APPENDICES

A. Platform technical description

A.1. Migration Context

OPHELIA was developed to replace a prototype initially hosted on Google Colab with a Gradio interface. This migration to a stand-alone, production-deployable solution made it possible to reduce the software footprint while improving performance and accessibility.

A.2. Data volume considerations

DataFrames provides the following.

- Efficient tabular data handling,
- Vectorized numerical computation,
- Built-in time-series indexing and resampling capabilities,
- Straightforward integration with scientific libraries.

For typical datasets (hourly or sub-hourly measurements over several years), memory usage remains moderate, and computational performance is adequate on standard personal computers.

Scalability perspective. If future use cases involve significantly larger datasets (e.g., multi-institutional databases, high-frequency monitoring, or centralized national repositories), the architecture could evolve toward more advanced storage solutions.

Future developments include the integration of **ArangoDB** (a graph-oriented database) for more flexible management of multi-sensor data, as well as centralized authentication via **Keycloak** (SSO) to facilitate multi-user deployment in institutional environments.

A.3. Unit tests

To ensure the reliability of the platform's computational outputs, a unit test suite was developed targeting the functions that constitute the analytical backbone of the platform. The guiding principle was to focus on functions called across multiple modules, where a silent numerical error a function that executes without exception but returns a physically inconsistent value — would propagate undetected through all downstream results. Functions responsible for interface rendering or external API calls were excluded, as their behavior depends on runtime conditions that are not amenable to unit testing. Tests were organized into six thematic groups: (1) thermodynamic formulations (dew point, absolute humidity, saturation vapor pressure, vapor pressure, mixing ratio), verified against tabulated ASHRAE reference values and physical invariants; (2) MAD-based outlier detection, tested on synthetic series with injected spikes and edge cases including NaN-containing series; (3) date parsing functions, covering all formats declared in the configuration mapping as well as degenerate inputs such as empty strings and None values; (4) metadata JSON serialization, including round-trip consistency and robustness to malformed entries; (5) utility functions for list normalization, whose failure would affect the entire sensor registration workflow; and (6) temporal aggregation functions, verified on synthetic series with analytically known statistics. The suite was implemented as a self-contained executable cell within the Google Colaboratory environment.

At the time of submission, 83 of 86 tests pass. The three discrepancies are not functional errors: one reflects the known reduced accuracy of the Magnus–Tetens formulation above 60°C, a temperature never encountered in heritage monitoring conditions, and two correspond to a degenerate MAD edge case on constant synthetic series, which does not arise with real sensor data.

A.4. User interaction

Dynamic and guided interface. The application leverages interactive components provided by the `Gradio` framework to facilitate intuitive navigation. User inputs (dropdown menus, multi-select lists, sliders, date selectors) are dynamically updated according to metadata-based filtering. For example, once a site is selected, only the corresponding sensors are displayed; once sensors are selected, only available variables are proposed. This dynamic updating prevents inconsistent selections and reduces user error.

Input fields are pre-filled with default or previously used values when relevant (e.g., standard aggregation periods, typical spectral parameters). Such preconfiguration lowers the entry barrier for new users while still allowing full parameter customization for advanced users.

Data verification and preview. To ensure transparency and trust in the data handling process, the platform integrates preview functionalities. Users can visualize selected datasets before launching analyses, either as tabular previews or quick-look graphs. This step allows verification of variable consistency, detection of missing values, and confirmation of the selected time window.

Interactive plotting options further support exploratory checking: zooming, curve hiding/showing, and tooltip value display facilitate rapid inspection prior to deeper analysis.

Data export and interoperability. The platform provides download options for both processed datasets and generated outputs (e.g., aggregated statistics tables). This ensures interoperability with external tools and allows users to integrate results into reports, publications, or complementary analyses.

Pedagogical support and onboarding. A dedicated homepage centralizes explanatory materials, including:

- Short tutorial videos,
- Conceptual diagrams explaining the workflow
- Example use cases.

This pedagogical layer is particularly important in heritage conservation contexts, where users may not have advanced training in data science or signal processing.

Visual coherence and usability. A consistent graphic charter has been implemented, with a distinct color identity for each analysis tab (visualization, statistics, advanced analysis, future projections). This visual coding reinforces cognitive orientation within the platform and improves usability during iterative exploration.

