Data Engineering · Term project proposal

AI-based highway icing risk map integrating weather and road conditions

Estimating icing risk, including black ice, where no road sensor exists

Team 1 · Taegyeong Kim, Minsoo Lee, Jian Choi, Heeryeong Park, Ingon Chu

POSTECH

Motivation

Road ice is invisible until it causes a pile-up



35

vehicles, 5 deaths: black ice

Seosan–Yeongdeok Expwy · Jan 10, 2026

38

vehicles; a 21-car pile-up a month earlier

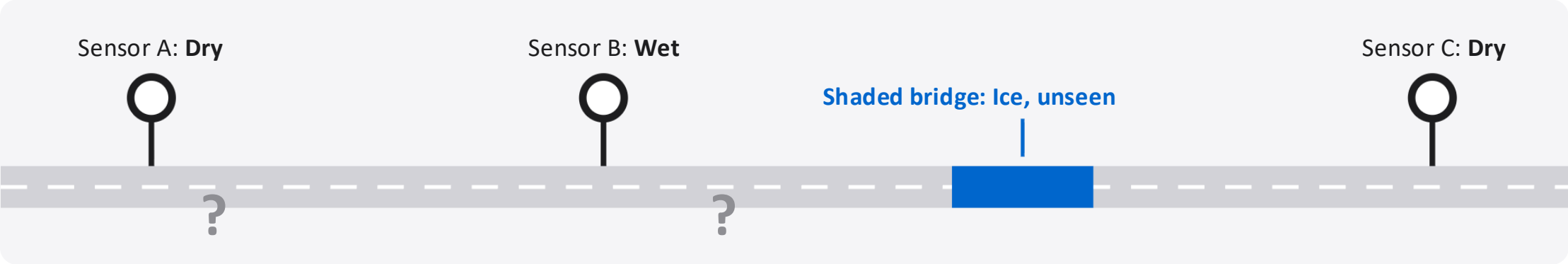
Jeonbuk Province, our study area · snow and ice · 2025

Why it is so dangerous

Black ice is **a thin, transparent film** that looks like a merely wet road. It forms **at night, on bridges and in shade**, often on otherwise dry days.

Problem

The gap: a sensor sees one point; ice forms in between



A few sensors

at selected points along the highway

~20 km

between regular road sensors on the highway

100s of m

length of a single bridge or shaded cut

Our goal

Fill every gap with a virtual sensor

Problem statement

Can we learn icing patterns where no sensor exists?

Input

Weather

temperature, dew point, rain, cloud

Terrain & sun

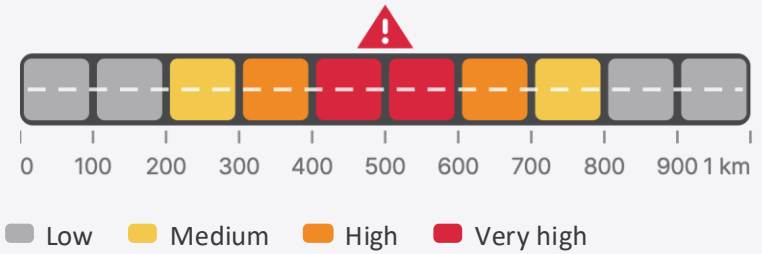
shade and slope from the elevation model

Road facilities

bridges, tunnels, curvature

Output

Probability of ice (0–1) per 100 m point



Key idea Learn how ice forms at the sensors, then apply it to every point between them. 100 m resolves a single bridge or shaded cut.

Why it matters

From "where ice formed last year" to "where ice forms, even without a sensor"

Operations

See which segments between sensors are likely to ice, and treat those first.

Road operators

Policy

Choose ice-prone sections from how weather and terrain combine, not from long-term averages.

Road safety agencies

Drivers

Segment-level warnings on navigation apps and road signs as risk rises.

Navigation and road signs

With weather forecasts See whether ice will form in the coming hours, and pre-treat before it does.

Novelty

What is new

01	**Dynamic prediction**	Predicted from recent observations or weather forecasts, not a fixed score sheet
02	**Virtual sensors**	From a sparse sensor network to every 100 m of road
03	**Terrain- and sun-aware features**	Shadow, sunlight and curvature at every point

Why this data: effect, cause and place

Road sensors (RWIS)

Road temperature and road state: the only direct measurement of ice.

Every 10 min

Weather stations

Ice needs a sub-zero surface and moisture: temperature, dew point, rain, snow and cloud explain both.

Every 10 min to 1 h

Terrain & facilities

Same weather, different exposure: elevation, shade, curvature, bridges, tunnels, rivers.

30 m elevation model · map layers

Weather + **Terrain & facilities** → **Road state**

Weather and place exist everywhere, so the model transfers to unsensed points.

