Multivariate Routes Through Traffic Anomalies

Erin Xu

December 1st, 2025

Introduction

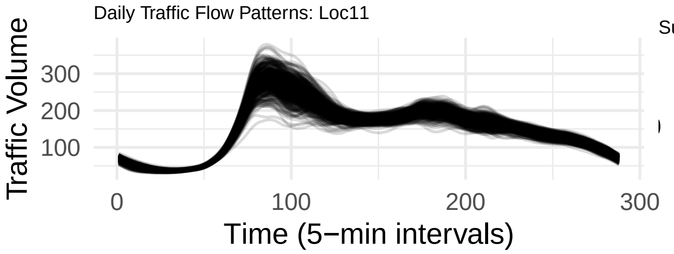
Due to the unavoidable nature of traffic congestion in urban locations, studying its patterns and underlying dynamics enables daily commuters and transportation authorities to transition from reactive management to proactive intervention, leading to overall reduced congestion, lower emissions and improved commuter safety around high-density areas. However, accurately detecting and predicting traffic anomalies that cause significant delays remains a challenge due to the inherent complexity of traffic dynamics, which are continuous, stochastic, spatiotemporally autocorrelated and cross-correlated (Columbia University Mailman School of Public Health, n.d.).

As existing prediction has evolved from interval-based pointwise in univariate time-series data to a functional approach at the network-level utilizing neural networks (Ma et al., 2024), this project offers a comparative assessment of common multivariate statistical techniques that are highly interpretable as benchmarks for further study. The goal of this paper is to identify and classify location-specific anomalies from the intraday patterns of traffic volume flow collected across 26 monitoring sites around the University of Toronto by comparing principal component analysis (PCA), factor analysis (FA), and independent component analysis (ICA) methods, which are selected for their ability to achieve dimension reduction, interpret latent regimes, and isolate mixed data components.

Data Description

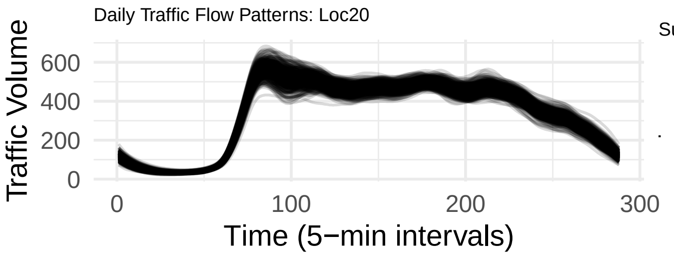
The dataset, synthetically modeled from Ma et al. (2024), has a natural tensor structure consisting of 26 locations (l), 384 (n) days each, and 288 (p) five-minute time points per day. Then each slice $X \in \mathbb{R}^{n \times p}$, called the daily traffic matrix, corresponds to one location and forms a 384×288 matrix, with each entry as volume in vehicles.

No data cleaning was required, as all locations share uniform dimensions and contain no missing observations. Exploratory visualizations were produced for two randomly selected locations to conserve space. Spaghetti plots by location and by day, accompanied by summary statistics, highlight clear daily peak structures. The first location exhibits lower median flow than the second, suggesting a less trafficked or more residential area. In contrast, the two randomly sampled days display similar median volumes, consistent with typical weekday patterns. These preliminary observations motivate the subsequent use of statistical methods to quantify temporal and spatial structure in the full dataset.



Summary for Loc11

1st Qu.	Median	Mean	3rd Qu.
94	163	150	192

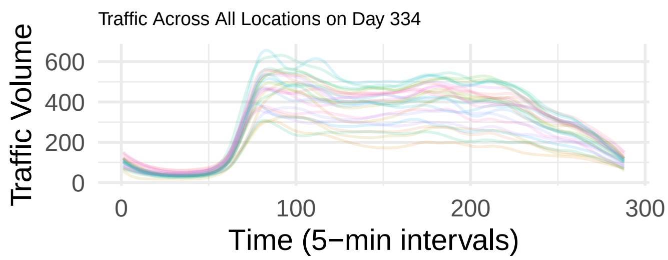


Summary for Loc20

1st Qu.	Median	Mean	3rd Qu.
167	432	341	482

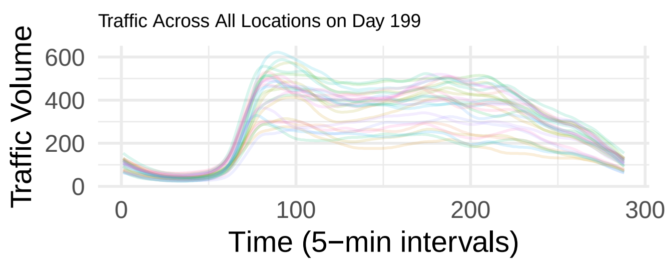
Summary for Day 334

1st Qu.	Median	Mean	3rd Qu.
137.0	299.0	284.9	423.0



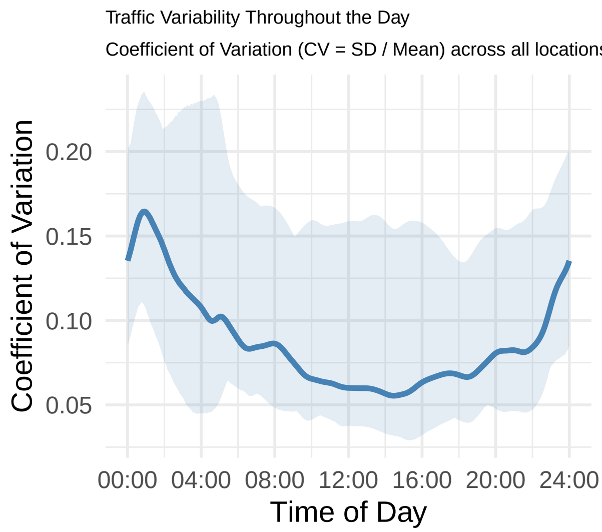
Summary for Day 199

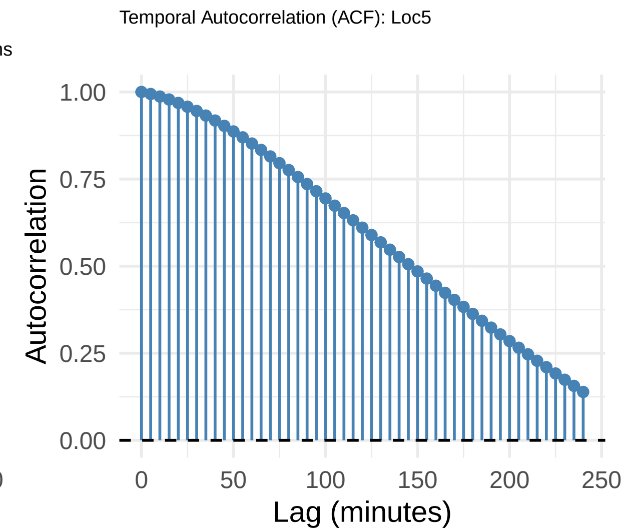
1st Qu.	Median	Mean	3rd Qu.
132.0	274.0	270.6	403.0



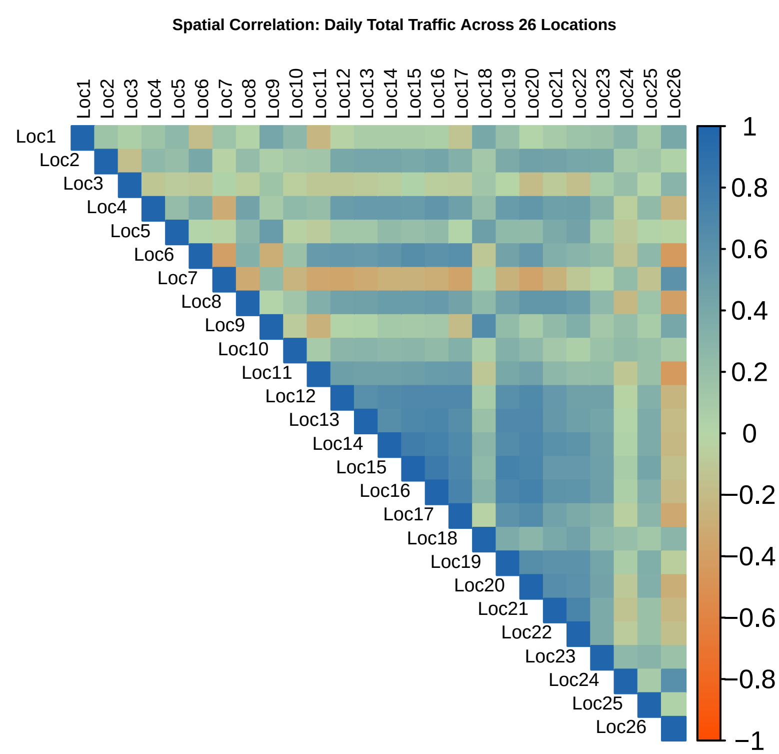
The CV and ACF analyses indicate that the traffic system exhibits a strong diurnal rhythm and persistent temporal dependence, supporting the use of dimension-reduction methods. CV quantifies day-to-day variability at each 5-minute interval. Averaged across locations, the CV curve shows highest variability overnight, a sharp decline during the morning, and a pronounced minimum around midday, followed by increasing variability toward the evening peak. The narrow CV shadow around noon suggests highly consistent midday traffic, whereas the wider shadow during peak hours reflects differing commuter patterns across locations.

Temporal autocorrelation was assessed at a representative site (Location 5), selected because its mean CV is closest to the median across all locations. The ACF reveals strong short-term persistence: autocorrelation decays gradually over several hours and remains positive even at four-hour lags, implying that intraday traffic evolves smoothly. The absence of negative correlations indicates a stable daily cycle.





The spatial correlation heatmap has shows that the center locations generally are correlated with each other, which suggests functional subregions where locations follow similar demand cycles. Most correlations fall in the moderate positive range, indicating coordinated but heterogeneous behavior across the network. Therefore location-specific anomaly detect is meaningful with location specific nuances, and network-wide patterns could also be coherent enough as the system has shared temporal dynamics.



Methodology

The traffic dataset is high-dimensional, functional in nature (with $p = 288$ time points per day), and exhibits substantial temporal dependence.

Consequently, dimension reduction constitutes a fundamental component of the anomaly detection framework with PCA serving as the primary dimension-reduction mechanism, as each daily traffic profile is represented by a matrix $X \in \mathbb{R}^{n \times p}$, and PCA assumes that although each curve x_i lies in a high-dimensional ambient space $\mathbb{R}^p$, its intrinsic variation is concentrated on a low-dimensional manifold. This representation is consistent with standard functional data analysis practice, where curves are expressed in a reduced basis and where PCA provides a computationally efficient empirical basis for large-scale datasets.

For each location, the 384×288 traffic matrix X was transposed so that rows corresponded to days and columns to time points. PCA was then conducted on the centered traffic matrix, $X_c = X - \mathbf{1}_n \bar{x}^\mathsf{T}$, where $\bar{x}$ denotes the sample mean profile. Let $V = [v_1, \dots, v_p]$ denote the orthonormal loading functions (the empirical eigenbasis), and let $S = X_c V$ denote the corresponding score matrix. The score of day i on component j is given by $s_{ij} = X_{c,i}^\mathsf{T} v_j$. The number of retained components k was selected via the 90% variance-explained criterion: $\frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{p} \lambda_j} \geq 0.90$, where $\lambda_j = \sigma_j^2$ denotes the variance explained by component j, which can be obtained through the squared singular values of the sample covariance matrix. This process highlights major regimes, like a morning peak, and discards higher-frequency noise.

Then anomalies were identified directly from the PCA score matrix $S \in \mathbb{R}^{n \times k}$ as PCA scores are uncorrelated and form an orthogonal basis. For each component j, a day i was flagged as anomalous if $s_{ij} < Q_{1,j} - 1.5\,\mathrm{IQR}_j$ or $s_{ij} > Q_{3,j} + 1.5\,\mathrm{IQR}_j$, where $Q_{1,j}$, $Q_{3,j}$, and IQR_j denotes the first quartile, third quartile, and interquartile range of component j respectively. In comparison to Mahalanobis-distance methods, this boxplot method is robust enough to not assume normality and is standard in functional data outlier detection, which is important as PC scores are not guaranteed to be multivariate normally distributed (MVN). Boxplots give interpretable, location-specific anomaly sets that can be classified by time-of-day (Shang & Hyndman, 2010).

FA provides an alternative dimension-reduction framework. FA postulates that an observed random vector x satisfies $x = \Lambda z + \varepsilon$, where z is a lower-dimensional latent vector, Λ is a loading matrix (factor scores and loadings), and ε represents idiosyncratic noise. FA requires a full-rank, invertible sample covariance matrix; estimation therefore proceeds via iterated PCA and not on the raw data, until uniqueness variances converge.

FA anomalies were also detected using the same $1.5 \times IQR$ rule applied to the factor score matrix F, where $f_{ij} < Q_{1,j} - 1.5\,\mathrm{IQR}_j$ or $f_{ij} > Q_{3,j} + 1.5\,\mathrm{IQR}_j$. Because FA captures deviations from latent structural factors rather than maximizing total variance, FA anomalies correspond to days whose patterns violate the inferred latent structure like shifted peaks, whereas PCA anomalies reflect variance-aligned distortions like spikes associated with incidents.

ICA was applied to the PCA scores to extract statistically independent latent signals embedded within the traffic profiles. ICA assumes whitened inputs with identity covariance and therefore requires PCA preprocessing, like in FA (Hyvärinen & Oja, 2000). ICA yields the decomposition $S_{\text{PCA}} = AS_{\text{ICA}}$, where S_{ICA} contains the source signals (scores) and A is the mixing matrix (loadings).

ICA anomalies were detected via the same $1.5 \times \text{IQR}$ rule applied to the independent source scores. Intuitively, ICA anomalies represent rare independent micro-events that sharply distort the traffic curve of a different location, for example a sudden dip/spike that only lasts a few intervals that PCA smooths out, or odd jumps that FA distributes across factors and randomness.

All anomalies identified across PCA, FA, and ICA were projected onto their two-dimensional principal component subspace and partitioned using k-means clustering (Piech & Ng, 2013). This method was selected for its computational efficiency and suitability for continuous Euclidean feature spaces. The number of clusters was chosen by maximizing the average silhouette coefficient: $k^* = \arg\max_k \left\{ \frac{1}{N} \sum_{i=1}^{N} \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \right\}$, where $a(i)$ is the average within-cluster distance and $b(i)$ is the minimum average between-cluster distance (Rousseeuw, 1987). This procedure yields interpretable groups of anomalies that reflect distinct structural perturbations in daily traffic dynamics. Taken together, PCA, FA, and ICA form a benchmark set: PCA captures global variance-driven deviations, FA captures structural inconsistencies relative to latent factors, and ICA captures independent localized perturbations. Using all three offers a comprehensive view of anomalous behavior from orthogonal interpretive perspectives: variance, latent structure, and independence. This ensures that anomalies detected are robust to modeling assumptions and interpretable in terms of their functional, temporal, and structural characteristics.

