Capacity Building in Seasonal Hydrological Forecasting

Visualization and Statistical Analysis

AGRHYMET, Climate Regional Center for West-Africa and Sahel

@Arsène KIEMA

2025-10-06

Pedagogical Objectives

Learning outcomes

By the end of this module, participants will be able to:

- Create static plots using ggplot2.
- Customize plots (titles, legends, themes, colors).
- Visualize time series, distributions, and comparisons.
- Combine multiple variables in one plot.
- Produce clear, publication-ready visualizations for hydrological datasets.

Data visualization in R

Data visualization helps transform raw numbers into insight.
In hydrology, visual analysis reveals **seasonal patterns**, **extremes**, and **model accuracy**.

R offers powerful visualization tools:

- **Base R graphics** → simple, quick.
- **ggplot2** → grammar of graphics, flexible and elegant.
- **plotly / tmap / leaflet** → interactive plots and maps.

Visualizing Time Series

```r
library(ggplot2)

# Example: daily flow data
flows <- read.csv("data/flows_station.csv")
flows$Date <- as.Date(flows$Date)

ggplot(flows, aes(x = Date, y = Qobs)) +
  geom_line(color = "steelblue") +
  labs(title = "Daily Streamflow", y = "Flow (m³/s)", x = "") +
  theme_minimal()
```

Visualizing Time Series

Tips

- Use `geom_line()` for time series.
- Use `geom_smooth()` to highlight long-term trends.

```
ggplot(flows, aes(x = Date, y = Qobs)) +
  geom_line(color = "grey60") +
  geom_smooth(color = "blue", se = FALSE) +
  labs(title = "Streamflow Trend", y = "Flow (m³/s)", x = "") +
  theme_minimal()
```

Distribution Analysis

```
ggplot(flows, aes(x = Qobs)) +
  geom_histogram(bins = 30, fill = "skyblue", color = "black") +
  labs(title = "Distribution of Streamflow", x = "Flow (m³/s)", y = "
  theme_minimal()
```

Distribution Analysis

```r
ggplot(flows, aes(y = Qobs)) +
  geom_boxplot(fill = "orange") +
  labs(title = "Flow Variability", y = "Flow (m³/s)") +
  theme_minimal()
```

Comparing Variables

```r
# Example: relationship between rainfall and flow
rain <- read.csv("data/precipitation_station.csv")
rain$Date <- as.Date(rain$Date)

merged <- merge(flows, rain, by = "Date")

ggplot(merged, aes(x = Precip, y = Qobs)) +
  geom_point(color = "darkgreen") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Relationship between Rainfall and Flow",
       x = "Rainfall (mm)", y = "Flow (m³/s)") +
  theme_minimal()
```

Comparing Variables

Interpretation

A positive relationship between rainfall and discharge indicates a direct hydrological response of the basin.

Statistical Performance Indicators

```r
# Observed vs simulated flows
obs <- c(10, 20, 30, 40, 50)
sim <- c(12, 18, 32, 37, 52)

# Coefficient of determination
R2 <- cor(obs, sim)^2

# Root Mean Square Error
RMSE <- sqrt(mean((obs - sim)^2))
```

Statistical Performance Indicators

```r
# Nash-Sutcliffe Efficiency
NSE <- 1 - (sum((obs - sim)^2) / sum((obs - mean(obs))^2))

# Kling-Gupta Efficiency
r <- cor(obs, sim)
alpha <- sd(sim)/sd(obs)
beta <- mean(sim)/mean(obs)
KGE <- 1 - sqrt((r - 1)^2 + (alpha - 1)^2 + (beta - 1)^2)
```

Statistical Performance Indicators

```r
data.frame(R2, RMSE, NSE, KGE)
```

```
        R2     RMSE   NSE       KGE
1 0.975418 2.236068 0.975 0.9857471
```

Statistical Performance Indicators

Hydrological model evaluation

- $\mathbf{R^2}$ $\rightarrow$ strength of linear relationship.
- **RMSE** $\rightarrow$ overall error magnitude.
- **NSE** and **KGE** $\rightarrow$ widely used in hydrology for model performance.

Visualizing Model Performance

```r
perf <- data.frame(obs, sim)

ggplot(perf, aes(x = obs, y = sim)) +
  geom_point(size = 3, color = "darkblue") +
  geom_abline(intercept = 0, slope = 1, color = "red", linetype = "da
  labs(title = "Observed vs Simulated Flows",
       x = "Observed (m³/s)", y = "Simulated (m³/s)") +
  theme_minimal()
```

Visualizing Model Performance

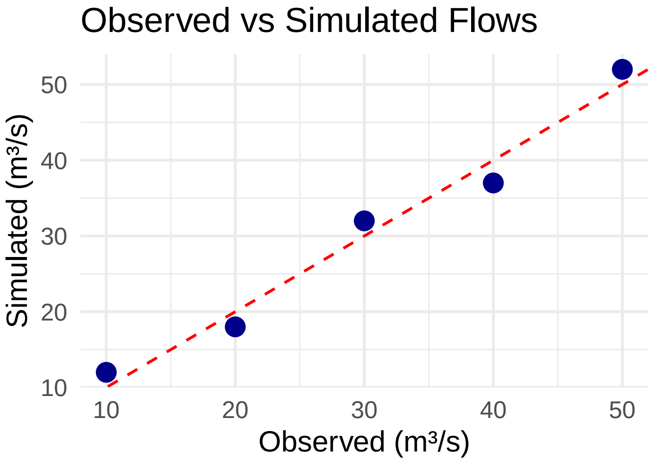

Figure 1: Model performance assesement

Practical Exercises

Exercise 1

- Plot the daily flow series for Station A.
- Add a smoothed trend line and a title.

Exercise 2

- Plot the histogram and boxplot of observed flows.
- Comment on the variability.

Practical Exercises

Exercise 3

- Merge rainfall and flow datasets.
- Plot rainfall vs flow scatterplot and compute correlation.

Exercise 4

- Given `Qobs` and `Qsim` columns in a CSV,
- calculate NSE, RMSE, and KGE using base R formulas.

THANK YOU FOR YOUR ATTENTATION