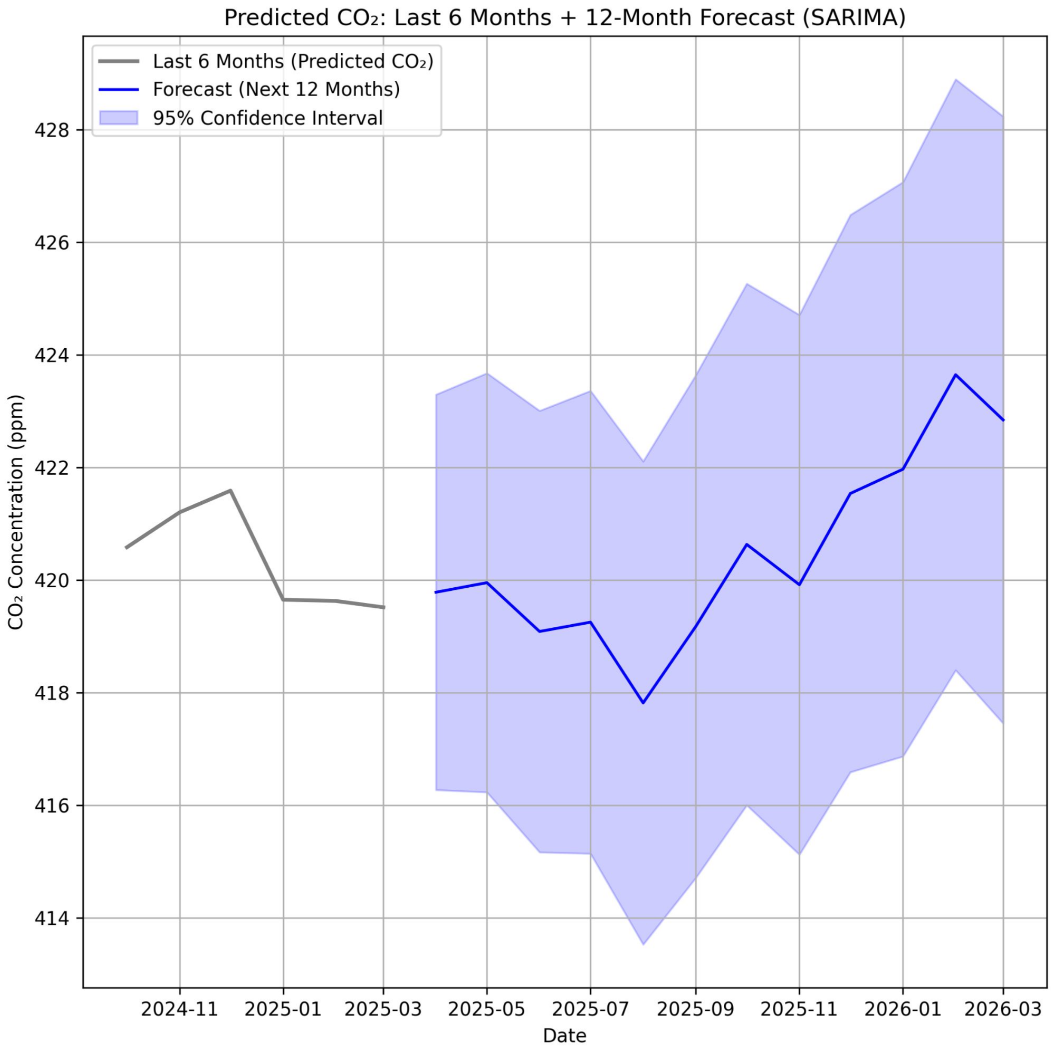
A Machine Learning and Time Series Approach to the Ground-Level CO_2 Forecast

The Ultimate Goal

To **predict and forecast** future monthly average ground-level CO_2 concentrations (in parts per million, ppm) based on **historical observations and related atmospheric and meteorological data**, using a **hybrid** machine learning (Random Forest) and time series (SARIMA) approach.