Sources: KMA, national spatial data

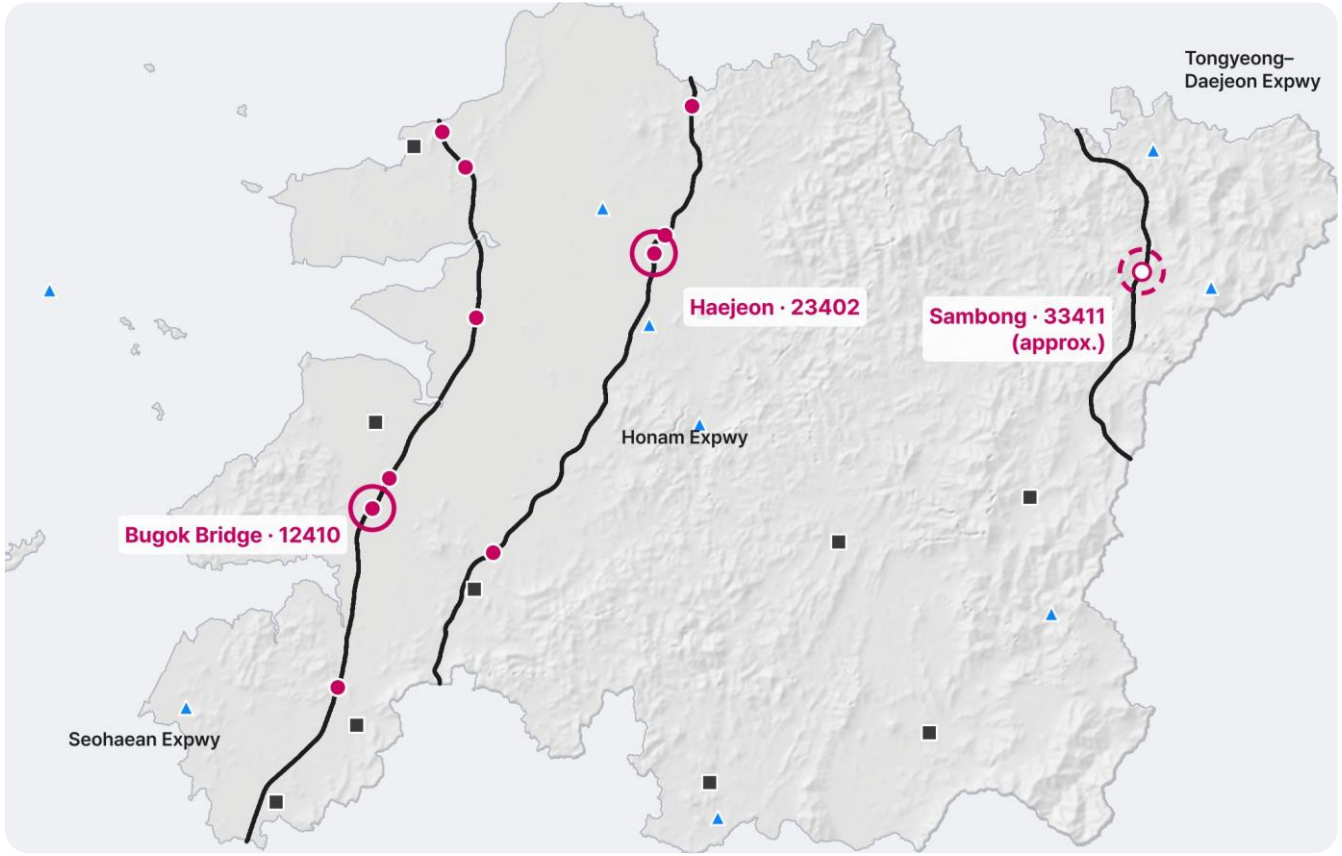
The core data: what each source tells us

Source	Every	What it gives	Used as
Road weather sensors (RWIS)	10 min	Road temperature and state: dry, wet, snow, ice	Label
Synoptic stations (ASOS)	1 h	Air/ground temp, dew point, cloud, sunshine, snow	Cause
Automatic stations (AWS)	10 min	Air temperature, dew point, humidity, rain, wind	Cause, denser network
Elevation model (DEM)	30 m grid	Terrain height	Sun, shade, terrain shape
Bridges & tunnels	per structure	Start and end of each structure	Place
Road curvature points	along road	How the road bends, point by point	Place
Road & river boundaries	polygons	Road outlines, bridges, tunnels, rivers	Place, moisture

We also collect Nationwide road sensors · 500 m weather grid · Highway camera weather · Accident & snow-removal logs · 5 m elevation

Source: Jeonbuk regional weather office · ASOS, AWS: KMA station networks

Where the sensors are, and where they are not



- **Road sensors (RWIS)** — label
- **Synoptic stations (ASOS)** — cause
- **Automatic stations (AWS)** — cause
- **Held-out test sensor**

