Visualisation and Analysis of Large Distributed Temperature Sensing (DTS) Datasets from High Voltage Cables

Sunny **CHAUDHARY**, George **CALLENDER,** Justin **DIX,** Paul **LEWIN**, University of Southampton, (United Kingdom),
sc.chaudhary@soton.ac.uk, g.m.callender@soton.ac.uk, jkd@soton.ac.uk, pll@soton.ac.uk

ABSTRACT

High-voltage cables are a critical part of a countries power infrastructure, whether transferring power from large-scale, onshore and offshore renewable projects, or expanding network capacity for more conventionally generated energy sources. Ensuring the integrity of these assets requires continuous thermal monitoring; distributed temperature sensing (DTS) is the standard tool but produces archives of extreme size at metre-scale and minute-scale sampling over ~100 KM and years, ≥100 GB. Analysis of such huge archives has up until now been impossible. The challenges are twofold: interactive out-of-core computation over full resolution and display-aware visualization that preserves local transients under hard pixel limits. This work delivers a preliminary approach: an out-of-core stack ordered tiling, zstd compression, memory-mapped streaming, and adaptive, pixel-aware aggregation; a layered analysis pipeline from coarse heatmaps to contours and fixed quantiles, perceptual composites, fixed-scale wavelet energy, and UMAP–HDBSCAN segmentation. Finally, a Bayesian-optimized weighted baseline correction combining TOPHAT, JBCD, MORMOL, and IARPLS is also presented. Results demonstrate efficient and robust visualisation of 100 GB/3.4 billion-points of data on 32 GB RAM; contours and quantiles make seasonal structure and thresholds explicit; composites separate temporal from spatial temperature spikes and highlight sparse sampling; and baseline removal suppresses drift to reveal load-correlated vertical banding in the residuals.

KEYWORDS

HVDC Cables, Distributed Temperature Sensing (DTS), Thermal Monitoring, Out-of-Core Processing, Spatiotemporal Visualization

INTRODUCTION

High-voltage direct (HVDC) and alternating (HVAC) current transmission, via buried cables, is central to electrical power transfer across the planet [1]. Safe loading and lifecycle management depend on a detailed understanding of thermal behaviour along the buried sections of both terrestrial and submerged routes [2], [3], [4]. Spatial and temporal variations in the thermal environments of the sub-surface along with localised cable variations (e.g. the presence of joints and HDD installations) can create thermal pinch points. These could either limit power delivery for the whole circuit or at worst result in failure. Distributed temperature sensing (DTS) using optical fibre provides quasi-continuous temperature (T) fields along such routes and has become a standard diagnostic tool, with industry guidance emphasising its role in operational decision-making [5]. These deployments produce archives of extreme volume and resolution. A typical system acquires *T* data at a metre-scale spatial resolution and minute-scale temporal sampling, extending several 10's of kilometres, and sustains this for years. Consequently, datasets of the order of 10^{11}–10^{12} samples, corresponding to ≥100 GB. Two distinct technical problems in analysing these data arise:

a) **Computation and processing** - Such archives cannot be browsed or processed interactively on consumer-grade hardware without partitioning or pre-aggregation. This disrupts context and limits statistical comparison across the full timeline. Out-of-core methods are required to maintain interactivity.

b) **Visualisation** - Any visual encoding must ultimately be projected onto a finite display, typically 10^6 pixels (px), each constrained to one colour value at a given time. When billions of data points are mapped directly, many-to-one collisions occur. Naïve aggregation reduces this collision but discards local variance and transients that may be operationally significant. Preserving spatiotemporal fidelity within the hard limits of display resolution requires adaptive aggregation strategies and linked statistical views.

Operational requirements extend beyond static summaries. Engineers must navigate complete routes without loss of resolution by panning and zooming in on arbitrary windows, and compute comparative metrics that expose outliers, recurrent patterns, and extreme events such as changes in depth of cover or third-party interference. Standard statistical packages and plotting libraries are not designed for these constraints or accustomed to DTS [6], [7]. Further, there is now significant interest in analysing DTS from extant systems to optimise cable design in the future.

This paper presents a preliminary framework that addresses both computational and visualisation bottlenecks. First, an out-of-core architecture enables interactive heatmaps, surface plots, and linked statistical views at full spatial and temporal extent. Second, adaptive visualisation techniques align data density with display capacity while retaining fine-scale information where it is most relevant. Third, a comparison layer computes user- or standard-defined metrics across arbitrary spatial or temporal subsets without pre-aggregation. A case study demonstrates application to HVDC cable monitoring and demonstrates the application of the framework to support decision-oriented workflows.

