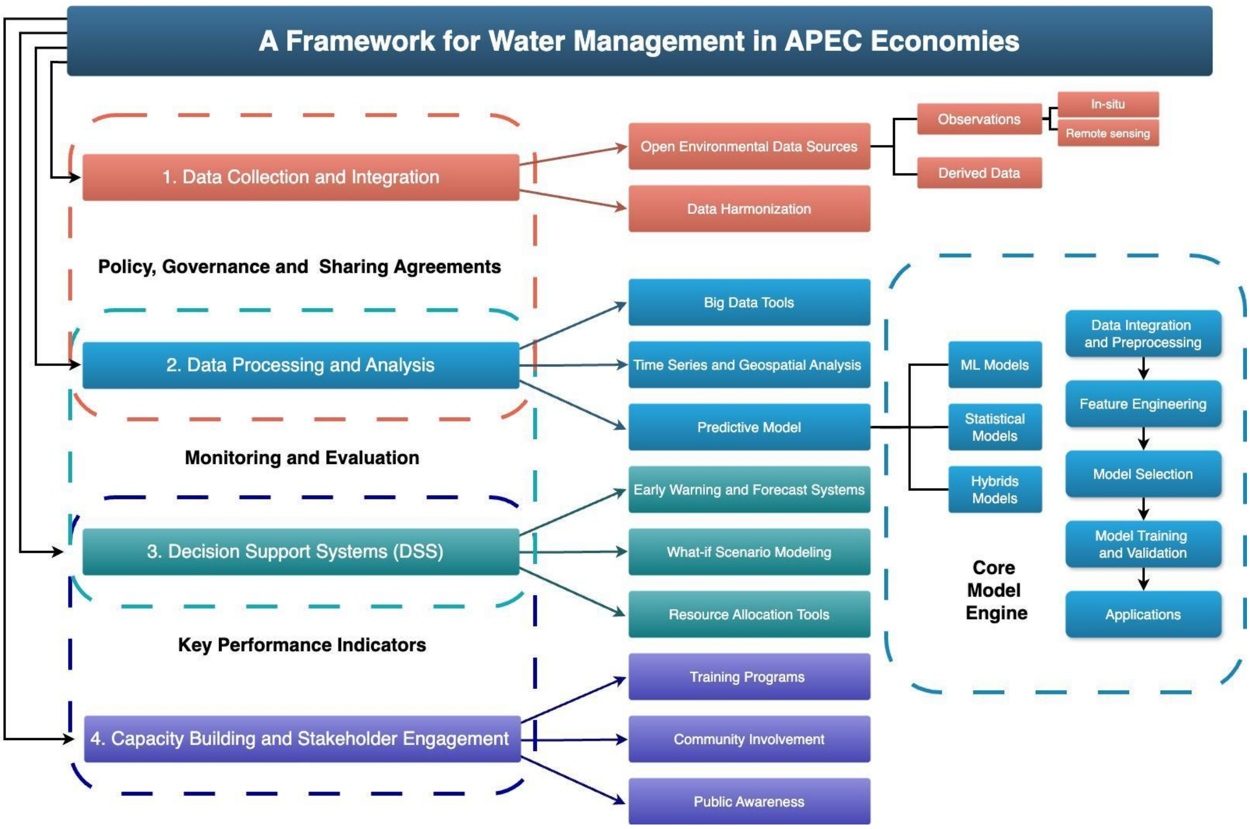


Figure 1 A Framework for enhancing sustainable and resilient water management in APEC economies

Key Components of the Framework

The framework consists of four core components, each of which addresses a critical aspect of water management and is guided by enabling factors that ensure successful implementation and long-term sustainability.

1. Data Collection and Integration

Reliable and comprehensive data is fundamental to effective water management. This component focuses on:

- **Utilizing open environmental data sources,** including **in-situ observations** (e.g., river gauges, weather stations) and **remote sensing data** (e.g., satellite imagery), to capture real-time and historical water-related data.
- **Data harmonization,** ensuring consistency, interoperability, and reliability by standardizing formats and integrating data from multiple sources to create a unified system.

Key Enablers: Policy, Governance, and Sharing Agreements

The success of data collection and integration depends on robust **policies, governance mechanisms, and data-sharing agreements**, which facilitate cross-border collaboration and promote ethical data use. Strong governance frameworks ensure standardized protocols, data security, and accessibility, fostering trust and transparency among APEC economies.

2. Data Processing and Analysis

Once data is collected, advanced analytical methods are required to derive actionable insights. This component includes:

- **Big data tools,** enabling efficient storage and processing of large datasets.
- **Time-series and geospatial analysis,** allowing for the identification of trends and spatial relationships in water resources.
- **Predictive modeling,** integrating **machine learning (ML), statistical methods, and hybrid approaches** to forecast water availability, demand, and risks.

Key Enablers: Policy, Governance, and Sharing Agreements; Monitoring and Evaluation (M&E)

Effective processing and analysis require clear governance structures to facilitate collaboration and data exchange. **Policy, governance, and sharing agreements** ensure interoperability; while **monitoring and evaluation** mechanisms assess the accuracy and reliability of data processing systems, providing continuous feedback to improve analysis methods.

3. Decision Support Systems (DSS)

Decision Support Systems (DSS) transform processed data into actionable insights, helping policymakers and stakeholders make informed decisions. This component includes:

- **Early warning and forecast systems,** offering timely alerts for floods, droughts, and other water-related risks.
- **What-if scenario modeling,** enabling stakeholders to simulate potential outcomes of different decisions or events, and assess their impacts.
- **Resource allocation tools,** optimizing water resource distribution to ensure equitable access and efficient usage.

Key Enablers: Monitoring and Evaluation; Key Performance Indicators (KPIs)

Monitoring and evaluation ensure that DSS tools remain effective and relevant, providing measurable insights into performance and efficiency. **KPIs,** such as response times, forecast accuracy, and decision-making effectiveness, provide quantitative metrics that drive continuous improvement.

4. Capacity Building and Stakeholder Engagement

A sustainable water management system requires active participation and knowledge dissemination. This component focuses on:

- **Training programs,** equipping officials, water managers, and local communities with skills to utilize data-driven tools effectively.
- **Community involvement,** ensuring local perspectives and knowledge are integrated into water management strategies.
- **Public awareness initiatives,** fostering a culture of water conservation and responsible use through outreach and communication efforts.

Key Enabler: Key Performance Indicators (KPIs)

Defining and tracking **KPIs** is crucial for assessing the effectiveness of capacity-building efforts. Metrics such as stakeholder participation rates, training completion levels, and public engagement outcomes provide tangible indicators of success and inform necessary adjustments to ensure long-term impact.

Key Focus Areas of the Framework

Core Model Engine Development

The Framework places significant emphasis on developing predictive models using a structured pipeline that includes data integration, especially the following components:

(1) **Data integration:** Integrating and preprocessing data from various sources. These include standardizing formats, cleaning incomplete or noisy data, and ensuring consistency and interoperability across datasets to create a unified and reliable input for subsequent steps.

(2) **Feature engineering:** Selecting relevant variable fields for predicting rainfall, considering meteorological regions such as tropical, subtropical, and midlatitude.

(3) **Model selection:** Choosing models suited to the physics of the region of interest. The use of machine learning, statistical techniques, or their hybrid ensures predictive models deliver accurate and reliable insights to support decision-making processes.

(4) Model training and validation: Training the model with the selected data and validating the model outputs using observations.

