# Shades of Green: Unveiling the Impact of Municipal Green Bonds on the Environment

Marta Campi [1], Gareth W. Peters[2], Kylie-Anne Richards[3]

[1] Institut Pasteur, Universite Paris Cite, Inserm, Hearing Institute

[2]Department of Statistics and Applied Probability, University of California Santa Barbara

[3]Finance Discipline Group, UTS Business School, University of Technology Sydney

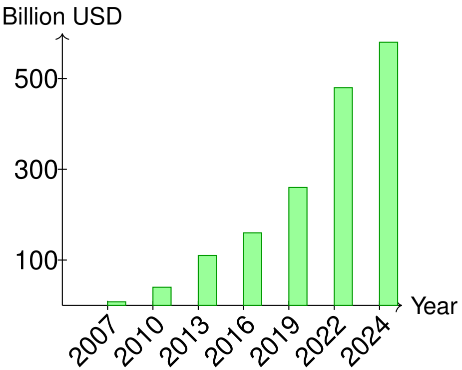
UTS, 15 May 2025

# Outline

# Outline

# Green Bonds: A Financial Response to Climate Change

- **Global climate challenges** require innovative financial solutions
- **Paris Agreement (COP21, 2015)** established need for global carbon market and transitions
- **Green bonds** have emerged as key financial instruments to:
  - Fund environmental initiatives through debt capital markets
  - Provide specific financing at reduced cost
- **First green bond** issued by European Investment Bank (EIB) in 2007
- **The Green Bond Principles (GBP)** provide guidelines for issuing green bonds, ensuring transparency and environmental integrity.
- **Growing investor demand for ESG** investments is fueling the expansion of the green bond market.




Global Green Bond Market Growth
Source: Climate Bonds Initiative, 2024

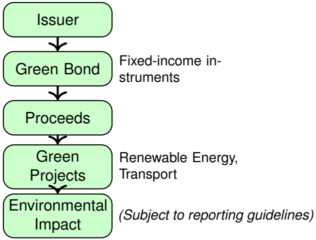
# The Green Bond Market

## Key Milestones

- 2007: EIB issues first Climate Awareness Bond
- 2008: World Bank issues first institutional green bond
- 2014: Green Bond Principles (GBP) introduced with four core components
  - Use of proceeds
  - Project evaluation and selection
  - Management of proceeds
  - Reporting
- 2021: GBP updated, 2022 appendix added

## Current Challenges

- Lack of detailed specificity about environmental initiatives
- Limited standardization in impact reporting *(inconsistent across issuers and regions)*
- Difficult to assess carbon reduction potential
- Limited monitoring frameworks
- **Risk of greenwashing** *due to lack of standardization and metrics*
- **Long timeframes:** *Environmental benefits often take decades to materialize*




Source: Green Bond Principles, ICMA

**Research Gap:** Limited quantitative assessment of green bond environmental impacts

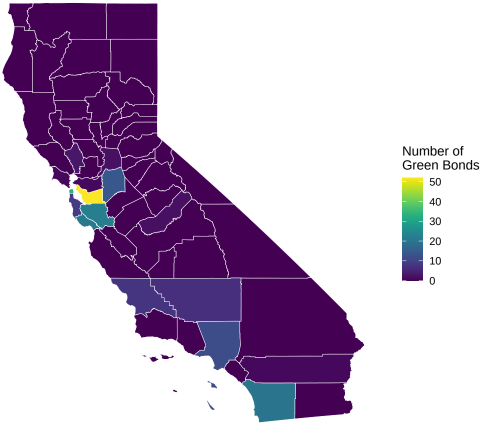
**Why U.S. Municipal Green Bonds?**

- Substantial U.S. carbon footprint *and significant potential for impact through local initiatives.*
- Transparent financial disclosure reporting
- Market size exceeding hundreds of billions *representing a significant pool of green financing.*
- Critical funding for public projects *with direct environmental consequences at the community level.*

**Why California?**

- Largest state issuer of municipal bonds
- Rich dataset of green bond issuance
- Severe pollution and climate impacts
- Dense monitoring station network
- A leader in environmental policy and sustainability initiatives