DATASET

Synthetic datasets were generated using Deepecho, a deep generative modelling library for time-series data, to protect confidentiality and anonymise sensitive operational records [8]. The synthetic data preserves the statistical and temporal characteristics of an original DTS archive while ensuring that no direct identifiers or raw measurements are disclosed. All results presented in this study are therefore based on the synthetic datasets, which were derived from but not traceable to the underlying operational data, which cannot be released due to commercial and security constraints. The spatial (s_p) resolution of the dataset is ~1.5 m, and the temporal (t_p) resolution is ~30 min, consisting of 100 KM over a period of one year, ~1.2 billion datapoints, equivalent to 36.8 GB.

FRAMEWORK DESIGN

Out-of-memory processing was designed to enable real-time analysis of DTS archives at scales beyond 100 GB, where naive loading and plotting are infeasible on standard workstations. The system combines a custom binary tiling format, adaptive caching, and display-aware rendering into a unified architecture that eliminates the need for fragmented datasets or pre-aggregated views. At the storage level, a custom binary container named DTSVTF partitions the spatiotemporal field into fixed 512×512 tiles arranged in Morton (Z-order) sequence [9], [10]. This provides hierarchical spatial locality and cache-friendly I/O, allowing tiles that are adjacent in distance and time to be retrieved together.

Each tile is compressed independently with zstd, enhanced by dictionaries that reflect the statistics of the input dataset [10]. Compression ratios of 8:1 to 12:1 are routinely achieved while ensuring decompression speeds suitable for interactive use. Integrity is maintained through per-tile SHA256 checksums, and dynamic encoding (i16 fixed-point or f32 floating-point), which is selected during ingestion to minimise storage overhead while retaining physical accuracy.

Access and computation are driven by memory-mapped streaming. Instead of loading the archive into RAM, only the tiles corresponding to the viewport or analytical query are materialised. A multi-gigabyte LRU cache retains recent tiles, sustaining >95% hit rates during sequential navigation of datasets with >100 million samples. Multi-threaded decompression and baseline correction, parallelised with Rayon [11], exploits all available CPU cores, and Single Instruction, Multiple Data (SIMD) vectorisation accelerates per-tile operations for statistical reduction [12]. Persistent caches store the results of computationally intensive processes, enabling sub-second reloads in subsequent sessions. Rather than producing fragmented plots for subsections of the dataset, the framework implements a pixel-aware adaptive aggregation scheme.

Data density is dynamically aligned to display resolution: when multiple samples map to the same pixel, the renderer computes statistical envelopes rather than discarding points. This prevents aliasing and preserves visibility of localised excursions such as hotspots or transients that would otherwise be suppressed by averaging. Adaptive level-of-detail rendering ensures that when zoomed in, cached high-resolution tiles are substituted seamlessly for coarse aggregates without recomputation. This enables the entire dataset to be explored in real time, with continuity across scales, and without loss of critical fine-scale information.

Metrics such as moving standard deviation, percentile envelopes, and counts are computed incrementally across tiles, ensuring that statistics reflect the full dataset even when only a fraction is resident in the memory. Query windows can be run interactively, with computations distributed across cores and merged in constant memory. This avoids the fragmentation that arises when archives are split into separate files for analysis or plotted by section ranges, while still allowing users to rank spans, detect anomalies, and compare temporal segments within a unified view.

BENCHMARK

Validated on a 100 GB DTS archive containing 3.4 billion data points. Despite the scale, the complete dataset could be navigated responsively on a standard workstation with 32 GB RAM, maintaining frame rates of 30–60 FPS during interactive panning and zooming. Benchmarks with smaller archives show ingestion of a 23.6 GB CSV into a 760 MB container with an additional 115 MB baseline cache, loading in under ~2 seconds and sustaining statistical plot generation in 50–200 ms. Peak memory usage for the 100 GB case was ~24.2 GB, well within the capacity of standard hardware.

RESULTS AND DISCUSSION

A coarse view of the dataset can be obtained through direct aggregation of T across s_p and t_p bins, without applying further statistical processing, as presented in Figure 1. The resulting heatmap displays T variation along the t_p axis (x-axis, one year) and the s_p axis (y-axis, 100 KM), with the colour scale representing the aggregated T distribution. In this case, the aggregation is performed using the mean; however, the method can be adapted to use other statistical measures such as minimum, maximum, median, 95th percentile, or 5th percentile, depending on the focus of analysis. This initial visualization serves as a baseline view of the data structure before applying further techniques.

