Enriching Motor Insurance Risk Models

with Satellite Weather Reanalysis Data

Evidence from UK Road Accidents

Luba Orlovsky | Principal Researcher, Earnix

Insurance Data Science Conference 2026 | Hannover

The Problem: Subjective Weather Reporting

UK insurers rely on police-reported weather from STATS19 — but these are subjective assessments with known limitations.

Temporal Mismatch

Police arrive after the accident; conditions may have changed

Subjectivity

Fog vs mist vs poor visibility depends on officer judgment

Reporting Priorities

Weather is secondary to documenting injuries and fault

Coarse Categories

'Rain' vs 7.3mm precipitation — no granularity for modelling

earnix

Data Sources & Enrichment

STATS19

UK official road accident database

- Location coordinates (lat/lon)
- Timestamps (date & time)
- Police-reported weather
- Severity & road characteristics

ERA5 Reanalysis

Copernicus satellite weather data

- Hourly measurements, ~31km grid
- Global coverage since 1940
- Objective & verifiable
- Continuous values (not categories)

Enrichment Process

Match each accident's GPS coordinates and exact hour → ERA5 grid cell → weather variables added

15,168 UK accidents (Jan–Feb 2025) enriched with objective weather measurements

earnix

Fog Threshold Calibration

The standard fog proxy (TD < 2.5°C) flags 59.7% of UK winter accidents — clearly too generous. We calibrate against police reports.

Threshold	ERA5 fog rate	Police recall	ERA5/Police
TD < 0.4°C	3.2%	32%	4.6x
TD < 0.8°C	**11.0%**	**60%**	**15.4x**
TD < 1.5°C	32.5%	92%	45.7x
TD < 2.5°C	59.7%	99%	83.9x

Key Insight

There is no single "correct" threshold. ERA5 provides a continuous humidity measurement that can be tuned to any desired sensitivity — a fundamental advantage over binary police reports. The threshold choice is a business decision for the insurer.

earnix

Finding #1: The Fog Detection Gap

Police Reports

108

fog cases (0.7%)

ERA5 (TD < 0.8°C)

1,664

fog cases (11.0%)

15.4x

more fog detected

Validation

99.1% of police fog reports are confirmed by ERA5 (at TD < 2.5°C)

When police do report fog, ERA5 always agrees. The gap is one of underreporting, not disagreement.

earnix

Finding #2: Weather Does Not Predict Severity

Once a collision has occurred, weather conditions show virtually no association with whether the outcome is Fatal/Serious or Slight.

Metric	Slight (n=11,463)	Fatal/Serious (n=3,705)
Mean temperature	4.91°C	4.91°C
% sub-zero	5.9%	5.3%
% near-freezing (<3°C)	27.8%	28.1%
Mean wind speed	15.4 km/h	15.6 km/h

Why This Matters for Actuaries

Weather affects whether an accident occurs (frequency), but once a collision happens, severity is determined by speed, road type, vehicle count, and infrastructure.

This cautions against over-weighting weather in severity models — a practically important negative finding.

earnix

Finding #3: Model Comparison

LightGBM, stratified 5-fold CV. number_of_casualties and number_of_vehicles excluded. PR-AUC is the primary metric.

Model	Feats	ROC AUC	PR-AUC
A: Baseline (no weather)	**13**	**0.610 ± 0.010**	**0.333 ± 0.010**
B: Baseline + Police	20	0.612 ± 0.011	0.332 ± 0.011
C: Baseline + ERA5 continuous	23	0.579 ± 0.008	0.297 ± 0.008
D: Baseline + ERA5 all	30	0.582 ± 0.007	0.300 ± 0.009
E: Combined (Police + ERA5)	37	0.577 ± 0.015	0.297 ± 0.011

Statistical Significance (paired t-tests on PR-AUC)

Comparison	ΔPR-AUC	p-value	Sig?
B (Police) vs A (Baseline)	+0.001	0.807	ns
C (ERA5) vs A (Baseline)	−0.009	0.183	ns
E (Combined) vs A (Baseline)	−0.009	0.087	ns

Random classifier PR-AUC baseline: 0.244. Models A & B shown after Optuna tuning (80 trials).

earnix

What Actually Drives Severity?

SHAP analysis (without number_of_casualties or number_of_vehicles): severity driven by road infrastructure.

1

Location (lat/lon)

Northern, rural areas have higher severity

2

Special conditions

Roadworks, defective surfaces at site

3

Road type

Single carriageways vs dual carriageways

4

Junction detail

Specific junction configurations matter

5

Junction control

Uncontrolled junctions more dangerous

6

Speed limit

Higher speed limits increase severity

All police weather flags (rain, fog, snow, high wind) contribute essentially zero to severity prediction. ERA5 weather features actively degrade model performance.

earnix

Where ERA5 Adds Value

Not severity prediction — but data quality, verification, and long-term analysis.

Fraud Detection

Verify claimed weather conditions against objective ERA5 records. Independent, timestamped verification.

Frequency Modelling

Weather affects how many accidents occur. ERA5's continuous variables enable granular frequency models.

Territorial Rating

Separate 'bad luck with weather' from 'genuinely risky area' using weather-adjusted factors.

Climate Trend Analysis

80+ years of ERA5 history enables climate change scenario modelling for reserving.

earnix

Limitations & Caveats

1

Missing precipitation

ERA5 precipitation/snowfall not extracted. Rain validation and ice risk indicators require re-extraction with temporal differencing.

2

Spatial resolution

ERA5's ~31km grid cannot capture valley fog, coastal effects, or urban micro-climates.

3

Fog threshold is a proxy

Temp-dewpoint spread approximates humidity, not actual visibility. Aerosols and terrain matter.

4

ERA5 isn't ground truth

It's a reanalysis — a model-based best estimate. Local conditions may differ.

5

Jan–Feb only

Winter-only analysis. Full-year extension would provide seasonal variation and larger sample.

earnix

Conclusions

15.4x
more fog detected (at TD < 0.8°C)

99.1%
police fog confirmed by ERA5

p=0.81
weather adds nothing to severity (ns)

Free
open data (STATS19 + ERA5)

Recommendations for Actuaries

→ Start with fraud detection — cross-reference claims against ERA5 for quick ROI

→ Use ERA5 for frequency models, not severity — weather affects if crashes happen, not how bad

→ Calibrate fog thresholds to your use case — tighter for fraud, broader for risk modelling

→ Document methodology for regulators — objective data is defensible

→ Think long-term — ERA5's 80+ year history enables climate trend analysis

Data: data.gov.uk + cds.climate.copernicus.eu | Code available on request

earnix

earnix

Luba Orlovsky

lubao@earnix.com