**Number of Green Bonds Issued by California Counties**




Source: Municipal Green Bond Data, 2015−2020

# Research Questions and Objectives

**Main Research Question:**

*How can we quantitatively assess the environmental impact of green bond use of proceeds?*

**Primary Objectives:**

- **Develop a methodological framework** for assessing the **environmental impact of green bonds**
- **Contribute to standardization and transparency** in green bond impact assessment

**Technical Objectives (to achieve the primary goals):**

- Identify **spatial-temporal relationships** between green bond attributes and pollution/climate patterns
- Quantify **correlations between green bond issuance and environmental outcomes**

# Outline

# Data and Methodology: Overview

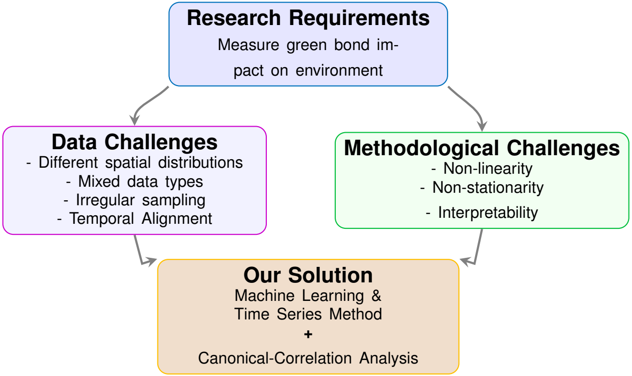
**Purpose of this section:**

- To explain how we obtained, prepared, and analyzed the data to answer our research question:

  *"How can we quantitatively assess the environmental impact of green bond use of proceeds?"*

**Key Stages:**

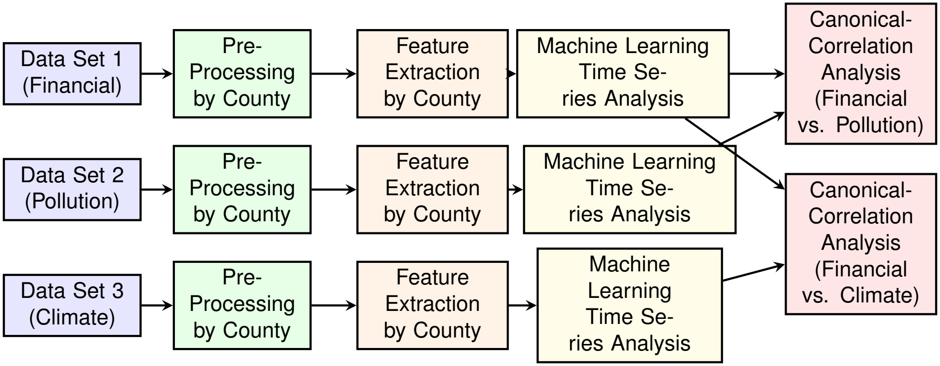
1. **Data Collection:** Gathering data from diverse sources
2. **Feature Engineering:** Cleaning and transforming the data
3. **Methodological Framework:** Developing a multivariate spatial-temporal method that can handle complex relationships and non-linear interactions in time series data




**Our approach addresses the challenges of integrating complex data, as well as the methodological challenges of dealing with non-stationarity, non-linearity, and interpretability.**

# Data and Methodology: Our Approach

**In Practice: A Multivariate Spatial-Temporal Analysis Framework**



# Why Multivariate Spatial-Temporal Analysis?

**Traditional Methods:**

- Linear regression models
- Single-variable time series
- Independent spatial analysis

**Limitations:**

- Cannot handle complex relationships across space and time simultaneously
- Struggle with integrating multiple data types
- Miss interactions between financial and environmental variables
- Poor performance with irregularly sampled data

**Our Multivariate Spatial-Temporal Approach:**

- **County-level analysis:** Captures local environmental impacts of green bonds
- **Canonical-correlation framework:** Reveals relationships between financial variables and environmental outcomes
- **Non-linear method:** Detects complex patterns that linear methods miss
- **Integrated analysis:** Combines financial, pollution, and climate data in a unified framework

**Key Advantage:** Our multivariate spatial-temporal approach allows us to quantify green bond environmental impacts across California counties by detecting non-linear relationships that traditional methods would miss.