Contour-Based Methods

Figure 2 illustrates an enhanced heatmap by incorporating T contours that expose the structure and dynamics of the thermal field across both t_p and s_p dimensions. T are discretized into fixed intervals such as 5 °C. These intervals are visualized in two ways: contour lines that trace isotherms with explicit labels and filled contour bands that provide continuous zones of equal T. This introduces topographic detail, clearly delineating gradients, boundaries, and transitions that would otherwise remain hidden. The contours also reveal seasonal effects based on T zone shifts and their evolution across the t_p axis. Warmer months can be identified as extended regions of higher T contours, while cooler periods emerge as expanded low-T zones. Warmer regions within these areas can further be correlated with the load profile. Hotspots become visible in time, indicating when elevated cable those conditions concentrate. They also provide quantitative information that can be directly linked to operational thresholds. By fixing the interval size, it enables consistent comparisons across datasets, such as between different years or different assets, without the distortion of variable scaling. In practical terms, this means anomalies

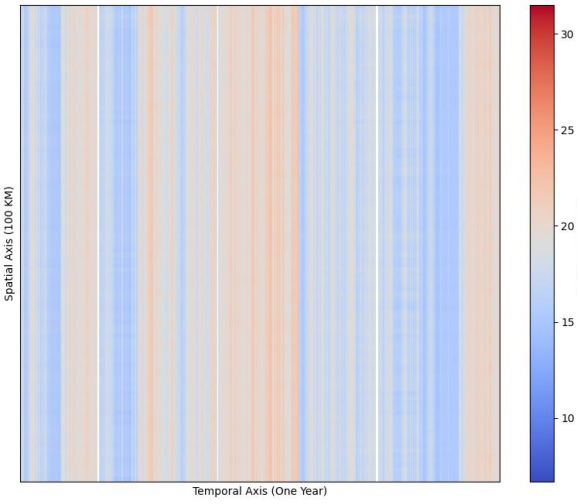

Figure 1. Coarse heatmap of T distribution aggregated across t_p (one year) and s_p (100 KM) axes. Colors represent mean T values, illustrating large-scale hot-cold variation in the dataset.

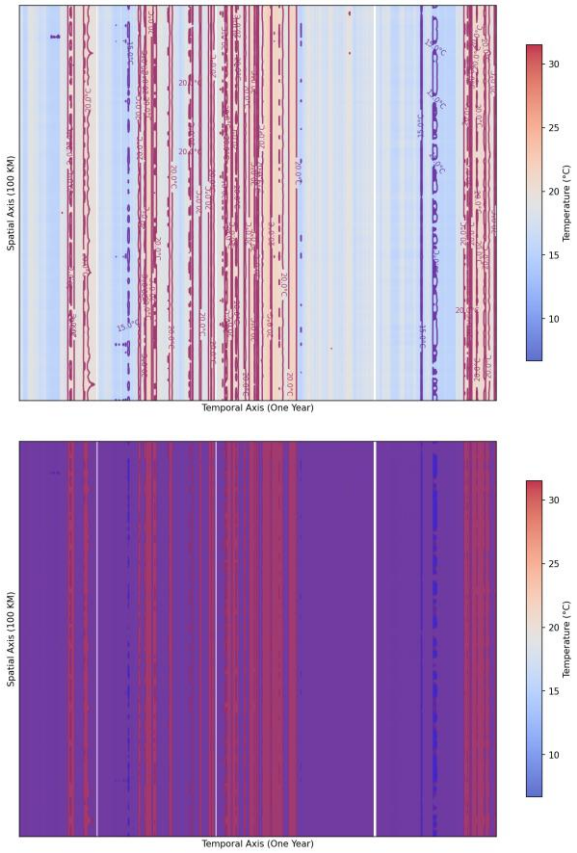

Figure 2. *T* distribution represented with contours (a) overlaid contour lines showing isotherms and labeled *T* levels, highlighting gradients and transitions (b) filled contour bands emphasizing zones of equal *T*, providing a view of seasonal patterns and hotspot regions.

can be detected and tracked against a stable reference framework. Figure 3 presents the *T* for three months and the full length of the hypothetical cable as an extension of contours presented in Figure 2., indicating where along the cable stress occurs, and in space. Measurements are aggregated onto a uniform grid, missing bins are interpolated via a radial bias function (RBF), and a Gaussian kernel (σ = 2 pixels) suppresses pixel-scale noise. The heatmap is overlaid with quantile isolines at the 5th, 50th, and 95th percentiles, which in this instance fall near 16.3 °C, 18.5 °C, and 21.7 °C, respectively.

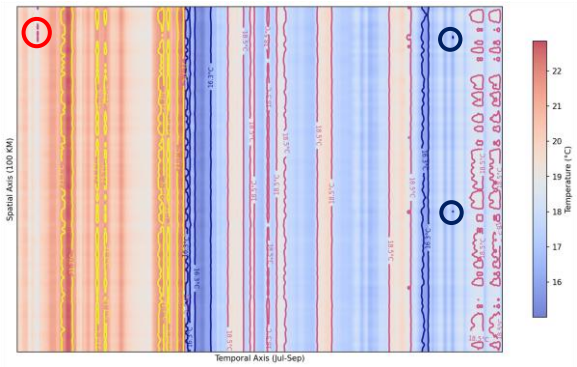

Figure 3. Spatiotemporal *T* field for three months with quantile isolines (Q5 (blue), Q50 (red), Q95 (yellow) ~ 16.3, 18.5, 21.7 °C) overlaid on a smoothed heatmap. Vertical banding indicates time-coherent forcing along 100 KM; localized departures highlight candidate anomaly zones. The red circles show *T* anomalies such as hot/cold spots.

