



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Federal Statistical Office FSO

Improving the quality of traffic statistics

A Data Science and AI project for the Federal Roads Office (FEDRO) conducted by the Data Science Competence Center (DSCC)

Prof. Dr. Bertrand Loison

*Vice director of the FSO
Head of the Data Science Competence Center (DSCC)*

Presentation for the United Nations

23.01.2024



Agenda

1. Federal Roads Office (FEDRO)
2. Traffic monitoring for statistics
3. Improving the quality of traffic statistics
4. Data pipeline: project challenges
5. Data Science and AI solution: tailoring algorithms to the task at hand
6. Application to interact with the algorithms
7. Accurate data is essential for political decision-making
8. Improving the quality of traffic statistics: project summary

Federal Roads Office (FEDRO)

“The Federal Roads Office (FEDRO) was established in 1998 as Switzerland's federal authority responsible for road infrastructure and private road transport. It belongs to the Federal Department of the Environment, Transport, Energy and Communications (DETEC), and focuses on securing sustainable and safe mobility on the country's roads.”

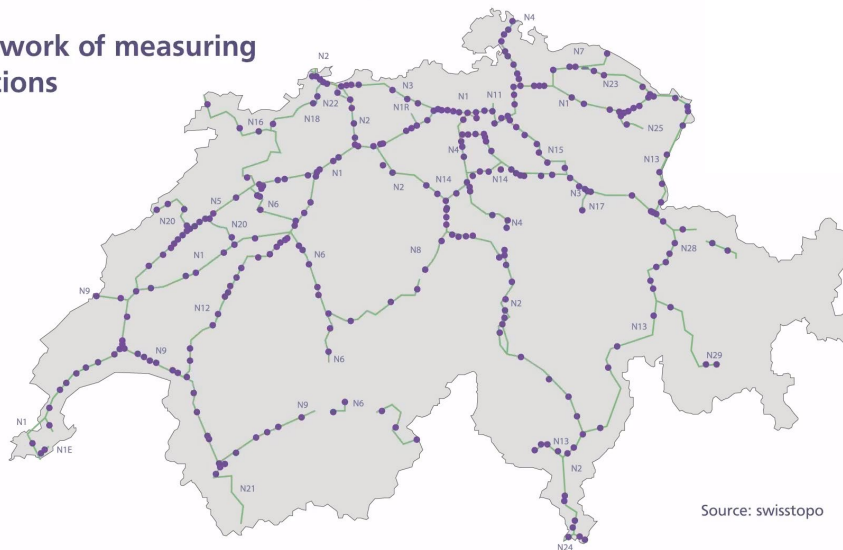


Source: FEDRO

Traffic monitoring for statistics by the Federal Roads Office (FEDRO)

“The traffic monitoring business application (VMON) is a system for controlling and validating traffic counting data. The VMON application manages around 500 counting stations in Switzerland.”

Network of measuring stations



Source: swisstopo

MISTRA - Fachapplikation Verkehrsmonitoring

AVZ Validierung | Messstellenverwaltung | About

Freigabe | Freigeben und schliessen | Validierung

Freigebe zurücksetzen | Gleich/Werte vert. | Alle Werte vert.

Systemparameter | Messstellenparameter | Kalender | Messstellengruppen

Parameter | Konfiguration | Navigation

04.01.2022

Tagesprotokoll

		Pendent	Januar 2022																			
			01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
523	MUENSINGEN (AB)	1	6	6	5	7	6	69	6	5	6	6	10	11	10	10	10	10	11	9	10	10
525	TENERO	1						N									R					
527	GONDO - VS40001	0		N							N		N					N				
528	GURTNELLEN (AB)	0																				
529	WASSEN (AB)	0																				
530	VERZWEIGUNG ALTDORF (AB)	0																				U
533	HENGGART (AB) - ZH0497	1						N														
534	LAUSANNE BLECHERETTE (AR)	1					NG		N													N
535	LAUSANNE BELMONT (AR)	1						N														
536	LAUSANNE LA CROIX (AR)	1																				
537	VEVEY (AR)	0							N													
538	MONTREUX (AR)	0	KU	KU	KU	NKRU	KU	NKU	KU	KU	KU	KU	KU	KU	KU	KU	KU	KU	KU	NKU	KU	KU

