LA-UR-25-23483

Functional Analysis of Variance (F-ANOVA): Tutorial

Grouping Data and Identifying Interactions Across Arbitrary Domains

Adam C Watts

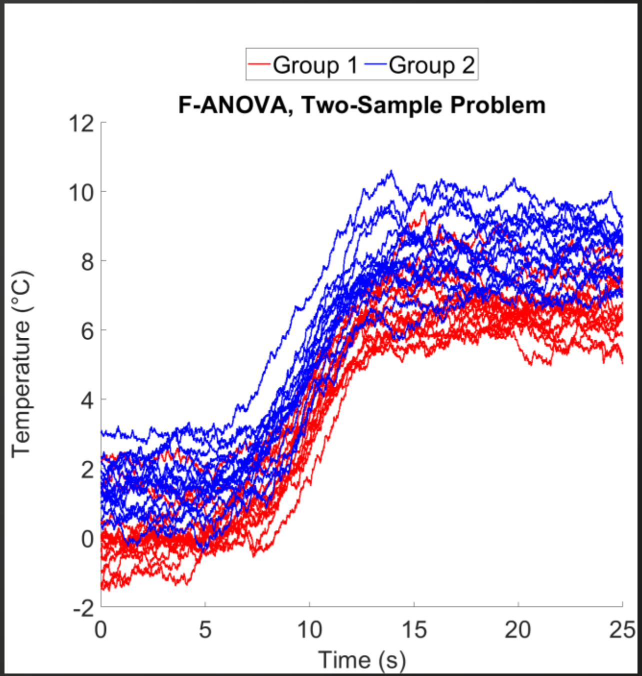
April 24, 2025

Managed by Triad National Security, LLC for the U.S. Department of Energy's NNSA

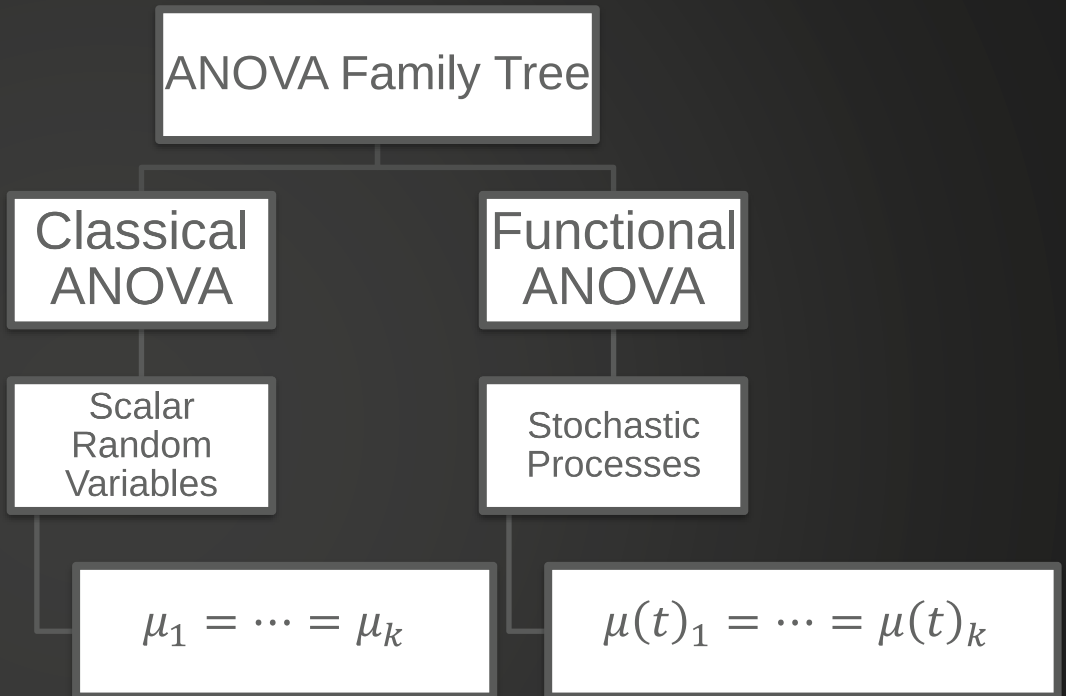
ANOVA and Stochastic Processes

- Classical ANOVA (Analysis of Variance) is used for testing whether there are significant differences among the means of a scalar random variable measured across multiple groups
  - Compares the variability between groups to the variability within groups.

- Many natural processes are *not* scalar random variables
  - They consist of a collection of random variables (stochastic process: $SP(\mu, \gamma)$)
  - Over a domain of interest: $\mathcal{T} = [t_0, t_f]$.
- Examples
  - Temperature measurements over time
  - Growth measurements over time
  - Shock response over frequencies



Functional-Analysis of Variance (F-ANOVA)



- Functional-ANOVA works on stochastic processes
- Global test, compared to a pointwise test
- For convenience, assume domain is over time, $t \in \mathcal{T}$
- F-ANOVA is used for testing whether there are significant differences among the mean functions of stochastic processes over some domain measured across multiple groups

F-ANOVA in Practice vs Theory

- F-ANOVA is based on functions, but all the samples we collect are vectors
- Numerical implementations of F-ANOVA are based on vectors
- The i^{th} functional samples are discretized accordingly:

$$y_i \overset{i.i.d}{\sim} \mathrm{SP}\big(\mu(t), \gamma(t,s)\big)$$

$$y_i(t) \;\rightarrow\; \boldsymbol{y}_i = [y_i(t_1), y_i(t_2), \dots, y_i(t_M)]$$