The field exhibits a pronounced warm regime at the start (dominant 95th-percentile band ~21.7–22.8 °C), followed by an abrupt whole cable cooling marked by the ~16.3 °C isotherm. Through the mid and late months, the contours stabilize around the median ~18.5 °C, with intermittent warm regions. The near-vertical alignment of contours across most of the domain indicates time-dominated forcing that is coherent along the route, whereas small, spatially confined deviations near the period end suggest localized anomalies superimposed on the seasonal decay. Analytically, the plot answers: (a) timing of peak conditions and the onset and duration of the cooling episode are explicit from the contour chronology, (b) spatial persistence or confinement of anomalies is visible from the lateral extent of high- or low-*T* bands, and (c) labelled isolines provide immediate threshold context for exceedance analysis without re-scaling. This makes the view operationally useful for scheduling inspections during hot phases, evaluating whether recurrent segments approach thermal limits, discriminating ambient seasonal trends from localized defects, and establishing comparable metrics for year-over-year benchmarking and calibration.

Multi-Colour Channel-Based Methods

A separate synthetic dataset is used here to show close to realistic features that could appear in DTS datasets, raw *T* heatmap is illustrated in Figure 4a. Figure 4b is a Hue, Saturation, Value (HSV) multi-channel view of the

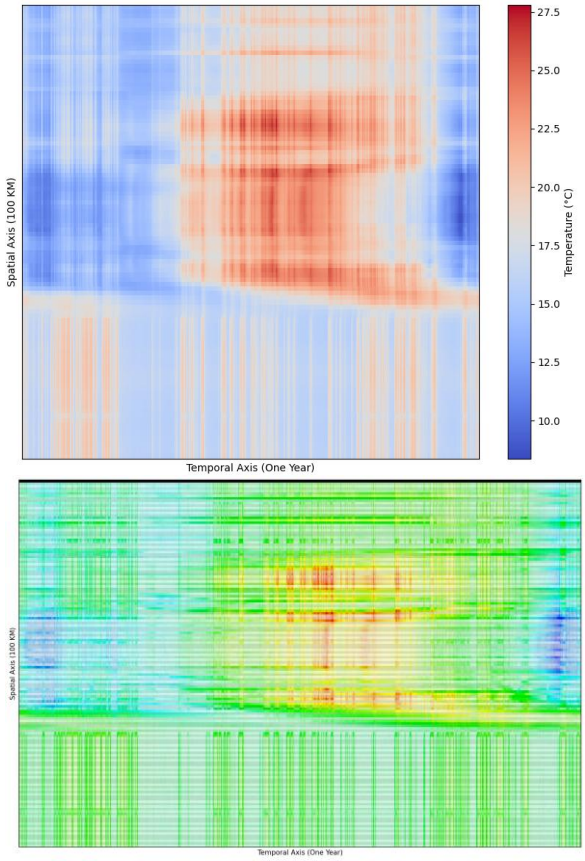

Figure 4. (a) Temperature heatmap of synthetic data with derived realistic features (b) HSV multi-channel heatmap of the spatiotemporal field: hue encodes mean *T* (blue→red), saturation encodes intra-pixel standard deviation, and value encodes *log (sample count)*.

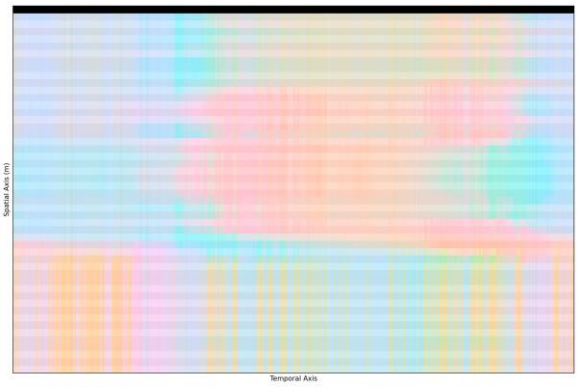

Figure 5. OKLab composite visualization of the spatiotemporal field. *log (sample count)* → brightness; temporal anomaly (z-score vs time-median) → red/green; spatial anomaly (z-score vs distance-median) → yellow/blue.

spatiotemporal field: hue encodes mean *T* (cold→blue, hot→red), saturation encodes the local standard deviation of the *T*, and value encodes the data density via *log(count)*. Saturation was clipped at the empirical 95th percentile to stabilize extremes; missing pixels were rendered black. The image exposes regime structure: hue shows baseline shifts over time, saturation highlights zones of elevated variability, and value reveals gaps or uneven coverage; composite view separates width-wide changes from localized anomalies and prioritizes segments for further analysis.