Clear instructional messages, contextual help texts, and example placeholders are integrated throughout the interface. Feedback messages inform users about successful operations, missing inputs, or potential inconsistencies. Such interface feedback mechanisms contribute to robustness, transparency, and ease of use.

A.4.1 Multi-graph workspace and customization

The visualization tab includes multiple graph panels, allowing users to display and compare several plots simultaneously. This multi-view configuration facilitates cross-comparison between raw and aggregated data, between different sensors, or between indoor and outdoor conditions.

Titles and legends are automatically generated based on selected variables and metadata but remain fully customizable by the user. This ensures clarity for both exploratory analysis and exportable figures intended for reporting or publication.

B. Preprocessing

B.0.1 Derived psychrometric variables

When air temperature T (in °C) and relative humidity RH (in %) are available, additional psychrometric variables are computed following standard formulations [1, 2].

Saturation vapor pressure. The saturation vapor pressure over liquid water is computed using the Magnus–Tetens approximation [1]:

$$e_s(T) = 6.112 \exp\left(\frac{17.62\,T}{243.12 + T}\right) \tag{1}$$

where $e_s(T)$ is expressed in hPa and T in °C.

Actual vapor pressure. The partial pressure of water vapor is:

$$e = \frac{RH}{100}\, e_s(T) \tag{2}$$

where e is in hPa.

Dew point temperature. Defining

$$\gamma(T, RH) = \ln\left(\frac{RH}{100}\right) + \frac{17.62\,T}{243.12 + T}, \tag{3}$$

the dew point temperature T_d (°C) is obtained as:

$$T_d = \frac{243.12\,\gamma(T, RH)}{17.62 - \gamma(T, RH)}. \tag{4}$$

Mixing ratio. The mixing ratio w ($\mathrm{kg\,kg^{-1}}$) is computed from:

$$w = 0.622\,\frac{e}{P - e} \tag{5}$$

where P is the atmospheric pressure (hPa) and 0.622 is the ratio of the specific gas constants of water vapor and dry air (R_v/R_d). If pressure measurements are unavailable, a standard value of $P = 1013.25$ hPa may be assumed with caution for elevated sites.

Absolute humidity. Absolute humidity ρ_v ($\mathrm{g\,m^{-3}}$) is derived from the ideal gas law:

$$\rho_v = \frac{100\,e}{R_v\,(T + 273.15)} \tag{6}$$

where $R_v = 461.5\ \mathrm{J\,kg^{-1}\,K^{-1}}$ is the specific gas constant for water vapor, e is in hPa, and T in °C (the factor 100 converts hPa to Pa).

C. Analysis modules: computation methods

C.1. Statistic module: General statistical indicators

For each selected variable and sensor, classical descriptive statistics are calculated over the indicated period :

- **Mean** Average value over the selected period:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- **Standard deviation** Measure of dispersion around the mean:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2}$$

- **Minimum and maximum** Extreme values observed during the monitoring period:

$$x_{\min}, \quad x_{\max}$$

- **Amplitude (Delta)** Total range of variation:

$$\Delta = x_{\max} - x_{\min}$$

C.2. Advanced analysis module: Correlation map

For two time series X and Y, the Pearson correlation coefficient is defined as [5]:

$$r_{XY} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{7}$$

where $\bar{x}$ and $\bar{y}$ are the mean values and σ_X, σ_Y the standard deviations.

Method. The correlation matrix is computed using the Pearson method as implemented in the `pandas` statistical framework.

Interpretation and interest. Correlation matrices are particularly useful to:

- Assess indoor–outdoor climatic coupling,
- Identify redundant sensors,
- Detect anomalous behaviour,

However, correlations are sensitive to shared seasonal cycles and long-term trends. Preprocessing (detrending or seasonal adjustment) may be required to isolate physically meaningful relationships.