Source: FEDRO

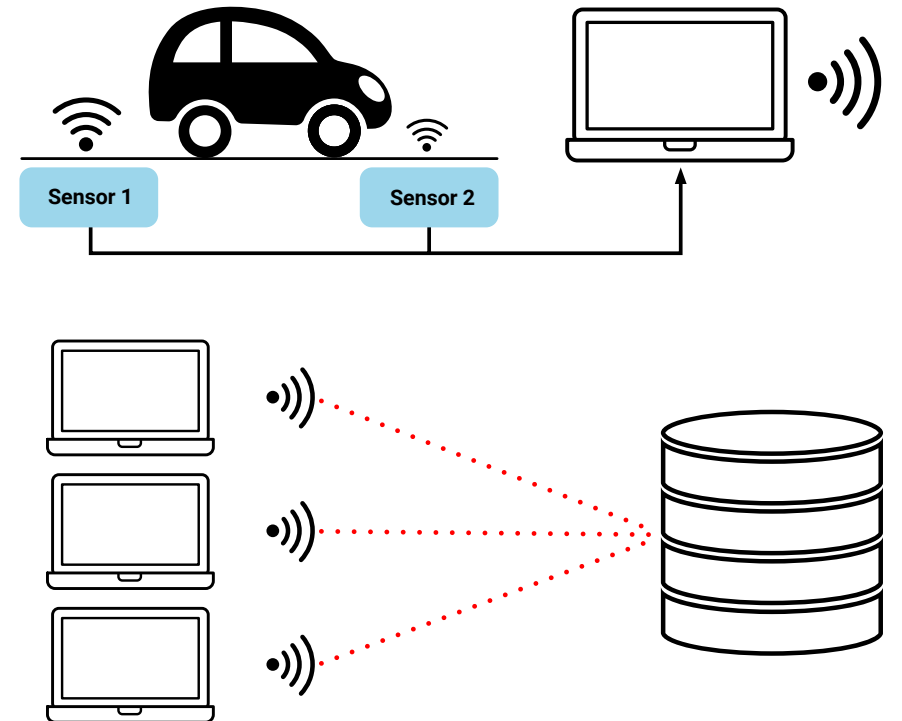
Improving the quality of traffic statistics for the Federal Roads Office (FEDRO)

Automated traffic monitoring project overview:

Objective: automatic detection of measurement errors and reconstruction of missing data.

Current practice: data cleansing is carried out by human experts.

New solution: Development of a toolbox in R for automatic anomaly detection, data reconstruction, deep statistics and visualisations.



Source: FSO

Data pipeline: project challenges

Sensors can send erroneous data:

- a vehicle may not be recognized or classified in the wrong category by the existing operating system.

Sensor communication breakdowns with the central server:

- communication is based on the telecommunications network, which is sometimes subject to disruption;
- missing data in the event of a communication breakdown;
- surplus data "backlog" after a certain period of time.

Individual time series distribution:

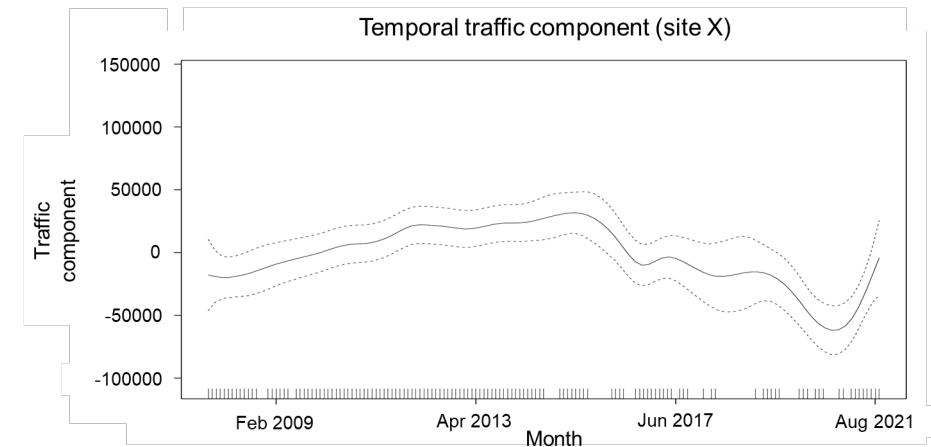
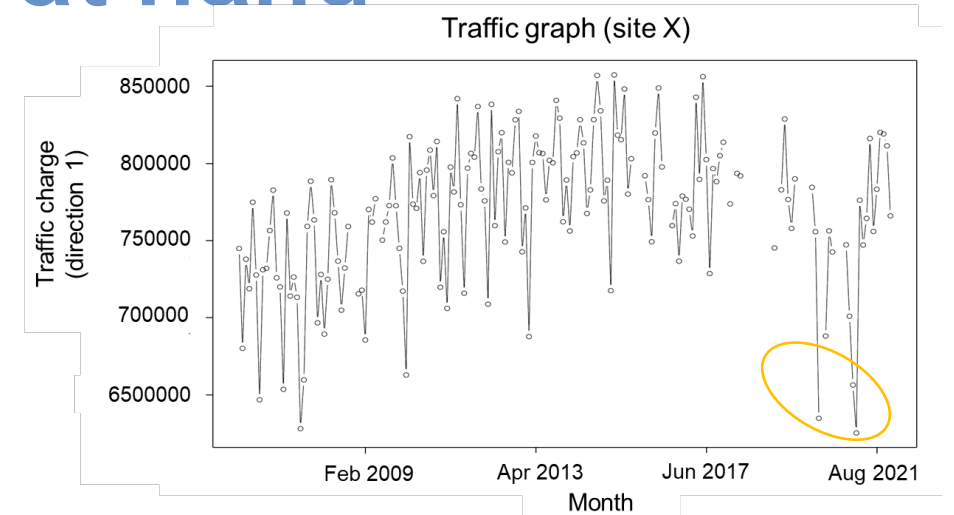
- the distribution of traffic data is influenced by the effects of seasonality and extreme events (outliers);
- the pattern of data distribution varies over time.