Figure 5 shows a perceptual multi-channel composite of the DTS field using the OKLab colour space [13]. Three independent diagnostics are encoded simultaneously: lightness (L) conveys data reliability as the normalized log of the sample count per pixel; the red–green opponent axis (a) encodes temporal anomalies computed as a column-wise z-score of the *T* relative to the time-median at each time step; the yellow–blue opponent axis (b) encodes spatial anomalies computed as a row-wise z-score relative to the distance-median at each distance. Chroma is deliberately limited, such that differences from nominal conditions appear as soft tints rather than saturating the display. Broad pastel bands aligned vertically indicate time-coherent shifts that are uniform along the span (temporal anomalies), while horizontal tinting signals distance-localized structure (spatial anomalies). Regions that appear near-neutral grey denote behavior consistent with the local medians, and dim areas indicate sparse sampling where conclusions should be discounted.

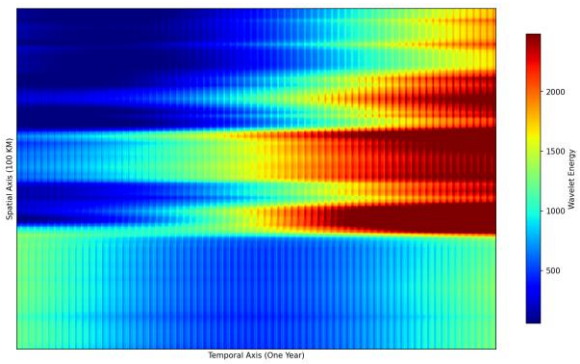

Figure 6. Wavelet energy map for a fixed analysis scale across one year (x) and 100 KM (y) for a period of one month. Warm colors mark intervals where the targeted periodic component is strong and spatially coherent; cool colors mark weak modulation.

Transforms and Clustering-Based Methods

Figure 6 is a fixed-scale continuous-wavelet energy map of the field. $x(y,t)$. For each distance y the time series $x_y(t)$ is transformed into a complex Morlet wavelet [14], $\varphi(t) = \pi^{-\frac{1}{4}} e^{i\omega_o t} e^{i\omega_o t} e^{-\frac{t^2}{2}}$. The continuous wavelet transform is: $W_x(a,b) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{a}} \varphi^* \left(\frac{t-b}{a}\right) dt$ with scale $a > 0$ and time shift b. Wavelet central frequency $f_c = c(\varphi)$, so the pseudo-period of scale a under sampling Δt is, $T(a) = \frac{a}{f_c} \Delta t$ and $f_a = \frac{1}{T(a)} = \frac{f_c}{a \Delta T}$. At the selected scale of $a = a_0$ (i.e., fixed period $T(a_o)$, The wavelet energy is $E_y = \left|W_{x_y}(a_o, b)\right|^2$ and the plotted images is $E(y,b)$ with colour indicating the magnitude. Broad, nearly horizontal warm bands indicate that the targeted periodic component grows in amplitude coherently across long spatial extents later in the year.

Three strata are evident: a lower zone of weak energy that ramps gradually, a mid-depth zone where energy strengthens early and saturates, and an upper zone with delayed amplification. Fine vertical combing reflects sub-periodic modulation of the carrier at this fixed scale and sampling, not spatial structure. Edge regions may be slightly biased by the cone of influence (finite-length effects).

The image shows when and where the selected scale becomes dominant, from which (a) time envelope. $A(b) = \int E(y,b)\, dy$ can be extracted, (b) spatial coherence can be measured $C(b) = 1 - Var_y[E_y(b)]/Mean_y[E_y(b)]^2$ and (c) different distance segments can be compared using either E_y^{max} or seasonal integrals $\int E_y(b) db$. Here y is the distance, t is the time, $x(y,t)$ is the *T* field, $x_y(t)$ is the time-series at y, $\varphi(t)$ is the Morlet mother wavelet, ω_o is the central angular frequency, $a > 0$ is the selected scale – dilates φ where larger a implies lower frequency, $W_x(a.b)$ is the CWT coefficient, $T(a)$ is the pseudo-period, $f(a)$ is the pseudo-frequency and $E(y,b)$ is the energy field.

Figure 7 is a Uniform Manifold Approximation and Projection (UMAP) combined with Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) segmentation back-projected onto the time–distance plane [14], [15]. Each pixel is colored by its cluster label, and the heatmap is partitioned into discrete regimes. One feature vector per sample is created – $[t^*, y^*, T]$ where t^* is $t/(W-1)$ and y^* is $y/(H-1)$ which are the time and distance indices for a given *T*. Following this, a k-nearest neighbors graph is built, and UMAP is used to convert this graph into a

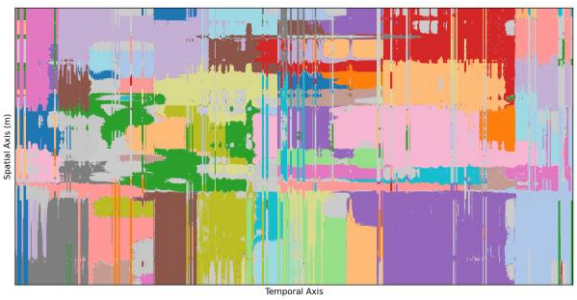

Figure 7. UMAP–HDBSCAN segmentation of the time–distance field. Colors label clusters obtained from a UMAP embedding. Horizontal blocks indicate distance-localized persistent regimes, vertical slabs indicate system-wide temporal transitions, and fine speckle marks indicate sparse or transient states.

fuzzy simplicial set for an optimised two-dimensional embedding by cross-entropy minimisation to preserve local neighborhoods. HDBSCAN is then applied to these embeddings to obtain density clusters; labels are predicted for all samples and back projected to (t, y) pixels to create a segmented heatmap.