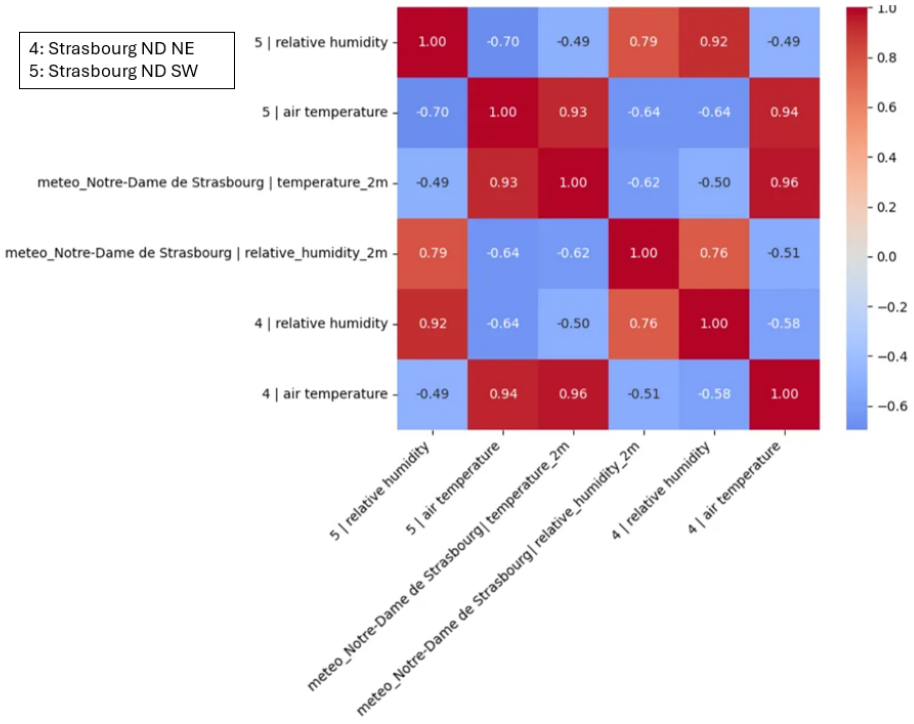

Figure S1: Correlation map

Case study

The correlation matrix highlights strong positive correlations between air temperature measurements from both sensors and the external meteorological data (up to 0.96), indicating good consistency and reliability of the sensors. Conversely, relative humidity is negatively correlated with temperature (around -0.6 to -0.7), reflecting expected physical relationships. High positive correlations between humidity sensors (up to 0.92) also suggest coherent moisture dynamics across orientations, although some variability remains due to microclimatic effects.

C.3. Advanced analysis module: Transfer Fourier Analysis

If $X(f)$ and $Y(f)$ denote the Fourier transforms of the reference and target signals, the transfer function $H(f)$ is defined as:

$$H(f) = \frac{Y(f)}{X(f)} \tag{8}$$

Spectral quantities and interpretation. Spectral quantities are estimated using Welch's averaged periodogram method [16], implemented in the `scipy.signal` spectral analysis module.

- **Power Spectral Density (PSD),** $P_{xx}(f)$ The PSD represents the distribution of signal variance across frequencies [3]:

$$P_{xx}(f) = \frac{1}{T}|X(f)|^2 \tag{9}$$

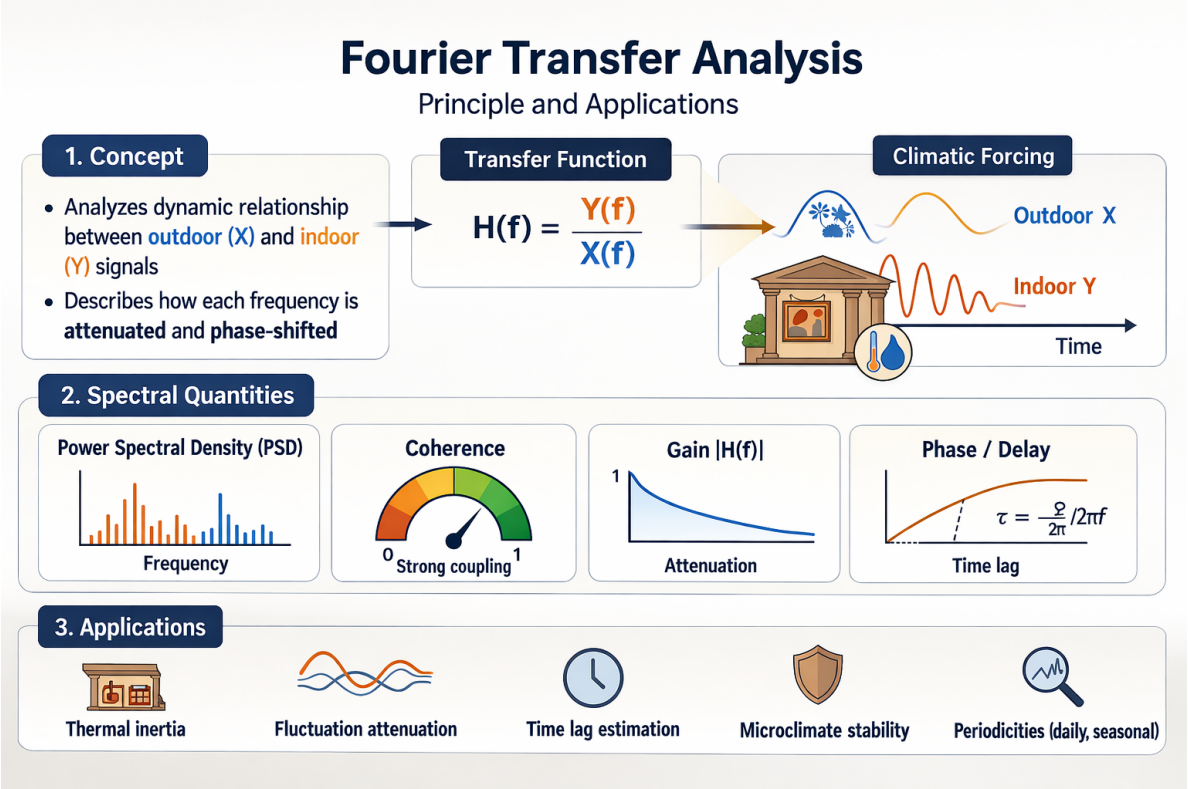

Figure S2: Transfer Fourier Analysis principle (generated by IA)

It indicates which periodicities (daily, seasonal, multi-day) dominate the variability of a climatic variable.

- **Cross Spectral Density (CSD),** $P_{xy}(f)$

$$P_{xy}(f) = \frac{1}{T} X(f) Y^*(f) \tag{10}$$

The CSD quantifies how much two signals co-vary at each frequency and forms the basis for gain, phase and coherence calculations.

- **Coherence,** $C_{xy}(f)$

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f) P_{yy}(f)} \tag{11}$$

Coherence ranges from 0 to 1 and measures the strength of linear coupling between the two signals at a given frequency. High coherence indicates reliable transfer function estimation.

- **Gain,** $G(f)$

$$G(f) = |H(f)| \tag{12}$$

The gain expresses the attenuation or amplification factor between the reference and the target signal at frequency f. In buildings, gain below 1 indicates damping of external fluctuations.

- **Phase,** $\phi(f)$

$$\phi(f) = \arg(H(f)) \tag{13}$$

The phase quantifies the phase shift between signals. A negative phase typically indicates a delayed indoor response relative to outdoor forcing.

- **Time delay,** $\tau(f)$

$$\tau(f) = -\frac{\phi(f)}{2\pi f} \tag{14}$$

The delay converts phase shift into a physically interpretable time lag (e.g., hours of thermal inertia).

- **Thermal and Hygrometric Indicators**

Building performance indicators are derived from the estimated transfer function by focusing on the diurnal excitation, which represents the dominant climatic forcing for both temperature and relative humidity in most free-running or weakly conditioned buildings. The central diurnal frequency is defined as

$$f_{day} = \frac{1}{24 * 3600} \tag{15}$$

and a tolerance band of $\pm 20\%$ around this frequency is considered to account for spectral resolution and leakage effects. Only spectral components within this band and exhibiting a coherence value above a predefined threshold are retained to ensure statistical robustness and physical interpretability.