Modeling Approach

Component	Method	Field
Random Forest	Supervised ML	Machine Learning
SARIMA	Parametric model	Statistics / Time Series
Combined Approach	ML + Forecasting	Hybrid (Data Science)

What They Do

Step	What It Does
Random Forest	Learns patterns from atmospheric and meteorological variables to estimate CO_2
SARIMA	Statistically extrapolates time series based on historical trends

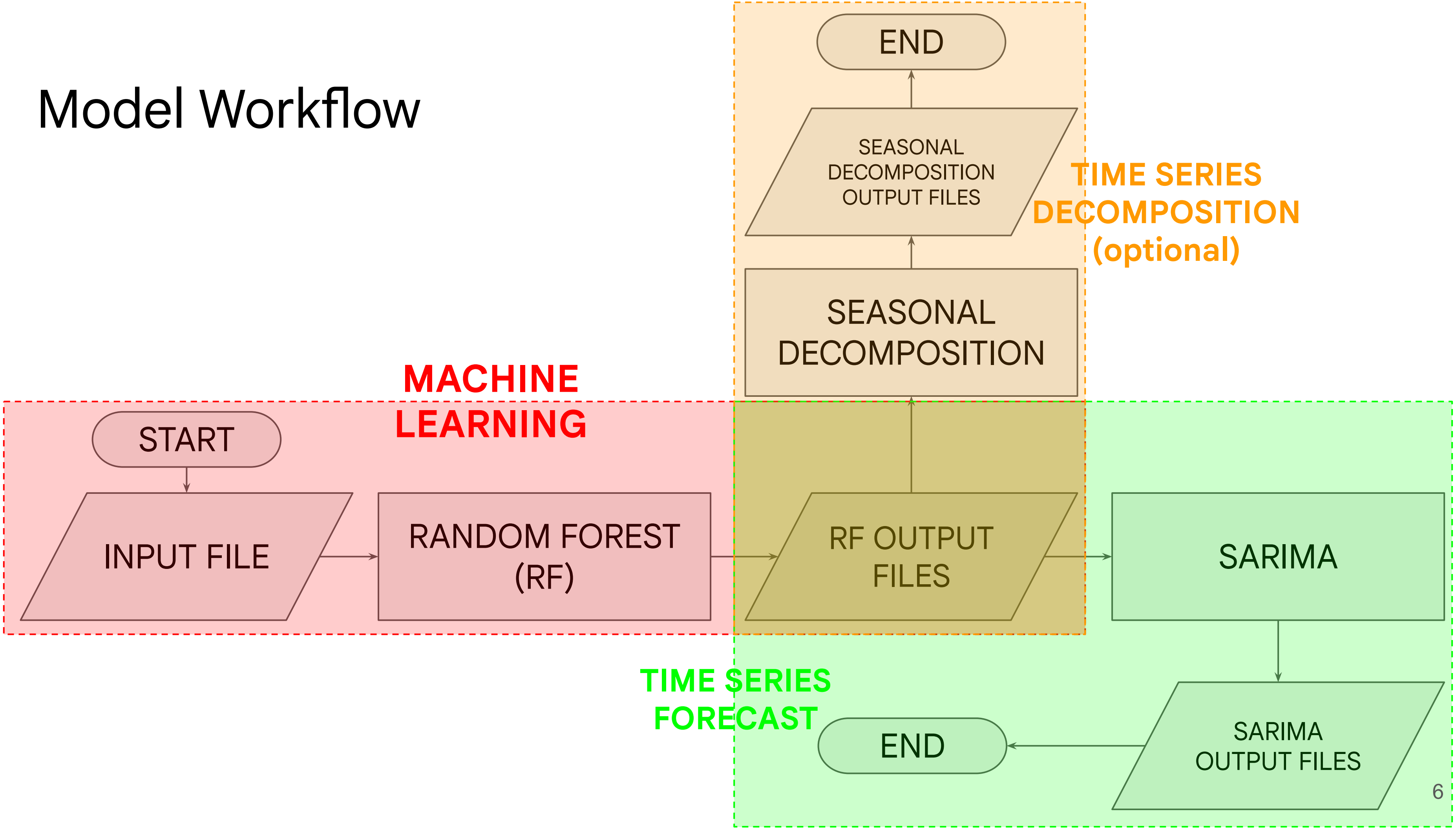
No human rule-based programming	All relationships are **learned from data**, not manually encoded

Why Hybrid

Limitation of Random Forest	How SARIMA Adds Value
Does not model time-dependent structure	SARIMA handles **lags, trends, seasonal cycles**
No built-in forecasting horizon	SARIMA **projects into the future** using temporal dynamics
Ignores autocorrelation in residuals	SARIMA models **serial correlation** for more accurate forecast

Limitation of SARIMA	How the Hybrid Approach Helps
Assumes linear relationships	Random Forest handles **non-linear** patterns and interactions
Cannot use many external predictors	Random Forest uses **multi-dimensional input features**
Struggles with sudden changes or shifts	Random Forest can adapt to **external drivers** (e.g., temperature, wind)
Poor at spatial or multivariate integration	Random Forest can ingest **diverse datasets** easily

