

## Dataset Metadata for Universal Complex Systems Predictive Theory (UCSPT)

This document describes the datasets used in the UCSPT submission to *Nature*, ensuring reproducibility and transparency as per *Nature*'s guidelines. All datasets are available in the public, read-only repository at [https://github.com/anonymous/UCSPT\\_GPA\\_Datasets](https://github.com/anonymous/UCSPT_GPA_Datasets). To comply with double anonymized review, author information has been removed from all reviewer-accessible files. Contact for data access is via the *Nature* submission system.

### 1. UCSPT\_GHA.csv

- **Description:** Historical global atmospheric data (1500–2025) for CO<sub>2</sub> concentration, temperature, population, and internet access.
- **Rows:** 526
- **Columns:** Year, CO<sub>2</sub> (ppm), Temp (°C), Population\_Billion, Internet\_Access (%), Source
- **Key Validation:** CO<sub>2</sub> for 2025  $\approx$  430.5 ppm ( $\pm 0.1$  ppm),  $R^2 \geq 0.999870$ , PCC  $\geq 0.965$
- **Source:** NOAA, IPCC, UN, World Bank, UCSPT General Predictive Algorithm (GPA) outputs

### 2. UCSPT\_FSA.csv

- **Description:** Future scenario analysis data (2025–2195) for CO<sub>2</sub>, emissions, and socioeconomic metrics.
- **Rows:** 171
- **Columns:** Year, CO<sub>2</sub> (ppm), Emissions (GtCO<sub>2</sub>), Education\_Access (%), Connectivity (%), Healthcare\_Access (%), Displacement (millions), Source
- **Key Validation:** 2030 CO<sub>2</sub>  $\approx$  425.0 ppm ( $\pm 0.3$  ppm), 2200 CO<sub>2</sub>  $\approx$  500.0 ppm ( $\pm 0.3$  ppm), 2030 Emissions  $\approx$  38.0 GtCO<sub>2</sub> ( $\pm 0.3$  GtCO<sub>2</sub>), 2200 Emissions  $\approx$  12.0 GtCO<sub>2</sub> ( $\pm 0.3$  GtCO<sub>2</sub>), CO<sub>2</sub>  $R^2 \geq 0.999978$ , PCC  $\geq 0.902$ ; Emissions  $R^2 \geq 0.999978$ , PCC  $\leq -0.999$

- **Source:** Friedlingstein [7], IPCC, UN, UCSPT GPA outputs

### 3. UCSPT\_GSS\_Sample.csv

- **Description:** Socioeconomic data (2020–2120) for education access.
- **Rows:** 101
- **Columns:** Year, Education\_Access (%)
- **Key Validation:** Education\_Access  $\approx$  50.0–80.0%,  $R^2 \geq 0.95$ , PCC  $\geq 0.85$
- **Source:** World Bank [17], Ostrom [12], UCSPT GPA outputs

### 4. UCSPT\_Tox21\_Sample.csv

- **Description:** Toxicological data (2020–2120) for chemical exposure levels.
- **Rows:** 101
- **Columns:** Year, Toxicity (arbitrary units)
- **Key Validation:** Toxicity  $\approx$  10.0–30.0,  $R^2 \geq 0.95$ , PCC  $\geq 0.85$
- **Source:** Allen [10], UCSPT GPA outputs

### 5. UCSPT\_1000Genomes\_Sample.csv

- **Description:** Genomic data (2020–2120) for allele frequency trends.
- **Rows:** 101
- **Columns:** Year, Allele\_Frequency
- **Key Validation:** Allele\_Frequency  $\approx$  0.1–0.3,  $R^2 \geq 0.95$ , PCC  $\geq 0.85$
- **Source:** 1000Genomes [16], UCSPT GPA outputs

## 6. UCSPT\_EarthBioGenome\_Sample.csv

- **Description:** Ecological data (2020–2120) for biodiversity trends.
- **Rows:** 101
- **Columns:** Year, Biodiversity\_Index
- **Key Validation:** Biodiversity\_Index  $\approx$  60.0–85.0,  $R^2 \geq 0.92$ , PCC  $\geq 0.85$
- **Note:** The  $R^2$  threshold is  $\geq 0.92$  due to higher variability in ecological data, consistent with Earth BioGenome [22].
- **Source:** Earth BioGenome [22], UCSPT GPA outputs

**Validation:** All datasets were validated using UCSPT\_Script.R (R v4.4+), producing  $R^2$  and PCC metrics aligning with manuscript claims. See UCSPT\_Validation\_Report.txt for details.