RainfallChloride

Rainfall chloride at 85 South African sites, with a fitted prediction model – in-browser app • Hydrogeological Repository

What it does

RainfallChloride puts chloride concentration in rainfall behind a map, at 85 South African sites across four sources, and behind a fitted model that predicts it anywhere in the country from distance to the coast and elevation. Find sites by quaternary catchment, or by distance from a town or a coordinate; shade the map by observed or predicted chloride; see the predicted distribution across whatever area you searched; and check the model's own leave-one-out record before trusting it. Built on model/, a separate research pipeline with its own README, validation scripts and a record of approaches tried and rejected.

The four sources do not give the same kind of number – DWS spot samples read 2.53× higher than Van Wyk's network over the same ground, and CSIR about 2× higher again, for reasons not yet resolved. The app tells every source apart on the map and in every export, rather than blending them into one undifferentiated dot, and rather than pooling them it fits a per-convention offset on top of shared geography – so a prediction is always *as* a stated convention, and the reader is told which.

Where the data comes from

Source	Sites	Convention	What it holds
DWS	57	Arithmetic mean, spot samples	Dissolved chloride from the Department of Water and Sanitation's rainwater-chemistry monitoring network, 2003–2008. 157 dated samples, 1–7 per site. Not used to fit the model.
Van Wyk	19	Harmonic mean, reweighted	Van Wyk's thesis network. Reweighted from an equal seasonal average to the true rainfall-depth fractions (Van Wyk et al. 2011) – trains the model.
Sandveld	4	Precip-weighted (3) / bulk (1)	West Coast rain gauges, coordinates and elevations printed exactly in the source table. Three train the model.
CSIR	5	Unknown	Digitised from a CSIR map figure. Shown for context, left out of the colour scale and the model.

What you get

Output	Detail
Map	Every site over the quaternary catchments, shaded by observed chloride (red) or predicted chloride (blue), or both together. The symbol says which source the figure comes from.
Search	By quaternary catchment, town or coordinate - all three resolve to a point and a 10 to 200 km radius (a quaternary's own centroid stands in for the point), so the same circle, zoom and predicted-chloride distribution apply to all three. In coordinate mode the map is the coordinate picker: click it to set the point.
Selection	Click a site, or a row in the list, to select it; clicking it again clears it. One site at a time - the detail panel, the report and the LOO highlight all follow it.

Output	Detail
Site detail	Name, coordinates, elevation, quaternary, source, measurement convention, the reported chloride figure, and the model's prediction under that *same* site's own convention with its 68% interval - a fair comparison, since the convention is what the model actually predicts. Plus the dated sample table where one exists.
Prediction	A “Predict as” control picks the convention; click anywhere on the map for a reading, or shade a national grid by predicted value - the grid is its own surface, not painted onto the quaternary catchments, so it stays on when catchment outlines are switched off and never implies chloride respects catchment boundaries.
Distribution	The current search radius - not just the sites inside it - sampled and histogrammed through the model under the active convention, whichever of the three search modes found it. Drawn twice over: how far the prediction moves across the area, and that *plus* the model's own scatter.
Recharge	The chloride mass balance, plotted: recharge = rainfall chloride ÷ groundwater chloride, swept across groundwater values to 1 000 mg/L and carrying the rainfall prediction's own uncertainty as a band. Enter your own groundwater chloride to mark the read-off on the curve.
Validation	A button that plots all 85 database sites against their genuine leave-one-out prediction, shaded by a $\pm 1\sigma$ band. The selected site is always highlighted and follows the selection live.
Charts	All three charts pan, box-zoom, autoscale and read values back on hover - the same charting component the other apps in the repository use.
Report	A standalone, printable page of the current view - map, selection, distribution and validation - opened in a new tab.
Exports	CSV of the whole site catalogue, of every dated sample, or of the current search results; PNG of the map.

Reading the map symbols

Symbol	Source	What it means
Circle	DWS	A monitoring point with dated, repeated samples. 57 sites.
Diamond	Van Wyk	A dedicated collector network, reweighted for the model. 19 sites.
Triangle	Sandveld	West Coast rain gauges, mostly precipitation-weighted. 4 sites.
Square (dimmed)	CSIR	A different, uncorrected convention – shown for context, off the colour scale. 5 sites.

