Fitting Multivariate Time Series Models: Analyzing Environmental and Chemical Data

Sajad Ahmad Mir (M21MA207), Himalaya (M21MA204), Nasrat Jahan (M21MA205), Raman Reshi (M21MA206)

Department of Mathematics
Indian Institute of Technology, Jodhpur

August 25, 2025

Department of Mathematics

Table of Contents

Introduction

- Time series analysis is essential in understanding temporal patterns and dependencies in environmental and chemical datasets.
- Multivariate time series models allow for simultaneous analysis of multiple interrelated variables.
- Objective: Model the relationships between environmental parameters using VAR and VARMA models.
- Data Source: India Meteorological Department (IMD) Pune, containing 11 variables such as rainfall, pH, conductivity, and more.

Data

	station	year	month	date	rainfall	ph	conductivity	sulphate	nitrate	chloride	ammonium	calcium	magnesium	sodium	potassium	ph
1	42027	2015	2	2	12	7.55	79.7	3.02	7.27	1.85	0.16	12.36	1.04	0.84	0.45	
2	42027	2015	2	3	4	6.86	18.4	0.26	2.24	0.88	0.14	2.81	0.14	0.08	0.09	
3	42027	2015	2	8	4	7.63	59.8	2.07	4.22	1.70	0.05	9.30	0.63	0.76	0.45	
4	42027	2015	2	16	11	6.83	21.1	1.10	3.16	1.33	0.08	2.87	0.24	0.45	0.18	
5	42027	2015	2	19	21	6.46	13.0	0.39	2.86	3.72	0.07	3.37	0.16	0.12	0.11	
6	42027	2015	2	25	45	4.83	31.9	2.50	7.71	1.59	0.12	2.93	0.34	0.54	0.25	
7	42027	2015	3	3	10	6.86	17.5	0.24	2.95	0.81	0.03	2.75	0.15	0.09	0.11	
8	42027	2015	3	4	1	6.92	53.3	3.03	16.23	1.48	0.07	6.91	0.75	0.39	0.55	
9	42027	2015	3	5	2	5.60	13.9	0.50	4.06	0.64	0.16	1.81	0.12	0.05	0.11	
10	42027	2015	3	9	54	6.45	19.6	1.65	4.53	0.78	0.02	2.84	0.19	0.32	0.12	
11	42027	2015	3	16	37	7.47	70.3	1.93	11.87	1.31	0.02	10.88	0.65	0.46	0.43	
12	42027	2015	3	25	11	7.25	34.3	0.97	3.67	3.05	0.05	6.31	0.42	0.70	0.28	
13	42027	2015	3	29	44	6.96	21.7	0.83	3.47	4.27	0.05	5.04	0.62	0.31	0.20	
14	42027	2015	4	2	27	6.32	15.2	0.35	4.32	0.83	0.16	2.35	0.13	0.11	0.24	