# Data Sources and Collection

To assess environmental impacts of green bonds, we integrated **three diverse datasets**:

| Aspect | Pollution Data | Climate Data | Green Bond Data |
|---|---|---|---|
| **Source** | US EPA | NOAA (GSOD) | Bloomberg Terminal |
| **Variables** | $CO_2$, $NO_2$, AQI, $PM_{2.5}$ | Temp Min, Temp Max, Temp Mean, Precipitation | Size, Maturity, Coupon, etc. |
| **Period** | 2010-2020 | 2010-2020 | 2015-2020 |
| **Spatial Coverage** | 9 counties | 9 counties | 9 counties |
| **Collection Points** | 84 monitors | 51 stations | 167 bonds |
| **Data Type** | Time Series | Time Series | Mixed |
| **Data Characteristics** | Dynamic | Dynamic | Static |

We selected pollution and climate monitors based on their distance from the cities (50km radius from cities with population bigger than 250,000 inhabitants).

# Data Processing: From Raw Data to Features

## Pollution Data

**Data Cleaning:**

- 84 monitors selected within 50km of major cities
- Removal of monitors with $>$30% missing data
- Cubic spline for remaining gaps
- Daily alignment of time series

**Feature Extraction:**

- Daily spatial averages of $CO_2$, $NO_2$, $PM_{2.5}$, AQI
- Time series pre-whitening by year

## Climate Data

**Data Cleaning:**

- 51 weather stations within 50km of cities
- Removal of monitors with $>$30% missing data
- Cubic spline for missing values
- Daily alignment of time series

**Feature Extraction:**

- Spatial mean temperature averages
- Parkinson volatility (high/low temperature)
- Total precipitation aggregation

## Green Bond Data

**Data Cleaning:**

- 5.2M bonds screened via Bloomberg
- Filtered to CA municipal green bonds issued between 2015-2020
- Excluded bonds with optionality
- Amount Outstanding $>$ \$10 million, Rating AAA-BBB
- Final selection: 167 bonds

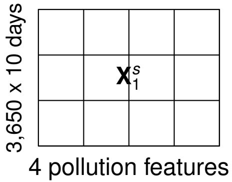
**Feature Extraction:**

- One-hot encoding for categorical variables
- Day-count transformation for date variables
- Combined Jaccard + RBF kernel for mixed data

# Data Structure: Input Matrices for kPCA

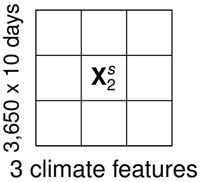
## Pollution Data Matrix

- **Structure:** Time series by county
- **Rows:** Days (2010-2020)
- **Columns:** averaged $CO_2$, $NO_2$, AQI, $PM_{2.5}$
- **Dimensions:** $(3{,}650 \times 10) \times 4$



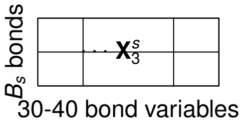

## Climate Data Matrix

- **Structure:** Time series by county
- **Rows:** Days (2010-2020)
- **Columns:** averaged Temp, Volatility, Total Precip
- **Dimensions:** $(3{,}650 \times 10) \times 3$



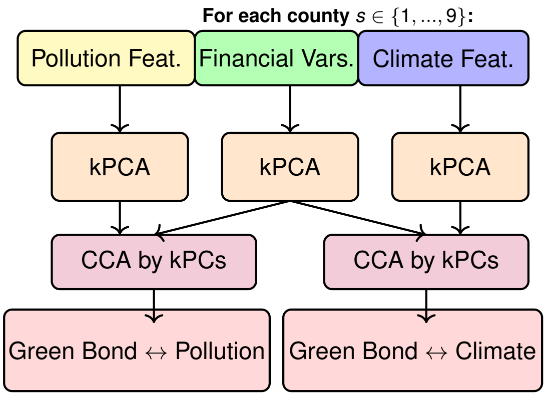

## Green Bond Data Matrix

- **Structure:** Bond-level data by county
- **Rows:** Individual bonds (varies)
- **Columns:** Numerical + categorical
- **Dimensions:** $B_s \times$ 30-40