Results

```r
file <- "traffic.xlsx"
sheet_names <- getSheetNames(file)
num_sheets <- length(sheet_names)

set.seed(67)
df <- lapply(sheet_names, function(sheet) {
  as.matrix(read.xlsx(file, sheet = sheet, colNames = TRUE))
})
names(df) <- sheet_names
```

Helper Functions

```r
choose_k_pca <- function(pca, threshold = 0.90) {
  var_expl <- pca$sdev^2 / sum(pca$sdev^2)
  cumvar <- cumsum(var_expl)
```

```r
  k <- which(cumvar >= threshold)[1]
  return(k)
}

find_anomalies_from_scores <- function(score_mat, multiplier = 1.5) {
  n_days <- nrow(score_mat)
  is_outlier <- rep(FALSE, n_days)

  for (j in seq_len(ncol(score_mat))) {
    x <- score_mat[, j]
    stats <- boxplot.stats(x, coef = multiplier)
    is_outlier[which(x %in% stats$out)] <- TRUE
  }

  which(is_outlier)
}
```

Main Analysis Function

```r
analyze_location <- function(loc_name,
                             X,
                             var_expl_threshold = 0.90,
                             max_factors = 3,
                             anomaly_coef = 1.5,
                             do_ica = TRUE,
                             n_ica_comp = 3) {

  message("Processing: ", loc_name)

  # ---- 1. Transpose and center data ----
  X_t <- t(X)  # rows = days, cols = timepoints
  X_centered <- scale(X_t, center = TRUE, scale = FALSE)

  # ---- 2. PCA ----
  pca <- prcomp(X_centered, center = FALSE, scale. = FALSE)
  k_pca <- choose_k_pca(pca, threshold = var_expl_threshold)

  pca_scores <- pca$x[, 1:k_pca, drop = FALSE]
  pca_loadings <- pca$rotation[, 1:k_pca, drop = FALSE]
  pca_anom_idx <- find_anomalies_from_scores(pca_scores, multiplier = anomaly_coef)
  pca_anom_days <- rownames(pca_scores)[pca_anom_idx]
```

```r
# ---- 3. Factor Analysis on PCA scores ----
fa_model <- NULL
fa_scores <- NULL
fa_loadings <- NULL
fa_anom_days <- character(0)

n_factors <- min(max_factors, k_pca, 5)
n_pcs_for_fa <- min(50, k_pca)
n_factors_fa <- min(n_factors, n_pcs_for_fa - 1)

if (n_factors_fa >= 1 && n_pcs_for_fa >= 3) {
  fa_model <- tryCatch(
    factanal(pca_scores[, 1:n_pcs_for_fa, drop = FALSE],
             factors = n_factors_fa,
             scores = "regression",
             rotation = "varimax"),
    error = function(e) NULL
  )

  if (!is.null(fa_model)) {
    fa_scores <- fa_model$scores
    fa_loadings <- pca_loadings[, 1:n_pcs_for_fa] %*% fa_model$loadings[, , drop = FAI
    fa_anom_idx <- find_anomalies_from_scores(fa_scores, multiplier = anomaly_coef)
    fa_anom_days <- rownames(fa_scores)[fa_anom_idx]
  }
}

# ---- 4. ICA on PCA scores ----
ica_result <- NULL
ica_scores <- NULL
ica_loadings <- NULL
ica_anom_days <- character(0)

if (do_ica && k_pca >= 2) {
  n_ica_to_use <- min(n_ica_comp, k_pca)

  ica_result <- tryCatch(
    fastICA(pca_scores, n.comp = n_ica_to_use, method = "C"),
    error = function(e) NULL
  )

  if (!is.null(ica_result)) {
    ica_scores <- ica_result$S
    if (is.null(rownames(ica_scores))) {
```

```r
        rownames(ica_scores) <- rownames(pca_scores)
      }

      ica_loadings <- pca_loadings[, 1:n_ica_to_use, drop = FALSE] %*% ica_result$A
      ica_anom_idx <- find_anomalies_from_scores(ica_scores, multiplier = anomaly_coef)
      ica_anom_days <- rownames(ica_scores)[ica_anom_idx]
    }
  }

  # ---- Return results ----
  list(
    location = loc_name,
    pca = pca,
    k_pca = k_pca,
    pca_scores = pca_scores,
    pca_loadings = pca_loadings,
    pca_anom_days = pca_anom_days,
    fa_model = fa_model,
    fa_scores = fa_scores,
    fa_loadings = fa_loadings,
    fa_anom_days = fa_anom_days,
    ica = ica_result,
    ica_scores = ica_scores,
    ica_loadings = ica_loadings,
    ica_anom_days = ica_anom_days
  )
}

# ---- Run analysis for all locations ----
location_results <- lapply(names(df), function(loc_name) {
  analyze_location(
    loc_name = loc_name,
    X = df[[loc_name]],
    var_expl_threshold = 0.90,
    max_factors = 3,
    anomaly_coef = 1.5,
    do_ica = TRUE
  )
})
```

```
## Processing: Loc1
```

```
## Processing: Loc2
```

```
## Processing: Loc3
## Processing: Loc4
## Processing: Loc5
## Processing: Loc6
## Processing: Loc7
## Processing: Loc8
## Processing: Loc9
## Processing: Loc10
## Processing: Loc11
## Processing: Loc12
## Processing: Loc13
## Processing: Loc14
## Processing: Loc15
## Processing: Loc16
## Processing: Loc17
## Processing: Loc18
## Processing: Loc19
## Processing: Loc20
## Processing: Loc21
## Processing: Loc22
## Processing: Loc23
## Processing: Loc24
## Processing: Loc25
## Processing: Loc26
```

```r
names(location_results) <- names(df)
```

Anomaly Summary by Location

```r
for (loc_name in names(location_results)) {
  res <- location_results[[loc_name]]
  cat("\n", rep("=", 60), "\n", sep = "")
  cat("Location:", loc_name, "\n")
  cat(rep("=", 60), "\n")
  cat("PCA components:", res$k_pca, "\n")
  cat("PCA anomalies (", length(res$pca_anom_days), "):",
      paste(res$pca_anom_days, collapse = ", "), "\n")
  cat("FA anomalies (", length(res$fa_anom_days), "):",
      paste(res$fa_anom_days, collapse = ", "), "\n")
  cat("ICA anomalies (", length(res$ica_anom_days), "):",
      paste(res$ica_anom_days, collapse = ", "), "\n")
}
```

```
##
## ============================================================
## Location: Loc1
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 5
## PCA anomalies ( 30 ): WkDay-1, WkDay-43, WkDay-46, WkDay-56, WkDay-77, WkDay-89, WkDa
## FA anomalies ( 0 ):
## ICA anomalies ( 17 ): WkDay-11, WkDay-38, WkDay-150, WkDay-189, WkDay-212, WkDay-213,
##
## ============================================================
## Location: Loc2
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 7
## PCA anomalies ( 20 ): WkDay-1, WkDay-37, WkDay-61, WkDay-79, WkDay-80, WkDay-118, WkD
## FA anomalies ( 10 ): WkDay-27, WkDay-37, WkDay-61, WkDay-91, WkDay-124, WkDay-147, Wk
## ICA anomalies ( 14 ): WkDay-1, WkDay-78, WkDay-79, WkDay-80, WkDay-120, WkDay-177, Wk
##
## ============================================================
## Location: Loc3
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 5
## PCA anomalies ( 24 ): WkDay-5, WkDay-23, WkDay-26, WkDay-28, WkDay-51, WkDay-72, WkDa
## FA anomalies ( 0 ):
## ICA anomalies ( 17 ): WkDay-17, WkDay-23, WkDay-28, WkDay-67, WkDay-76, WkDay-99, WkD
```

```
##
## ============================================================
## Location: Loc4
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 7
## PCA anomalies ( 20 ): WkDay-5, WkDay-27, WkDay-29, WkDay-34, WkDay-73, WkDay-91, WkDa
## FA anomalies ( 11 ): WkDay-5, WkDay-27, WkDay-34, WkDay-73, WkDay-95, WkDay-117, WkDa
## ICA anomalies ( 10 ): WkDay-27, WkDay-36, WkDay-159, WkDay-214, WkDay-241, WkDay-261,
##
## ============================================================
## Location: Loc5
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 7
## PCA anomalies ( 18 ): WkDay-1, WkDay-21, WkDay-117, WkDay-138, WkDay-201, WkDay-203,
## FA anomalies ( 5 ): WkDay-21, WkDay-201, WkDay-223, WkDay-228, WkDay-332
## ICA anomalies ( 8 ): WkDay-5, WkDay-6, WkDay-21, WkDay-203, WkDay-222, WkDay-223, WkD
##
## ============================================================
## Location: Loc6
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 6
## PCA anomalies ( 6 ): WkDay-111, WkDay-232, WkDay-267, WkDay-309, WkDay-311, WkDay-336
## FA anomalies ( 4 ): WkDay-232, WkDay-267, WkDay-309, WkDay-336
## ICA anomalies ( 3 ): WkDay-267, WkDay-309, WkDay-336
##
## ============================================================
## Location: Loc7
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 2
## PCA anomalies ( 4 ): WkDay-20, WkDay-220, WkDay-256, WkDay-259
## FA anomalies ( 0 ):
## ICA anomalies ( 15 ): WkDay-7, WkDay-8, WkDay-10, WkDay-20, WkDay-219, WkDay-220, WkD
##
## ============================================================
## Location: Loc8
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 8
## PCA anomalies ( 23 ): WkDay-1, WkDay-7, WkDay-48, WkDay-84, WkDay-86, WkDay-98, WkDay
## FA anomalies ( 10 ): WkDay-48, WkDay-86, WkDay-109, WkDay-113, WkDay-128, WkDay-132,
## ICA anomalies ( 6 ): WkDay-1, WkDay-7, WkDay-230, WkDay-232, WkDay-327, WkDay-382
##
## ============================================================
## Location: Loc9
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 4
```

```
## PCA anomalies ( 14 ): WkDay-24, WkDay-40, WkDay-50, WkDay-60, WkDay-127, WkDay-248, W
## FA anomalies ( 0 ):
## ICA anomalies ( 12 ): WkDay-24, WkDay-37, WkDay-50, WkDay-206, WkDay-248, WkDay-289,
##
## ============================================================
## Location: Loc10
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 5
## PCA anomalies ( 13 ): WkDay-46, WkDay-62, WkDay-86, WkDay-100, WkDay-169, WkDay-195,
## FA anomalies ( 0 ):
## ICA anomalies ( 7 ): WkDay-9, WkDay-10, WkDay-51, WkDay-62, WkDay-169, WkDay-171, WkD
##
## ============================================================
## Location: Loc11
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 6
## PCA anomalies ( 24 ): WkDay-1, WkDay-4, WkDay-7, WkDay-13, WkDay-15, WkDay-26, WkDay-
## FA anomalies ( 10 ): WkDay-1, WkDay-4, WkDay-26, WkDay-54, WkDay-66, WkDay-76, WkDay-
## ICA anomalies ( 9 ): WkDay-1, WkDay-2, WkDay-13, WkDay-14, WkDay-54, WkDay-222, WkDay
##
## ============================================================
## Location: Loc12
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 8
## PCA anomalies ( 23 ): WkDay-1, WkDay-4, WkDay-7, WkDay-19, WkDay-55, WkDay-131, WkDay
## FA anomalies ( 14 ): WkDay-7, WkDay-19, WkDay-55, WkDay-70, WkDay-77, WkDay-152, WkDa
## ICA anomalies ( 19 ): WkDay-1, WkDay-7, WkDay-19, WkDay-70, WkDay-77, WkDay-121, WkDa
##
## ============================================================
## Location: Loc13
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 9
## PCA anomalies ( 20 ): WkDay-1, WkDay-25, WkDay-31, WkDay-33, WkDay-34, WkDay-98, WkDa
## FA anomalies ( 9 ): WkDay-1, WkDay-50, WkDay-98, WkDay-139, WkDay-269, WkDay-271, WkD
## ICA anomalies ( 17 ): WkDay-1, WkDay-13, WkDay-16, WkDay-27, WkDay-37, WkDay-98, WkDa
##
## ============================================================
## Location: Loc14
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 8
## PCA anomalies ( 21 ): WkDay-1, WkDay-4, WkDay-24, WkDay-64, WkDay-106, WkDay-163, WkD
## FA anomalies ( 8 ): WkDay-24, WkDay-163, WkDay-205, WkDay-217, WkDay-274, WkDay-330,
## ICA anomalies ( 6 ): WkDay-81, WkDay-106, WkDay-175, WkDay-230, WkDay-232, WkDay-364
##
## ============================================================
```

```
## Location: Loc15
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 8
## PCA anomalies ( 20 ): WkDay-1, WkDay-5, WkDay-33, WkDay-66, WkDay-80, WkDay-83, WkDay
## FA anomalies ( 10 ): WkDay-1, WkDay-5, WkDay-33, WkDay-161, WkDay-232, WkDay-308, WkD
## ICA anomalies ( 8 ): WkDay-1, WkDay-161, WkDay-184, WkDay-214, WkDay-316, WkDay-317,
##
## ============================================================
## Location: Loc16
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 9
## PCA anomalies ( 29 ): WkDay-1, WkDay-9, WkDay-36, WkDay-45, WkDay-62, WkDay-77, WkDay
## FA anomalies ( 10 ): WkDay-1, WkDay-178, WkDay-230, WkDay-232, WkDay-260, WkDay-317,
## ICA anomalies ( 11 ): WkDay-9, WkDay-78, WkDay-191, WkDay-213, WkDay-222, WkDay-229,
##
## ============================================================
## Location: Loc17
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 8
## PCA anomalies ( 31 ): WkDay-1, WkDay-2, WkDay-6, WkDay-13, WkDay-46, WkDay-54, WkDay-
## FA anomalies ( 12 ): WkDay-1, WkDay-6, WkDay-13, WkDay-66, WkDay-80, WkDay-106, WkDay
## ICA anomalies ( 8 ): WkDay-2, WkDay-24, WkDay-54, WkDay-109, WkDay-212, WkDay-232, Wk
##
## ============================================================
## Location: Loc18
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 5
## PCA anomalies ( 15 ): WkDay-1, WkDay-26, WkDay-51, WkDay-93, WkDay-184, WkDay-199, Wk
## FA anomalies ( 0 ):
## ICA anomalies ( 5 ): WkDay-1, WkDay-239, WkDay-258, WkDay-259, WkDay-325
##
## ============================================================
## Location: Loc19
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 9
## PCA anomalies ( 21 ): WkDay-1, WkDay-14, WkDay-74, WkDay-76, WkDay-80, WkDay-114, WkD
## FA anomalies ( 10 ): WkDay-14, WkDay-80, WkDay-104, WkDay-114, WkDay-117, WkDay-176,
## ICA anomalies ( 13 ): WkDay-1, WkDay-24, WkDay-28, WkDay-74, WkDay-82, WkDay-137, WkD
##
## ============================================================
## Location: Loc20
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 6
## PCA anomalies ( 19 ): WkDay-1, WkDay-27, WkDay-29, WkDay-31, WkDay-49, WkDay-80, WkDa
## FA anomalies ( 6 ): WkDay-1, WkDay-29, WkDay-31, WkDay-49, WkDay-354, WkDay-377
```

```
## ICA anomalies ( 4 ): WkDay-1, WkDay-205, WkDay-206, WkDay-212
##
## ============================================================
## Location: Loc21
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 5
## PCA anomalies ( 10 ): WkDay-83, WkDay-133, WkDay-188, WkDay-215, WkDay-223, WkDay-224
## FA anomalies ( 0 ):
## ICA anomalies ( 7 ): WkDay-1, WkDay-52, WkDay-232, WkDay-264, WkDay-327, WkDay-328, W
##
## ============================================================
## Location: Loc22
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 6
## PCA anomalies ( 13 ): WkDay-1, WkDay-11, WkDay-69, WkDay-70, WkDay-127, WkDay-140, Wk
## FA anomalies ( 10 ): WkDay-11, WkDay-69, WkDay-70, WkDay-140, WkDay-159, WkDay-221, W
## ICA anomalies ( 13 ): WkDay-1, WkDay-2, WkDay-69, WkDay-71, WkDay-127, WkDay-222, WkD
##
## ============================================================
## Location: Loc23
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 6
## PCA anomalies ( 20 ): WkDay-1, WkDay-33, WkDay-35, WkDay-47, WkDay-137, WkDay-156, Wk
## FA anomalies ( 16 ): WkDay-1, WkDay-33, WkDay-37, WkDay-38, WkDay-47, WkDay-137, WkDa
## ICA anomalies ( 19 ): WkDay-1, WkDay-33, WkDay-37, WkDay-38, WkDay-137, WkDay-176, Wk
##
## ============================================================
## Location: Loc24
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 3
## PCA anomalies ( 5 ): WkDay-97, WkDay-162, WkDay-251, WkDay-278, WkDay-340
## FA anomalies ( 0 ):
## ICA anomalies ( 12 ): WkDay-57, WkDay-63, WkDay-100, WkDay-103, WkDay-162, WkDay-239,
##
## ============================================================
## Location: Loc25
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
## PCA components: 7
## PCA anomalies ( 22 ): WkDay-1, WkDay-31, WkDay-33, WkDay-35, WkDay-36, WkDay-37, WkDa
## FA anomalies ( 6 ): WkDay-95, WkDay-102, WkDay-251, WkDay-257, WkDay-295, WkDay-384
## ICA anomalies ( 19 ): WkDay-35, WkDay-36, WkDay-37, WkDay-47, WkDay-52, WkDay-63, WkD
##
## ============================================================
## Location: Loc26
## = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
```

```
## PCA components: 2
## PCA anomalies ( 6 ): WkDay-1, WkDay-33, WkDay-226, WkDay-227, WkDay-235, WkDay-239
## FA anomalies ( 0 ):
## ICA anomalies ( 7 ): WkDay-18, WkDay-221, WkDay-222, WkDay-226, WkDay-231, WkDay-234,
```