Large horizontal blocks indicate distance-localized regimes persistent over time; large vertical slabs indicate system-wide temporal transitions; thin vertical streaks reflect rapid events; small, speckled regions usually arise from sparsity or brief transients. The result is a compact annual state map that enables change detection via boundary motion and label switching, isolates noisy fragments, and compresses for downstream modeling of cross-span or cross-year comparisons via cluster statistics or similarity analysis.

Baseline Removal

Finally, Figure 8 demonstrates a baseline subtraction of the DTS synthetic dataset using multiple established methods in combination with Bayesian optimization. This dataset has been particularly designed to replicate the *T* profiles in a Horizontal Direct Duct (HDD) where the *T* is mainly dominated by the ambient *T* to replicate the optical fibre being placed in a separate duct to the [DC] cable and therefore temperature influence of the cable is significantly reduced.

Figure 8a shows the raw *T* profile (grey) together with baselines derived from Top-hat Transformation (TOPHAT), Joint Baseline Correction and Denoising (JBCD), Morphological, and Mollified Baseline (MORMOL), and Improved Asymmetrically Reweighted Penalized Least Squares (IARPLS) [16], [17], [18], [19]. Instead of relying on a single method, a weighted combined baseline was constructed. The optimal weights and method parameters were selected automatically using Bayesian optimization, which systematically explores the parameter space to maximize the quality of the baseline fit. The individual methods serve complementary roles: TOPHAT, a morphological filter, removes broad background fluctuations; JBCD decomposes the signal into baseline and structural components; MORMOL adapts morphological operators to variable signal trends; and IARPLS applies penalized least squares with adaptive weighting to suppress noise and curvature.

Figure 8b presents the residual (original *T* minus weighted combined baseline), which isolates the short-term fluctuations and anomalies while suppressing systematic drifts. Figure 9a presents the spatiotemporal distribution of the synthetic dataset DTS measurements as a heatmap. Figure 9b presents the same heatmap with the baseline *T* subtracted. Compared to the raw *T* field, the baseline representation removes long-term drift and noise. After baseline removal, the influence of the simulated synthetic load profile becomes more apparent. The periodic vertical banding along the temporal axis corresponds to load-driven heating and cooling cycles, presenting a much clearer perspective on the thermal response and its correlation with operational conditions. It is important to note, however, that this baseline-corrected field does not represent the true ambient *T* at depth. Rather, it serves as an approximation designed to visualise a cleaner thermal response by suppressing background effects. This provides an effective way to highlight load-related variations, but it should be interpreted as a processed signal rather than a direct measurement of the ambient.