How a run works

Step	What happens
1. Find sites	Search by catchment, town or coordinate, then select a site on the map or from the list. Every search also draws the predicted-chloride distribution for its radius.
2. Export	CSV of the catalogue, the raw samples, or the current search results.
3. Report	A standalone page of the current view, ready to print or save as PDF.

Predicting chloride from distance to coast and elevation

chloride.db carries two fitted models (model/README.md); this app predicts with the **conventional** one, not the narrower physical one, because it is the one that can answer “what would this measurement convention report here” - the question a map, a click anywhere, or an area distribution all actually ask. Fitted on all 85 database sites by Huber robust regression, geography plus a per-convention offset: **Cl = exp(1.866 – 0.192·log(1+distance_km) – 0.000446·elevation_m + offset[convention])**. Leave-one-out R^2(log) 0.537, MAE 2.78 mg/L, median fold error 1.32× – noisier than a single-convention fit only because mixing five conventions is inherently noisier, not because the model is worse; adding the convention offset alone lifts R^2 from 0.291 to 0.537 (see the validation panel).

The other model (model_coefficients, trained on the 22 sites measuring the precipitation-weighted convention only) answers a different, narrower question - the physical quantity a chloride mass balance needs - and is not what this app predicts with; see model/README.md if that is the question you have.

The map's own shading is a grid of its own, not painted onto the quaternary catchments - chloride is continuous with distance and elevation and has nothing to do with catchment boundaries, so colouring catchment-shaped patches would have implied a precision the model does not have. It stays visible whether or not the catchment outlines themselves are switched on. Each elevation-grid cell is split three ways per axis for the shading, to about 9 km cells, with elevation and distance-to-coast interpolated within each cell: a smoother surface off the same ~28 km source data, not finer data.

Predicted-chloride distribution

Selecting a quaternary catchment, a town or a coordinate samples 250 points across that search's own radius – the circle itself, not just the sites inside it (a quaternary centres the circle on the catchment's own centroid) – looks up distance-to-coast and elevation for each, and histograms the model's prediction under the active “Predict as” convention. This is a Monte Carlo summary of what the model says about the area, not a measurement.

Two uncertainties are drawn, because the first badly understates the answer on its own. The **bars** are how far the model's central estimate moves from place to place across the radius. The wider **shaded area** folds in the model's own fitted scatter as well (±1σ in log space, the same figure the validation band and the site detail's 68% interval quote), and is the range a new measurement at that location would actually fall in. Over a small radius the model's own uncertainty is much the larger of the two: a 50 km search near the West Coast moves the central estimate over roughly 7–11 mg/L, while the honest predictive range is nearer 3–21. The bars alone would read far more confident than the model is.

Recharge by chloride mass balance

The reason rainfall chloride is worth compiling at all: in the chloride mass balance, **recharge = chloride in rainfall ÷ chloride in groundwater**. This app supplies the numerator, so the chart sweeps the denominator – groundwater chloride, out to 1 000 mg/L – and reads the recharge percentage off it, carrying the rainfall prediction's own ±1σ through as a band. Enter a groundwater chloride of your own and the read-off is marked on the curve and stated as a figure, with its uncertainty range; a table beside the chart gives the same at round values.

Both axes are logarithmic: the sweep spans three orders of magnitude, and on log-log the relation is a straight line, so the uncertainty band keeps a constant width instead of appearing to collapse as the curve flattens. The curve is capped at 100% and starts at the predicted rainfall value, because groundwater fresher than the rain that recharged it is outside what the mass balance can describe.

> **This is the plain mass balance, and it is an upper bound.**
>
> It makes no allowance for dry deposition, for chloride lost to runoff rather than infiltrating, or for evaporative concentration in the soil before recharge - each of which would lower the true recharge fraction. Nor does it carry any uncertainty in your groundwater chloride: the band shown is the rainfall term's alone. Treat the result as a first estimate to be checked against local evidence, not a finished recharge figure.