Summary Tables

```r
pca_summary <- data.frame(
  Location = names(location_results),
  k_pca = sapply(location_results, function(res) res$k_pca),
  Variance_Explained = sapply(location_results, function(res) {
    pca <- res$pca
    var_expl <- pca$sdev^2 / sum(pca$sdev^2)
    cumvar <- cumsum(var_expl)
    round(cumvar[res$k_pca], 4)
  }),
  PCA_Anomaly_Count = sapply(location_results, function(res) length(res$pca_anom_days))
)
rownames(pca_summary) <- NULL
print(pca_summary)
```

```
##    Location k_pca Variance_Explained PCA_Anomaly_Count
## 1      Loc1     5             0.9145                30
## 2      Loc2     7             0.9192                20
## 3      Loc3     5             0.9138                24
## 4      Loc4     7             0.9119                20
## 5      Loc5     7             0.9058                18
## 6      Loc6     6             0.9127                 6
## 7      Loc7     2             0.9369                 4
## 8      Loc8     8             0.9054                23
## 9      Loc9     4             0.9011                14
## 10    Loc10     5             0.9153                13
## 11    Loc11     6             0.9024                24
## 12    Loc12     8             0.9176                23
## 13    Loc13     9             0.9161                20
## 14    Loc14     8             0.9106                21
## 15    Loc15     8             0.9186                20
## 16    Loc16     9             0.9187                29
## 17    Loc17     8             0.9153                31
## 18    Loc18     5             0.9117                15
## 19    Loc19     9             0.9006                21
## 20    Loc20     6             0.9132                19
```

```
## 21     Loc21     5            0.9107                  10
## 22     Loc22     6            0.9168                  13
## 23     Loc23     6            0.9027                  20
## 24     Loc24     3            0.9181                   5
## 25     Loc25     7            0.9176                  22
## 26     Loc26     2            0.9404                   6
```

```r
fa_summary <- data.frame(
  Location = names(location_results),
  FA_Success = sapply(location_results, \(res) !is.null(res$fa_model)),
  Factors_Extracted = sapply(location_results, function(res) {
    if (is.null(res$fa_model)) return(0)
    ncol(res$fa_model$loadings)
  }),
  FA_Anomaly_Count = sapply(location_results, function(res) length(res$fa_anom_days))
)
rownames(fa_summary) <- NULL
print(fa_summary)
```

```
##    Location FA_Success Factors_Extracted FA_Anomaly_Count
## 1      Loc1      FALSE                 0                0
## 2      Loc2       TRUE                 3               10
## 3      Loc3      FALSE                 0                0
## 4      Loc4       TRUE                 3               11
## 5      Loc5       TRUE                 3                5
## 6      Loc6       TRUE                 3                4
## 7      Loc7      FALSE                 0                0
## 8      Loc8       TRUE                 3               10
## 9      Loc9      FALSE                 0                0
## 10    Loc10      FALSE                 0                0
## 11    Loc11       TRUE                 3               10
## 12    Loc12       TRUE                 3               14
## 13    Loc13       TRUE                 3                9
## 14    Loc14       TRUE                 3                8
## 15    Loc15       TRUE                 3               10
## 16    Loc16       TRUE                 3               10
## 17    Loc17       TRUE                 3               12
## 18    Loc18      FALSE                 0                0
## 19    Loc19       TRUE                 3               10
## 20    Loc20       TRUE                 3                6
## 21    Loc21      FALSE                 0                0
## 22    Loc22       TRUE                 3               10
## 23    Loc23       TRUE                 3               16
## 24    Loc24      FALSE                 0                0
```

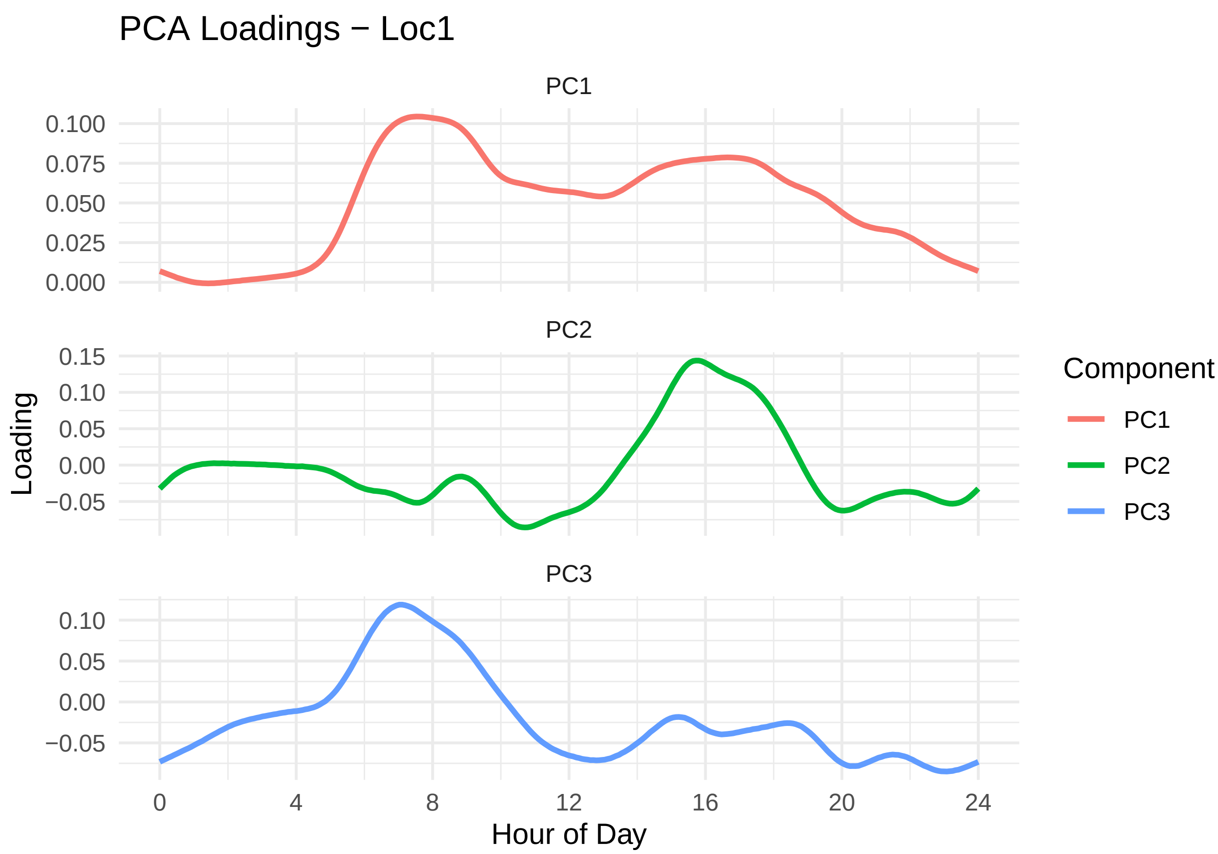
```
## 25     Loc25       TRUE                  3                6
## 26     Loc26      FALSE                  0                0
```

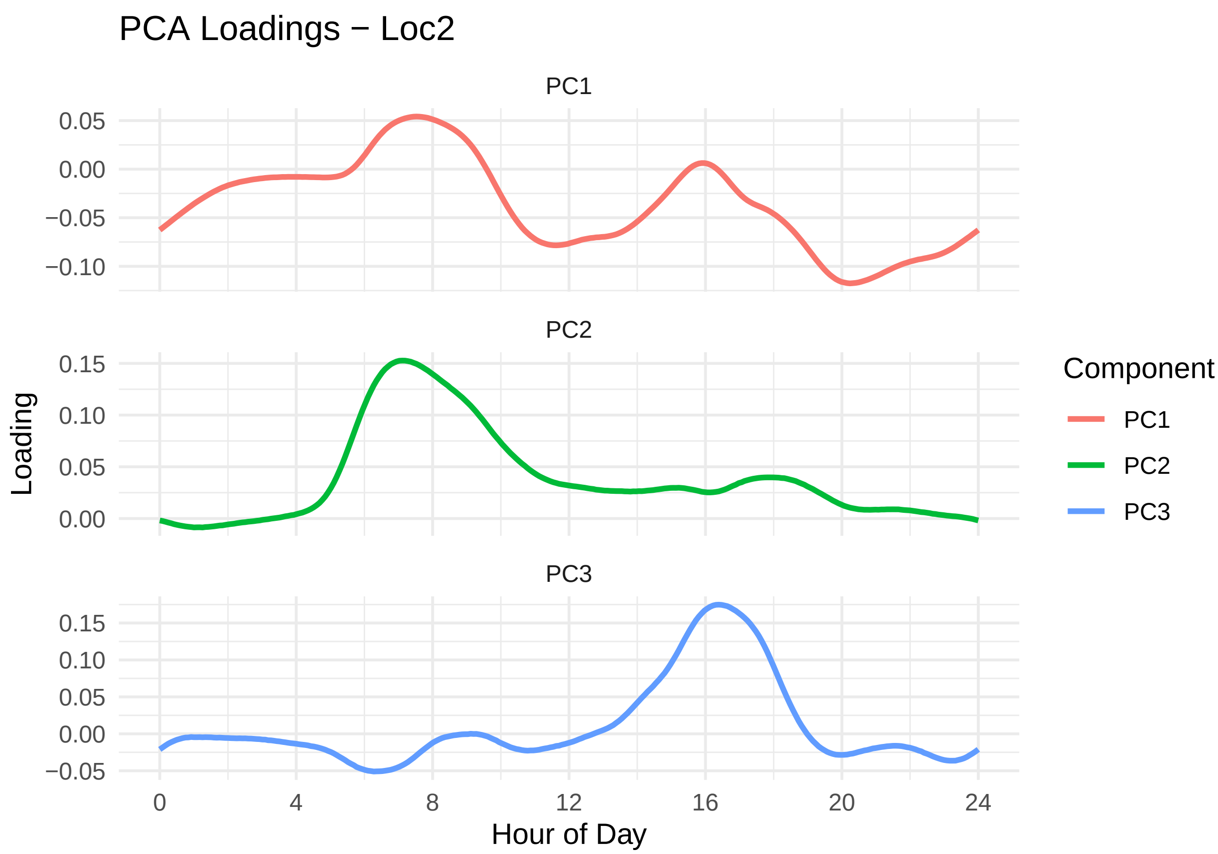
```
ica_summary <- data.frame(
  Location = names(location_results),
  ICA_Success = sapply(location_results, \(res) !is.null(res$ica_scores)),
  ICA_Components = sapply(location_results, function(res) {
    if (is.null(res$ica_scores)) return(0)
    ncol(res$ica_scores)
  }),
  ICA_Anomaly_Count = sapply(location_results, function(res) length(res$ica_anom_days))
)
rownames(ica_summary) <- NULL
print(ica_summary)
```

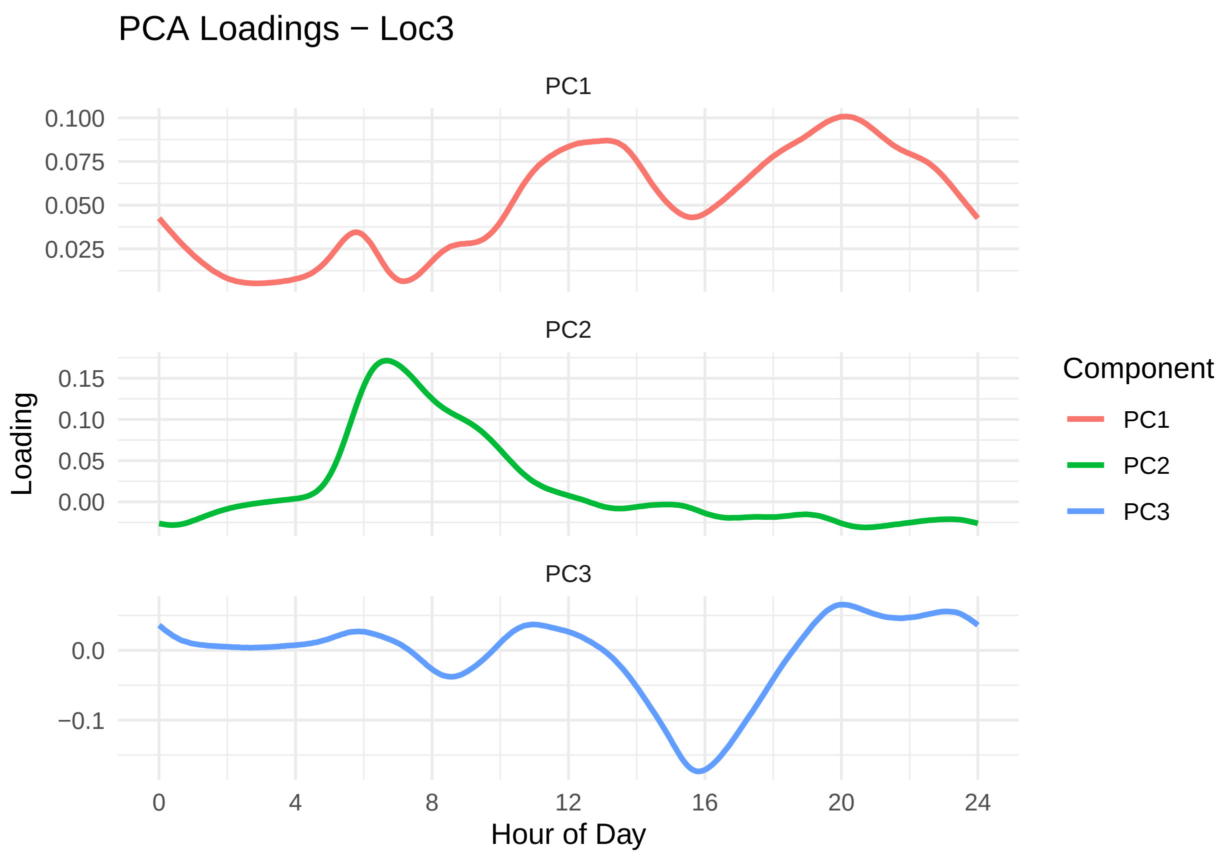
```
##    Location ICA_Success ICA_Components ICA_Anomaly_Count
## 1      Loc1        TRUE              3                17
## 2      Loc2        TRUE              3                14
## 3      Loc3        TRUE              3                17
## 4      Loc4        TRUE              3                10
## 5      Loc5        TRUE              3                 8
## 6      Loc6        TRUE              3                 3
## 7      Loc7        TRUE              2                15
## 8      Loc8        TRUE              3                 6
## 9      Loc9        TRUE              3                12
## 10    Loc10        TRUE              3                 7
## 11    Loc11        TRUE              3                 9
## 12    Loc12        TRUE              3                19
## 13    Loc13        TRUE              3                17
## 14    Loc14        TRUE              3                 6
## 15    Loc15        TRUE              3                 8
## 16    Loc16        TRUE              3                11
## 17    Loc17        TRUE              3                 8
## 18    Loc18        TRUE              3                 5
## 19    Loc19        TRUE              3                13
## 20    Loc20        TRUE              3                 4
## 21    Loc21        TRUE              3                 7
## 22    Loc22        TRUE              3                13
## 23    Loc23        TRUE              3                19
## 24    Loc24        TRUE              3                12
## 25    Loc25        TRUE              3                19
## 26    Loc26        TRUE              2                 7
```