(5) Application: Implementing the trained model in operational systems to provide real-time predictions and support decision-making.

Public Awareness

Public engagement plays a pivotal role in ensuring the successful implementation of the Framework. Training programs, workshops, and outreach initiatives aim to educate stakeholders on best practices, increase community participation, and create a sense of ownership in water management efforts.

Implementation Guideline for the Water Management Framework

Effective water management is critical for ensuring sustainable development, mitigating climate change impacts, and addressing growing water-related challenges such as scarcity, flooding, and pollution. This **Framework Implementation Guideline** provides a structured, step-by-step approach to assessing water resource issues, building robust data infrastructure, developing decision-support systems, and scaling solutions economy-wide. By leveraging modern technologies such as the Internet of Things (IoT), machine learning models, and real-time monitoring, this framework enables informed decision-making and proactive water resource management. Furthermore, capacity-building efforts ensure that stakeholders including policymakers, technical experts, and local communities are equipped with the necessary skills to implement and sustain these strategies. Through this guideline, economies can build resilience, enhance efficiency, and foster long-term sustainability in water management.

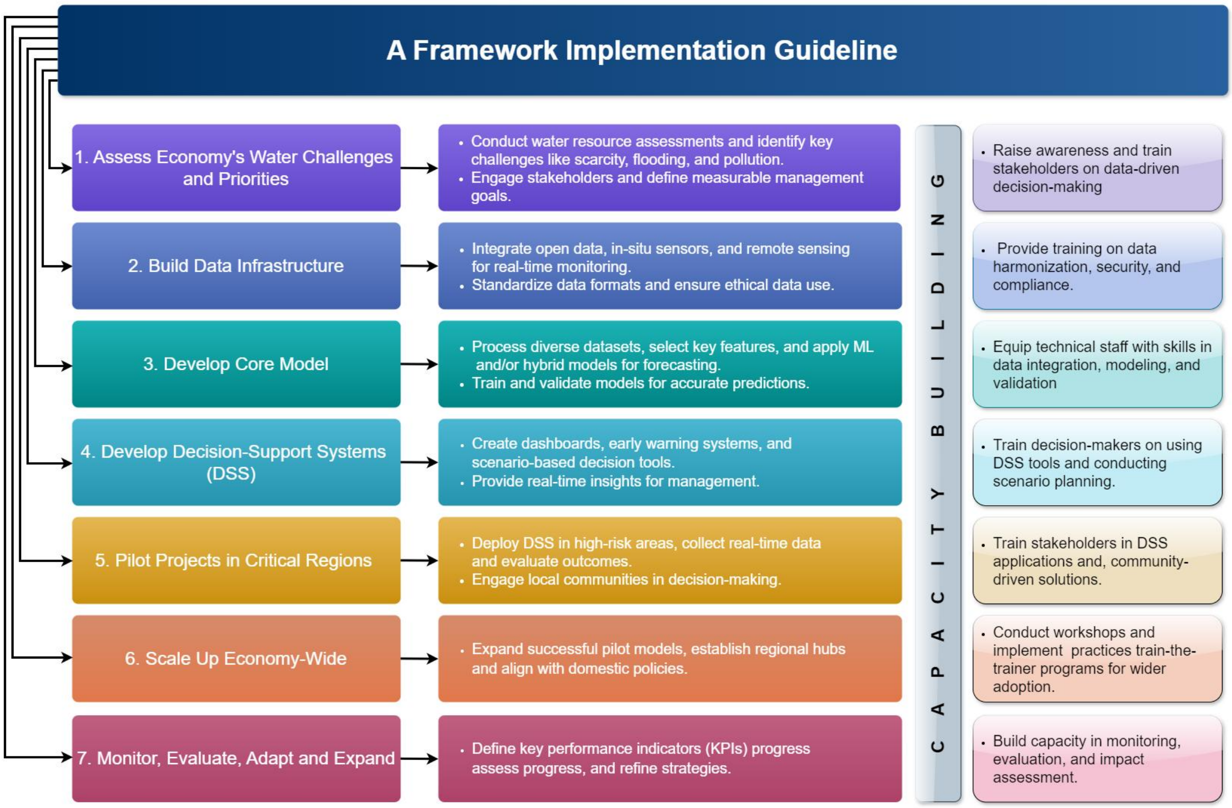


Figure 2 A Framework Implementation Guideline for enhancing sustainable and resilient water management in APEC economies

The framework implementation consists of the following steps:

1. Assessing Economy’s Water Challenges and Priorities

The first step in implementing the framework is to assess the specific water challenges and priorities of the economy. This requires conducting water resource assessments using Geographic Information Systems (GIS) and hydrological models. The assessment should identify key challenges such as water scarcity, flooding, and pollution. Stakeholder engagement is crucial in this phase to ensure diverse perspectives are included. Additionally, clear and measurable water management goals should be defined, aligning them with domestic priorities.

Key Actions:

- Conduct water resource assessments using GIS and hydrological models.
- Identify key challenges, including water scarcity, flooding, and pollution.
- Engage stakeholders to integrate their perspectives into the framework.
- Define measurable goals aligned with domestic and regional water management priorities.

Key Actions for Capacity Building:

- ○ Engage stakeholders in water resource assessment and raise awareness of water challenges.
- ○ Train stakeholders on open-data usage and inclusive decision-making processes.

2. Building Data Infrastructure

A strong data infrastructure is essential for efficient water management. This involves leveraging open data sources, IoT sensors, and remote sensing technologies to monitor water resources. Standardizing data formats for seamless integration across platforms is necessary. Ensuring data privacy, security, and ethical use is also a priority. Stakeholders must be trained in data collection, processing, and analysis tools to maximize the benefits of the infrastructure.

Key Actions:

- Utilize open data sources, IoT sensors, and remote sensing technologies.
- Standardize data formats for easy integration and interoperability.
- Ensure data privacy, security, and ethical data use.
- Provide training to stakeholders on data collection and analysis tools.

Key Actions for Capacity Building:

- ○ Provide technical training on IoT sensors, remote sensing, and cloud platforms.
- ○ Develop skills in data harmonization, secure data sharing, drafting data-sharing agreements, and effective compliance monitoring through workshops.

3. Developing Core Model

Developing a reliable core model is key to the success of the framework. This phase includes integrating and preprocessing data from multiple sources, engineering features, and selecting suitable models. These models may include machine learning algorithms, statistical techniques, or hybrid approaches to generate actionable insights for water management. Proper training, validation, and deployment of these models are required to ensure their effectiveness.

Key Actions:

- Integrate and preprocess data from diverse sources.
- Perform feature engineering to enhance model accuracy.
- Select the most appropriate models, including machine learning or hybrid approaches.
- Train, validate, and deploy models to provide useful water management insights.

Key Actions for Capacity Building:

- ○ Train technical staff in data integration, feature engineering, and model development.
- ○ Build expertise in validating and deploying machine learning, statistical, and hybrid models.

Feature engineering and **selecting high-quality data sources** are critical to developing effective decision-support systems. This workshop focuses on these key aspects, ensuring that the data integrated into dashboards and early warning systems is accurate, relevant, and actionable. The results will be reported in the following section.

4. Developing Decision-Support Systems (DSS)

Decision-Support Systems (DSS) play a crucial role in converting data into actionable insights. This includes creating dashboards with visual representations such as maps, graphs, and real-time alerts. Additionally, implementing early warning systems for floods and droughts, as well as scenario modeling for resource allocation, will help stakeholders make informed decisions.

