VegterAnalysis

Vegter’s geohydrological statistics, over any study area – in-browser app • Hydrogeological Repository

What it does

VegterAnalysis runs the statistical analysis of WRC project K5/2745, *Geo-statistical analysis and sub-delineation of all Vegter Regions*, over any study area you choose: one of Vegter’s **65** groundwater regions, a quaternary catchment, or a radius around a town or a coordinate. It draws on **227 217** boreholes from the National Groundwater Archive that carry at least one of the parameters the methodology uses – between them **1.30 million** static water levels, **176 308** water strikes and **12 031** chemical analyses.

Vegter reports almost everything per geological unit. This does the same, and adds the other groupings the methodology allows: quaternary catchment, Vegter region, aquifer rating, borehole depth band, DRASTIC vulnerability and the mapped water type. Changing the grouping regroups every table and chart on the page without going back to the server, so the same area can be read seven ways in a minute.

The statistics

Statistic	How it is calculated
Distribution and density	Boreholes per class and per square kilometre. Class areas are measured on the unsimplified covers in an equal-area projection. A grouping with no footprint, such as a depth band, gets counts and shares but no density, and the column says so rather than showing a zero.
Summary statistics	Mean and sample standard deviation per class for the water level, water strike depth, blow yield, EC, borehole depth, production discharge and pump-test transmissivity. The median and quartiles are shown beside them: these distributions are heavily right-skewed and a standard deviation larger than its own mean describes nothing.
Frequency analysis	Histograms binned from zero whatever the smallest observation is, so two areas can be compared. For a parameter measured downwards the right-hand axis carries the number of boreholes that reached each depth – which is what stops a tall bar at 150 m being read as common when only nine holes went that deep. With many classes on a phone the chart grows taller to fit its legend, rather than letting it squeeze the plot.
Strike frequency	Strikes per depth range divided by the boreholes passing through it. The app reads the peak off it as the depth band water is most often struck in, and the share of all strikes within each depth.
Yield against strike	One point per borehole or one per water strike, with Pearson's correlation coefficient on it.
Yield against dyke	A dyke intersection is a **DLRT** or **DIBS** unit in the lithology log. A borehole with no log has an *unknown* dyke status rather than none and is excluded.
Water level correlation	Water level against surface elevation, both in metres above sea level, with the 1:1 line drawn.
Water character against the map	Vegter mapped the hydrochemical facies he expected across the country, in the four classes that are the quadrants of a Piper diamond. Every borehole is stamped with the type it stands on, so the prediction can be held against the facies the archive's own analyses fall in – per mapped type, with the number of analyses behind each verdict beside it.
Water character	A Piper diagram, drawn by the same code as ChemGraphs so the two apps' diagrams can be laid over each other, with the electrical conductivity classed against the DWAF (1996) domestic guideline.

Beyond the published tables

Four statistics are calculated here that the original analysis does not, each because the data supports it and the decision needs it:

- **Borehole prospects.** The share of boreholes that blew at least 0.5, 2 and 5 l/s, per class – Vegter maps the equivalent as a probability index of a successful borehole and of one yielding 2 l/s or more. Read with the denominator beside it: this is the chance among boreholes that have a recorded yield, not among holes drilled.
- **A test on the dyke comparison.** Two bar charts side by side invite a conclusion that a few tenths of a litre a second between samples this scattered will not support. Welch's test is run on the two yield samples and the app says plainly when the difference is within what chance produces.

- **The correlation that is not circular.** Water level against elevation is near 1 almost whatever the data does, because both axes carry the collar elevation and it dominates both. The correlation between elevation and the *depth* to water, which shares no term, is given beside it – that is the question the plot is standing in for.
- **Medians and quartiles** alongside every mean and standard deviation.

Where the data comes from

Dataset	What it supplies
National Groundwater Archive	Boreholes and every observation: positions, collar elevations, static water levels, water strikes and blow yields, lithology, production discharge, pump-test transmissivity and chemistry. The archive's numeric code columns are resolved to words once, when the app's database is built.
Simplified geological map	The 1:1 000 000 map from the Council for Geoscience – 13 522 polygons over 366 units, each with its lithology and its aquifer rating from 1 to 10. The simplified map is used rather than the 1:50 000 sheets because it is the resolution the borehole data can support.
Vegter's groundwater regions	The 65 regions the study areas are built around.
Vegter's water quality map	The hydrochemical facies expected over each part of the country – 184 polygons over 17 classes, each naming one or two of the four Piper facies and the share of the water it holds. It covers 99.8 % of the boreholes in the database.
Quaternary catchments, DRASTIC	Catchment boundaries, and the national groundwater vulnerability grid sampled at each borehole.

What comes out

- A standalone printable report of the whole analysis – the map, what the area holds, the density and summary tables, the charts, what the analysis found in words, and the caveats that have to travel with the numbers. A chart the page shows one class at a time becomes one figure per class, so the charts are chosen before the report is built and it names the ones left out – an absent chart should not read as an empty study area.
- PNG and SVG export of every chart, and of the Piper diagram as the vector figure itself rather than a picture of one.
- PNG export of the map, drawn from its own geometry without the zoom buttons, the base-map selector or the scale bar.
- CSV of the summary statistics, the histogram bins as drawn, the strike frequency, the borehole prospects and the Piper coordinates – each a long-form table with its units, nothing rounded.
- CSV of the per-borehole parameters every statistic was computed from, and of the raw archive records behind them.

Before the numbers are used

- **Most positions are farm centroids, not drill sites.** Before GPS a borehole was recorded at the centroid of the farm it fell on, so which geological unit or catchment a borehole is credited to is only as good as that position.
- **The archive records no dry boreholes** – six of 280 994 records carry a "dry" status. Every yield figure is conditional on a yield having been recorded, and no drilling success rate can be calculated from this archive at all.
- **Chemistry and discharge are regional.** In this copy of the archive both exist only in the A, B and X primary catchments; elsewhere those panels are empty for want of data, not because the water is clean.

- **Two archive placeholders are excluded.** A blow yield of 0.0001 l/s appears at 51 163 strikes – 29 % of every strike in the archive at one value – and a depth in the 999–1000 m band at 2 526 boreholes. Both are “not measured” written as a number, and kept they would set the mode and the standard deviation of every statistic they land in.
- **Geology is surface geology.** A borehole is credited to the unit its collar stands on; it may be producing from something else entirely a hundred metres down.