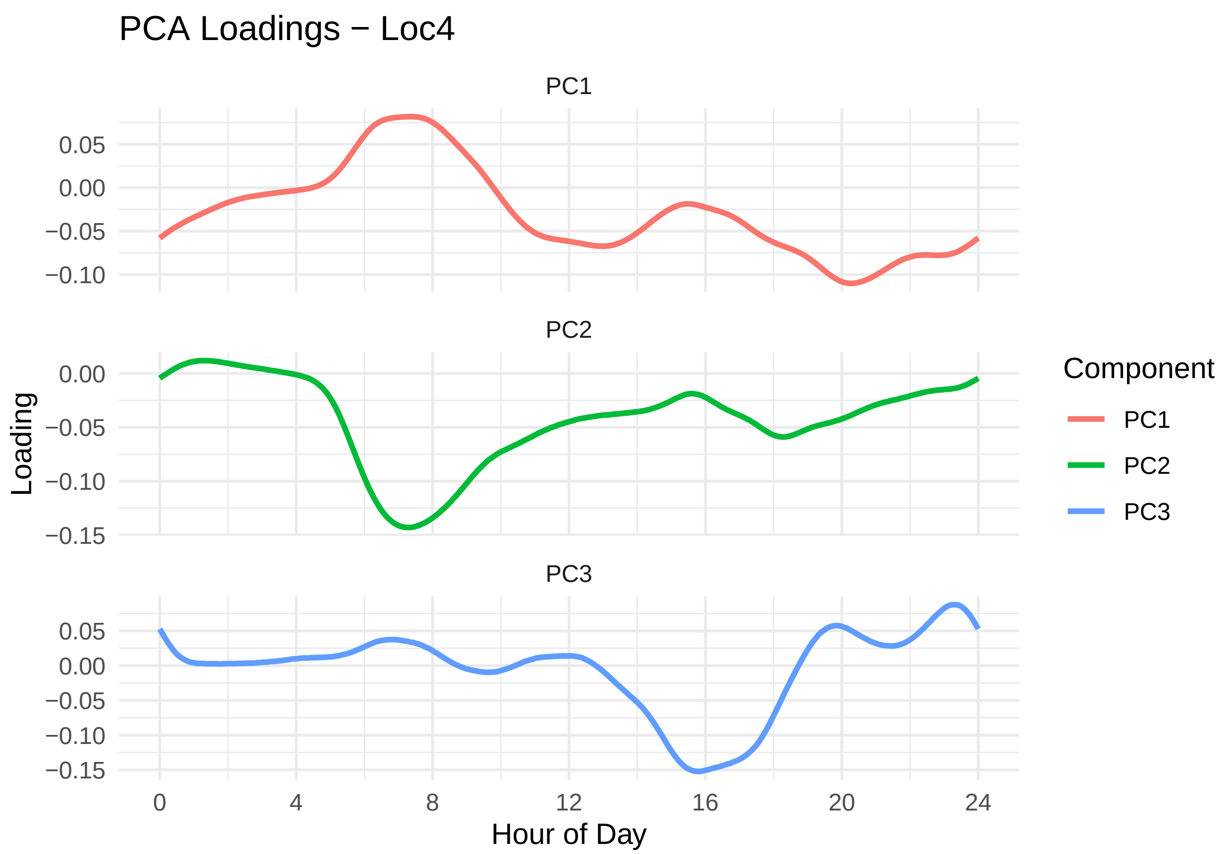
PCA Loadings Visualization

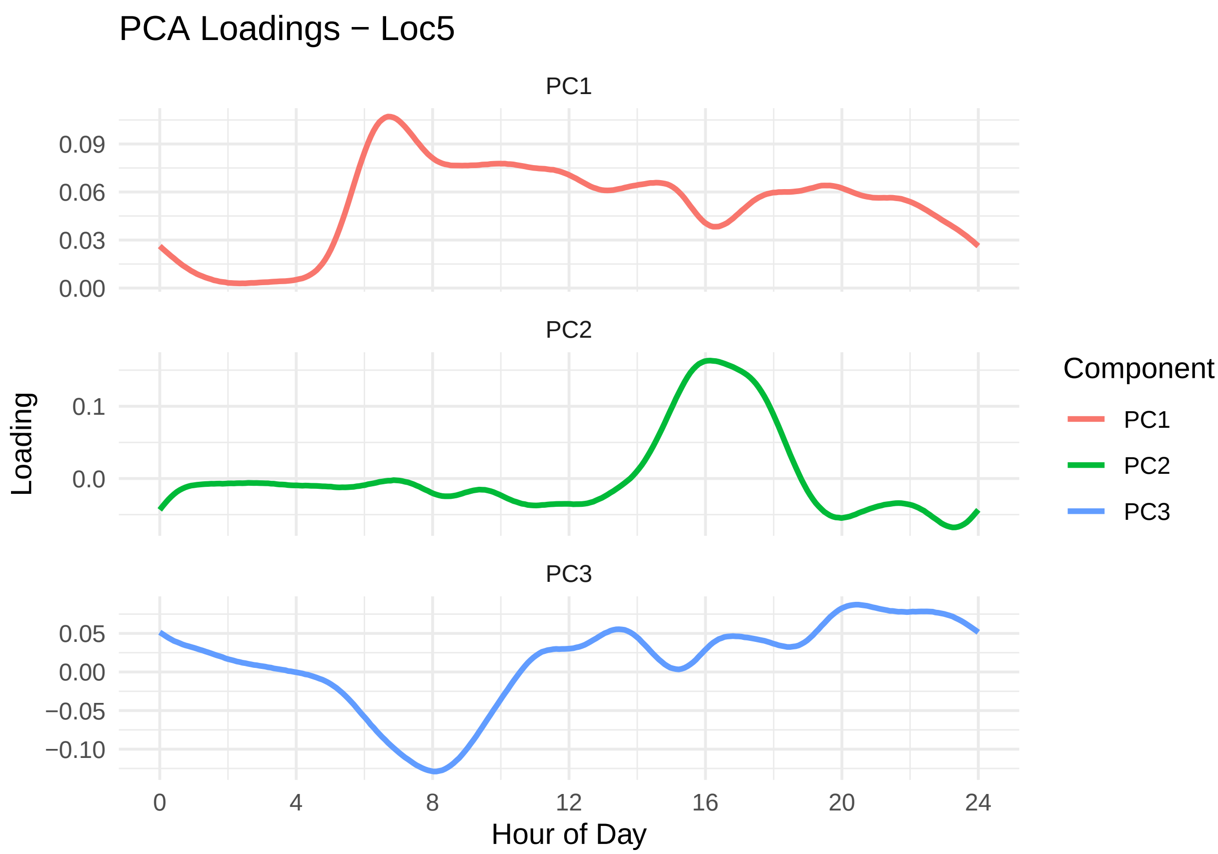
```
plot_pca_loadings <- function(res, loc_name, n_comp = 3) {
  loadings <- res$pca_loadings[, 1:min(n_comp, ncol(res$pca_loadings)), drop = FALSE]
  time_hours <- seq(0, 24, length.out = nrow(loadings))

  df <- data.frame(Time = time_hours, loadings)
  colnames(df) <- c("Time", paste0("PC", 1:ncol(loadings)))

  df_long <- pivot_longer(df, cols = -Time, names_to = "Component", values_to = "Loading

  ggplot(df_long, aes(x = Time, y = Loading, color = Component)) +
    geom_line(size = 1) +
    facet_wrap(~Component, ncol = 1, scales = "free_y") +
    labs(title = paste("PCA Loadings -", loc_name),
         x = "Hour of Day", y = "Loading") +
    theme_minimal() +
    scale_x_continuous(breaks = seq(0, 24, 4))
}

for (loc_name in names(location_results)) {
  print(plot_pca_loadings(location_results[[loc_name]], loc_name))
}
```

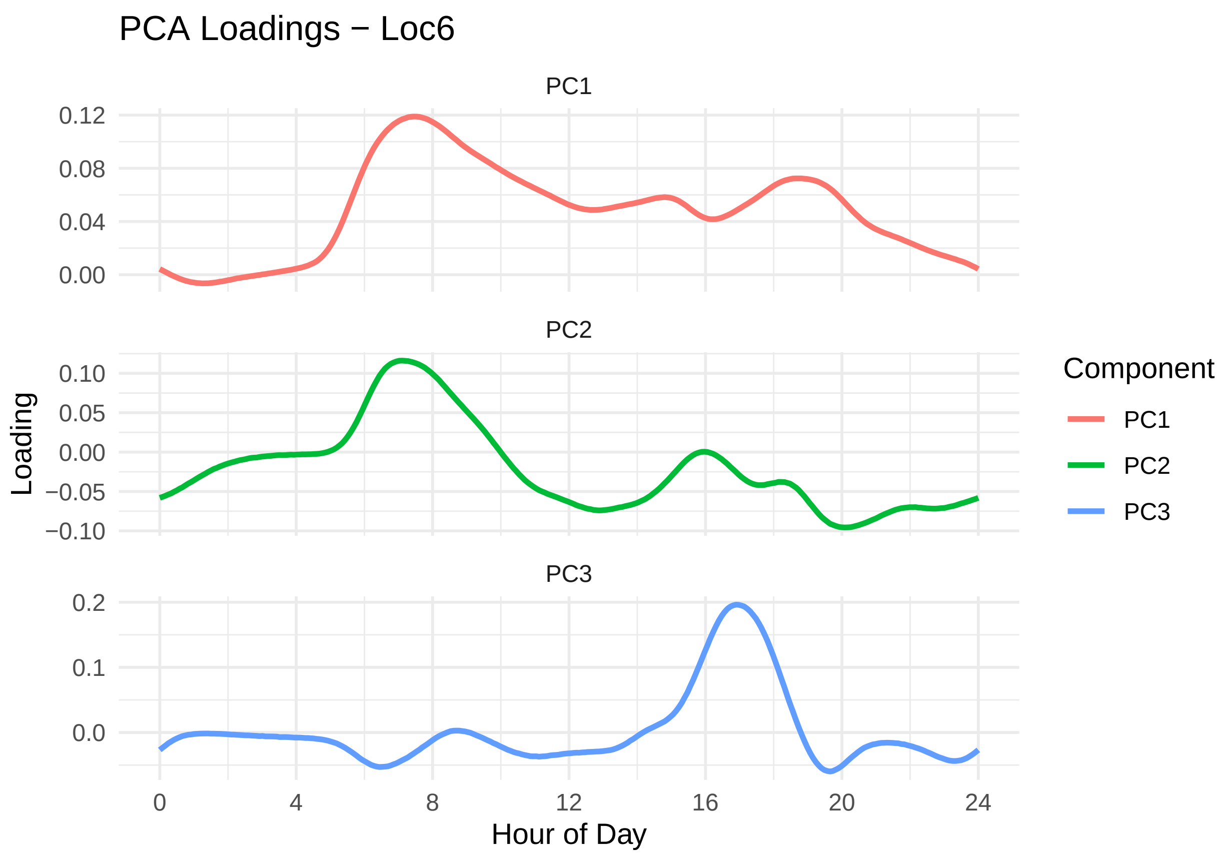

PCA Loadings − Loc1
PC1
0.100
0.075
0.050
0.025
0.000
PC2
0.15
0.10
0.05
0.00
−0.05
PC3
0.10
0.05
0.00
−0.05
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

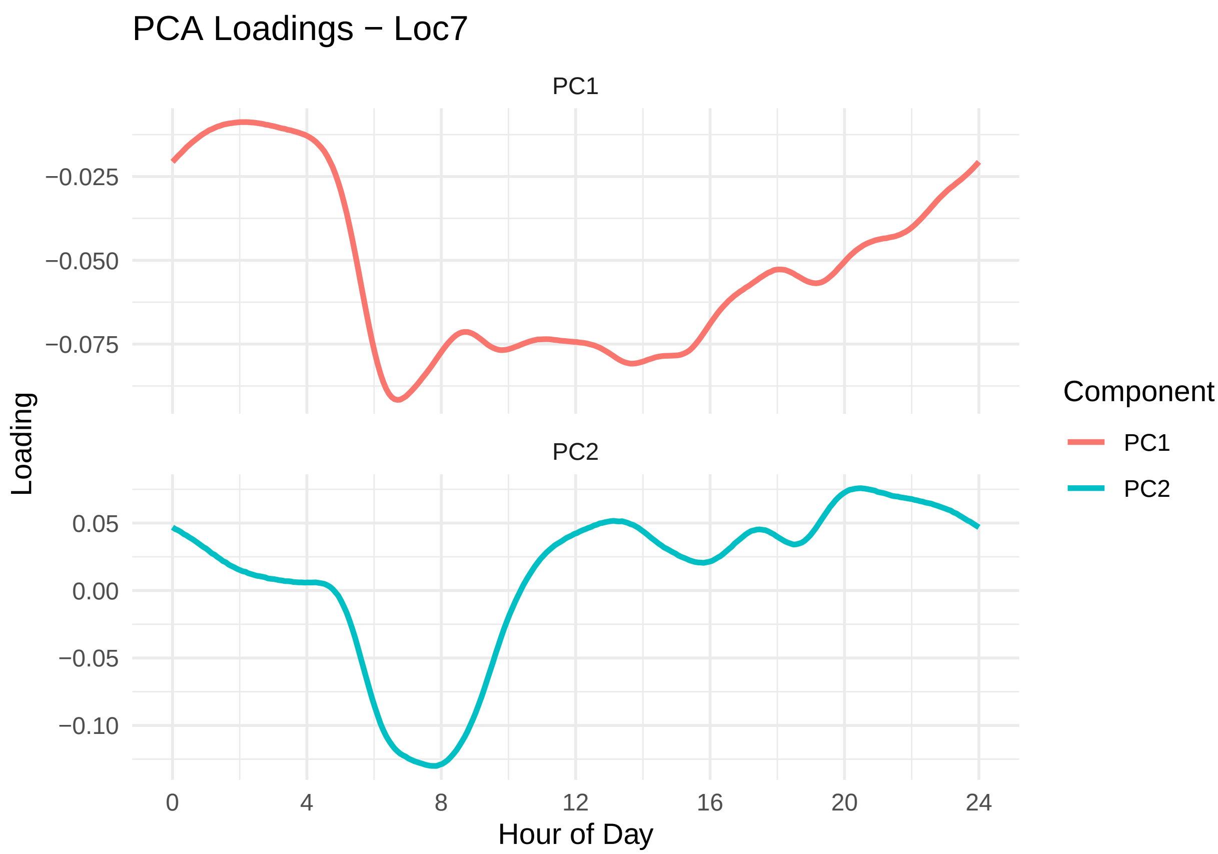

PCA Loadings − Loc2
PC1
PC2
PC3
Loading
Hour of Day
Component
PC1
PC2
PC3
0.05
0.00
−0.05
−0.10
0.15
0.10
0.05
0.00
0.15
0.10
0.05
0.00
−0.05
0
4
8
12
16
20
24


PCA Loadings − Loc3
PC1
0.100
0.075
0.050
0.025
PC2
0.15
0.10
0.05
0.00
PC3
0.0
−0.1
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

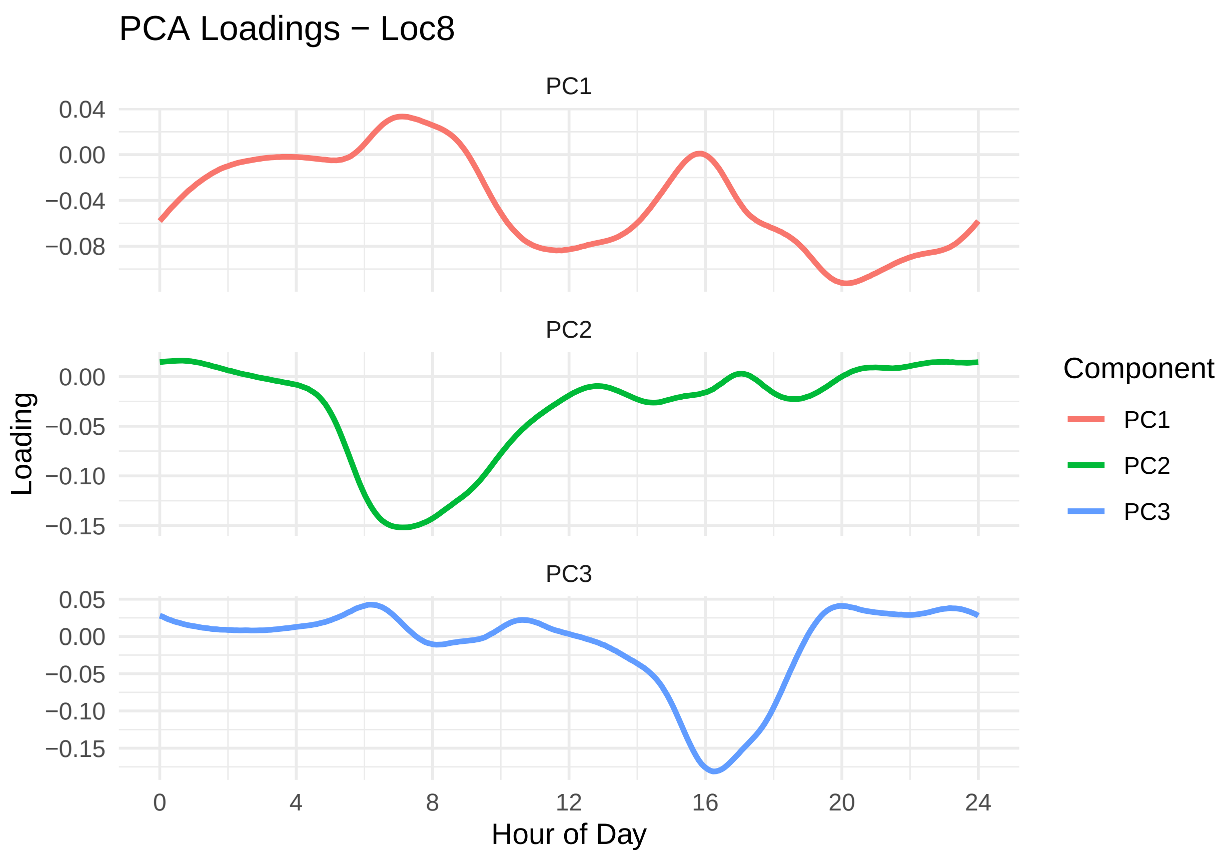

PCA Loadings − Loc4
PC1
0.05
0.00
−0.05
−0.10
PC2
0.00
−0.05
−0.10
−0.15
PC3
0.05
0.00
−0.05
−0.10
−0.15
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