Held out for testing, one per route

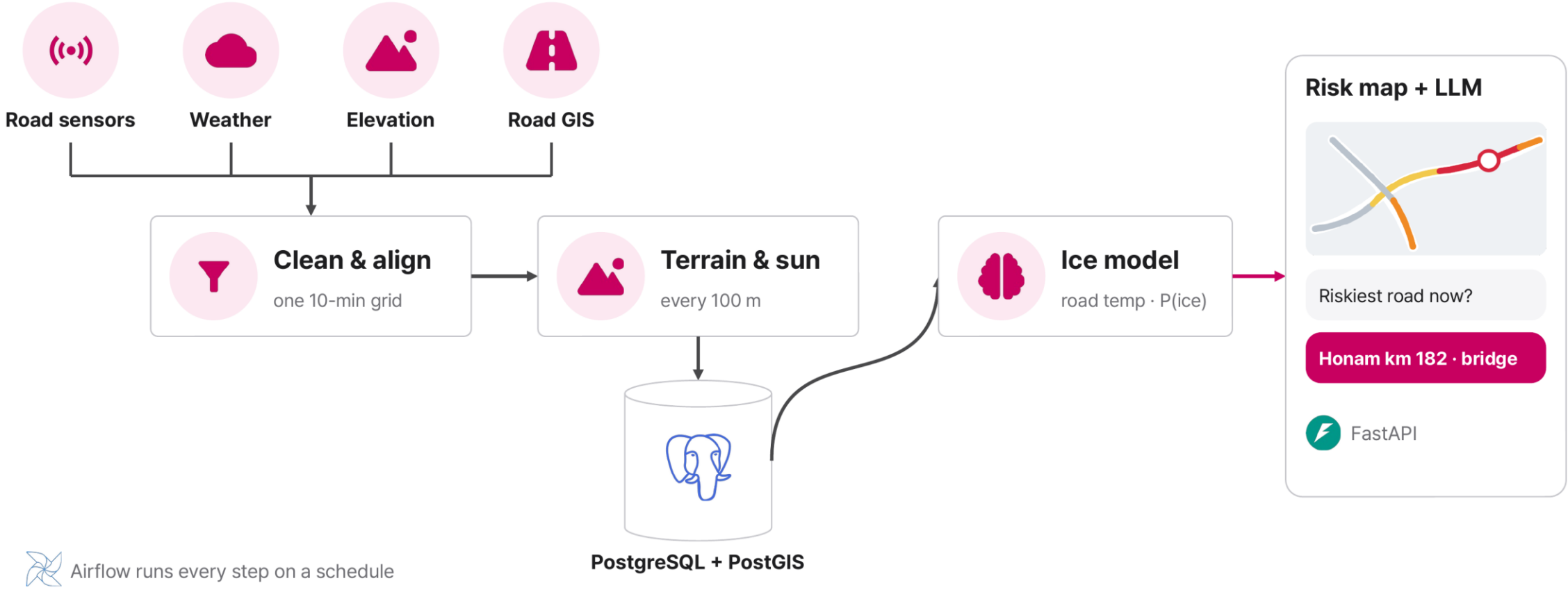
Bugok Bridge · 12410
Seohaean · coastal plain, bridge deck