Model Workflow

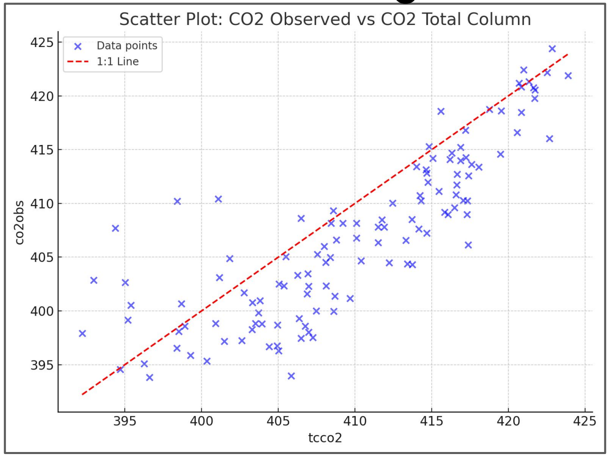


Input Parameters

Parameter	Description	Role	Source
co2obs	CO_2 concentration (ground-based)	Target	BMKG
tcco2	Total column CO_2 (satellite)	Feature	NASA
tcco_1e4	Total column CO	Feature	ECMWF
tcch4_1e4	Total column CH_4	Feature	ECMWF
u10	10m U-component of wind	Feature	ECMWF
v10	10m V-component of wind	Feature	ECMWF
t2m	2m temperature	Feature	ECMWF
mslp	Mean sea level pressure	Feature	ECMWF

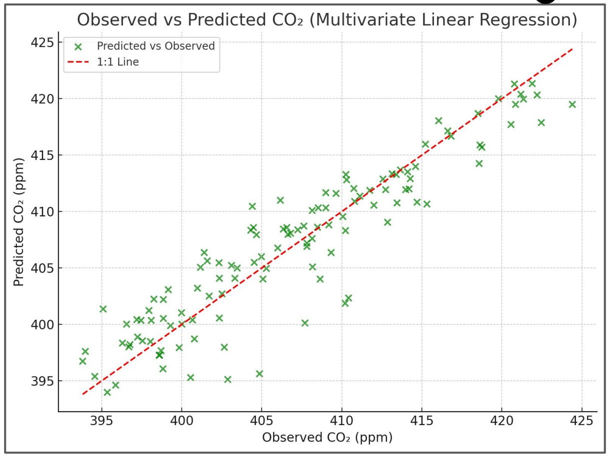
Comparing Correlations

Linear Reg.



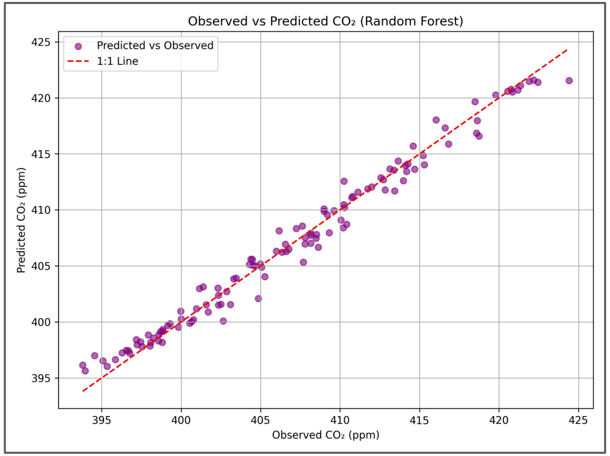
R: 0.85

Multivariate Linear Reg.