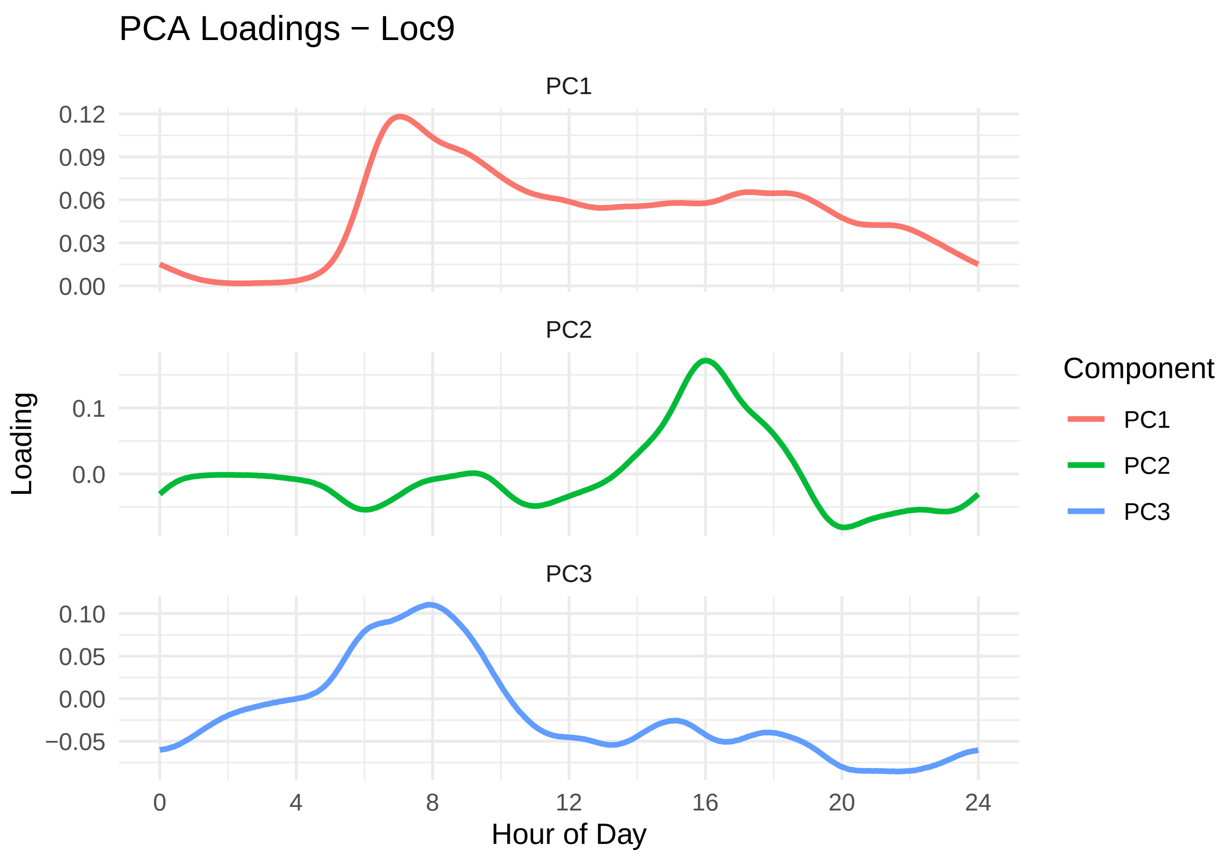

PCA Loadings − Loc5
PC1
PC2
PC3
Loading
Hour of Day
Component
0.09
0.06
0.03
0.00
0.1
0.0
0.05
0.00
−0.05
−0.10
0
4
8
12
16
20
24

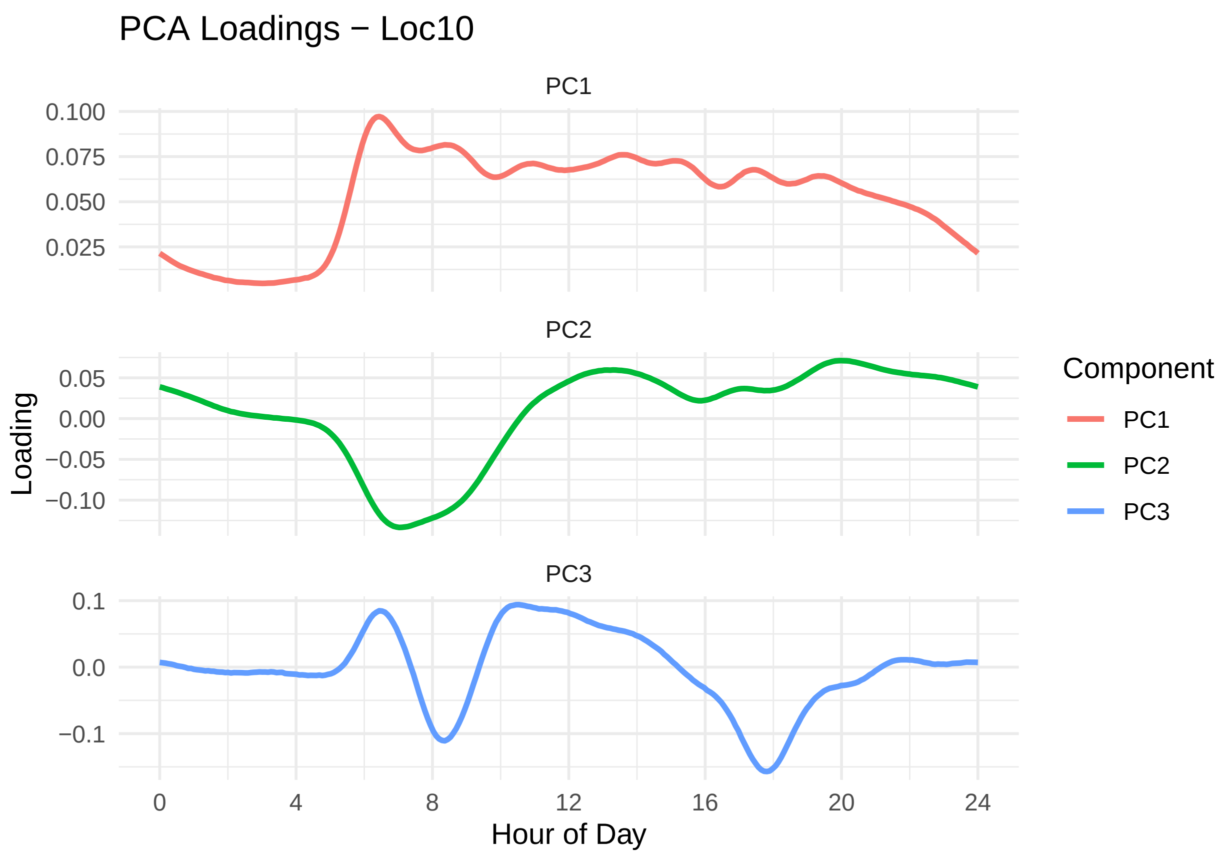

PCA Loadings − Loc6
PC1
0.12
0.08
0.04
0.00
PC2
0.10
0.05
0.00
−0.05
−0.10
PC3
0.2
0.1
0.0
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

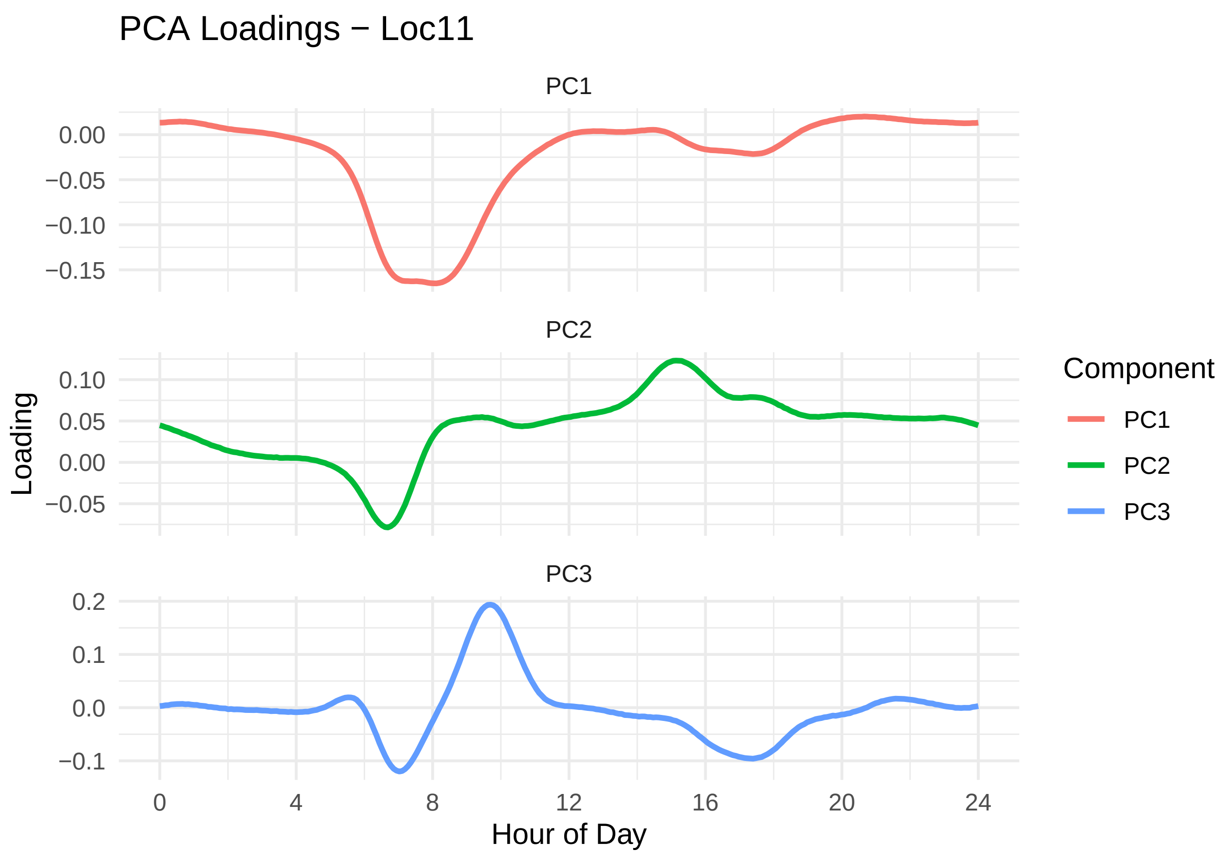


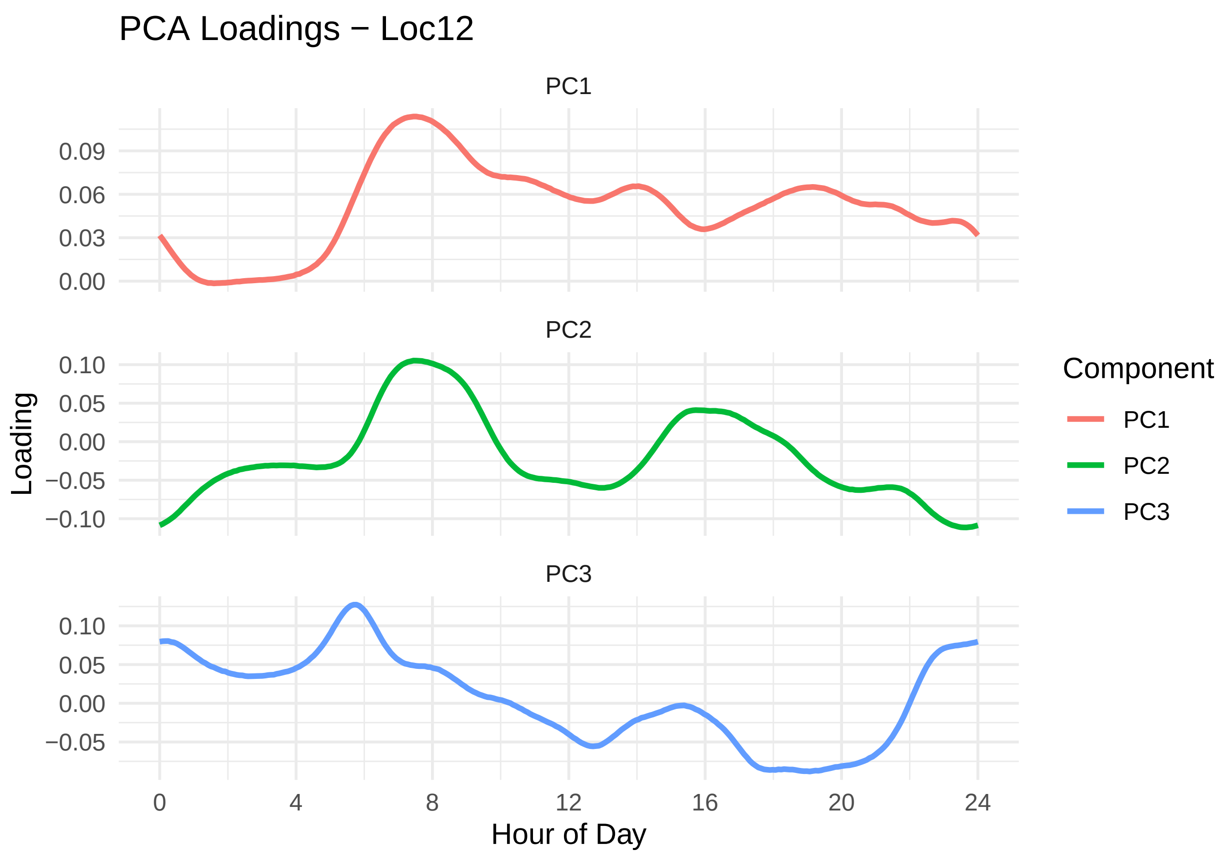
PCA Loadings − Loc7
PC1
PC2
−0.025
−0.050
−0.075
0.05
0.00
−0.05
−0.10
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2


PCA Loadings − Loc8
PC1
0.04
0.00
−0.04
−0.08
PC2
0.00
−0.05
−0.10
−0.15
PC3
0.05
0.00
−0.05
−0.10
−0.15
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

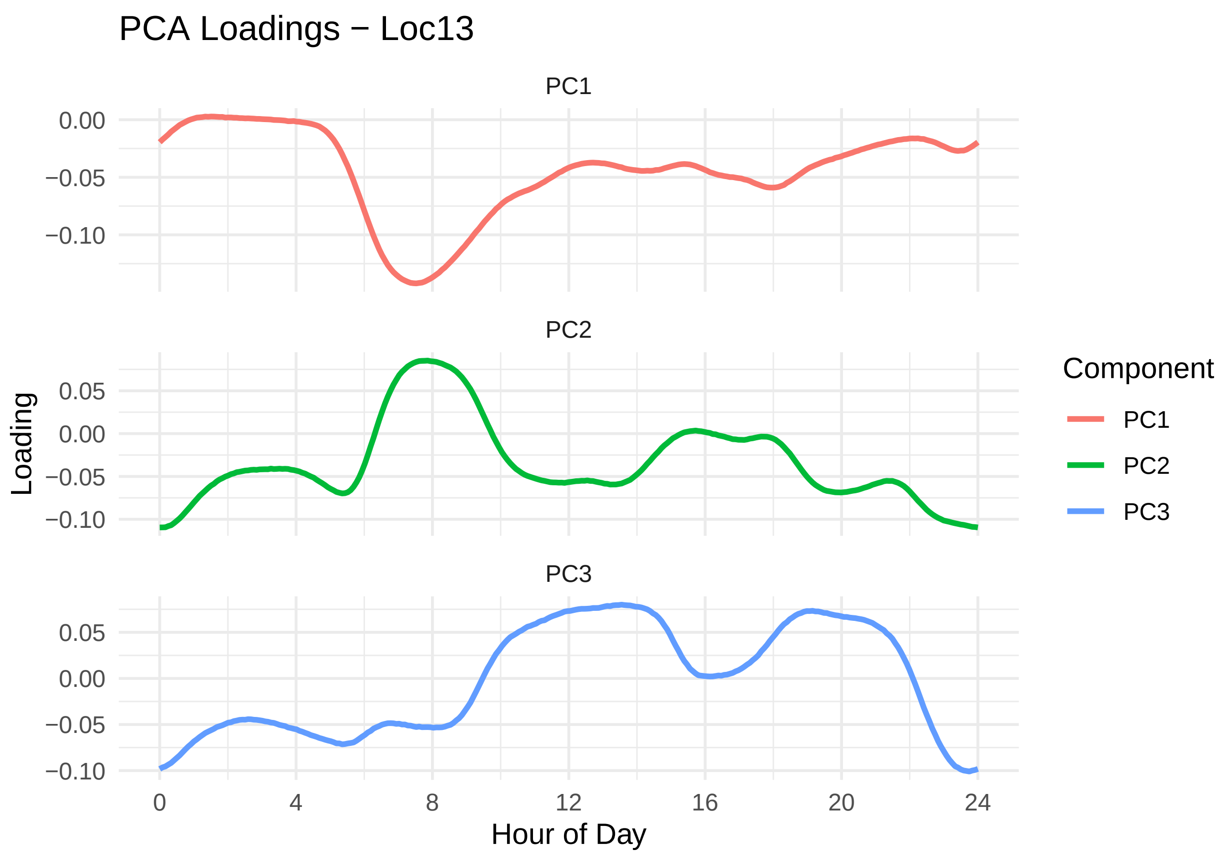

PCA Loadings − Loc9
PC1
0.12
0.09
0.06
0.03
0.00
PC2
0.1
0.0
PC3
0.10
0.05
0.00
−0.05
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