Haejeon · 23402
Honam · inland plain

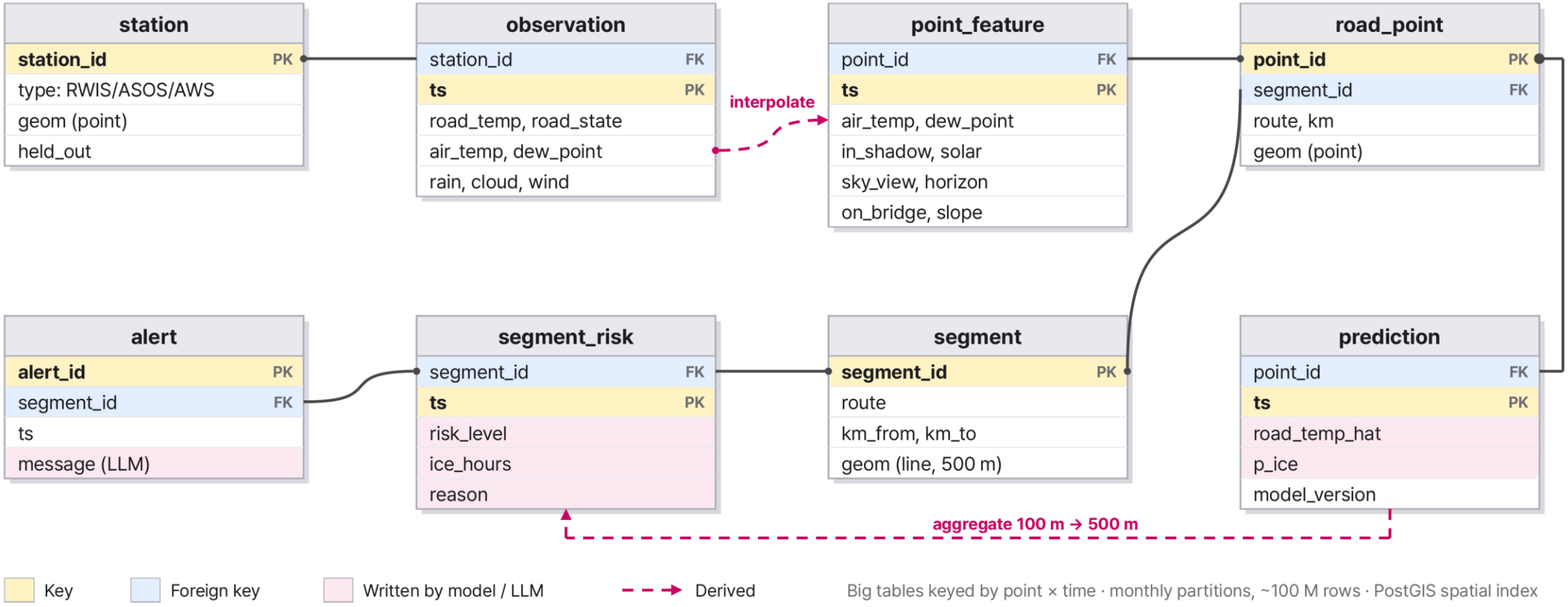
Sambong · 33411
Tongyeong–Daejeon · mountains