CONCLUSIONS

The study provides a preliminary approach to interactive

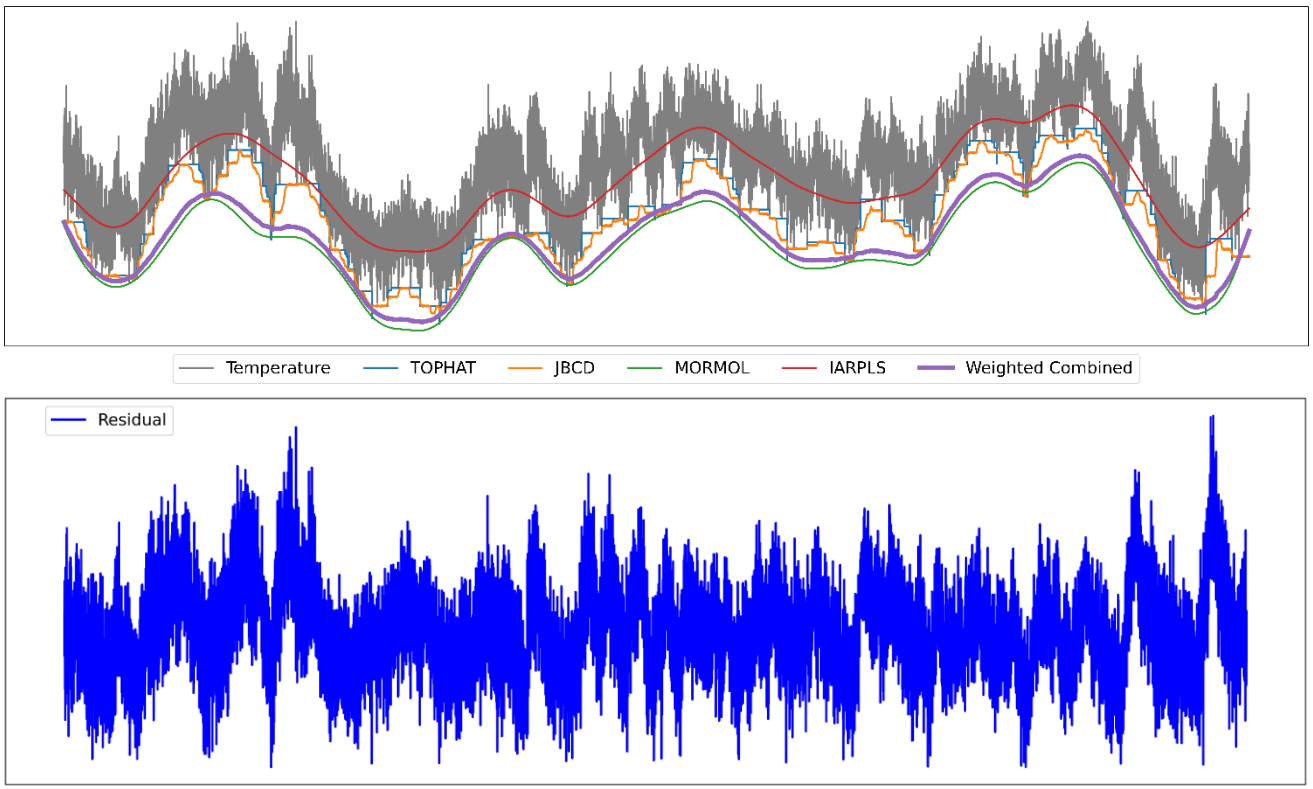

Figure 8. Comparison of baseline correction methods applied to distributed *T* measurements (x-axis – Date-Time; y-axis – Temperature) (a) Original *T* profile (grey) with fitted baselines from TOPHAT, JBCD, MORMOL, IARPLS, and the optimized weighted combination (b) Residual signal obtained by subtracting the weighted combined baseline.

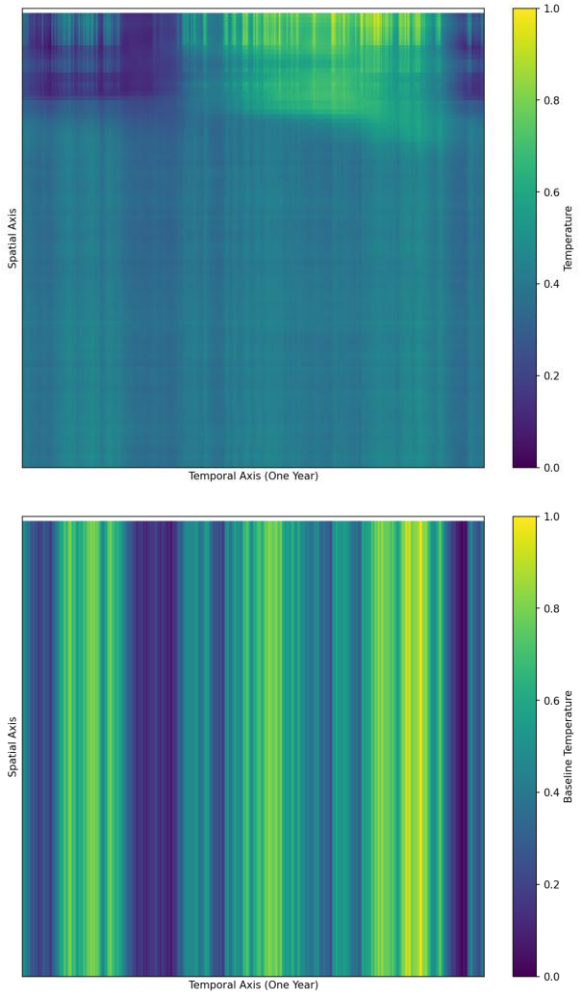

Figure 9. Heatmap of (a) synthetic dataset from the HDD. The *T* is dominated by the baseline/ambient *T* (b) Heatmap at the HDD after removal of the baseline demonstrating that there are still higher frequency load dependent temperature components despite the full temperature being primarily driven by baseline conditions.

analysis of ≥100 GB DTS archives from HVDC cables and separates load-related structure from baseline drift using statistical analysis. Synthetic data (~1.5 m spatial, 30 min temporal, 100 KM, ~1.2 b points) was utilised for the demonstration. An out-of-core stack 512×512 Morton-ordered tiling, zstd compression, memory-mapped streaming, multi-core processing, and display-aware aggregation delivers responsive full-route visualization on commodity hardware. Analysis proceeds from coarse heatmaps to contour and quantile overlays that expose gradients, thresholds, seasonality, and persistence. Perceptual composites (HSV, OKLab) encode mean, variability, sampling density, and decouple temporal from spatial anomalies to guide exploration of data. Transform-domain and learning views add diagnostics: fixed-scale wavelet energy maps localize coherent periodic content, and UMAP–HDBSCAN segments operational regimes for change detection. Baseline correction then combines TOPHAT, JBCD, MORMOL, and IARPLS under Bayesian-optimized parameters and weights; residuals (original minus weighted baseline) suppress drift and potentially reveal load-correlated banding and localized anomalies. This work provides a preliminary way for a detailed analysis of DTS data in much more depth.