Key Actions:

- Develop dashboards with visual elements such as maps and graphs.
- Implement early warning systems for floods and droughts.
- Use scenario modeling to simulate and analyze potential outcomes.
- Optimize resource allocation tools for equitable and efficient water usage.

Key Actions for Capacity Building:

- ○ Train policymakers and water managers to use dashboards and DSS tools.
- ○ Conduct scenario-planning workshops and build local communities's capacity to interpret and act on early warnings.

5. Pilot Projects in Critical Regions

Before scaling the framework across the economy, pilot projects should be implemented in high-risk regions. These projects will test DSS, collect real-time data, engage local communities, and evaluate outcomes. The insights from these pilot projects will guide broader scalability efforts.

Key Actions:

- Identify high-risk regions for initial pilot implementations.
- Deploy DSS tools to these regions for real-world testing.
- Collect real-time data and involve local communities in the process.
- Evaluate project outcomes to refine the framework for larger-scale deployment.

Key Actions for Capacity Building:

- ○ Train stakeholders in data collection, monitoring, and DSS tools.
- ○ Empower communities to engage in local decision making and provide feedback to improve pilot plans.

6. Scaling Up Economy-Wide

After successful pilot projects, the framework can be expanded across the economy. A domestic rollout plan should be developed, incorporating lessons learned from the pilots. Regional hubs should be established for data collection and decision-making. Additionally, aligning water management strategies with broader domestic policies and development goals will ensure a coordinated approach to long-term sustainability.

Key Actions:

- Develop a rollout plan based on pilot project outcomes.
- Establish regional hubs for data collection and decision-making.

- Align water management strategies with domestic policies and sustainable development goals.

Key Actions for Capacity Building:

- Train regional staff in data systems and DSS tools.
- Host domestic workshops to share best practices and lessons from pilot plans.
- Expand capacity through train-the-trainer programs to scale up efforts.

7. Monitoring, Evaluating, Adapting, and Expanding the System

To ensure continuous improvement, the framework requires ongoing monitoring and evaluation. Defining key performance indicators (KPIs) such as water efficiency and disaster reduction helps track progress. Regular assessments and refinements based on collected data ensure that strategies remain effective. Incorporating stakeholder feedback will further enhance the framework's adaptability.

Key Actions:

- Define KPIs to measure success (e.g., water efficiency and disaster reduction).
- Conduct regular assessments and refine strategies based on findings.
- Incorporate feedback from stakeholders to improve the framework continuously.

Key Actions for Capacity Building:

- Train stakeholders in monitoring and evaluation methods, focusing on tracking KPIs.
- Build capacity to assess water management impacts and iteratively refine frameworks using feedback.

By following this structured implementation guideline, economies can ensure a comprehensive, data-driven, and sustainable approach to water management. Furthermore, this guideline can also be applied to other environmental related problems. The Framework Implementation similarly begins with identifying the key issues of each economy, determining which datasets and variable fields are necessary for building a model, building a decision-support system, scaling it up, and monitoring as well as evaluating the model. These steps lead to enhancing resilience against water-related challenges and other environmental related issues in the economies.

Understanding the Similarities and Differences between Tropical and Midlatitude Meteorology for Framework Development

The Importance of Feature Engineering in Developing Localized Rainfall Prediction Models

A fundamental component of the framework's success stems from building a robust core model, which relies heavily on feature engineering. The first step in this process is selecting appropriate variables for rainfall prediction models. These variables originate from diverse data sources, including observational data, remote sensing data, reanalysis datasets, and model outputs. The necessity of careful feature engineering and model selection arises because existing model outputs do not always yield accurate rainfall predictions for localized regions. Many global models operate at a relatively coarse resolution (2–5 km), which is insufficient to resolve the intricate atmospheric processes necessary for cloud formation which is an essential precursor to rainfall.

While increasing the resolution of these models would improve accuracy, doing so requires extensive computational resources, posing a significant challenge for developing economies. Furthermore, many of the parameterization schemes used in existing models were developed in midlatitude regions, where atmospheric dynamics differ significantly from those in tropical climates. These parameterizations attempt to estimate small-scale atmospheric processes using mathematical equations, but their effectiveness varies across different climatic zones. Additionally, each economy or region presents unique challenges due to variations in land use, topography, and localized weather patterns. Therefore, it is crucial to develop a localized model that utilizes region-specific data and is efficient in both computational resources and processing time.

To improve the accuracy of localized rainfall predictions, it is essential to understand the similarities and differences in meteorological processes across different climate zones. One key similarity between tropical and midlatitude meteorology is that cloud formation is fundamental to rainfall prediction. Clouds are associated with updrafts and condensation, and their presence can serve as a critical predictor for rainfall. Furthermore, there are multiple types of clouds, including warm clouds, ice clouds, and mixed-phase clouds, each playing a unique role in atmospheric processes. Warm clouds, typically found at altitudes of 2–3 km, consist of condensed water droplets that reflect sunlight but produce minimal precipitation. Ice clouds, located in the upper troposphere typically 10–15 km above the surface, influence the atmospheric radiation budget and can impact convective cloud formation over time. Finally, mixed-phase clouds, which contain both liquid droplets and ice particles, are often associated with deep convection and complex interactions among hydrometeors.

Despite these universal cloud processes, significant differences exist between tropical and midlatitude meteorology. One key distinction lies in the thermodynamics and large-scale atmospheric environment influencing cloud development. Cloud processes generally occur over short timescales, on the order of days, while seasonal rainfall prediction operates on longer timescales. However, understanding the dynamics and thermodynamics of mixed-phase clouds remains crucial for selecting the most relevant variables for accurate seasonal predictions. By incorporating these considerations into the feature engineering process, localized models can better capture atmospheric conditions unique to each region, ultimately improving the accuracy and efficiency of rainfall forecasts.

Distinguishing Tropical and Midlatitude Meteorology for Improved Rainfall Prediction

A fundamental component of the success in a core model development is the recognition of the distinct characteristics of the region for which to build the model. In other words, distinguishing the differences between tropical and midlatitude meteorology is crucial in shaping how rainfall prediction models should be developed and refined, as the governing atmospheric processes vary significantly between the two zones. Understanding these distinctions is essential for designing accurate, region-specific models that enhance the reliability of precipitation forecasts.

The midlatitude, or the area poleward of 30°N/S, is characterized by a strong horizontal temperature gradient, with warmer air near the equator and colder air near the poles. This temperature contrast gives rise to jet streams, the narrow bands of strong winds in the upper atmosphere (around 300 hPa) that drive large-scale weather systems. Because midlatitude weather is predominantly influenced by synoptic-scale processes, numerical weather prediction models tend to perform well in these regions. The predictability of midlatitude weather arises from the well-defined frontal systems and large-scale dynamic forcing mechanisms that govern atmospheric circulation, making standard numerical and statistical modeling approaches relatively effective.

In contrast, the tropics, located equatorward of 30°N/S, operates under vastly different conditions. The tropics exhibits a weak temperature gradient (WTG) due to more uniform solar heating, meaning that temperature contrasts are not the primary driver of weather systems. Instead, tropical weather is governed by continuous interactions between clouds, moisture availability, and large-scale circulations, leading to complex feedback mechanisms that make rainfall prediction significantly more challenging. The role of cloud microphysics is much more pronounced in tropical systems, requiring different modeling approaches than those used in midlatitude regions.