Data pipeline

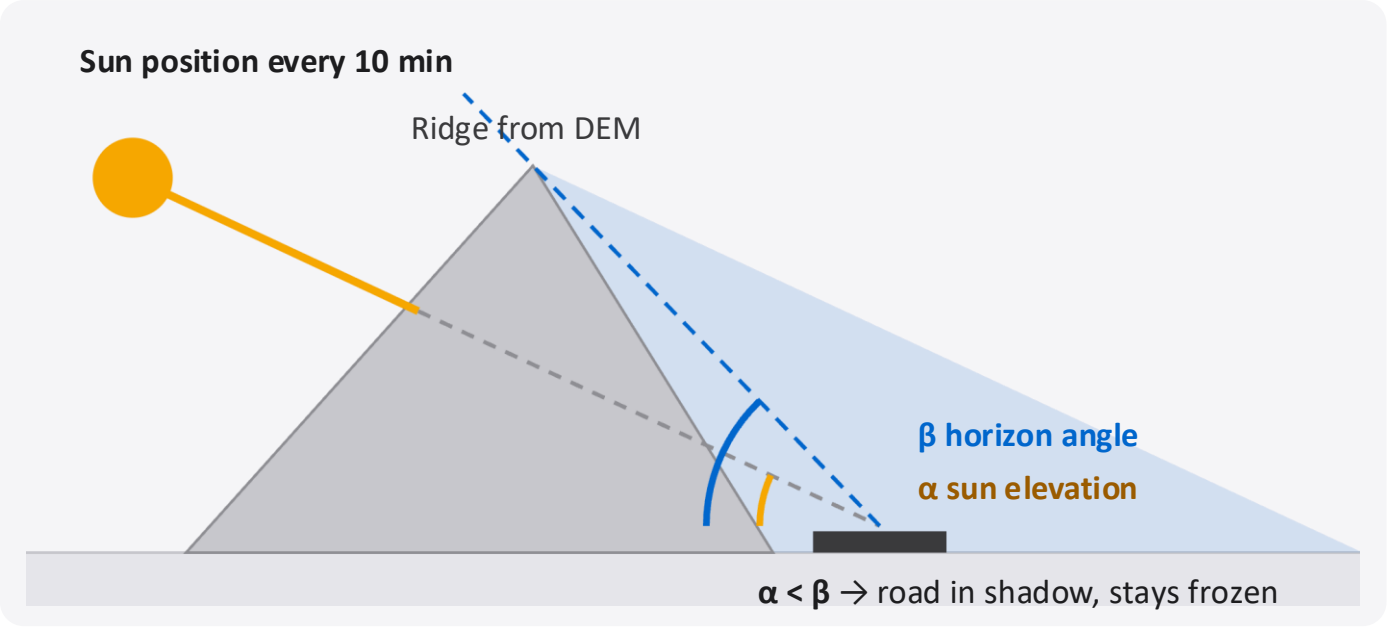
From raw files to a live risk map



One schema: stations, road points and time steps



Where the sun never reaches



1 · Once
36-direction horizon per point

2 · Every 10 min
Sun elevation & azimuth

3 · Compare
Table lookup, no ray-tracing

Features per 100 m point

In shadow now
yes / no, every 10 min

Sunshine hours today
how long the surface was warmed

Solar energy received
clear-sky × slope and aspect × cloud cover

Time since last sun
how long it has been cooling

Sky view factor
night-time heat loss to open sky

Shadow mask from the real DEM

08:30 · sun 7.7° above horizon



10:00 · sun 21°



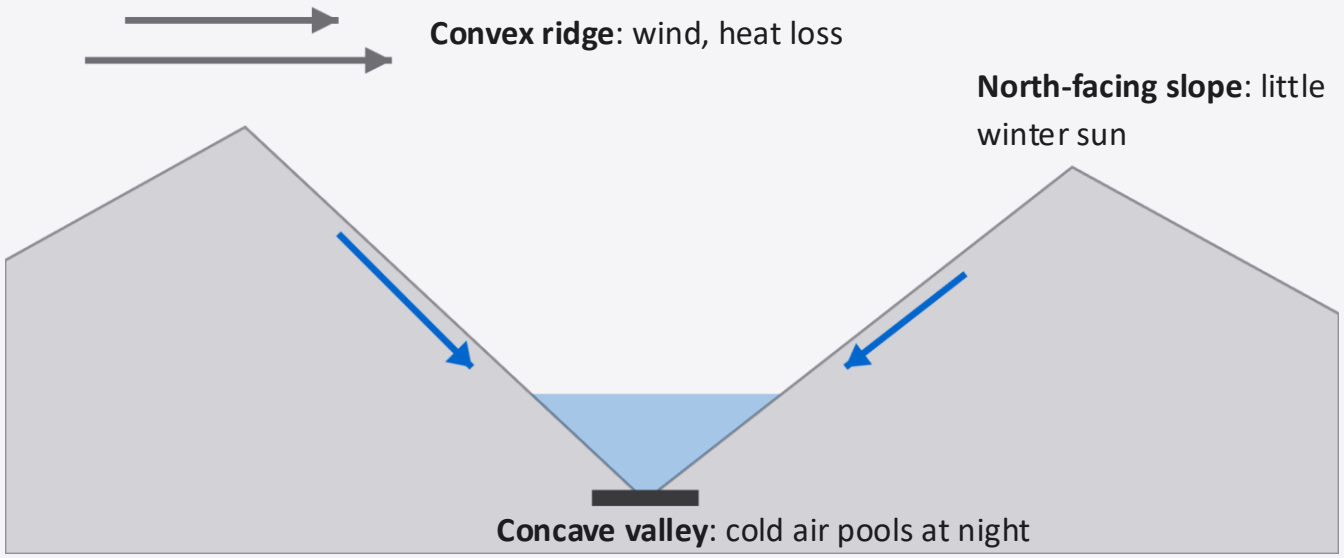
Terrain shadow
Road point in shadow
Road point in sun

Share of route in shadow

Time	T–D	Honam	Seohaean
08:30	61%	15%	< 1%
09:00	28%	2%	0%
09:30	6%	< 1%	0%
10:00	1%	< 1%	0%

Source: computed from the DEM, Jan 10 · T–D = Tongyeong–Daejeon Expwy

Terrain shape decides where ice forms first



Two road points under the same weather can freeze at different times.
Terrain shape tells the model which one first.

Features per 100 m point

Terrain curvature
concave valley vs convex ridge

Relative elevation
valley floor vs hilltop nearby

Slope and aspect
north-facing cuts get little sun

Road curvature
sharp bends where ice hurts most

Bridges, tunnels, rivers
cold from below, sudden transitions, moisture

Evaluation

How we will know it works

Road ice (road state = Ice)

Daily precision · recall · F1

from 10-min estimates, per sensor and overall

Road surface temperature

MAE · RMSE · bias (°C)