Figure: Multivariate Environmental and Chemical Dataset

Data

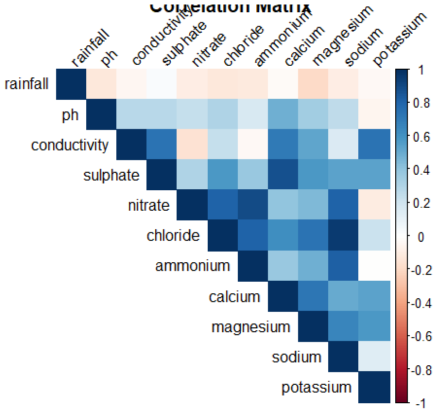
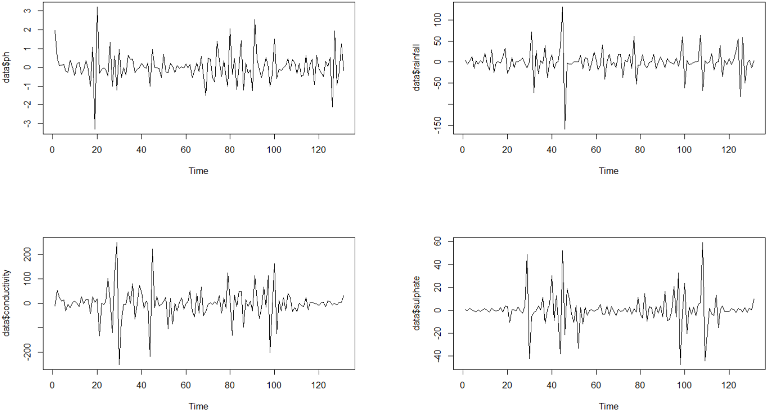
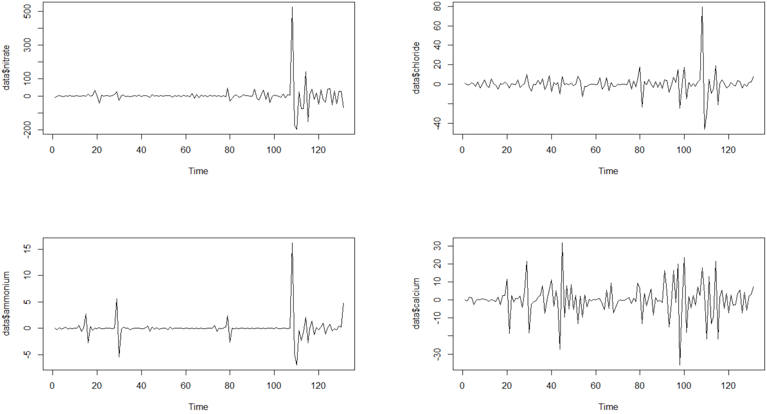


Figure: Multivariate Environmental and Chemical Dataset

Basic Analysis



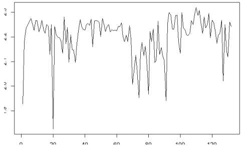

data$nitrate
Time
data$chloride
Time
data$ammonium
Time
data$calcium
Time

GARCH Modelling

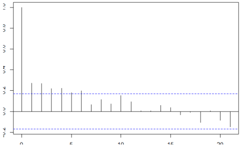
After exploring ARIMA models, we found that they were not able to capture the volatility in the data effectively. Hence, we switched to the GARCH model, which is specifically designed to model volatility. We applied this model to the log-transformed pH data ($\log(1 + \mathrm{ph})$) to better understand fluctuations and volatility in the data.

Log Transformation and ACF/PACF

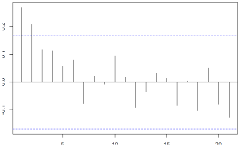
The log transformation $\log(1 + \text{ph})$ was applied to the pH data to stabilize the variance. Below are the plots for the log-transformed data, along with its ACF and PACF:



log(1+ph)



ACF of log(1+ph)



PACF of log(1+ph)

Table: Log-transformed data and its ACF/PACF.

Residual Diagnostics

We performed diagnostic tests on the residuals of the transformed series $\log(1 + \text{ph})$:

Box-Ljung Test:

$$X^2 = 9.9072, \quad df = 1, \quad p - value = 0.001646.$$

This indicates significant autocorrelation in the residuals.

ARCH LM Test:

$$\chi^2 = 24.758, \quad df = 12, \quad p - value = 0.01601.$$

This indicates the presence of ARCH effects in the residuals.

GARCH Model Specification

The GARCH(2,1) model was applied to the log-transformed data. The model equations are as follows:

$$
\begin{aligned}
y_t &= \mu + \epsilon_t, \\
\epsilon_t &= \sigma_t z_t, \\
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \beta_1 \sigma_{t-1}^2,
\end{aligned}
$$

where $y_t = \log(1 + \text{ph})$ and $z_t \sim N(0, 1)$.

Model Output

The fitted GARCH(2,1) model has the following estimated coefficients:

$$\mu = 2.0896, \quad \omega = 0.00079265, \quad \alpha_1 = 0.0664, \quad \alpha_2 = 0.00000001, \quad \beta_1 = 0.8210.$$

These coefficients were estimated using maximum likelihood estimation.