- For each county $s \in \{1, \ldots, 9\}$, we apply kPCA separately to $\mathbf{X}_1^s$, $\mathbf{X}_2^s$, and $\mathbf{X}_3^s$
- Extract leading kernel principal components (kPCs) from each dataset
- These matrices become the input for our kPCA analysis

# Methodological Framework: kPCA-CCA Pipeline




**Key Implementation Details: Per-County**

**Analysis:**

- Analysis performed for nine California counties
- Each county processed separately
- kPCA applied to each dataset independently

**Kernel Functions:**

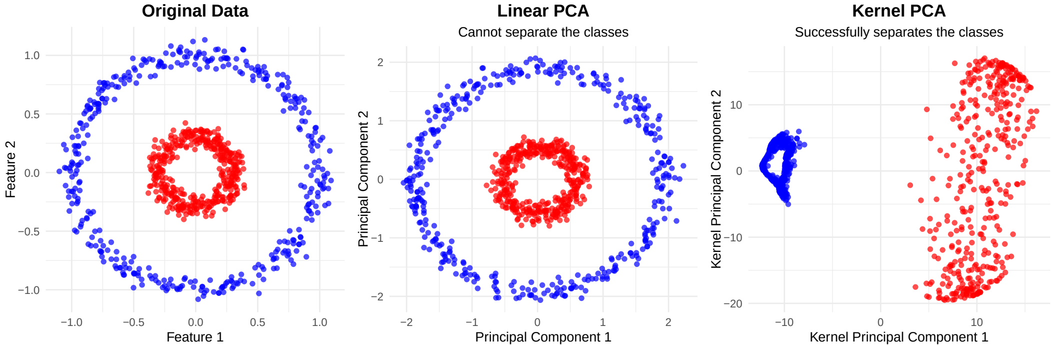
- For Financial Data:
  - RBF kernel for numerical variables
  - Jaccard kernel for categorical variables
- For Pollution & Climate Data:
  - RBF kernel for time series features

**Two Parallel CCA Analyses:**

- Financial kPCs $\leftrightarrow$ Pollution kPCs
- Financial kPCs $\leftrightarrow$ Climate kPCs

# Power of kPCA vs PCA: Visual Comparison

**Definition:** *Kernel PCA extends linear PCA to capture non-linear patterns by mapping data into a high-dimensional feature space.*

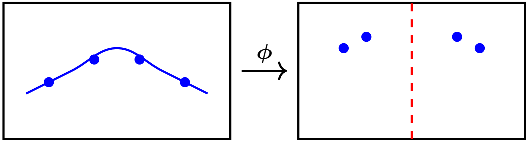



**Linear PCA:** Fails with non-linear data **Kernel PCA:** Captures complex patterns

**Key insight:** Our data contains non-linear relationships that only kPCA can effectively detect.

# Mathematical Formulation: Kernel PCA

## From PCA to kPCA:




Input Space $\mathcal{X}$      Feature Space $\mathcal{F}$

## PCA vs. kPCA Comparison:

| **Linear PCA** | **Kernel PCA** |
|---|---|
| $\mathbf{S}_X = \mathbf{X}^\top \mathbf{X}$ | $\mathbf{K} = \mathbf{\Phi}\mathbf{\Phi}^\top$ |
| $\mathbf{S}_X\mathbf{W} = \mathbf{W}\mathbf{\Lambda}$ | $\mathbf{K}\mathbf{Z} = \mathbf{Z}\mathbf{\Lambda}$ |
| $\mathbf{L} = \mathbf{X}\mathbf{W}$ | $\mathbf{A} = \mathbf{Z}\mathbf{\Lambda}^{1/2}$ |
| $\mathbf{\Lambda}$ contains variance along linear directions | $\mathbf{\Lambda}$ contains variance along non-linear manifolds |