The **thermal attenuation coefficient** is defined as the mean gain within the coherent diurnal band :

$$\overline{G}_{day} = \frac{1}{N} \sum_{f \in \mathcal{B}} G(f) \tag{16}$$

It represents the amplitude ratio at the daily frequency. The thermal attenuation coefficient and time lag are analogous to the decrement factor and time shift defined in ISO 13786, but derived here from in-situ monitoring under real climatic excitation [8]. For example, when comparing exterior and interior signals, values lower than unity indicate damping of external fluctuations due to conduction resistance and heat storage effects within the building envelope and internal mass.

For hygrometric analysis, a **moisture buffering index** is defined as the inverse of the mean gain:

$$BI = \frac{1}{\overline{G}_{day}} \tag{17}$$

This index can quantify the ability of indoor air and materials to attenuate outdoor humidity variations. High buffering values indicate strong sorption capacity and effective moisture storage, analogous to thermal inertia but governed by vapor diffusion and hygroscopic phenomena. The proposed hygrometric buffering index is conceptually related to the Moisture Buffer Value (ISO 24353) [13] [14], but expressed in the frequency domain to characterize dynamic response under natural climatic forcing.

Finally, the mean coherence within the diurnal band is reported as a reliability metric. High coherence indicates that fluctuations are predominantly driven by forcing at the considered frequency, supporting interpretation of gain and delay as meaningful transfer properties. Low coherence suggests significant internal loads, occupancy effects, or

nonlinear processes, limiting strict transfer function interpretation.

$$\overline{C}_{day} = \frac{1}{M} \sum_{f \in [0.8 f_{day}, 1.2 f_{day}]} C_{xy}(f) \tag{18}$$

Quantity	**PSD**	**CSD**	**Coherence**	**Gain**	**Phase**	**Delay**
Formula	$P_{xx}(f) = \frac{1}{T}\lvert X(f)\rvert^2$	$P_{xy}(f) = \frac{1}{T}X(f)Y^*(f)$	$C_{xy}(f) = \frac{\lvert P_{xy}(f)\rvert^2}{P_{xx}(f)P_{yy}(f)}$	$G(f) = \lvert H(f)\rvert$	$\phi(f) = \arg(H(f))$	$\tau(f) = -\frac{\phi(f)}{2\pi f}$
Description	Variance distribution across frequencies	Covariance between signals at each frequency	Strength of linear coupling (0–1)	Attenuation or amplification factor	Phase shift between input and response	Time lag derived from phase

Table S1: Main spectral quantities used in Fourier transfer analysis.

Interpretation and limitations. Fourier transfer analysis is particularly relevant for quantifying building thermal inertia, assessing damping of outdoor fluctuations, or estimating phase lag between forcing and response. However, interpretation requires caution: the method assumes approximate linearity and stationarity over the analysed period. Low coherence values invalidate gain and delay estimation at corresponding frequencies. Non-linear processes may limit strict transfer function interpretation.

Preprocessing. Prior to spectral estimation, signals may undergo:

- Mean removal,
- Detrending (linear, polynomial, Savitzky–Golay, or LOESS),
- Removal of daily or annual seasonal components.

These steps reduce non-stationarity and spectral leakage.

C.4. Future projection module: different climate projection models

Climatic models originate from major international climate modeling centers and were developed within the CMIP6 and HighResMIP initiatives. Differences in spatial resolution, atmospheric dynamics, and parameterization schemes result in variability between projections. Considering several models allows exploration of inter-model variability and provides a first-order estimate of structural uncertainty when interpreting future climate impacts on heritage materials and structures [9]. The selected models for the OPHELIA module are :

- CMCC_CM2_VHR4: very high-resolution coupled climate model developed by the Euro-Mediterranean Center on Climate Change [4].
- FGOALS_f3_H: high-resolution coupled model developed by the Chinese Academy of Sciences within CMIP6 HighResMIP [6].
- HiRAM_SIT_HR: high-resolution atmospheric model derived from the GFDL HiRAM configuration [12].

- MRI_AGCM3_2_S: atmospheric general circulation model developed by the Meteorological Research Institute (Japan) [11].
- EC_Earth3P_HR: high-resolution configuration of the EC-Earth coupled climate model [7].
- MPI_ESM1_2_XR: extra-high-resolution version of the Max Planck Institute Earth System Model [10].
- NICAM16_8S: non-hydrostatic atmospheric global climate model based on an icosahedral grid [15].