Error Analysis

The parameter estimates, standard errors, t-values, and significance levels are summarized in the table below:

Table: Error Analysis of GARCH(2,1) Model.

Parameter	Estimate	Std. Error	t-value	p-value
μ	2.0896	0.004983	419.333	$< 2 \times 10^{-16}$
ω	0.00079265	0.0005166	1.534	0.125
α_1	0.0664	0.04679	1.419	0.156
α_2	0.00000001	0.05872	0.000	1.000
β_1	0.8210	0.09434	8.703	$< 2 \times 10^{-16}$

Information Criteria and Conclusion

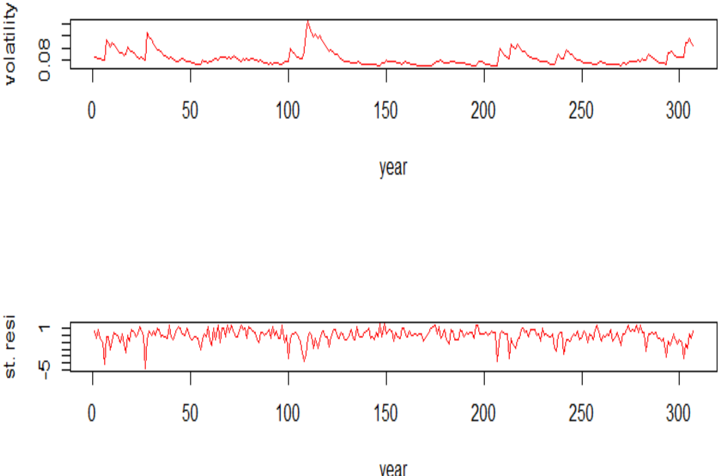
The information criteria for model selection are as follows:

$$\begin{aligned} \text{AIC} &= -2.1158, \\ \text{BIC} &= -2.0551, \\ \text{SIC} &= -2.1163, \\ \text{HQIC} &= -2.0915. \end{aligned}$$

Based on the diagnostic tests and model fitting, the GARCH(2,1) model provides a good fit to the log-transformed pH data, capturing volatility and autocorrelation.

Volatility Plot

Below is the volatility plot generated from the fitted GARCH model, showing the time-varying volatility over the period:



Theory Behind VAR

- VAR models capture the dynamic relationships between multiple time series variables.
- The model includes lags of both dependent and independent variables.
- Estimated coefficients represent the influence of past values on future outcomes.

Final VAR Model Results

- After applying the GARCH model to log(1+ph), we explored the Vector Autoregressive (VAR) model.
- The VAR model captures interdependencies between multiple time series variables.
- It provides a comprehensive analysis of dynamic interactions within the data.

Final VAR Model Coefficients

- **Overview:** VAR models analyze interdependencies and dynamic relationships between multiple time series variables.
- **Multiple Time Series:** Allows simultaneous analysis of multiple interrelated variables.
- **Lag Structure:** Captures the effect of past values (lags) on future outcomes.
- **Model Equations:**

$$\text{Rainfall}_t = c_1 + \phi_{11}\text{Rainfall}_{t-1} + \phi_{12}\text{pH}_{t-1} + \ldots + \phi_{12}\text{Rainfall}_{t-2} + \epsilon_1$$

VAR(2) Model Equations

The Vector Autoregression (VAR) model for the variables is given below. Each equation represents the evolution of one variable based on its own lags. For simplicity, only lagged terms of the same variable are shown.