## Solving the equations:

- In PCA: Solve using the explicit covariance matrix
- In kPCA: Solve without ever computing $\phi$ explicitly

## Mathematical Framework:

For points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^D$:

1. **Compute Gram matrix K** where $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$
2. **Eigendecomposition:** $\mathbf{K}\mathbf{Z} = \mathbf{Z}\mathbf{\Lambda}$
3. **Extract kPCs:** $\mathbf{A} = \mathbf{Z}\mathbf{\Lambda}^{1/2}$

## The "Kernel Trick":

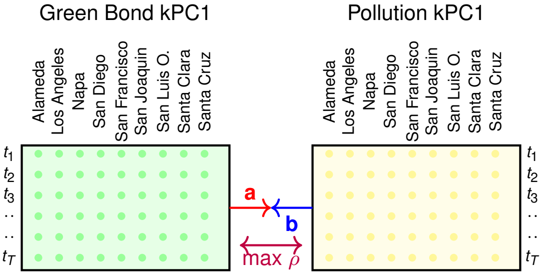
$$\begin{aligned} k(\mathbf{x}_i, \mathbf{x}_j) &= \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \\ &= \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j) \end{aligned}$$

## Our Custom Kernel Functions:

- **For numerical and time series variables:**
  $k^{(RBF)}(\mathbf{x}_i, \mathbf{x}_j) = \sigma^2 \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\gamma^2}\right)$
- **For categorical variables:**
  $k^{(J)}(A, B) = \frac{|A \cap B|}{|A \cup B|}$
- **Combined kernel for Green Data:**
  $k(\mathbf{x}_i, \mathbf{x}_j) = k^{(RBF)}(\mathbf{x}_i^{num}, \mathbf{x}_j^{num}) + k^{(J)}(\mathbf{x}_i^{cat}, \mathbf{x}_j^{cat})$

# Mathematical Formulation: CCA

**Definition:** CCA finds linear combinations of variables from two datasets that have maximum correlation with each other.




$U = \mathbf{a}^\intercal \mathbf{X}$ $\qquad$ $V = \mathbf{b}^\intercal \mathbf{Y}$

**CCA Objective:** Find vectors **a** and **b** to maximize

$$\rho(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^\intercal \Sigma_{XY} \mathbf{b}}{\sqrt{\mathbf{a}^\intercal \Sigma_{XX} \mathbf{a}} \sqrt{\mathbf{b}^\intercal \Sigma_{YY} \mathbf{b}}}$$

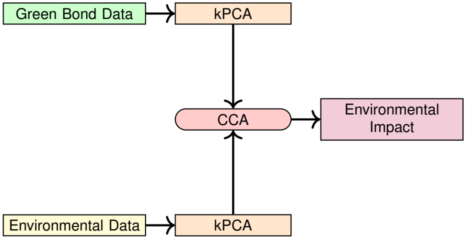
**Key Components:**

1. **Canonical variates:** $U_i = \mathbf{a}_i^\intercal \mathbf{X}$ and $V_i = \mathbf{b}_i^\intercal \mathbf{Y}$
   - Synthetic variables that maximize correlation
   - Each pair $(U_i, V_i)$ is uncorrelated with other pairs
2. **Assessment metrics:**
   - **Canonical correlation ($\rho_i^*$):** Strength of relationship
   - **Squared correlation ($\rho_i^{*^2}$):** Shared variance
   - **Structure coefficients:** $\mathrm{Corr}(X_j, U_i)$ and $\mathrm{Corr}(Y_j, V_i)$
     - Show how original variables contribute to canonical variates
     - Critical for interpreting results
   - **Redundancy index:** $(\sum_{j=1}^{d'} \mathrm{Corr}^2(Y_j, V_i)/d')\rho_j^{*^2}$
     - Measures explanatory power
3. **Statistical testing:**
   - Rao's F-approximation test for significance
   - Tests $H_0 : \rho_1 = \rho_2 = \ldots = \rho_k = 0$
   - Identifies which canonical functions are significant