ACKNOWLEDGEMENTS

The authors would like to acknowledge National Grid Electricity Transmission for supporting the project and agreeing to the publication of the results. Analysis of the Thermal Influence of Cable Surroundings (AnTICs NIA2_NGET0042) was made possible through the Network Innovation Allowance.

REFERENCES

[1] A. Stan, S. Costinaș, and G. Ion, “Overview and Assessment of HVDC Current Applications and Future Trends,” *Energies (Basel)*, vol. 15, no. 3
[2] S. Ahmad, Z. H. Rizvi, and F. Wuttke, “Unveiling soil thermal behavior under ultra-high voltage power cable operations,” *Sci Rep*, vol. 15, no. 1
[3] M. Quercio, J. C. Del Pino Lopez, S. Grasso, and A. Canova, “Numerical and experimental analysis of thermal behaviour of high voltage power cable in unfilled ducts,” *Sci Rep*, vol. 14, no. 1
[4] P. Y. Wang *et al.*, “Dynamic Thermal Analysis of High-Voltage Power Cable Insulation for Cable Dynamic Thermal Rating,” *IEEE Access*, vol. 7, pp. 56095–56106
[5] A. Ukil, H. Braendle, and P. Krippner, “Distributed Temperature Sensing: Review of Technology and Applications,” *IEEE Sens J*, vol. 12, no. 5, pp. 885–892, 2012,
[6] J. A. Cottam, A. Lumsdaine, and P. Wang, “Abstract rendering: out-of-core rendering for information visualization,” *Visualization and Data Analysis 2014*, vol. 9017, p. 90170K
[7] “DTSGUI https://www.usgs.gov/data/dtsgui
[8] “GitHub - sdv-dev/DeepEcho: Synthetic Data Generation for mixed-type, multivariate time series.”
[9] L. J. Stocco and G. Schrack, “On Spatial Orders and Location Codes,” *IEEE Transactions on Computers*, vol. 58, no. 3, pp. 424–432
[10] L. A. Walker, Y. Li, M. McGlothlin, and D. Cai, “A Comparison of Lossless Compression Methods in Microscopy Data Storage Applications,” *bioRxiv*, Jan. 2023,
[11] J. Abdi, G. Posluns, G. Zhang, B. Wang, and M. C. Jeffrey, “When Is Parallelism Fearless and Zero-Cost with Rust?” *Proceedings of the 36th ACM Symposium on Parallelism in Algorithms and Architectures*, pp. 27–40
[12] O. Polychroniou and K. A. Ross, “VIP: A SIMD vectorized analytical query engine,” *The VLDB Journal*, vol. 29 no. 6, pp. 1243–1261
[13] O. A. Basova, A. S. Grigoryev, and D. P. Nikolaev, “Calibration Model for Perceptual Compensation of Defective Pixels of Self-Emitting Display,” *Proceedings - European Council for Modelling and Simulation, ECMS*, vol. 2022-May, pp. 150–158
[14] L. McInnes, J. Healy, N. Saul, and L. Großberger, “UMAP: Uniform Manifold Approximation and Projection,” *J. Open Source Softw.*, vol. 3, no. 29, p. 861
[15] L. McInnes, J. Healy, and S. Astels, “hdbscan: Hierarchical density-based clustering,” *J. Open Source Softw.*, vol. 2, no. 11, p. 205
[16] S.-J. Baek, A. Park, Y.-J. Ahn, and J. Choo, “Baseline correction using asymmetrically reweighted penalized least squares smoothing.,” *Analyst*, vol. 140 1, no. 1, pp. 250–7
[17] I. Johansen, “Scenario modelling with morphological analysis,” *Technol Forecast Soc Change*, vol. 126, pp. 116–125
[18] H. Liu, Z. Zhang, S. Liu, L. Yan, T. Liu, and T. Zhang, “Joint Baseline-Correction and Denoising for Raman Spectra,” *Appl Spectrosc*, vol. 69, no. 9, pp. 1013–1022
[19] H. Zhu, J. Zhang, G. Xu, and L. Deng, “Balanced Ring Top-Hat Transformation for Infrared Small-Target Detection with Guided Filter Kernel,” *IEEE Trans Aerosp Electron Syst*, vol. 56, no. 5, pp. 3892–3903