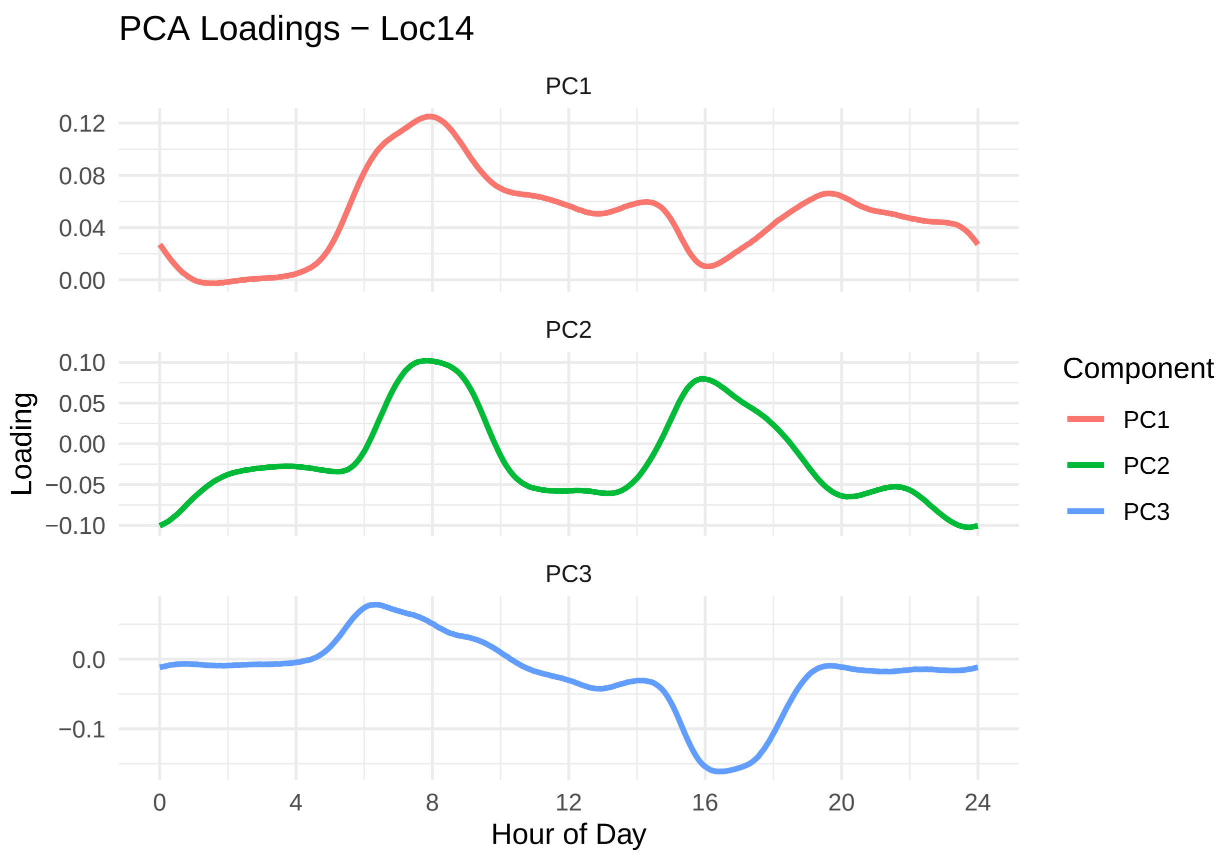

PCA Loadings − Loc10
PC1
0.100
0.075
0.050
0.025
PC2
0.05
0.00
−0.05
−0.10
PC3
0.1
0.0
−0.1
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

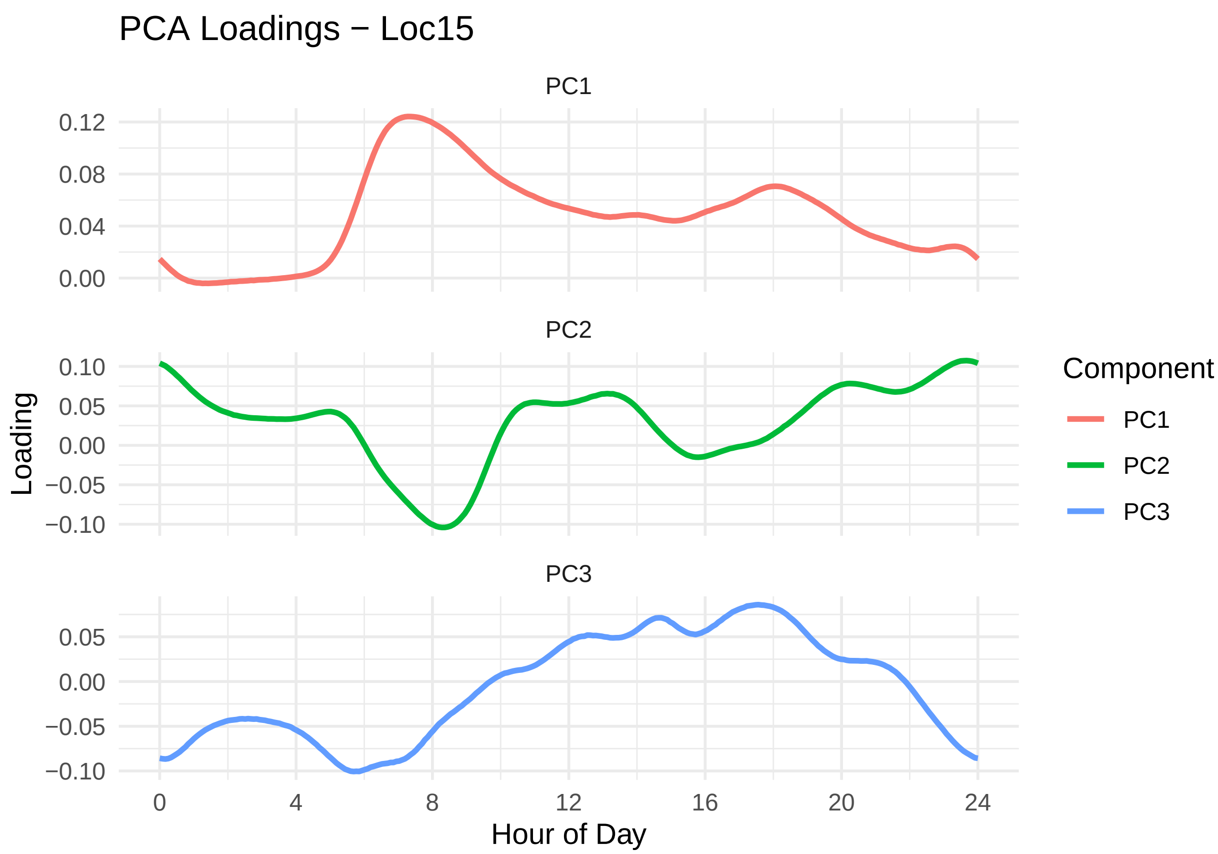

PCA Loadings − Loc11
PC1
0.00
−0.05
−0.10
−0.15
PC2
0.10
0.05
0.00
−0.05
PC3
0.2
0.1
0.0
−0.1
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3



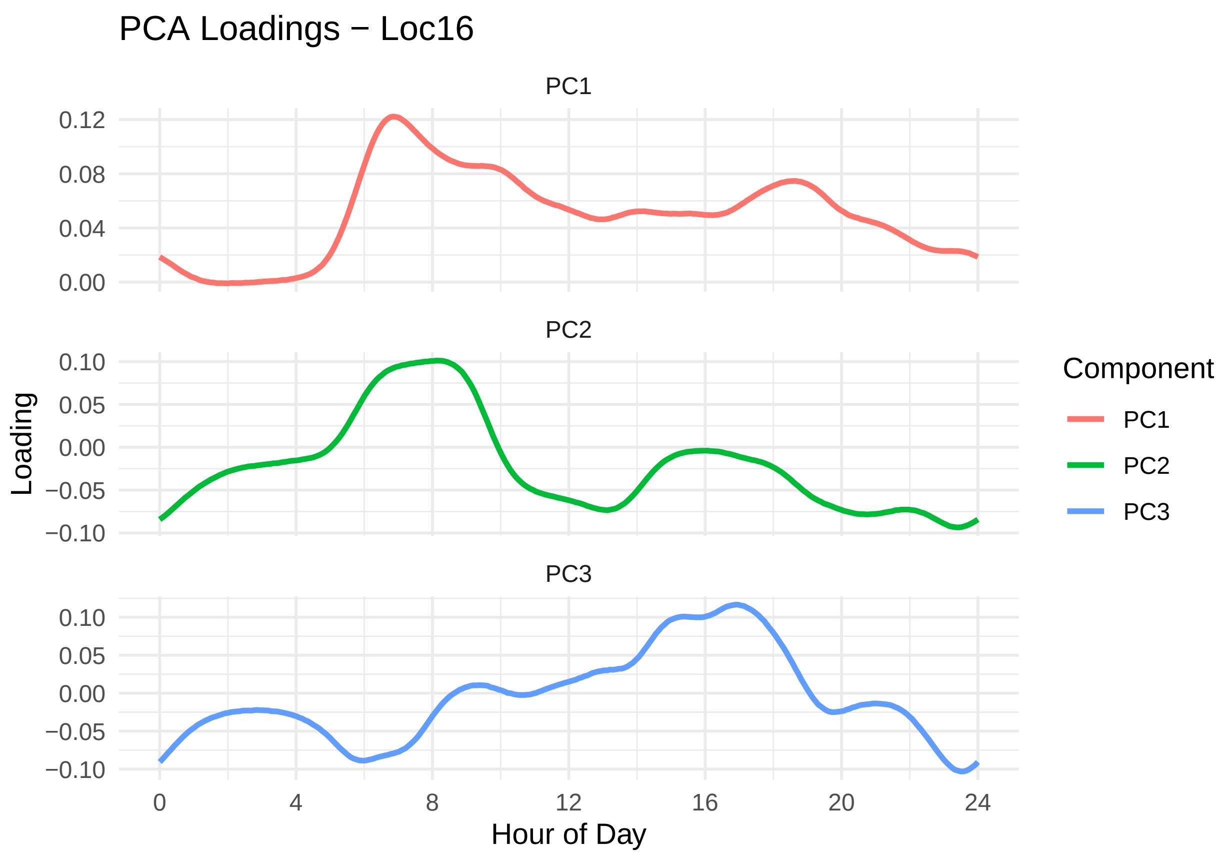
PCA Loadings − Loc12
PC1
0.09
0.06
0.03
0.00
PC2
0.10
0.05
0.00
−0.05
−0.10
PC3
0.10
0.05
0.00
−0.05
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

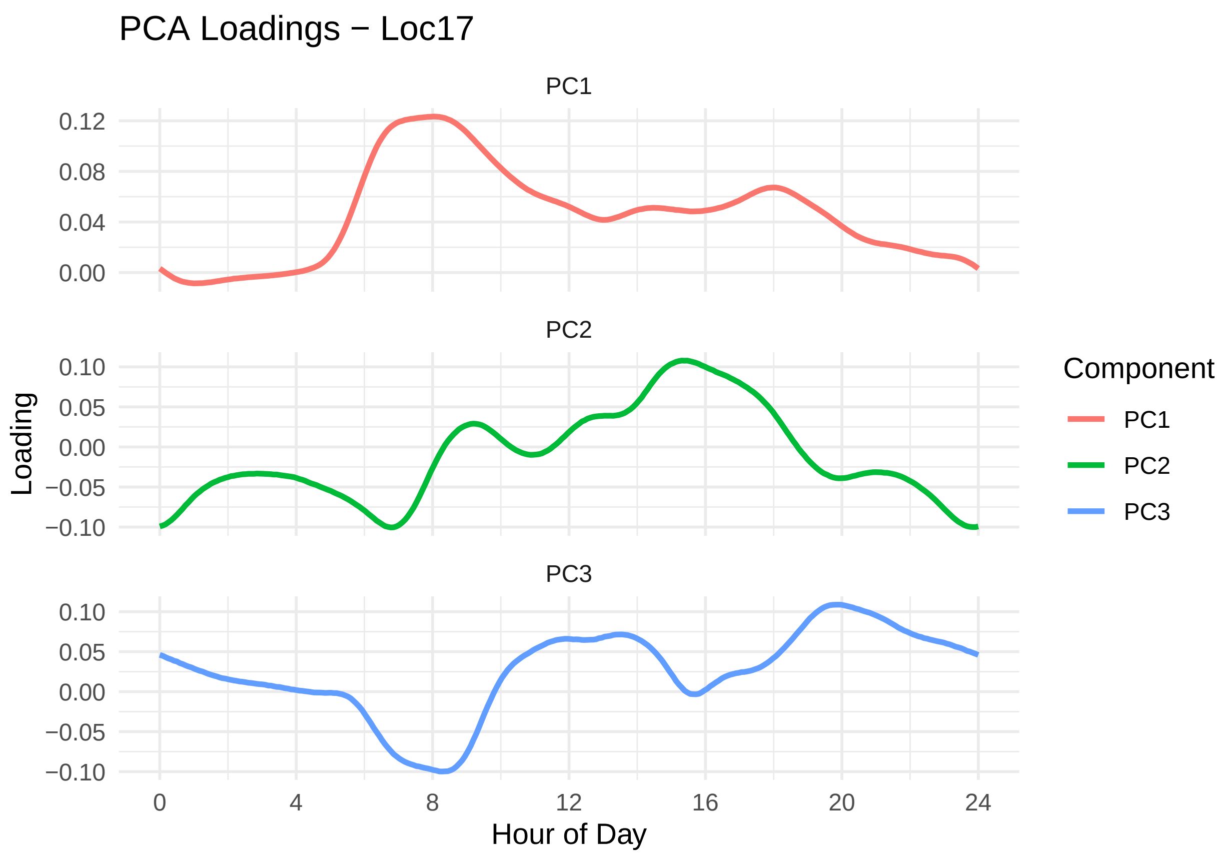

PCA Loadings − Loc13
PC1
0.00
−0.05
−0.10
PC2
0.05
0.00
−0.05
−0.10
PC3
0.05
0.00
−0.05
−0.10
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