Entirely in your browser

Unlike RainfallArchive, this dataset is small enough – 85 sites, 157 dated samples, 1 943 towns, a 1 805-point elevation grid and a simplified coastline – to download once at boot and hold in memory. Every search, the model, the distribution and the site-detail panel all run locally; there is no per-site request to the server once the app has loaded. The leave-one-out diagnostic is the one exception: refitting Huber-robust regression 85 times needs statsmodels, so it is precomputed at build time and only fetched, not recomputed, when the validation button is used.

What says the compilation is sound

> **Each DWS site's figure was checked against Van Wyk's own compiled average.**
>
> For 56 of the 57 overlapping sites, the mean computed here from the raw dated samples matches the chloride average already compiled in Van Wyk's own spreadsheet to within 0.008 mg/L. The one exception, site 1000009151, differs by about 1 mg/L – closer to that site's single most recent sample than to the mean of all seven – and is reported here as the plain mean, reproducible from the samples table rather than carried through unexplained.

> **The convention offset is doing real work, not just absorbing noise.**
>
> The validation panel compares three fits, and explains each one beneath it. **Geo only** - distance and elevation with all 85 sites pooled as though measured alike - scores leave-one-out R^2 0.291, because it has to explain the up-to-5× differences between conventions as if they were geography. **Geo + convention**, what the app predicts with, lifts that to 0.537. **Per-convention**, each convention fitted alone with no pooling at all, is the ceiling and only reaches 0.557 - and cannot predict for a convention at a location with none of its own sites nearby, which is why it is a benchmark rather than the model.

Limitations worth knowing

- **R^2 0.537 is real, not a small-sample artefact,** but it means a single prediction is often well off - the fold-error table in the validation panel shows several sites out by 5× or more, mostly the handful of sites whose convention has too few examples (n=1 for cumulative bulk) to fit a reliable offset from.
- **The CSIR and cumulative-bulk offsets rest on very few sites** (5 and 1) and model_convention_offsets.reliable flags the latter 0 for exactly that reason - treat predictions under those two conventions as indicative only.
- **Measurement convention, not geography, is the dominant signal in this dataset.** DWS spot samples read 2.53× higher than Van Wyk's network at the same distance-to-coast and elevation - which is precisely why this app predicts under a chosen convention rather than pretending one number fits every source.
- **One DWS point returned no chloride result.** Monitoring point 1000009148 (Mispah) was sampled but its chloride result is blank, so it has nothing to show and is left out entirely.
- **One Van Wyk site falls outside every quaternary polygon.** Eddie_A, near Cape Point, keeps a blank quaternary rather than a wrong one.
- **The elevation grid is coarse,** about 28 km, cached from a public elevation API rather than a high-resolution DEM – adequate next to the model's own elevation coefficient, but not survey grade. That spacing was inherited from the research pipeline's own national map, where it was a practicality (1 805 API calls), not a resolution the model justifies. The map shading is interpolated to about 9 km within it, which is smoothing, not new data.
- **Lesotho and Eswatini are blank on the shaded map.** The grid is clipped to a South Africa-only national boundary, in which Lesotho is an interior hole and Eswatini falls outside altogether. Nothing in the model stops at those borders – distance-to-coast and elevation are defined everywhere, and both countries sit well inside its range – so this is a boundary artefact, not a statement about them. Filling them needs real elevation data for those cells: interpolating across the Lesotho gap from the lower ground surrounding it would badly understate a country that rises above 3 400 m, and so overstate its chloride.
- **Only chloride is carried through.** The rest of the DWS rainwater-chemistry suite, and the wider thesis material on seasonal and regional patterns, is not represented here.

Credit

The records are the work of the Department of Water and Sanitation's rainwater-chemistry monitoring programme, C.J. van Wyk's thesis and its companion paper (van Wyk, van Tonder & Vermeulen 2011, Water SA 37(2)), and the CSIR. Bredenkamp & Vogel (2007, WRC KV 177/07) supplied four measured values used to externally validate model/README.md's other, physical model - not the conventional one this app predicts with, but built from the same database and worth the same care. This app compiles and presents that work and claims no authorship of it.

> **Beta.**
>
> RainfallChloride is published for evaluation and feedback. Check any figure you rely on against the original record before using it in an assessment.