Many variables that serve as reliable precipitation proxies in the midlatitudes, such as Convective Available Potential Energy (CAPE), are far less useful in the tropics. Instead, tropical rainfall is more strongly correlated with relative humidity (RH) and moisture convergence, necessitating an adjustment in the choice of predictive variables for ML-based models. The differences between tropical and midlatitude meteorology are summarized in Table 1.

Feature	Midlatitude	Tropical
Location	Poleward of 30°N/S	Equatorward of 30°N/S
Temperature Gradient	Strong horizontal gradient (warmer near the equator, colder near the poles)	Weak temperature gradient (WTG) due to uniform solar heating
Dominant Weather Drivers	Jet streams drive large-scale weather systems	Continuous interactions between clouds, thermodynamics, and circulations
Predictability	More predictable due to synoptic-scale processes	Less predictable due to complex feedback mechanisms
Rainfall Modeling	- Traditional numerical models perform well; - Higher resolution is better; - Alternative approaches such as ML-based models are helpful and most development is for this region	- Requires higher resolution; - Requires developments of dynamical, thermodynamic and cloud microphysics schemes targeted to the region; - Alternative approaches such as ML-based models are helpful but localized development is required

Table 1: Key differences between tropical and midlatitude meteorology that are important for rainfall prediction

Understanding these differences is crucial for developing accurate models. Many variables that serve as strong proxies for precipitation in midlatitude regions, such as Convective Available Potential Energy (CAPE), are not reliable in the tropics. Instead, relative humidity (RH) and moisture convergence are stronger indicators of precipitation probability in tropical climates. The workshop emphasized the need for incorporating region-specific meteorological variables into ML-based post-processing models to improve rainfall prediction accuracy.

Mathematically, CAPE is given by:

$$CAPE = \int_{LFC}^{EL} g \frac{T_{parcel} - T_{env}}{T_{env}} dz$$

where LFC is the level of free convection, EL is the equilibrium level, g is gravitational acceleration, T_{parcel} is the temperature of the air parcel, and T_{env} is the environmental temperature. While this is a strong predictor in midlatitude systems, tropical convection is more influenced by moisture availability and entrainment.

In other words, another key takeaway is the role of buoyancy and dilution in cloud formation. In midlatitude systems, buoyancy largely determines whether convection occurs, making CAPE a useful predictive variable. However, in tropical systems, dilution of clouds, or the entrainment, is equally important, requiring additional considerations such as entraining CAPE and mid-tropospheric humidity levels. The implication of this finding is that tropical rainfall models should incorporate parameters that reflect these interactions rather than relying solely on energy-based predictors.

A simplified equation for the "actual buoyancy" responsible for convective systems in the tropical atmosphere atmosphere can be described as:

$$Actual\ Buoyancy\ =\ Undiluted\ Buoyancy\ -\ Entrainment.$$

The **undiluted buoyancy** is a proxy for CAPE. It is proportional to the moist static energy (MSE) at the surface minus the saturated MSE in the mid-troposphere:

$$Undiluted\ Buoyancy\ \propto\ MSE_{sfc} - satMSE_{mid-trop}$$

Moist static energy is a combination of the potential energy, heat, and moisture content of the air at a level of interest. Mathematically, it can be written as:

$$MSE\ = gz\ +\ C_p T\ +\ L_v\, q$$

where z is the geopotential height above sea level, C_p is the specific heat at constant pressure, T is the absolute air temperature, L_v is the latent heat of vaporization, and q is water vapor specific humidity. To compute the saturated MSE, the water vapor specific humidity is simply replaced by the saturated water vapor specific humidity, which expresses how much water vapor the air would have under saturation, a quantity depending on the temperature. Simply put together,

$$Undiluted\ Buoyancy$$
$$\propto\ C_p(T_{sfc} - T_{500hPa}) + L_v\,(q_{sfc}\ - q_{sat}(T_{500hPa}) - gz_{500hPa}$$

where the subscript *sfc* denotes the near-surface air, 500 hPa is a representative level of the mid-troposphere, roughly 5 km above the surface.

The air will rise when the undiluted buoyancy is positive. To satisfy this condition often requires the following:

(1) the surface is warmer than the mid-troposphere,

(2) the surface is moister than the saturated mid-troposphere, and

(3) the sum of the two quantities above exceeds the potential energy of the air in the mid-troposphere.

In the midlatitude, these three terms are generally sufficient to cause positive undiluted buoyancy and hence convection. However, the **entrainment**, or the dilution, is also very important in the tropical atmosphere. Excessive dilution of the buoyant updrafts can simply “kill” the clouds, leading to an environment with high CAPE but no rain as in the heatwave.

The entrainment is inversely proportional to relative humidity:

$$Undiluted\ Buoyancy \propto 1 - RH$$

where RH is relative humidity, a measure of how much water vapor is present compared to the maximum possible amount.

Research has shown that in the deep tropics, precipitation is more directly related to RH rather than CAPE (Bretherton et al., 2004). Furthermore, CAPE itself is inversely proportional to RH (Raymond et al., 2015), indicating that traditional CAPE-based approaches to predicting rainfall may be ineffective in tropical regions. Instead, an entraining CAPE formulation should be considered:

$$Entraining\ CAPE = CAPE - Integrated\ Dilution$$

This highlights the importance of both the vertical temperature gradient and the mid-tropospheric humidity for effective cloud and rain formation.

Additionally, in tropical meteorology, weak temperature gradients (WTG) play a crucial role. Unlike in the midlatitudes, where strong temperature contrasts drive weather systems, the tropics experiences a more uniform temperature distribution due to weak Coriolis force. Gravity waves redistribute solar heating, leading to overturning circulations that cause air to rise, condense, and form clouds, as shown in Figure 3.

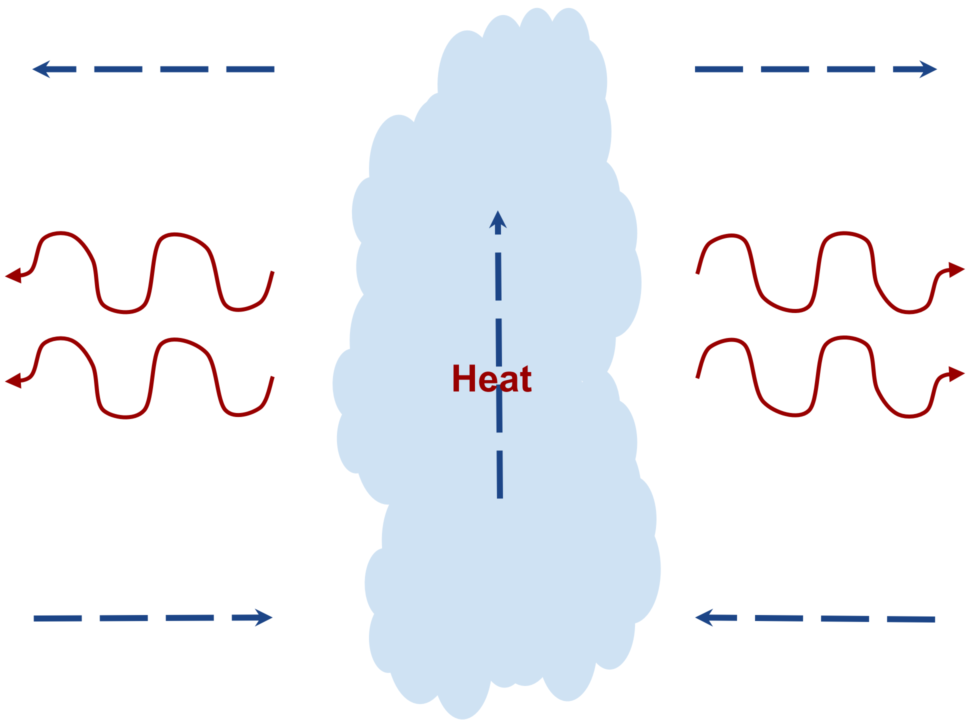


Figure 3: A schematic showing the relationship between a convective cloud, an overturning circulation (dashed blue arrows) or the air movement, and the heat generation and dissipation (red arrows).