Figure S3 represents the comparison of models for the evolution of relative humidity in Bibracte between 1950 and 2050.

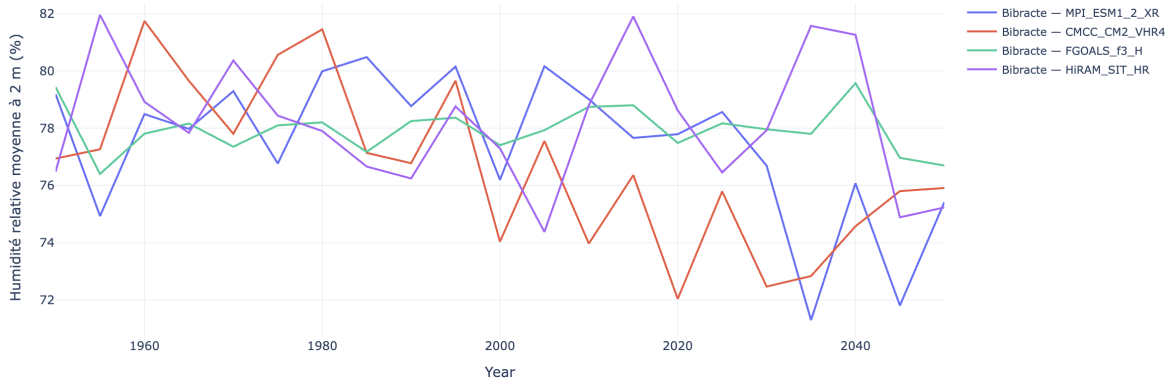

Figure S3: Comparison of climatic models for the evolution of relative humidity at 2m heigt in Bibracte between 1950 and 20250.

References

[1] Oleg A Alduchov and Robert E Eskridge. Improved magnus form approximation of saturation vapor pressure. *Journal of Applied Meteorology (1988-2005)*, pages 601–609, 1996.

[2] ASHRAE. *ASHRAE Handbook – HVAC Applications: Museums, Galleries, Archives and Libraries*. American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2011.

[3] Julius S. Bendat and Allan G. Piersol. *Random Data: Analysis and Measurement Procedures*. Wiley, Hoboken, NJ, 4 edition, 2010.

[4] Centro Euro-Mediterraneo sui Cambiamenti Climatici. Cmcc-cm2 high resolution climate model, 2020.

[5] Christopher Chatfield. *The Analysis of Time Series: An Introduction*. Chapman & Hall/CRC, Boca Raton, 6 edition, 2003.

[6] Chinese Academy of Sciences. Fgoals-f3 climate model description, 2020.

[7] EC-Earth Consortium. Ec-earth3 model description. *Geoscientific Model Development*, 2020.

[8] International Organization for Standardization. Iso 13786:2017 – thermal performance of building components – dynamic thermal characteristics – calculation methods, 2017.

[9] IPCC. Ipcc sixth assessment report, working group i. *Cambridge University Press*, 2021.

[10] Max Planck Institute for Meteorology. Mpi-esm1.2 earth system model, 2019.

[11] Meteorological Research Institute Japan. Mri-agcm3 atmospheric general circulation model, 2017.

[12] NOAA GFDL. Gfdl hiram high resolution atmospheric model, 2019.

[13] Andrés José Prieto Ibáñez, Juan Manuel Macías Bernal, María José Chávez de Diego, and Francisco Javier Alejandre Sánchez. Expert system for predicting buildings service life under iso 31000 standard. application in architectural heritage. *Journal of Cultural Heritage*, 18:209–218, March 2016.

[14] Carsten Rode, Ruut H. Peuhkuri, L. H. Mortensen, K. K. Hansen, B. Time, K. Svennberg, J. Arfvidsson, L. E. Harderup, J. Holme, and T. Ojanen. Moisture buffer value of building materials. *Journal of Thermal Envelope and Building Science*, 29(1):31–46, 2005.

[15] Satoh et al. Nonhydrostatic icosahedral atmospheric model (nicam). *Journal of the Meteorological Society of Japan*, 2014.

[16] Peter D. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.