R: 0.92

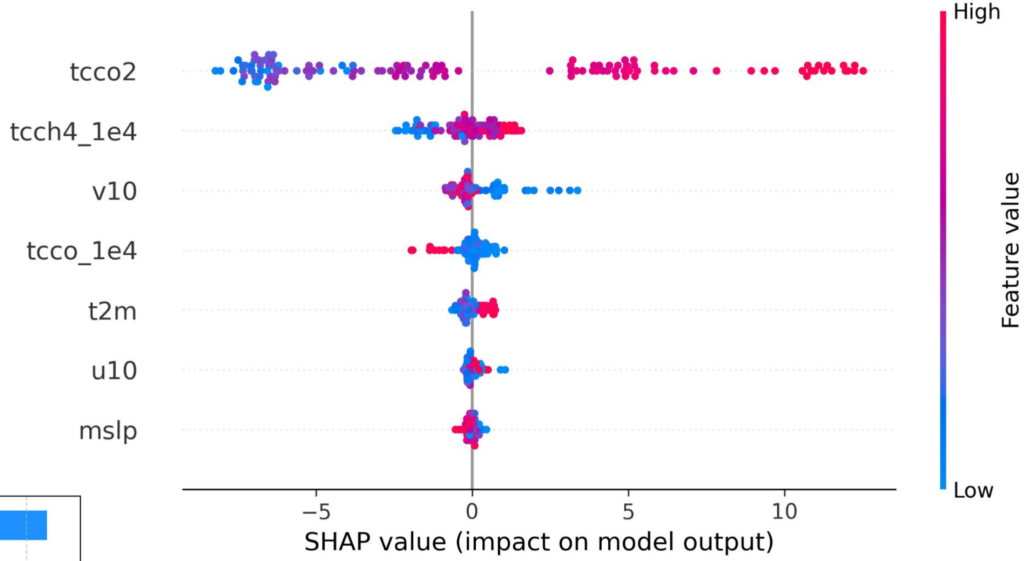
Random Forest

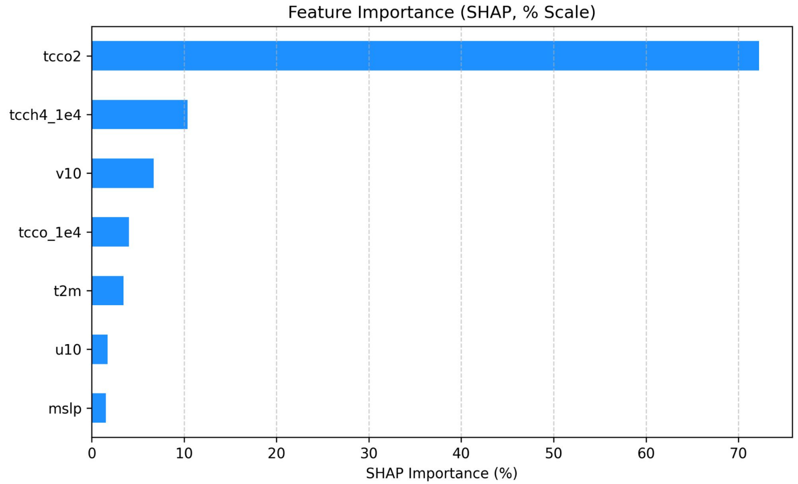


R: 0.99

Influence from other features? Seasonal patterns? Non-linearity?

Feature Importance





Feature Contribution

Feature	Definition	Relevance at Bukit Kototabang
tcco2	Total column CO_2 (from satellite OCO-2), in ppm	Captures the regional CO_2 burden above BKT. High columns often coincide with large-scale CO_2 enhancements (e.g. biomass-burning outflows or continental plumes) that, once mixed down, raise surface CO_2 at the elevated plateau site.
tcch4_1e4	Total column CH_4 (scaled by 1×10^4), in molecules/cm^2	Acts as a co-indicator of wetland or combustion sources around Sumatra. Elevated CH_4 columns frequently accompany CO_2 enhancements from tropical wetlands or fire events, improving model sensitivity to joint GHG variability.
v10	Meridional wind at 10 m (north–south component), in m s^{-1}	Reflects seasonal shifts in monsoonal transport: SSE winds (DJF–MAM) versus NNW winds (JJA–SON). Positive v10 (northward flow) can bring CO_2-rich air from northern land sources, while negative v10 (southward) brings cleaner maritime air.
tcco_1e4	Total column CO (scaled by 1×10^4), in molecules/cm^2	Traces combustion (biomass burning, local agriculture) around the station. Peaks in column CO often coincide with fire emissions in West Sumatra, serving as a proxy for CO_2 co-emissions that elevate surface mixing ratios.
t2m	2 m air temperature, in °C	Governs boundary-layer dynamics at the ~865 m elevation of BKT. Warmer days (up to ~25 °C) enhance thermal mixing and CO_2 draw-down, while cooler nights (down to ~16 °C) suppress mixing, allowing CO_2 to accumulate near the surface.
u10	Zonal wind at 10 m (east–west component), in m s^{-1}	Controls east–west advection of air masses: westerlies (−u10) bring maritime-boundary air from the Indian Ocean (low CO_2), easterlies (+u10) can transport continental plumes from Sumatra or beyond, modulating surface CO_2.
mslp	Mean sea-level pressure, in hPa.	Encodes synoptic-scale stability: high pressure promotes subsidence and shallow mixing (raising surface CO_2), whereas low pressure enhances turbulence and vertical mixing (diluting CO_2). Seasonal pressure patterns influence local CO_2 signals.