The column moisture budget equation describes how moisture changes in a vertical atmospheric column:

$$\left\langle \frac{\partial q}{\partial t} \right\rangle + \langle V \cdot \nabla_h q \rangle + \left\langle \omega \cdot \frac{\partial q}{\partial p} \right\rangle = E - P$$

where:

- q is specific humidity,
- V is horizontal wind (with zonal and meridional components),
- ω is vertical pressure velocity,
- E is evaporation,
- P is precipitation,
- ∇_h denotes the horizontal gradient,
- t and p represent time and pressure, respectively, and
- $\langle \; \rangle$ represents the vertical integration over the atmospheric column.

This equation highlights that local precipitation results from either local evaporation or the horizontal convergence of moisture within the atmospheric column. In general, increased horizontal moisture convergence leads to enhanced precipitation. Moisture advection from wetter to drier regions also contributes to rainfall probability. In the tropics, equatorial waves play a significant role in transporting large-scale moisture

(Mayta and Adames-Coralliza, 2024), alongside other large-scale circulations such as the El Niño-Southern Oscillation (ENSO), the Madden-Julian Oscillation (MJO), and the Indian Ocean Dipole (IOD). Given the critical influence of these processes on tropical precipitation, incorporating total moisture flux convergence or key components of the associated moisture fluxes into predictive models can significantly improve rainfall forecasting in tropical climates.

Observational Data and Model Outputs for Core Model Development

A successful core model for rainfall prediction and water management requires an understanding of the region’s specific meteorological characteristics and the use of accurate, comprehensive data. Various data sources are crucial for model development, including observational data, remote sensing products, reanalysis datasets, and high-resolution model outputs.

Observational Data

Observational data provide direct measurements from weather stations, river gauges, and ground-based radar systems. These in-situ measurements are the most accurate and accessible forms of data, often serving as a reference for evaluating satellite-based data and model outputs. While relatively inexpensive, they require regular maintenance, and data disturbances or shortages can impact model training. Installing such measurement tools in remote or mountainous regions is particularly beneficial for topographical rainfall monitoring and prediction.

Additionally, vertical profile data from radiosondes or weather balloons offer valuable atmospheric soundings, measuring temperature, humidity, and wind at different altitudes. These are essential for calculating moist static energy (MSE) and relative humidity (RH), both fundamental to atmospheric buoyancy. Although radiosonde data collection is costlier than in-situ measurements, launching balloons multiple times per day is crucial for capturing diurnal atmospheric variability.

Remote sensing data

Remote sensing data, which can be ground- or satellite-based, provide broader coverage than in-situ measurements but require calibration using station-based or radiosonde data. This calibration process often involves radiative transfer codes, which must be obtained from data providers when necessary. Remote sensing techniques are classified as passive or active. **Passive remote sensing** relies on naturally occurring energy, such as sunlight, to record reflected radiation. This category includes satellite imagery, infrared sensors, and water vapor sensors. In contrast, **active remote sensing** systems generate their own energy source to illuminate a target and measure the reflected radiation. Examples include lidars, radars, and Doppler radars.

Lidar, which stands for ‘Light Detection and Ranging,’ is a remote sensing technique that employs laser pulses to measure atmospheric properties such as wind speed, direction, and aerosol profiles. These data are essential for weather forecasting, climate modeling, and air quality management. Lidar works by emitting laser pulses and analyzing the scattered or absorbed light to gather information about the atmosphere. It measures aerosol profiles to assess the vertical distribution of dust, smoke, and pollutants, which are crucial for understanding air quality and climate. *Doppler lidar* utilizes the Doppler effect to measure wind speed and direction, while Raman lidar can provide temperature and water vapor profiles. Lidar is also valuable for studying cloud height, thickness, and composition. An example of an open lidar dataset is the U.S. National Aeronautics and Space Administration's (NASA) MPLNET, a federated network of Micro-Pulse Lidar (MPL) systems designed to measure aerosol and cloud vertical structure, and boundary layer heights. The sensors are located at various locations globally, mostly through collaborations with local research institutes who share the data through the MPL network system.

Radar, which stands for 'Radio Detection And Ranging,' is another form of active remote sensing. Radar systems come in various types based on wavelength. *Millimeter-wave (mmWave) radar*, with a wavelength of 1 to 10 mm, is highly sensitive to small particles like cloud droplets, making it useful for detecting fog and studying cloud properties. However, its short wavelengths result in strong attenuation by rain, limiting its range. On the other hand, *centimeter-wave (cmWave) radar*, with wavelengths of 5 to 10 cm, penetrates precipitation better, making it ideal for storm warnings and precipitation studies. *S-band radars (10 cm)* are particularly effective for storm tracking, whereas *C-band radars (5 cm)* are useful for precipitation monitoring but can experience attenuation in heavy rain.

Doppler radar, a specialized type, measures the velocity and direction of precipitation particles using the Doppler effect. It provides critical data on storm structure, rainfall intensity, and severe weather detection. Additionally, *dual-polarization radar* enhances precipitation classification by distinguishing between rain, hail, snow, and ice pellets. This technology transmits and receives signals in both horizontal and vertical polarization, offering more detailed insights into precipitation characteristics.

Satellite remote sensing datasets provide another valuable source of atmospheric data. The Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG) algorithm, developed by NASA, combines satellite observations and ground-based data to estimate precipitation over the Earth's surface at a high temporal resolution. Similarly, the Clouds and the Earth's Radiant Energy System (CERES) dataset measures Earth's radiation budget and cloud properties, which influence precipitation patterns. Advances in remote sensing technology now allow the measurement of atmospheric states from space, reducing reliance on ground-based observations.

Traditionally, vertical velocity measurements required ground-based instruments such as Doppler radar or mmWave radar. However, a recent study showed that satellite-based techniques can now enable the measurement of clear-air vertical motions (Poujol and Bony, 2024), though further research is needed to assess their reliability for cloud-scale processes.

Reanalysis Data

Reanalysis datasets, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) Atmospheric Reanalysis version 5 (ERA5), the National Centers for Environmental Prediction (NCEP), the Japanese 55-year Reanalysis (JRA-55), and the Modern-Era Retrospective analysis for Research and Applications (MERRA), provide long-term atmospheric data by combining historical observations with model simulations. They have great spatial and temporal coverage, but may lack the resolution needed for complicated topography. These datasets are invaluable for climate trend analysis, historical evaluations, and machine learning-based predictive models. However, their accuracy depends on the availability and quality of input observations, underscoring the importance of improving ground-based data collection.

High resolution models

High-resolution weather models, though computationally expensive, provide detailed forecasts used by meteorological services such as ECMWF and the Icosahedral Nonhydrostatic (ICON) model. Some high-resolution model outputs are publicly available in near real-time, but accessing and storing them requires significant resources.