▶ **Topic areas: plausibility check, search and imputation of missing data**

Data Science and AI solution: tailoring algorithms to the task at hand

**Anomaly detection:
rule based approach for outlier detection**

**Missing data imputation:
Generalised Additive Models (GAMs)**



Application to interact with the algorithms

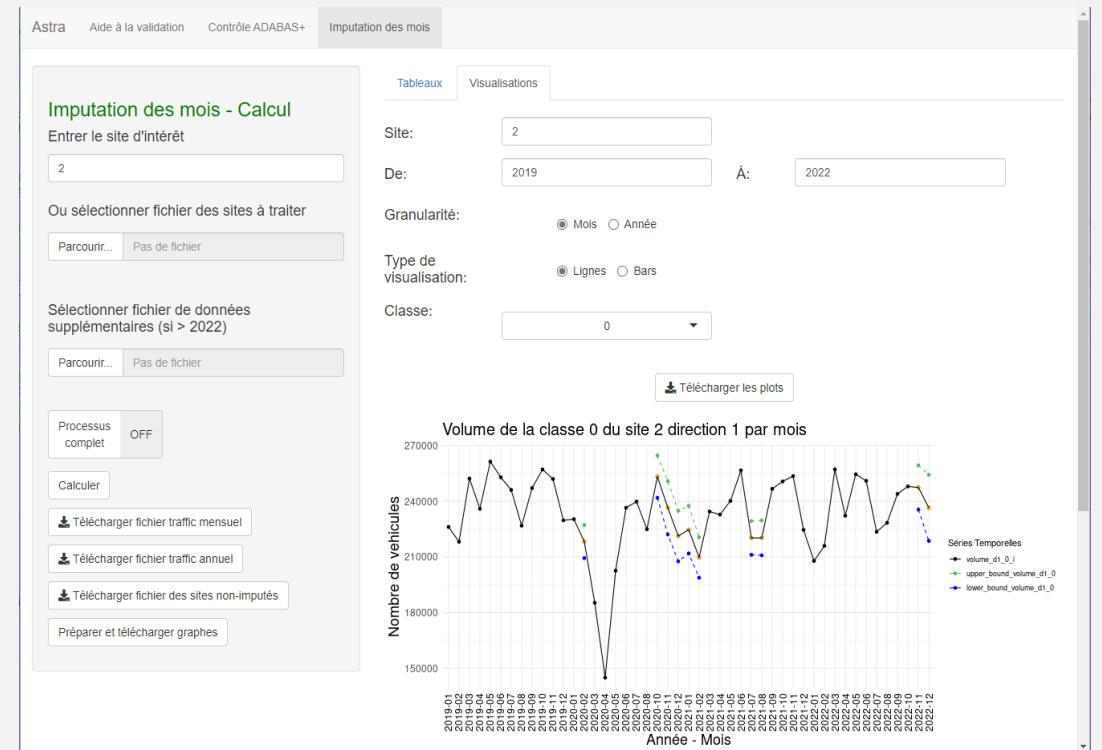
What data scientists/engineers do

```
#' Find the site at which at the granularity, when there is an 'N' not all
#' rows and columns of status are 'N'.
#'
#' @param melted_df data.table in long format (all status in one column)
#' @param granularity site and time (month, day, or day and class) at which
#' to check for inconsistent N status
#'
#' @return df with site and granularity columns where N status are inconsistent
get_any_and_all <- function(melted_df, granularity) {
  value <- any_and_all <- NULL
  no_data <- melted_df[, any(value == NO_DATA), by = granularity]
  names(no_data) <- c(granularity, "any")
  no_data$all <- melted_df[, all(value == NO_DATA), by = granularity]$V1
  no_data$any_and_all <- no_data$any == no_data$all
  no_data[any_and_all == FALSE, granularity, with = FALSE]
}

#' Script to validate N status distribution at different time interval
#' The zone and time with inconsistent N are saved in AD_INTEGRITY excel
#'
#' @param df data.table of raw AD data
validate_no_data <- function(df) {
  # Keep and melt status (all in one column)
  status_columns <- colnames(df)[stringr::str_detect(colnames(df), "^ST_R")]
  df <- df[, c("site", "day", "month", "KL", status_columns), with = FALSE]
  melted_df <- data.table::melt(df, measure.vars = status_columns)

  # Compute N irregularities in month, day and per class (all information)
  granularity <- c("site")
```