at 3 held-out sensors, one per route

Baseline to beat

Copy the nearest sensor

shows what terrain, sun and weather actually add

Validation design

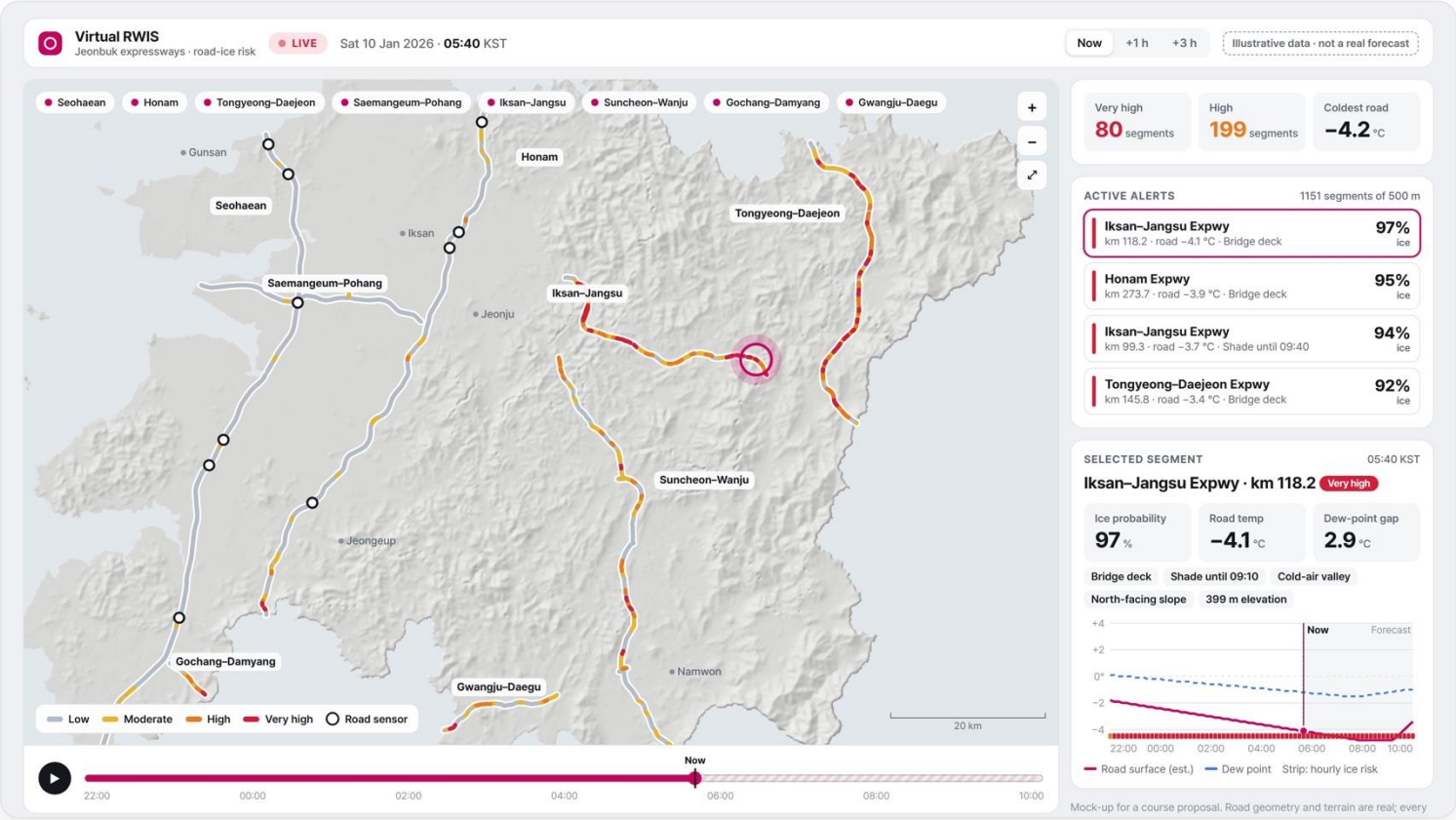
Leave-one-sensor-out CV

plus leave-one-winter-out; held-out sensors scored once, at the end

Success Beat "copy the nearest sensor" at sensors the model has never seen.

Prototype · Real-time alerts

From model to a live risk map



Live map

Every 500 m segment colored by ice probability, refreshed every 10 min.

Scrub the night

Replay the past hours, or look up to +3 h ahead with weather forecasts.