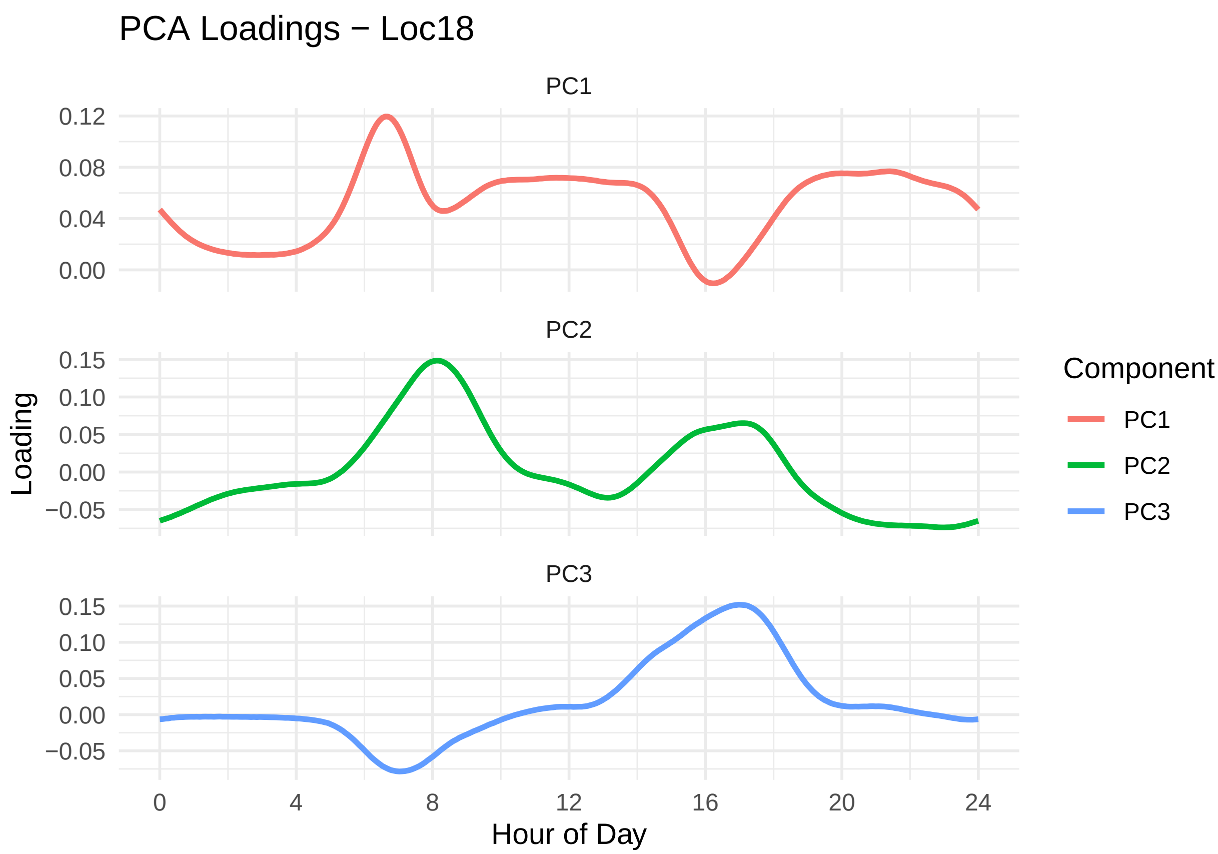

PCA Loadings − Loc14
PC1
0.12
0.08
0.04
0.00
PC2
0.10
0.05
0.00
−0.05
−0.10
PC3
0.0
−0.1
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3



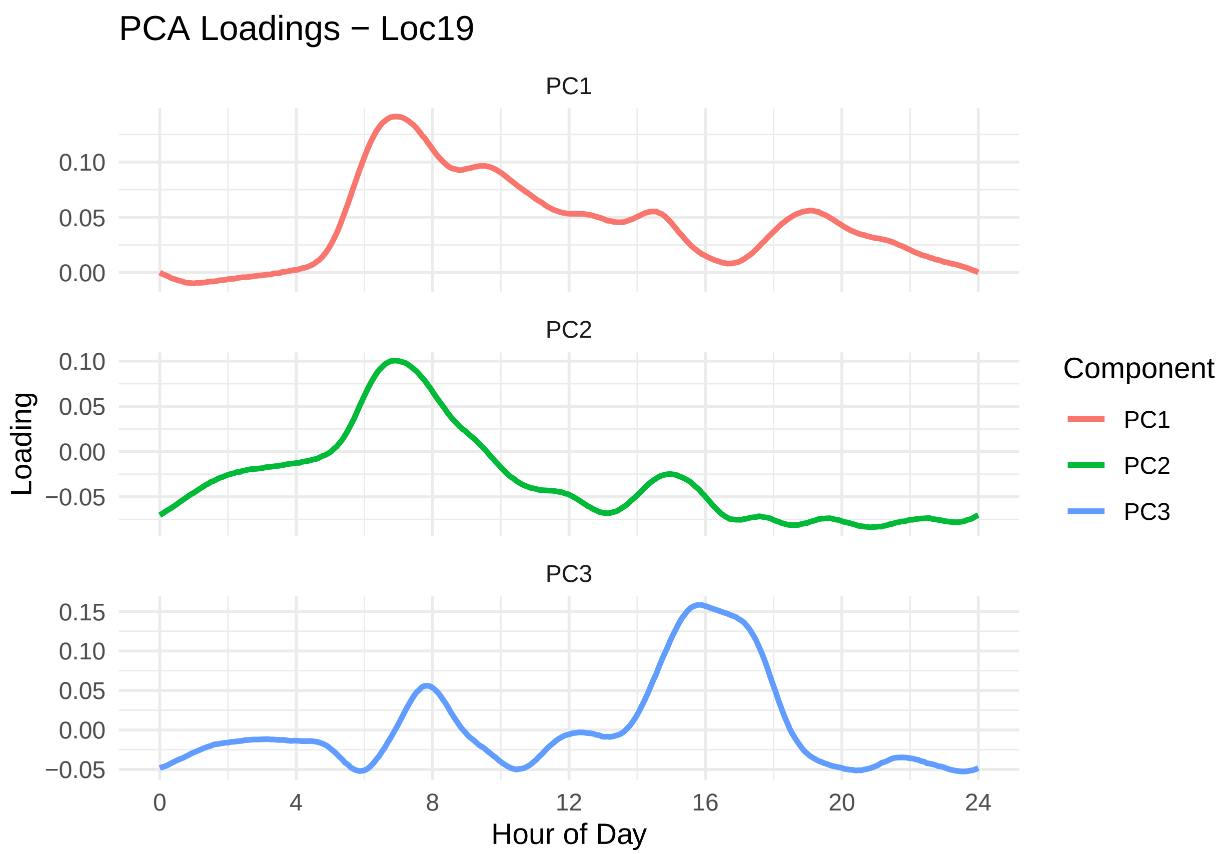
PCA Loadings − Loc15
PC1
0.12
0.08
0.04
0.00
PC2
0.10
0.05
0.00
−0.05
−0.10
PC3
0.05
0.00
−0.05
−0.10
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3


PCA Loadings − Loc16
PC1
PC2
PC3
Loading
Hour of Day
0
4
8
12
16
20
24
0.12
0.08
0.04
0.00
0.10
0.05
0.00
−0.05
−0.10
Component
PC1
PC2
PC3

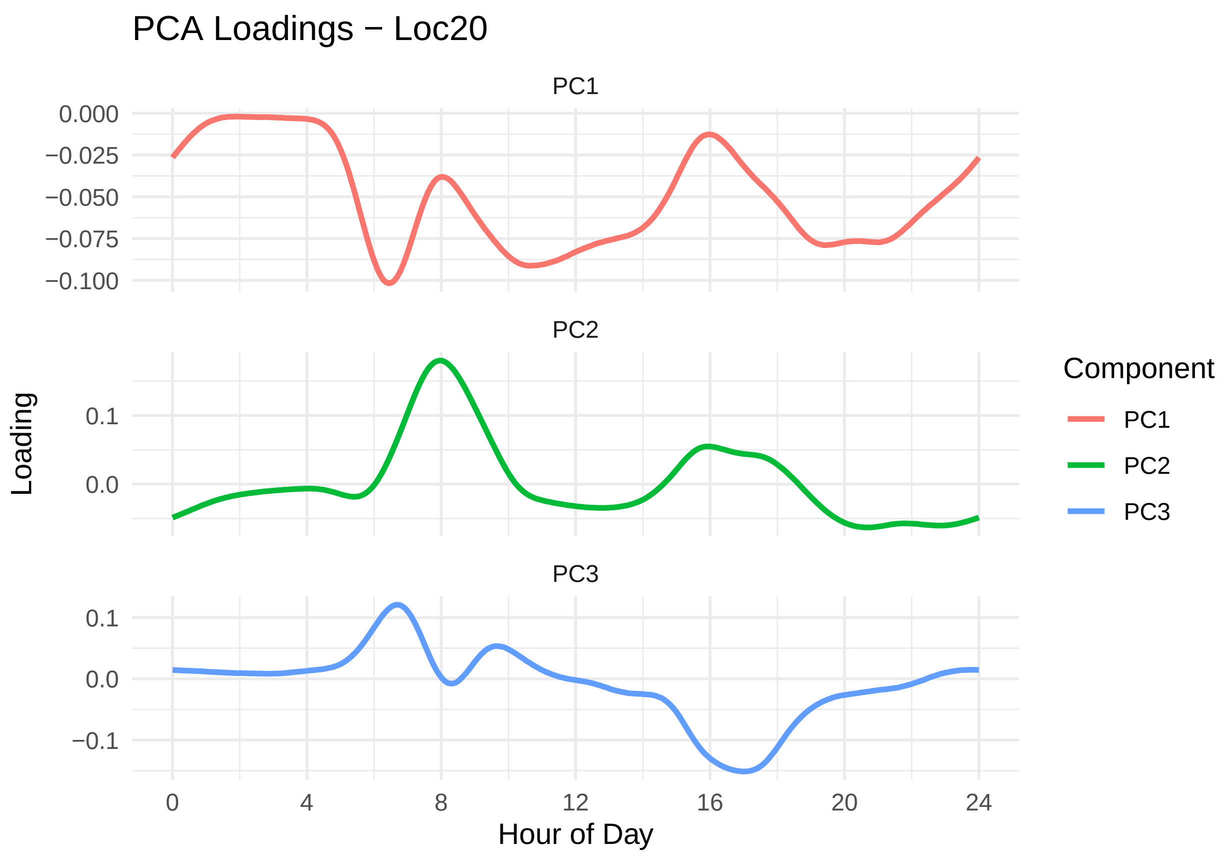

PCA Loadings − Loc17
PC1
0.12
0.08
0.04
0.00
PC2
0.10
0.05
0.00
−0.05
−0.10
PC3
0.10
0.05
0.00
−0.05
−0.10
0
4
8
12
16
20
24
Hour of Day
Loading
Component
PC1
PC2
PC3

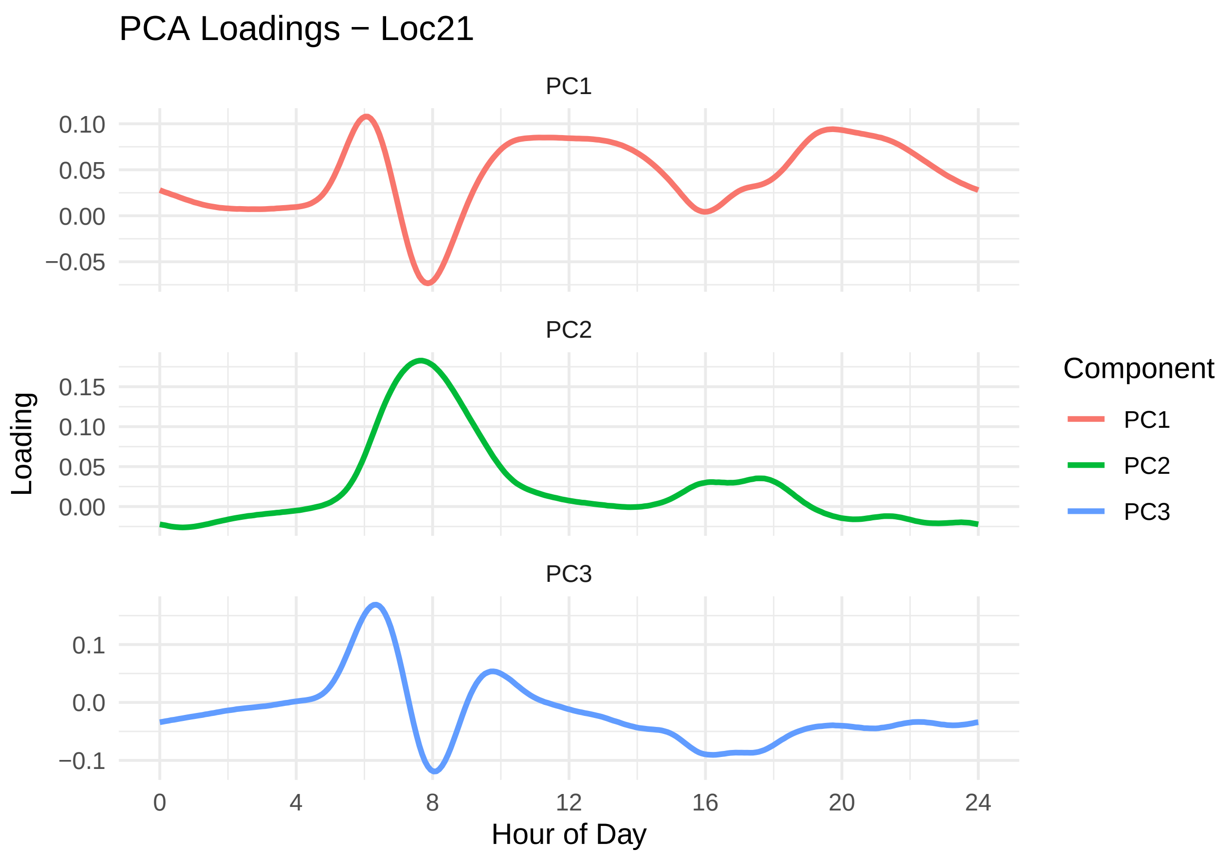


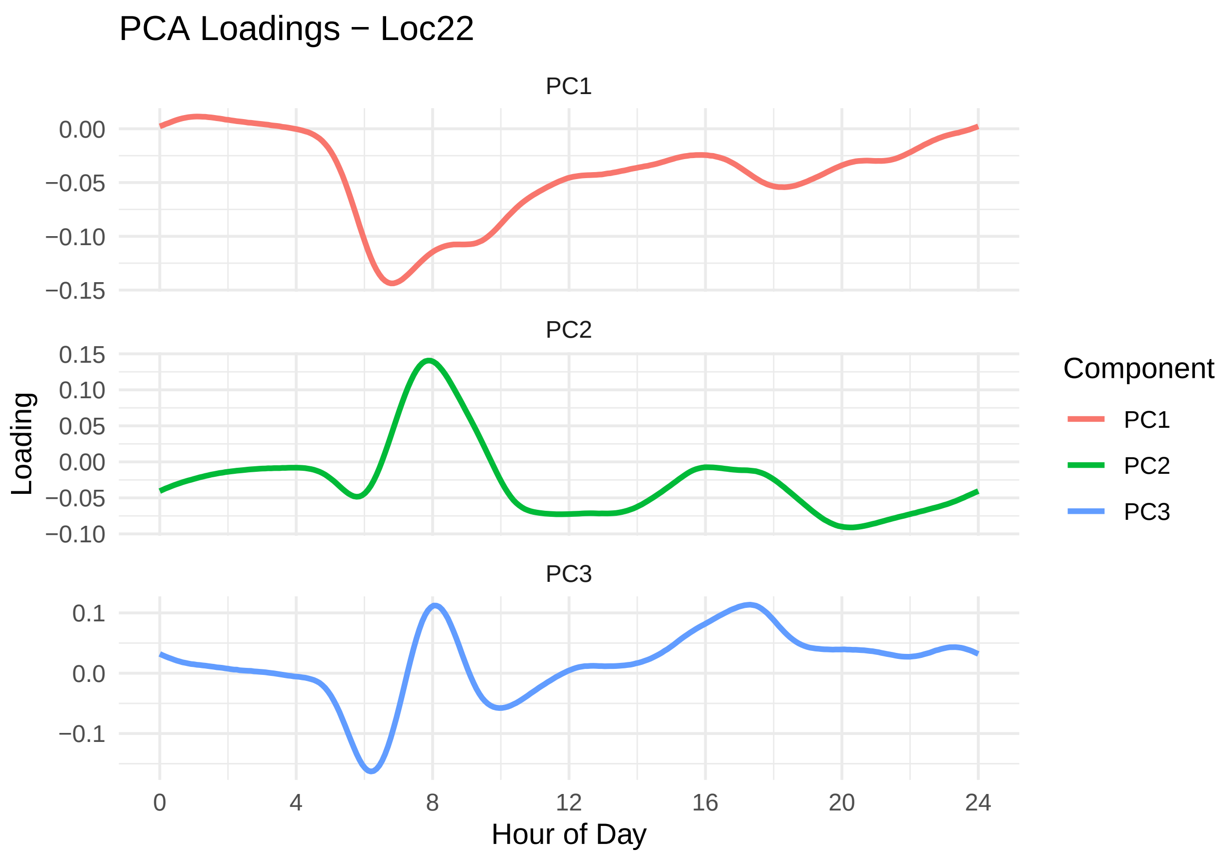
PCA Loadings − Loc18
PC1
PC2
PC3
0.12
0.08
0.04
0.00
0.15
0.10
0.05
0.00
−0.05
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