AR(1) Matrix

The AR(1) coefficients matrix is as follows:

$$
\begin{bmatrix}
0.1516 & 3.4544 & -0.1654 & -0.2187 & -0.1184 & -0.6534 & 5.7524 & -0.0727 & 8.4613 & -0.0976 & 0.6856 \\
0.0048 & 0.0649 & 0.0012 & -0.0194 & 0.0002 & 0.0057 & 0.0512 & 0.0338 & -0.0352 & 0.0045 & -0.0129 \\
0.1312 & -4.8985 & 1.1531 & -0.2044 & 0.6859 & 0.8863 & -13.7250 & -3.5664 & -19.0668 & -1.4476 & -3.3183 \\
0.1024 & 0.8773 & 0.0978 & 0.5145 & 0.1377 & 0.2184 & -3.4267 & -0.8161 & -2.1351 & -0.4074 & -0.2015 \\
0.6903 & 4.5272 & 0.0892 & 0.7612 & 0.5673 & 1.7614 & -2.9333 & -3.1043 & 3.8381 & -0.7153 & -0.4340 \\
0.1015 & 0.7773 & 0.0416 & 0.0392 & 0.0619 & 0.4743 & -0.0117 & -0.4704 & 0.0601 & -0.4082 & 0.4191 \\
0.0190 & 0.4190 & 0.0092 & 0.0301 & 0.0223 & 0.0713 & -0.1590 & -0.1590 & 0.0761 & -0.0490 & -0.0416 \\
0.0658 & -0.3858 & 0.0697 & 0.1749 & 0.1131 & 0.0237 & -2.0973 & -0.3956 & -0.7058 & -0.0698 & -0.0789 \\
0.0084 & -0.0742 & 0.0101 & 0.0103 & 0.0106 & 0.0310 & -0.0532 & -0.0664 & 0.0816 & -0.0498 & -0.0167 \\
0.0790 & 0.1754 & 0.0232 & -0.0227 & 0.0354 & 0.1577 & -0.0230 & -0.1846 & 0.1429 & -0.0309 & 0.3472 \\
-0.0002 & -0.0756 & 0.0255 & -0.0129 & 0.0271 & -0.0278 & -0.4930 & -0.1208 & -0.1594 & -0.0232 & 0.0947
\end{bmatrix}
$$

AR(2)-Matrix

$$
\begin{bmatrix}
-0.04125 & -0.51380 & 0.01409 & 0.44433 & -0.12600 & -0.5631 & 3.55120 & 0.5173 & -2.0554 & -0.1315 & 2.2177 \\
-0.00506 & 0.03261 & 0.00375 & -0.01006 & 0.00455 & -0.0052 & -0.10851 & 0.0172 & -0.0657 & -0.0064 & -0.0929 \\
0.20417 & 3.96806 & -0.13174 & 1.39086 & -0.30119 & -2.5116 & -3.12179 & 0.7216 & 5.7326 & 4.7923 & -5.4692 \\
0.03718 & -0.05538 & -0.05477 & 0.16165 & -0.07139 & -0.9388 & -0.05907 & 0.8499 & -0.3539 & 1.4627 & -0.2528 \\
-0.10532 & 8.02827 & -0.14893 & -1.20296 & 0.15182 & -3.6381 & 0.42683 & 3.7576 & -9.9728 & 3.6866 & 4.4782 \\
-0.02233 & 0.13124 & -0.04189 & -0.07012 & -0.03445 & -0.5664 & -0.22309 & 0.6741 & -0.9833 & 0.7713 & 0.3733 \\
0.00114 & 0.00404 & -0.00727 & -0.01005 & -0.00276 & -0.0640 & -0.00099 & 0.0881 & -0.1172 & 0.0827 & 0.0194 \\
0.01324 & 0.09584 & -0.02971 & 0.02057 & -0.02910 & -0.6606 & 0.13732 & 0.6754 & 0.1452 & 0.7871 & -0.4617 \\
-0.00335 & -0.04861 & -0.00653 & -0.00964 & -0.00254 & -0.0778 & -0.09612 & 0.0618 & 0.1239 & 0.1092 & 0.0768 \\
-0.02363 & -0.02779 & -0.03304 & -0.05080 & -0.02540 & -0.3720 & -0.28834 & 0.4230 & -0.4586 & 0.5547 & 0.3822 \\
0.00230 & 0.18762 & -0.01779 & 0.04284 & -0.01277 & -0.1069 & -0.04129 & 0.0378 & 0.1685 & 0.1952 & 0.2775
\end{bmatrix}
$$