# Methodological Innovation: kPCA-CCA

**Advantages of our kPCA-CCA approach:**

- Handles non-linear and non-stationary relationships
- Accommodates different data types
- Processes irregular sampling in time and space
- Maintains interpretability through two-stage process
- Detects robust correlations between multimodal datasets



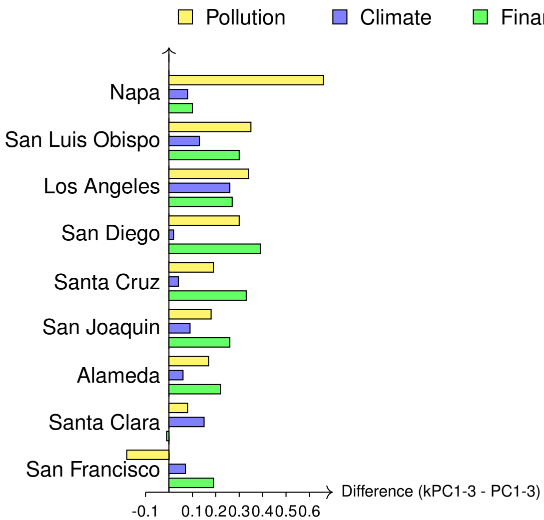

**Key innovations over existing methods:**

- Separate treatment of non-linearity (kPCA) and correlation (CCA)
- Integration of different kernel types for mixed-attribute data
- Out-of-sample evaluation on common spatial-temporal mesh
- Interpretability via contribution analysis

**Research Connection:** This methodology directly addresses our question of quantitatively assessing environmental impact of green bonds by detecting subtle non-linear relationships between financial instruments and environmental outcomes.

# Outline

# Comparative Performance of kPCA vs PCA




- **Performance Summary:**
  - 25 of 27 cases show positive differences
  - kPCA outperforms PCA by:
    - Pollution: avg. +24%
    - Climate: avg. +10%
    - Financial: avg. +23%
  - Greatest advantages in pollution data for Napa (+66%) and financial data for San Diego (+39%)
- **cKTA Formula:**

$$\hat{\rho}(\mathbf{K}_1, \mathbf{K}_2) = \frac{\langle \mathbf{K}_1^c, \mathbf{K}_2^c \rangle}{\sqrt{\langle \mathbf{K}_1^c, \mathbf{K}_1^c \rangle \langle \mathbf{K}_2^c, \mathbf{K}_2^c \rangle}} \in [-1, 1]$$

where $\mathbf{K}^c$ is the centered kernel matrix:

$$\mathbf{K}^c = \mathbf{K} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top \mathbf{K} - \frac{1}{N}\mathbf{K}\mathbf{1}\mathbf{1}^\top + \frac{1}{N^2}(\mathbf{1}^\top \mathbf{K}\mathbf{1})\mathbf{1}\mathbf{1}^\top$$


(a) kPC1-CV1 (b) kPC2-CV1

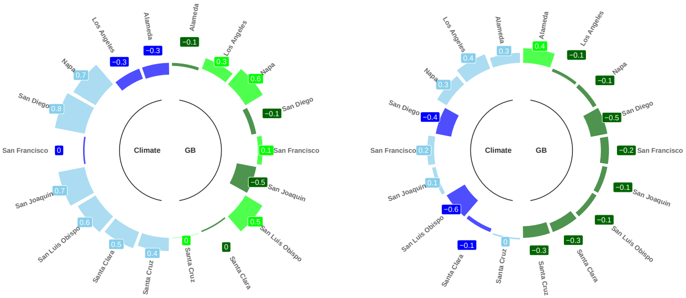
Figure: Canonical correlation plots showing county loadings

## Strong Correlations:

- **Using kPC1:**
  - First canonical correlation: 1.000
  - Second: 0.790
  - Squared: 0.999, 0.724
  - p-value $<$ 2.2e-16
- **Using kPC2:**
  - First canonical correlation: 1.000
  - Second: 0.881
  - Squared: 0.999, 0.776
  - p-value $<$ 2.2e-16