Projects like the Dynamics of the Atmospheric General Circulation Modeled on Non-hydrostatic Domains (DYAMOND) offer high-resolution simulations validated against observational data (Stevens et al., 2019). But because of the computational expenses, they are run only for 40-day (1 August–10 September 2016). Nonetheless, they offer very high resolution outputs that allow one to understand atmospheric processes important for rainfall formation and thus are very useful for research purposes.

Take-home points:

- One of the key discussions in the workshop was the importance of calibrating satellite-derived data with in-situ measurements. Proper calibration ensures that satellite data serve as a consistent and reliable source for long-term monitoring and predictive modeling.
- Maintaining a dense and well-calibrated network of in-situ stations is essential for accurate modeling, particularly in complex terrains such as mountainous regions. The integration of radar, lidar, and radiosonde data further enhances rainfall observation and prediction models.
- Observed data can also be used in data assimilation, a technique that merges observational data with numerical model predictions to improve weather forecasting accuracy.
- While models are powerful tools for understanding and predicting atmospheric processes, they are not infallible. Proper interpretation requires a strong

foundation in meteorology and access to high-quality observational data for verification and refinement.

Ultimately, the effectiveness of predictive models hinges on integrating diverse data sources, improving ground-based observations, and advancing computational techniques. The more comprehensive and region-specific the data, the better the model performance, leading to improved decision-making in water resource management and disaster preparedness.

Use Cases of the Framework: Feature Engineering in Different Economies

Feature engineering for machine learning (ML) models varies by region due to differences in climate systems, geography, and atmospheric dynamics. One key consideration agreed upon by most economies is that the **time frame** of prediction plays a crucial role in selecting the appropriate variables. As highlighted by an expert speaker, time-dependent variables influence both short-term and long-term weather forecasting.

In general:

- **Temperature fluctuations** are critical for large-scale and fast-moving systems, such as tropical cyclones (hurricanes, typhoons, and depressions), El Niño-Southern Oscillation (ENSO), and the Indian Ocean Dipole (IOD).
- **Water vapor fluctuations** are more relevant for intermediate-scale and slow-moving weather systems.

Below are examples of feature engineering tailored to specific regions, as discussed by expert speakers and participants from each economy:

United States (USA)

Feature engineering in the USA incorporates a range of atmospheric and meteorological variables, particularly for severe weather events such as convective storms, hurricanes, and seasonal precipitation anomalies.

- Convective Available Potential Energy (CAPE)
- Humidity
- Vertical velocity
- Cloud-top height (or pressure)
- Cloud optical thickness
- Past precipitation

These variables help in predicting thunderstorms, tornadoes, and large-scale climate patterns that impact different parts of the economy.

China; Hong Kong, China; and Korea (East Asia / Mid-Latitude Regions)

These economies experience a mix of monsoon systems, typhoons, and large-scale atmospheric interactions. Baroclinicity, or the presence of strong temperature gradients, plays a key role in shaping weather patterns. Feature engineering includes:

- Sea surface temperature
- Air temperature
- Pressure
- Humidity

- Geopotential height
- 850 hPa wind
- Soil moisture
- Climate indices such as ENSO, IOD, and MJO (Madden-Julian Oscillation)
- Other factors influencing monsoon and typhoon activity

These features allow models to capture both short-term atmospheric variations and long-term climate influences affecting weather predictability.

The Philippines and Indonesia (Tropical Islands)

As tropical island economies, the Philippines and Indonesia face challenges related to extreme rainfall, tropical cyclones, and seasonal weather variability. Feature engineering focuses on ocean-atmosphere interactions and convective processes.

- Sea surface temperature
- Upper air temperature
- Convective Available Potential Energy (CAPE)
- Low-level humidity
- Mid-level humidity
- Wind profiles throughout the troposphere

These variables are particularly useful for modeling tropical cyclone intensity, monsoon rains, and localized convective storms. One prominent difference between the Philippines and Indonesia is the season of rainfall as they are located in different hemispheres. While the variables for feature engineering may be similar for both economies, separate models are required to train the data for rainfall prediction.

Viet Nam (South China Sea Coastal Region)

Viet Nam's climate is influenced by monsoon activity, ENSO, and tropical cyclones in the South China Sea. Feature engineering in this region integrates multiple factors affecting seasonal and extreme weather events.

- Sea surface temperature
- Upper air temperature
- Humidity
- Wind
- ENSO Index
- Monsoon index

These features help capture the interactions between large-scale climate oscillations and regional weather patterns.

Chile (Mountainous, Mid-Latitude Climate)

Chile presents a unique challenge for numerical weather modeling due to its long and narrow geopolitical boundary stretching north-south across the South American continent. This geographical constraint requires numerical simulations to be performed over a long and narrow domain. Additionally, Chile's complex topography, with the Andes mountains to the east and the Pacific Ocean to the west, makes it challenging to develop high-resolution weather models that accurately capture the interaction between atmospheric processes and terrain-driven weather patterns.

To account for these complexities, ML-based models are helpful for Chile. For this climate pattern, feature engineering in Chile relies on:

- Total precipitable water
- Horizontal wind
- Vertical wind
- High-resolution topography (mountainous terrain influence)

These variables help improve the accuracy of precipitation forecasts, particularly for orographic rainfall and extreme weather events like atmospheric rivers, which are common in the region.

Thailand (Tropical Monsoon Climate)

Thailand experiences diverse weather patterns influenced by monsoon circulations, topography, and large-scale climate indices. Feature engineering includes a mix of atmospheric and surface-related variables.

- Sea level pressure
- Temperature
- Humidity
- Wind
- Topography
- Soil moisture and runoff
- ENSO index

These variables are critical for forecasting seasonal flooding, droughts, and monsoon-driven rainfall variability.