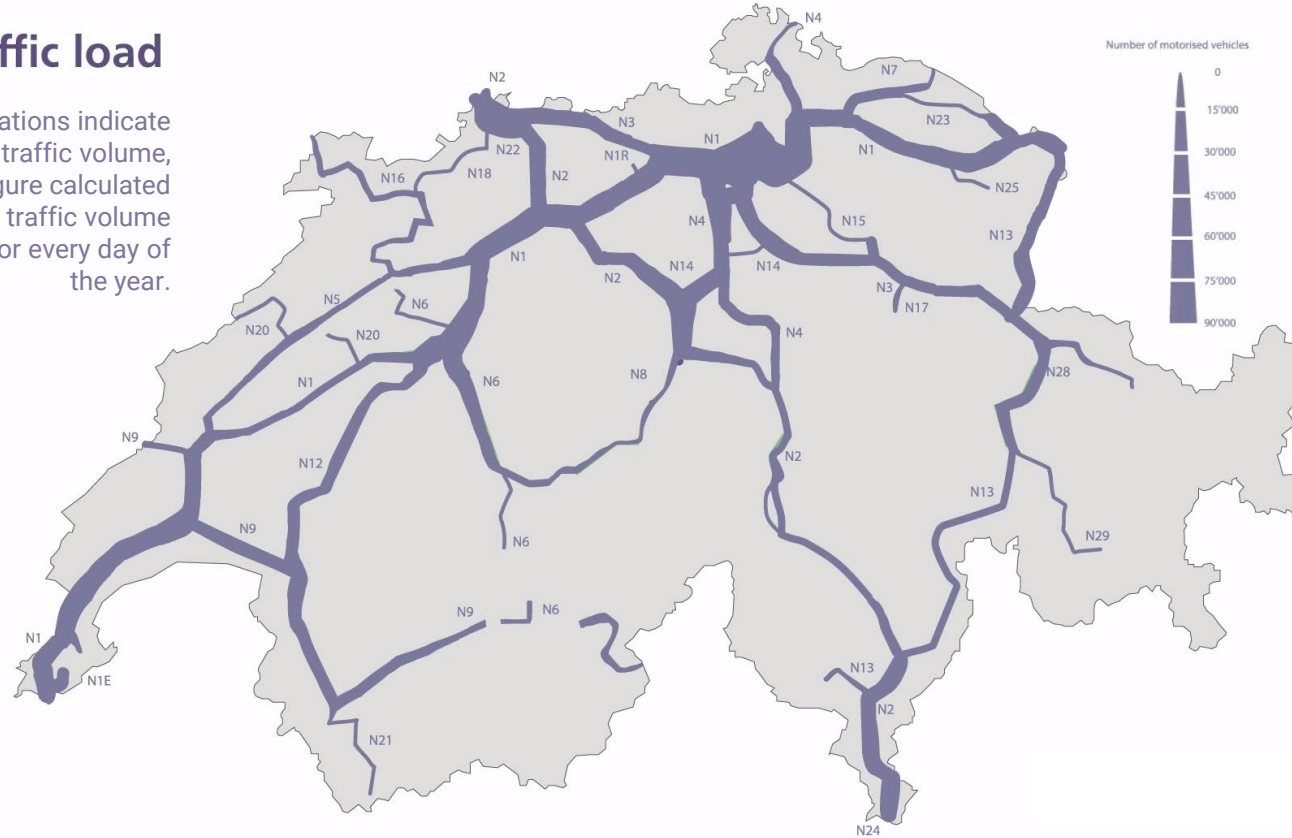
What users see



Accurate data is essential for political decision-making

Traffic load

The measuring stations indicate the average daily traffic volume, i. e. the mean figure calculated from all the 24-hour traffic volume figures measured for every day of the year.



“More than 85 per cent of all traffic jam hours were attributable to congestion. This high figure indicates that so many stretches of the network are so congested that even minor interruptions to traffic flow can result in lengthy traffic jams and prompt large numbers of drivers to use local roads instead of the motorways.”

Source: FEDRO

Improving the quality of traffic statistics: project summary

Service provided by the DSCC	Project execution
Language	English
Implementing organisation	Federal Roads Office (FEDRO)
Topic areas	Plausibility check, search and imputation of missing data
Start date / End date	2021 / ongoing
Project management	Federal Roads Office (FEDRO)
Data type	Structured sensor data
Components of machine learning	Supervised learning: Generalised additive models



Source: FSO