**Conclusion:** Clear association between green bond variables and pollution metrics

# kPCA-CCA Results: Financial vs. Climate Data




(a) kPC1-CV1 (b) kPC2-CV1

Figure: Canonical correlation plots showing county loadings

## Weaker Correlations:

- **Using kPC1:**
  - First canonical correlation ($\rho^*$): 0.815
  - Second correlation below 0.700 threshold
  - Squared canonical correlation: 0.712
  - Significant ($p < 2.2e\text{-}16$)
- **Using kPC2:**
  - No correlation exceeded 0.700 threshold
  - Maximum correlation ($\rho^*$): 0.664
  - Still statistically significant ($p < 2.2e\text{-}16$)

**Conclusion:** Detectable but weaker association between green bond variables and climate metrics, likely due to longer time frames for climate effects

# Variable Contribution Analysis

**Most influential categorical variables:**

- **San Diego:** Muni Source (bond issuer type)
- **Santa Clara:** Muni Issue Type (bond structure)
- **Santa Cruz:** Muni Source, Muni Offering Type
- Issuer Industry

**Most influential numerical variables:**

- Amount outstanding
- Maturity size
- Bid Option-Adjusted Spread (OAS)
- Dated date

**Interpretation:**

- **Municipal source matters:** Entities like transportation or energy authorities show stronger associations with pollution reduction, reflecting targeted project funding
- **Bond structure affects success:** Proper structuring influences market reception, potentially increasing funding for green initiatives
- **Sector classification matters:** Issuer's industry classification correlates with pollution metric associations, especially in counties like Santa Cruz

# Outline

# Key Findings

1. **Strong green bond-pollution relationship:**
   - Clear and interpretable correlation detected using non-linear methods
   - Significant association directly linked to green bond issuance amounts
   - Evidence that green bonds are having measurable impact on pollution reduction
2. **Weaker green bond-climate relationship:**
   - Detectable but less pronounced associations
   - Likely due to longer time frames needed for climate effects
   - Green bond market's relative youth (15-20 years) insufficient for climate impact detection
3. **Methodological insights:**
   - Non-linear methods (kPCA-CCA) significantly outperform linear approaches
   - Financial data particularly requires non-linear treatment
   - Different kPCs capture different variation frequencies

# Implications and Applications

**For Investors:**

- Improved transparency in assessing green bond environmental impacts
- Quantitative framework for evaluating potential investments
- Evidence supporting effectiveness of green bonds for pollution reduction

**For Issuers:**

- Insights on bond structures and attributes that maximize environmental impact
- Framework for impact reporting and monitoring
- Support for green bond market expansion

**For Regulators:**

- Statistical basis for standardizing environmental impact assessment
- Monitoring framework for evaluating market effectiveness
- Evidence to inform policy decisions on green finance

# Limitations and Future Research

**Limitations:**

- Limited time frame (particularly for climate effects)
- Geographic focus on California may not generalize globally
- Complex confounding factors challenging to fully resolve
- Early stage of green bond market development

**Future Research Directions:**

- Extend time frame as green bond market matures
- Expand geographic scope beyond California
- Incorporate additional environmental metrics
- Refine methodology to address specific green bond categories
- Develop predictive models for environmental impact
- Compare corporate vs. municipal green bonds

# Conclusions

- **Methodological contribution:** Novel kPCA-CCA framework for analyzing spatial-temporal relationships between financial and environmental data
- **Empirical contribution:** Quantitative evidence of green bond effectiveness in pollution reduction with measurable impacts
- **Market insight:** Climate effects likely require longer timeframes than current green bond market age (15-20 years)
- **Practical application:** Framework provides basis for monitoring, reporting, and decision-making in green bond market
- **Overall significance:** Our research reinforces the role of green bonds as integral financial tools for promoting environmental sustainability

**The green bond market is having a measurable positive environmental impact, particularly on pollution reduction, and offers significant potential for climate mitigation over longer timeframes.**

# Thank you for your attention!

GitHub Repository

Interactive Shiny App

# Questions?