Suggestions for the Framework

1. Validation Using Precipitation Anomalies

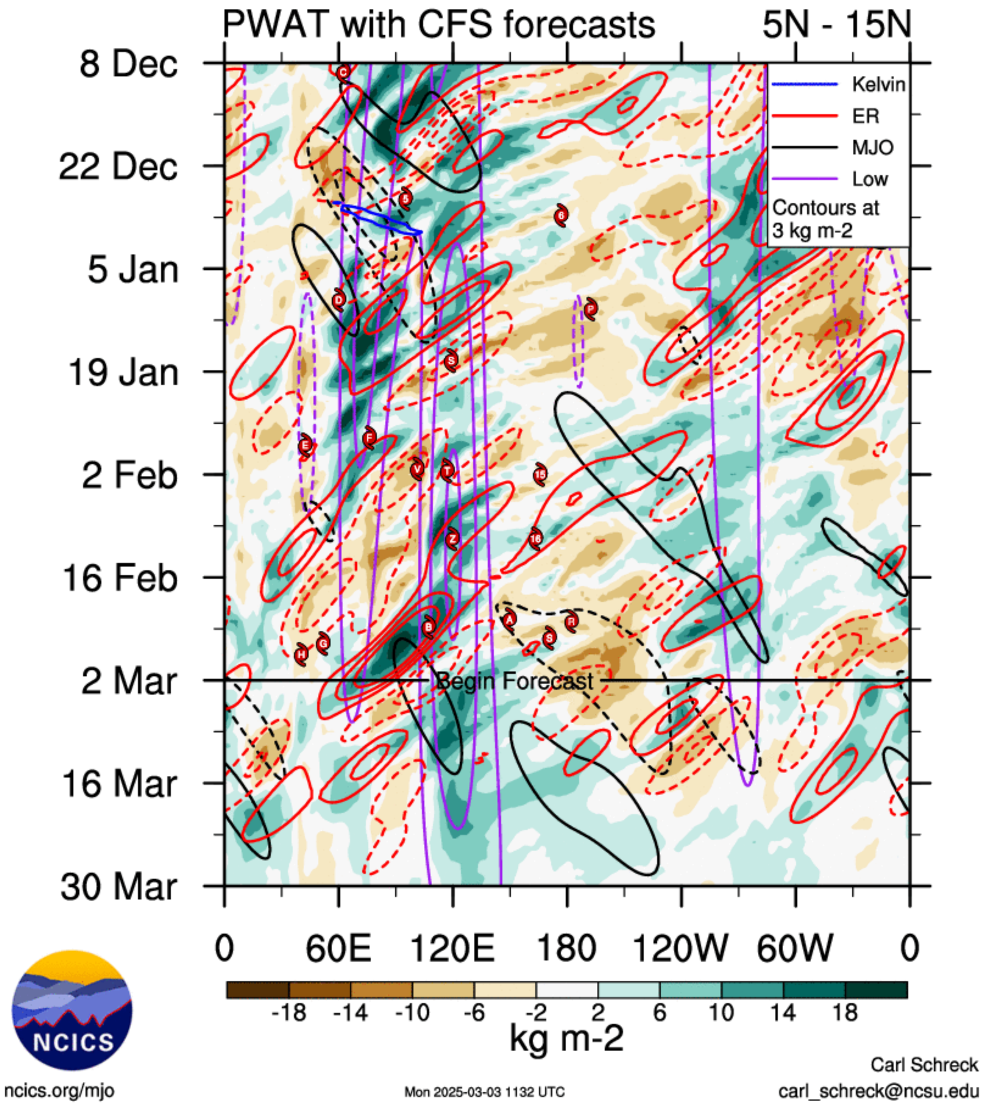
When developing a machine learning (ML) model for precipitation forecasting, it is essential to validate the model using precipitation anomalies rather than absolute precipitation values. Precipitation anomalies provide a measure of deviations from the long-term average, allowing for a clearer assessment of how well the model captures variability and extreme weather patterns. By focusing on anomalies, the model evaluation becomes less sensitive to biases in absolute precipitation amounts and better suited for detecting seasonal and interannual changes. Additionally, using anomaly-based validation helps improve model generalization across different regions and time periods, making it more robust for climate applications.

2. Importance of Climate Zones in ML Model Development

Climate zones significantly impact the development and performance of ML models for precipitation. Different geographic regions have distinct climate characteristics that pose unique challenges for model training and validation. For example, Chile's elongated north-south orientation, combined with its mountainous terrain, results in sharp climate gradients, from arid conditions in the north to temperate and wet conditions in the south. This variability requires an ML model that can account for both local and large-scale atmospheric drivers of precipitation. Similarly, Thailand’s diverse geography, with mountainous terrain in the north, a flat plateau in the central region, and a coastal peninsula in the south, necessitates an ML model that can adapt to multiple precipitation regimes. Developing region-specific models or incorporating climate zone classification into ML training can improve prediction accuracy across different geographic settings.

3. Equatorial Waves and Seasonal Precipitation Forecasting

Equatorial waves play a crucial role in seasonal precipitation forecasting, particularly in tropical regions. These large-scale atmospheric waves influence moisture transport, convective activity, and storm development. Hovmöller diagrams, which visualize the propagation of equatorial waves over time, are particularly useful for tracking their impact on precipitation. For example, analysis of the Equatorial Rossby wave during the workshop (Feb 24–25, 2025) revealed that it coincided with high precipitable water in the atmosphere (PWAT), directly contributing to a rainfall event in Bangkok, Thailand. By integrating equatorial wave dynamics into ML models, it is possible to improve seasonal forecasts by capturing these large-scale atmospheric interactions.



4. Expanding Ground-Based Observational Coverage

Higher coverage and resolution of ground-based observational data are essential for improving precipitation models. While sophisticated weather stations provide high-quality data, even inexpensive instruments can contribute valuable measurements, particularly in regions with sparse observational networks. Indonesia's use of ATHUS and ModATHUS rain gauges serves as a strong example of how cost-effective solutions can enhance data availability. A well-distributed observational network helps with model calibration, improves data assimilation in numerical weather prediction, and allows for better validation of satellite and ML-based estimates. Expanding observational networks should be a priority, particularly in regions with complex topography or variable rainfall patterns.

5. Community Engagement in Disaster Preparedness

Community engagement is a critical component of effective weather and climate modeling, particularly in relation to disaster prevention and response. Engaging communities in early warning systems, disaster preparedness, and map literacy ensures that scientific advancements translate into real-world benefits. Local stakeholders must be involved in developing and disseminating forecast products, making them accessible and actionable for decision-making. Training on how to interpret weather maps and hazard forecasts can empower communities to take proactive measures, reducing the impact of extreme weather events. Successful community engagement requires collaboration between scientists, policymakers, and local leaders to tailor solutions to specific regional needs.

6. Role of Radiative Heating in Cloud Formation

Radiative heating is a fundamental process in cloud formation, particularly in the tropics. Cloud radiative heating leads to moisture convergence, where heating inside convective clouds causes air to rise. Due to mass continuity, this results in convergence at the surface and divergence at higher altitudes, reinforcing convective circulation. In humid tropical environments, such overturning circulations sustain cloud development, leading to sustained precipitation. ML models aiming to predict tropical rainfall should incorporate radiative heating effects to better simulate convective processes and their impact on moisture distribution.

7. Surface Energy Flux and Land-Ocean Differences

Surface energy flux plays a crucial role in shaping weather patterns and precipitation, and modern models resolve these processes with increasing accuracy. One of the key distinctions between land and ocean is the role of the boundary layer, the lowest few kilometers of the atmosphere. Over land, the boundary layer exhibits strong diurnal variation, driven by solar heating during the day and cooling at night. This variability influences the depth of convective clouds and the intensity of rainfall. In contrast, oceanic regions experience weaker diurnal variations because of the ocean’s thermal inertia, leading to more continuous but less intense convection. ML models should account for these land-ocean differences to improve precipitation predictions, particularly in coastal regions where interactions between land and sea drive complex weather patterns.

8. Aerosols and Their Influence on Precipitation

Aerosols influence precipitation by acting as cloud condensation nuclei, facilitating the transition from water vapor to liquid droplets. However, their impact is secondary compared to other atmospheric factors such as moisture availability and large-scale dynamics. Small aerosol particles can enhance cloud formation but may suppress rainfall by preventing droplet growth into raindrops. Conversely, larger particles can lead to more efficient rainfall production by promoting coalescence. Notably, fine particulate matter (PM2.5) is too small to effectively seed rain clouds, meaning its presence does not directly lead to increased precipitation. While aerosols are an important consideration, ML models should prioritize moisture transport, radiative heating, and surface energy fluxes as primary drivers of precipitation variability.

Summary

This report outlines a comprehensive Framework for Enhancing Sustainable and Resilient Water Management across APEC economies, emphasizing the critical role of open environmental data and region-specific adaptations. The Framework provides a structured, scalable approach designed to foster sustainability and resilience in water resource management amidst challenges such as climate change, rapid urbanization, and increasing water demand. It comprises four key components including Data Collection and Integration, Data Processing and Analysis, Decision Support Systems (DSS), and Capacity Building and Stakeholder Engagement each supported by essential enabling factors, including robust governance structures, clearly defined policies, comprehensive monitoring and evaluation systems, and strategic use of Key Performance Indicators (KPIs).