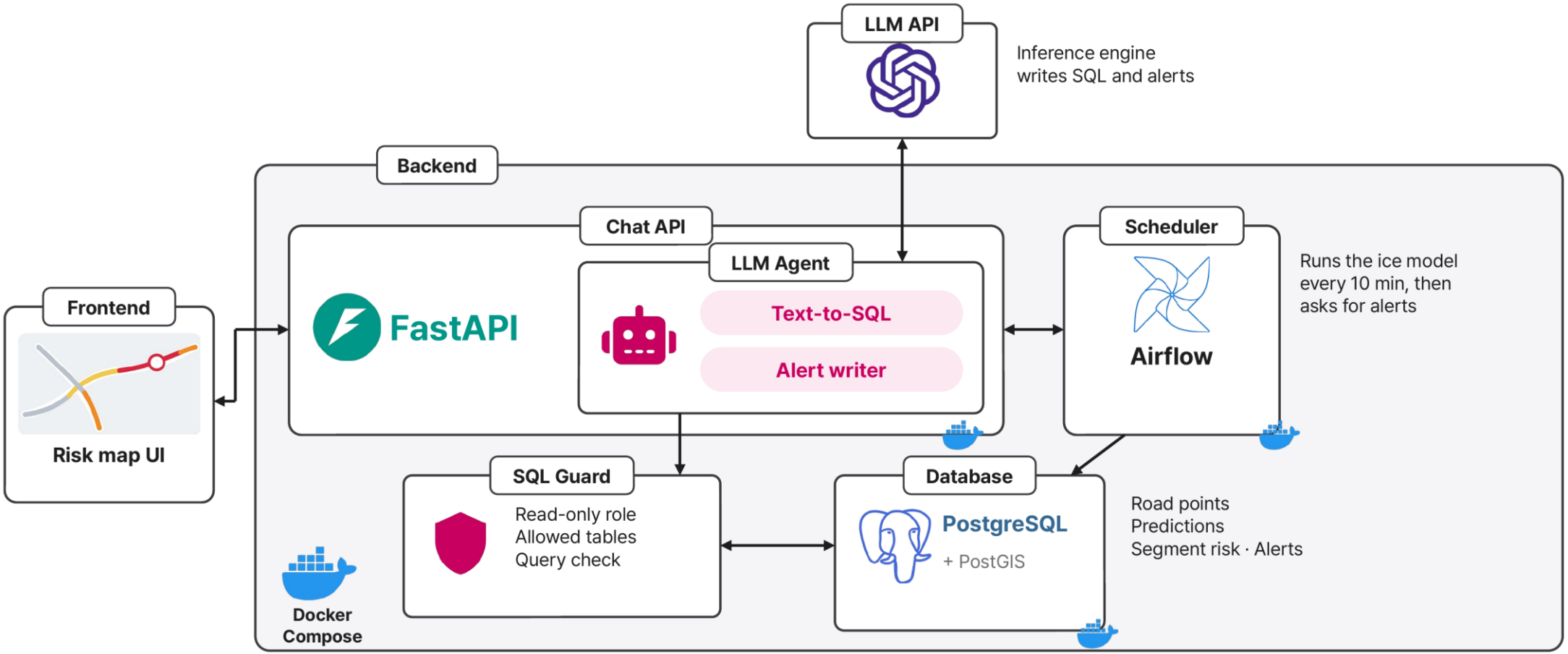
Ranked alerts

Worst segments first, each with its reason: bridge, shade, cold-air valley.

Segment detail

Road temperature against dew point through the night, so the risk is explainable.

Service architecture



LLM on top of the risk map: ask, explain, plan

1 · TEXT-TO-SQL

Ask in plain language

"Icy Honam segments before 7 a.m.?"

```
SELECT route, km_from, risk_level, reason
FROM segment_risk
WHERE route = 'Honam' AND risk_level >= 3
    AND ts BETWEEN '05:00' AND '07:00';
```

Segment	Time	Risk	Why
km 182	05:00–06:30	Very high	bridge
km 175	05:30–07:00	High	shade

2 · ALERT EXPLANATION

Explain each alert

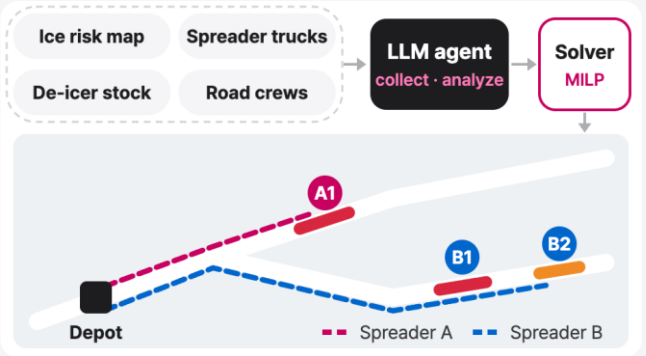
P(ice) 0.82 · road −1.4 °C · dew-point gap 0.6 °C · shade until 10:20 · bridge deck

VERY HIGH · km 182.5–183.0

"Black ice likely on the bridge deck: the surface is below freezing, almost at the dew point, and stays in hill shade until 10:20."

3 · PLANNING AGENT

Plan the de-icing run



REPORT TO MANAGER

Spreader A → km 182 bridge, by 05:40
Spreader B → 2 mountain cuts, by 06:30

Guardrails Read-only DB role · SQL checked before it runs · numbers come from the database · routes come from the solver, not the model

Thank you.

Questions?

AI-based highway icing risk map integrating weather and road conditions

Team 1 · Taegyeong Kim, Minsoo Lee, Jian Choi, Heeryeong Park, Ingon Chu

Backup slides for Q&A

Appendix

A1 · The messy reality we already found in the samples

A2 · One schema: stations, road points and time steps

The messy reality we already found in the samples

cp949

Legacy Korean encoding: default UTF-8 readers crash

4 coordinate systems

EPSG:5179, 5186, Korean 1985, WGS84, one file with no .prj

DEM in 0–0.887

Elevation normalized, not meters: restore it or every shadow vanishes

–99.9 · –999 · #NUM!