Standard Error

The standard error matrix for the AR(1) coefficients:

$$\begin{bmatrix}
0.0956 & 4.448 & 0.0897 & 0.4626 & 0.1485 & 0.9053 & 3.791 & 0.9268 & 4.621 & 1.2014 & 2.5827 \\
0.0028 & 0.132 & 0.0027 & 0.0137 & 0.0044 & 0.0269 & 0.113 & 0.0275 & 0.137 & 0.0357 & 0.0767 \\
0.2547 & 11.857 & 0.2391 & 1.2330 & 0.3957 & 2.4130 & 10.105 & 2.4702 & 12.318 & 3.2021 & 6.8839 \\
0.0533 & 2.480 & 0.0500 & 0.2579 & 0.0828 & 0.5048 & 2.114 & 0.5167 & 2.577 & 0.6698 & 1.4400 \\
0.2394 & 11.144 & 0.2248 & 1.1589 & 0.3720 & 2.2680 & 9.498 & 2.3217 & 11.578 & 3.0097 & 6.4701 \\
0.0408 & 1.897 & 0.0383 & 0.1973 & 0.0633 & 0.3862 & 1.617 & 0.3953 & 1.971 & 0.5124 & 1.1016 \\
0.0077 & 0.360 & 0.0073 & 0.0375 & 0.0120 & 0.0733 & 0.307 & 0.0750 & 0.374 & 0.0973 & 0.2091 \\
0.0341 & 1.589 & 0.0320 & 0.1652 & 0.0530 & 0.3234 & 1.354 & 0.3310 & 1.651 & 0.4291 & 0.9225 \\
0.0043 & 0.199 & 0.0040 & 0.0207 & 0.0067 & 0.0406 & 0.170 & 0.0416 & 0.207 & 0.0539 & 0.1158 \\
0.0314 & 1.464 & 0.0295 & 0.1522 & 0.0489 & 0.2979 & 1.248 & 0.3050 & 1.521 & 0.3954 & 0.8500 \\
0.0070 & 0.325 & 0.0066 & 0.0338 & 0.0108 & 0.0661 & 0.277 & 0.0677 & 0.337 & 0.0877 & 0.1886
\end{bmatrix}$$

Residual Covariance Matrix

The residual covariance matrix is:

$$
\boldsymbol{\Sigma} = \begin{bmatrix}
358.12 & -1.15 & 41.12 & 11.63 & -98.54 & -20.21 & -2.97 & 1.03 & -2.84 & -12.58 \\
-0.57 & & & & & & & & & \\
-1.15 & 0.32 & 9.31 & 1.48 & 2.99 & 1.00 & 0.04 & 1.76 & 0.15 & 0.66 \\
-0.03 & & & & & & & & & \\
41.12 & 9.31 & 2544.21 & 377.18 & 80.68 & 116.34 & 8.76 & 272.90 & 27.63 & 62.81 \\
53.76 & & & & & & & & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & & & & & & & & &
\end{bmatrix}
$$

$$
\begin{aligned}
\det(\mathrm{SSE}) &= 94864561636 \\
\mathrm{AIC} &= 28.94238 \\
\mathrm{BIC} &= 34.22752 \\
\mathrm{HQ} &= 31.09002
\end{aligned}
$$

Key Points of VAR

- **Model Estimation:** Coefficients estimated using Ordinary Least Squares (OLS).
- **Model Interpretation:**
  - Coefficients represent the strength and direction of relationships between variables at different lags.
  - Statistical significance is assessed to understand the impact of each variable's past values.
- **Applications:** VAR models are widely used in economics, environmental science, and social sciences to analyze dynamic systems.

Key Takeaways

- **Influence of Past Lags:** The AR(1) and AR(2) coefficients show the temporal dependence of each variable.
- **Statistical Significance:** Some variables show weak relationships with past values.
- **Dynamic Relationships:** Different variables exhibit varying temporal dependencies.

Conclusion

- The VAR model reveals significant temporal dependencies, but some relationships need refinement.
- Further model optimization is needed to improve forecast accuracy.

Thank You

Thank You!