To implement the framework, a detailed Implementation Guideline has been developed, providing a step-by-step roadmap for economies to effectively assess their unique water challenges, build robust data infrastructures, develop reliable predictive models, and establish decision-support tools tailored to their specific needs. This guideline emphasizes the importance of localized data infrastructure leveraging IoT, remote sensing, and machine learning technologies. It further advocates for pilot projects in high-risk regions, followed by a methodical scale-up strategy to ensure comprehensive domestic adoption. Continuous monitoring, evaluation, and stakeholder feedback mechanisms are embedded throughout the guidelines to facilitate adaptive management, enabling economies to iteratively refine and enhance their water management strategies.

Within this strategic context, the two-day APEC Workshop conducted in Bangkok provided a collaborative platform for participants to explore and refine critical technical components essential to the Framework's success, notably focusing on localized rainfall prediction through advanced feature engineering. Recognizing the unique meteorological characteristics of different APEC economies, discussions emphasized the importance of localized data collection and model customization to improve predictive accuracy.

Building upon this foundation, two key themes guided the workshop: (1) **feature engineering**, where participants explored how meteorological differences between tropical and midlatitude regions influence rainfall prediction, and (2) **the role of open environmental data**, which supports the development and refinement of predictive models. Experts highlighted that **regional meteorology must be well understood** to ensure successful model development, as climate zones, topography, and atmospheric dynamics all play critical roles in precipitation forecasting.

A major takeaway was the necessity of **high-resolution, high-coverage observational data**. While advanced remote sensing products and reanalysis

datasets are valuable, ground-based measurements remain crucial, particularly in complex terrains. Participants discussed cost-effective observational tools, such as the ATHUS and ModATHUS rain gauges used in Indonesia, and emphasized the importance of integrating **radar, lidar, and radiosonde data** to enhance model performance.

The workshop also addressed **regional challenges in numerical weather modeling**. For example, **Chile's long and narrow geographical domain** presents computational challenges for high-resolution simulations, while Thailand's diverse landscape, from mountainous terrain to coastal regions, requires tailored approaches to capture localized weather patterns. Furthermore, discussions on **equatorial waves, surface energy fluxes, and radiative heating** highlighted their influence on seasonal precipitation patterns and their integration into ML models.

By combining **meteorological expertise with ML-driven post-processing techniques**, the framework aims to provide **real-time, data-informed insights** for water resource management, disaster preparedness, and climate adaptation. The workshop reinforced that **a one-size-fits-all approach is inadequate**. Each economy must refine its models based on region-specific meteorological factors and observational capabilities. Future research should focus on improving **feature selection techniques, expanding observational networks, and leveraging open environmental data** to enhance the accuracy and reliability of predictive models.

Appendix:

Open Environmental Data (OED) Sources for Rainfall Prediction and Water Management

Data Source	Type	Description	Region of Use	URL
European Centre for Medium-Range Weather Forecasts **(ECMWF)** Open Data	Model output	Numerical weather prediction models with precipitation and atmospheric data	Global	https://www.ecmwf.int/en/forecasts/dataset/open-data
NOAA Global Forecast System (**GFS**)	Model output	Weather prediction model providing global atmospheric forecasts	Global	https://nomads.ncep.noaa.gov/
Fifth generation of ECMWF Atmospheric Reanalysis **(ERA5)**	Reanalysis	High-resolution global climate data, including precipitation, temperature, and wind fields	Global	https://cds.climate.copernicus.eu/
Japanese 55-year Reanalysis (**JRA-55**)	Reanalysis	Long-term atmospheric reanalysis dataset	Global	https://jra.kishou.go.jp/JRA-55/index_en.html
Integrated Multi-satellitE Retrievals for GPM **(IMERG)**	Remote Sensing	High-resolution precipitation estimates from NASA’s GPM mission	Global	https://gpm.nasa.gov/data/imerg
APHRODITE Precipitation Dataset	In-situ & Gridded	Long-term, high-resolution daily precipitation data based on rain gauge observations	Asia	http://www.chikyu.ac.jp/precip/

Data Source	Type	Description	Region of Use	URL
North Carolina Institute for Climate Studies (**NCICS**) **Tropical Monitoring**	Remote Sensing & Reanalysis	Monitoring equatorial waves for seasonal rainfall forecasting	Global – Focused on the tropics	https://ncics.org/portfolio/monitor/mjo/
Climate Hazards Group InfraRed Precipitation with Station Data (**CHIRPS**)	Remote Sensing & In-situ	Precipitation dataset combining satellite and station data	Global	https://www.chc.ucsb.edu/data/chirps
Clouds and the Earth's Radiant Energy System (**CERES**) Data Products	Remote sensing	Satellite-based observations of Earth's radiation budget and cloud properties, including hourly, daily, and monthly averages of radiative fluxes and cloud properties at various spatial scales	Global	https://ceres.larc.nasa.gov
Coupled Model Intercomparison Project Phase 6 (**CMIP6**)	Model output	The most recent global climate model outputs available (as of the date of the workshop); various variables including precipitation, temperature, humidity, wind, etc.	Global	https://pcmdi.llnl.gov/CMIP6/
The World Climate Research Programme (WCRP) Coordinated Regional Climate Downscaling	Model output	High-resolution regional climate downscaling outputs from CMIP models	Regional (multiple regions available)	https://cordex.org/

Data Source	Type	Description	Region of Use	URL
Experiment (**CORDEX**)				
Centre for Environmental Data Analysis (**CEDA**)	Data Repository	Environmental data, including climate and weather observations	Global	https://www.ceda.ac.uk
DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains (**DYAMOND**)	Model output	Very high-resolution (storm-resolving) global model outputs, available for 40 days during the boreal summer and boreal winter; various variables including precipitation, temperature, humidity, wind, etc. Boreal summer: 1 August 2016 - 10 September 2016 Boreal winter: 20 January 2020 - 1 March 2020	Global	https://easy.gems.dkrz.de/DYAMOND/index.html
International Centre for Water Hazard and Risk Management (**ICHARM**)	Hydrology & Climate Data	Water-related disaster risk management and forecasting	Asia-Pacific	https://www.pwri.go.jp/icharm/special_topic/20211029_aogeo_awci.html

Data Source	Type	Description	Region of Use	URL
ATHUS and **ModATHUS** Rain Gauges	In-situ Observations	Low-cost rain gauge network for precipitation monitoring	Indonesia	No public URL
Web-based Hydrological Assessment System (**WHAS**)	Hydrology & Weather Observational Data	Real-time hydrological and weather monitoring	Indonesia	https://www.whas.web.id
Flash Flood Warning System (**FFWS**)	Hydrology & Flood Monitoring	Global flood forecasting and early warning system	Global	https://wmo.int/media/update/early-warnings-all-developments-hydrology
Center for Climate and Resilience Research Meteorological dataset (**CR2MET**)	Observational data	A high-resolution precipitation and temperature dataset for the period of 1960-2021 in continental Chile	Chile	https://zenodo.org/records/7529682
Southeast Asia Limnology Network (**SEALNET**)	Remote sensing data	Collaboration platform for environmental and hydrological research	Southeast Asia	https://www.sealnetforum.org