Sentinels and spreadsheet artifacts inside numeric columns

1 h vs 10 min

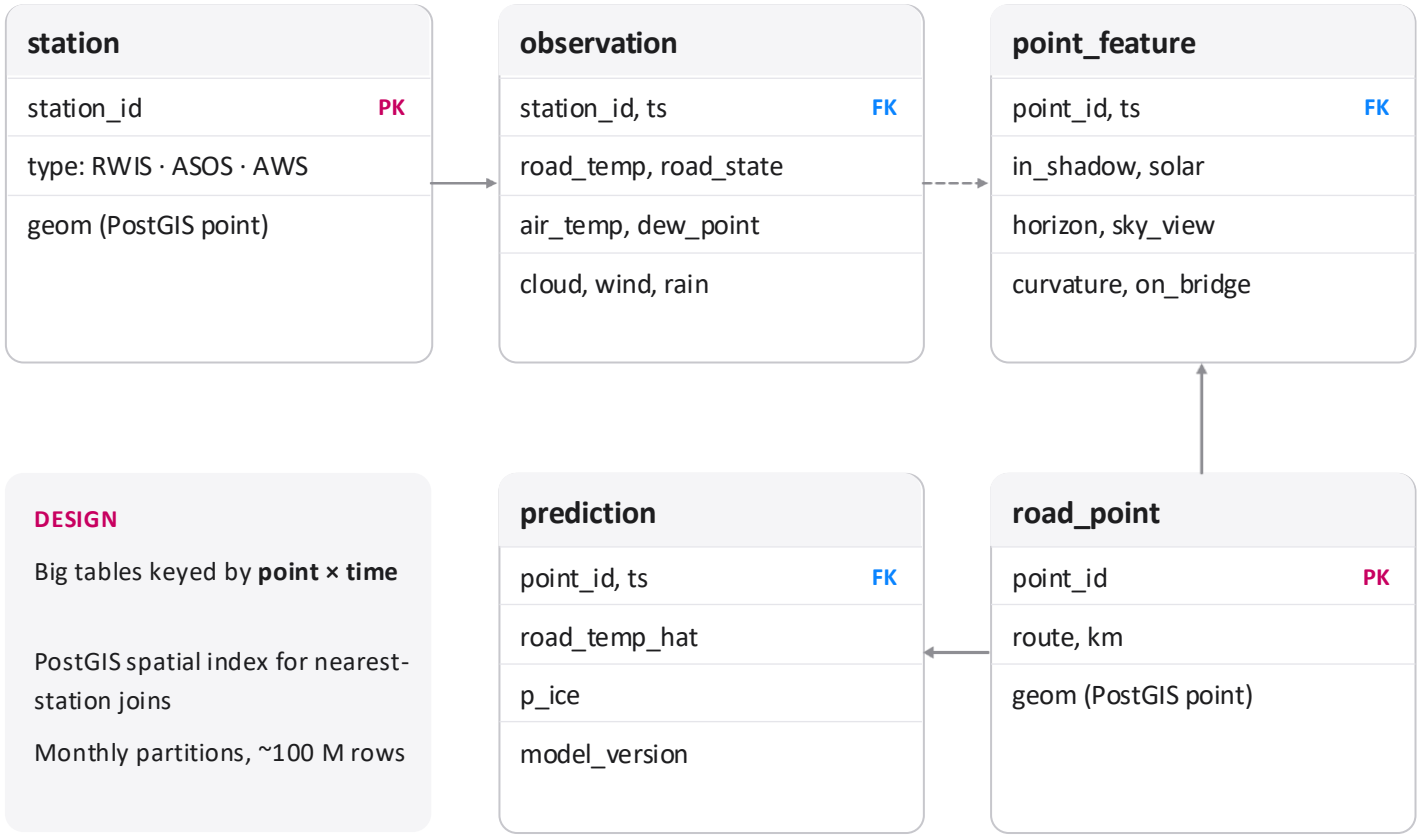
Hourly stations, 10-minute sensors, gaps from sensor outages

start = end

Bridge and tunnel rows with zero length and duplicates

Our job We clean and align all of it into one dataset: the foundation our road-ice model is built on.

One schema: stations, road points and time steps



DESIGN

Big tables keyed by **point × time**

PostGIS spatial index for nearest-station joins

Monthly partitions, ~100 M rows

EXAMPLE QUERY

“Where is the riskiest road right now?”

```
SELECT r.route, r.km,
    p.p_ice, p.road_temp_hat
FROM prediction p
JOIN road_point r USING (point_id)
WHERE p.ts = (
    SELECT MAX(ts) FROM prediction)
ORDER BY p.p_ice DESC
LIMIT 5;
```

Written by the LLM, run read-only through FastAPI