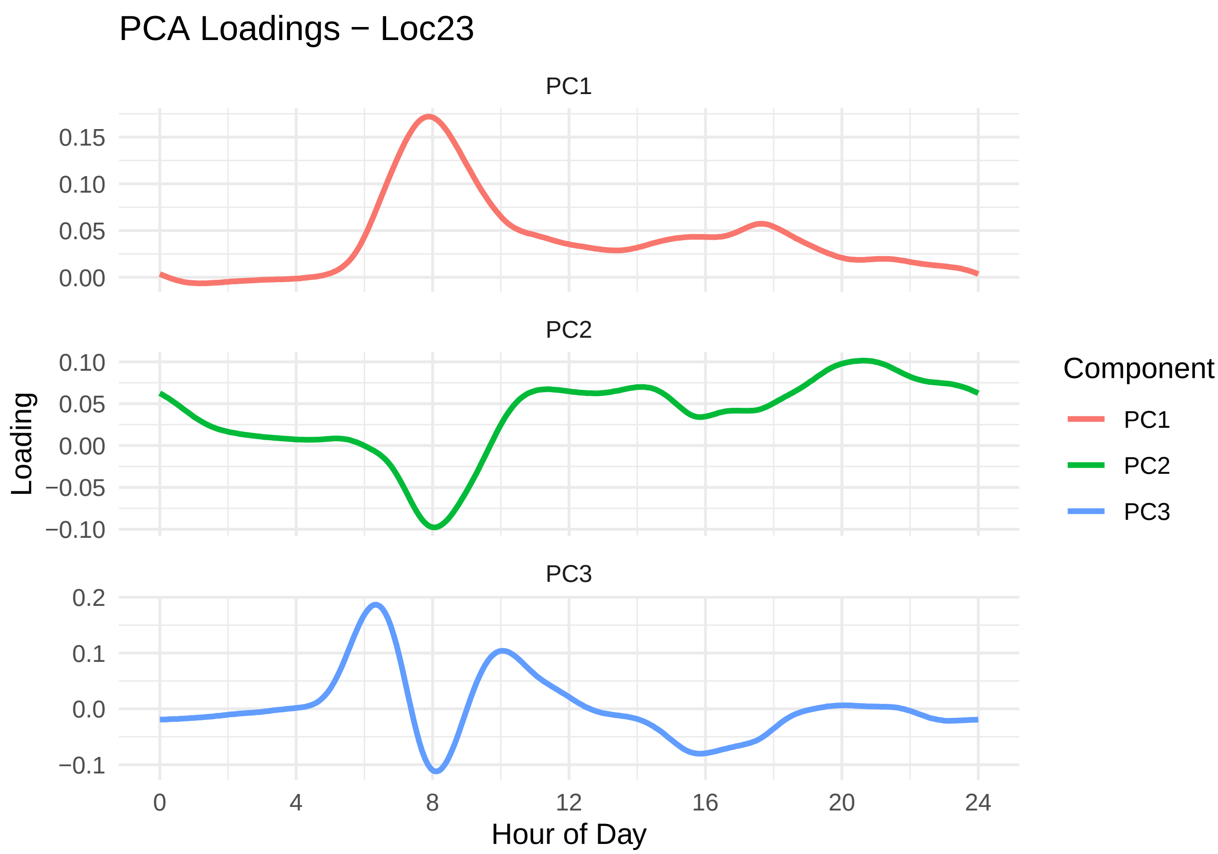


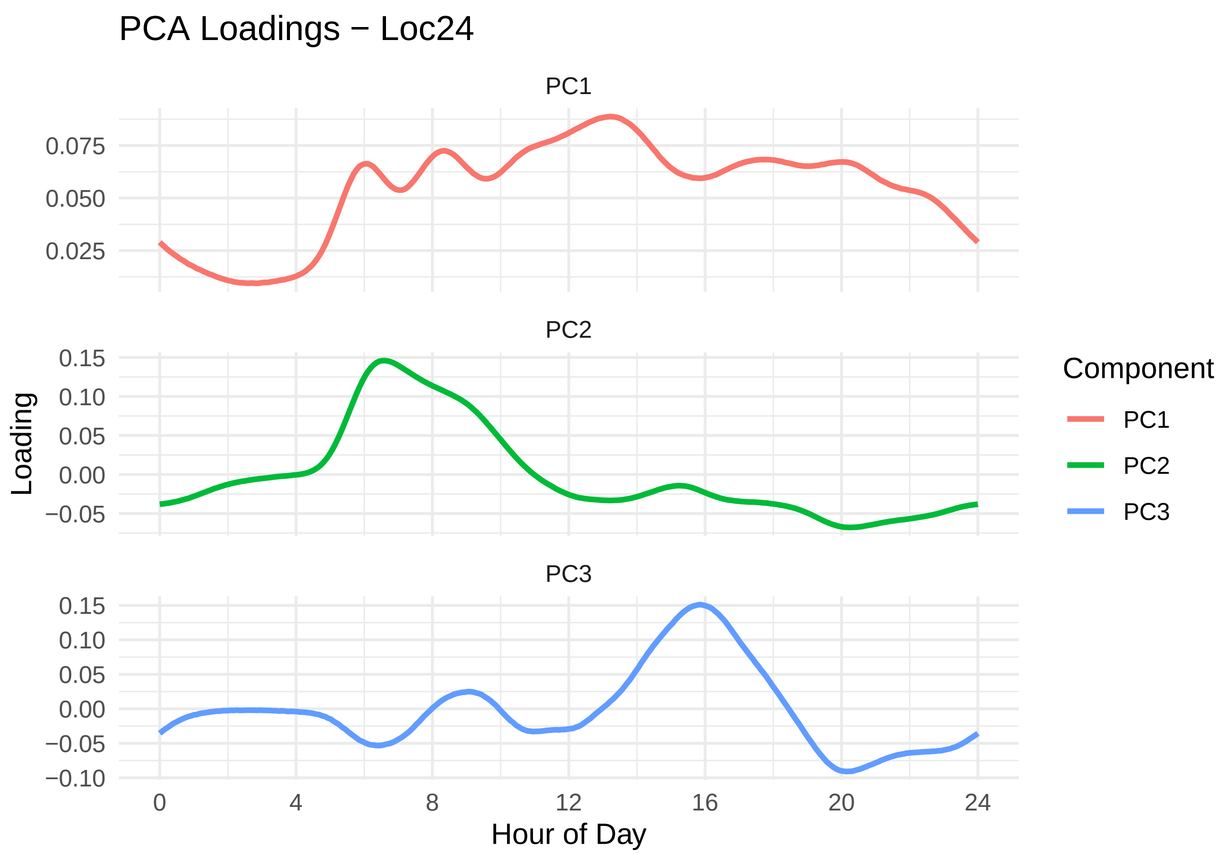
PCA Loadings − Loc19
PC1
PC2
PC3
Loading
0.10
0.05
0.00
−0.05
0.15
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

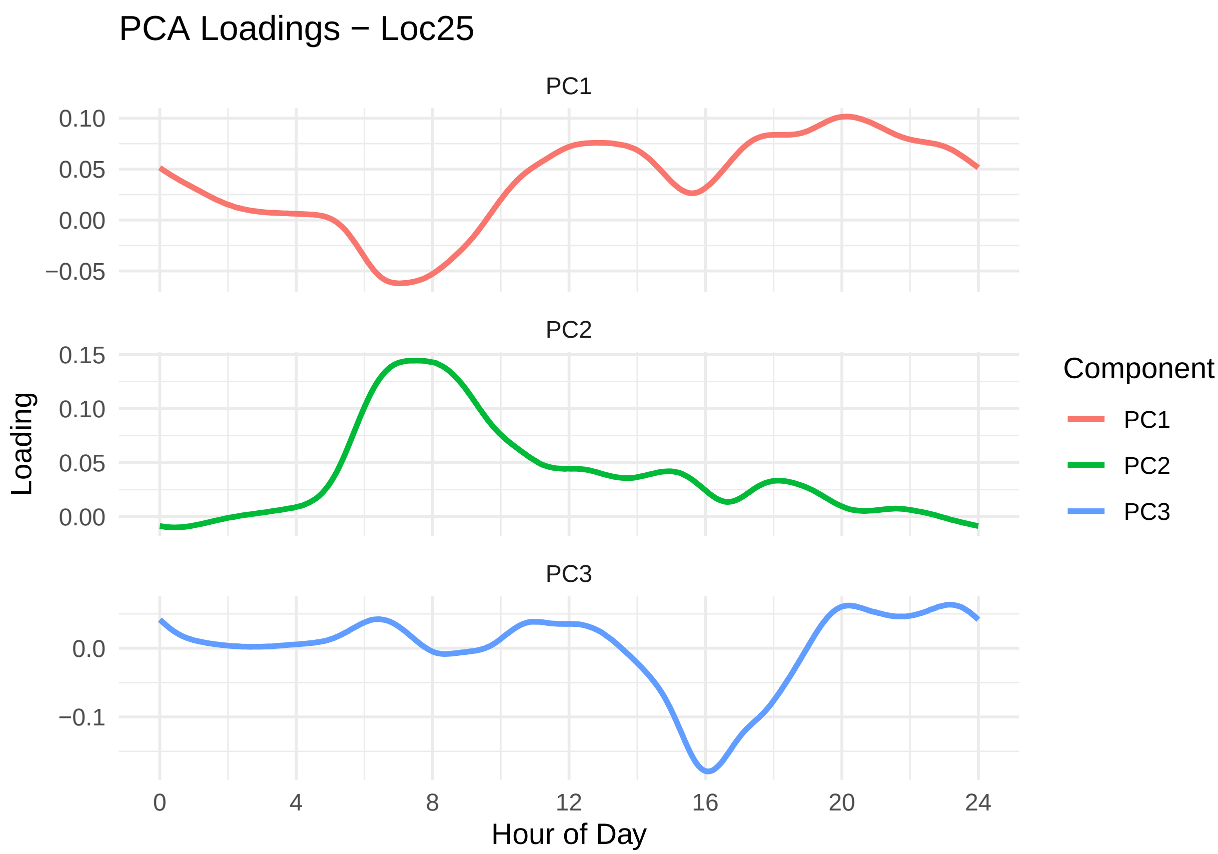

PCA Loadings − Loc20
PC1
0.000
−0.025
−0.050
−0.075
−0.100
PC2
0.1
0.0
PC3
0.1
0.0
−0.1
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3

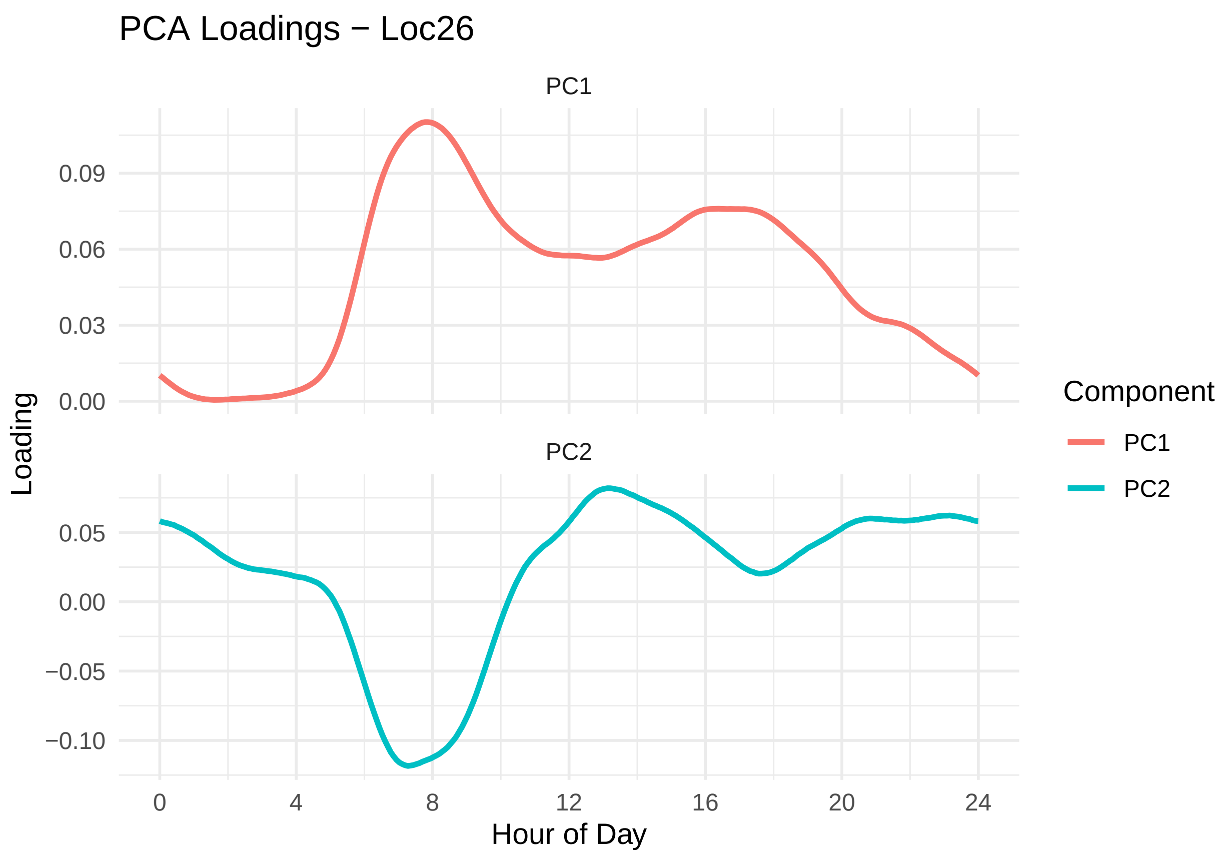

PCA Loadings − Loc21
PC1
PC2
PC3
Loading
Hour of Day
Component
PC1
PC2
PC3
0.10
0.05
0.00
−0.05
0.15
0.10
0.05
0.00
0.1
0.0
−0.1
0
4
8
12
16
20
24


PCA Loadings − Loc22
PC1
0.00
−0.05
−0.10
−0.15
PC2
0.15
0.10
0.05
0.00
−0.05
−0.10
PC3
0.1
0.0
−0.1
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3


PCA Loadings − Loc23
PC1
0.15
0.10
0.05
0.00
PC2
0.10
0.05
0.00
−0.05
−0.10
PC3
0.2
0.1
0.0
−0.1
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3


PCA Loadings − Loc24
PC1
PC2
PC3
0.075
0.050
0.025
0.15
0.10
0.05
0.00
−0.05
−0.10
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3


PCA Loadings − Loc25
PC1
0.10
0.05
0.00
−0.05
PC2
0.15
0.10
0.05
0.00
PC3
0.0
−0.1
Loading
0
4
8
12
16
20
24
Hour of Day
Component
PC1
PC2
PC3



Across locations, PCA retained between 2–4 components, explaining at least 90% of variance. PC1 typically represented overall daily volume, with high loadings throughout the day, while PC2 captured morning vs. evening imbalance. PCA identified 6–15 anomalous days per location, typically representing days with unusually high or low scores on a specific component

For FA

For ICA

Overall anomaly patterns across methods

Overlap among methods

Clustering of anomalous days

Discussion

Traffic dynamics can be intuitively viewed by their diurnal patterns, such as the sharp volume peaks observed during morning and evening rush hours. Conversely, traffic anomalies include deviations from these established norms, such as a sudden isolated car accident, holiday traffic, or systemic events like severe weather closures. Distinguishing these routine fluctuations from true anomalous events requires multivariate methods capable of decomposing the aggregate traffic flow into its underlying normal and abnormal source signals. Different multivariate decompositions reveal The three dimension-reduction methods—PCA, FA, and ICA—were employed jointly because they provide complementary and interpretable decompositions of the daily traffic profiles. PCA serves as a variance-maximizing benchmark: it extracts the dominant modes of variation and therefore identifies anomalies aligned with the primary directions along which traffic curves typically fluctuate. In contrast, FA models the covariance structure through latent common factors, yielding anomalies that deviate from the inferred low-rank dependence structure even when their total variance is modest. FA therefore acts as a structural benchmark that captures shifts in temporal pattern, peak timing, or shape-based distortions not necessarily associated with high variance. ICA provides an independent-signal benchmark by separating statistically independent micro-events embedded within the curves. Because ICA isolates localized or abrupt disturbances that PCA smooths and FA distributes across factors, it is particularly sensitive to short-lived spikes, dips, and irregularities. ## Appendix

References

The Gemini Flash 2.5 model was used to assist with the formatting of this section.

Columbia University Mailman School of Public Health. (n.d.). *Spatiotemporal analysis*.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*(4–5), 411–430.

Ma, T., Yao, F., & Zhou, Z. (2024). Network-level traffic flow prediction: Functional time series vs. Functional neural network approach. *The Annals of Applied Statistics*, *18*(1), 424–444.

Piech, C., & Ng, A. (2013). *K-means*.

Ringberg, H., Soule, A., Rexford, J., & Diot, C. (2007). Sensitivity of PCA for traffic anomaly detection. *ACM SIGMETRICS Performance Evaluation Review*, *35*(1), 109–120. https://doi.org/10.1145/1269899.1254895

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Shang, H. L., & Hyndman, R. J. (2010). *Exploratory graphics for functional data*.