- Where $t_1, t_2, \dots, t_M$ are discretized time points in our domain
- In practice:
  - The mean functions, $\mu_1(t)$, are instead mean vectors, $\boldsymbol{\mu}_1$
  - The covariance function, $\gamma(t,s)$, is instead a covariance matrix, Σ

Data Preprocessing For F-ANOVA

1. Data is sampled over a common domain
   - Just consistent between all realizations (Uniformly or nonuniformly sampled)
2. Can be a stationary or a non-stationary stochastic process
3. Resolution of each sample is sufficiently long
   - Best practice is 1000 discretized elements per realization [1]
4. If the data doesn't meet these requirements, functionalize it using a reconstruction method.
   - Kernel Smoothing
   - Splines

[1] Zhang, J.-T. (2013). Analysis of Variance for Functional Data (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b15005

Trivial Example

- Suppose we collect room temperature data from 3 different rooms over the course of 25 hours.
- 15 samples collected from each room

$$\boldsymbol{y}_{1:15,\,A}(t), \qquad \boldsymbol{y}_{1:15,\,B}(t), \qquad \boldsymbol{y}_{1:15,\,C}(t)$$

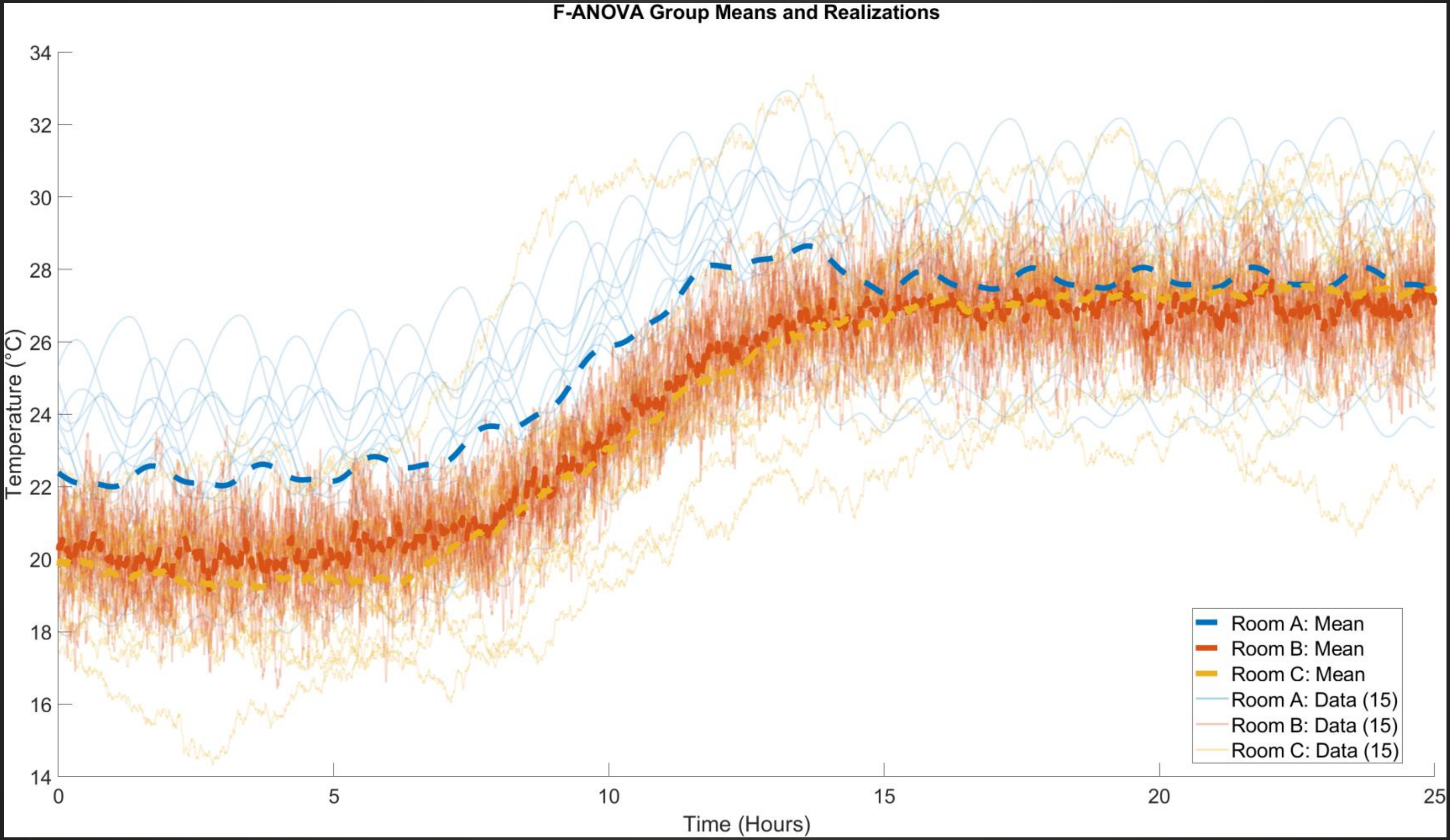
- 1 Factor (rooms) with 3 levels/groups (Rooms A, B, and C)
- Each group has a different covariance function
- Data is sampled uniformly, but more importantly, consistent across the domain

$$\Delta t = \frac{25}{3000}, \qquad \mathcal{T} = [0, 25]$$

- Each realization has $M = 3000$ elements.

Do the mean temperatures statistically differ across the three rooms? If so, which ones?

Example Data Plotted



How does F-ANOVA Work?

Central Limit Theorem of IID stochastic processes

- The test statistics throughout F-ANOVA are distributed as zero mean Gaussian processes: $\mathrm{GP}(0, \gamma)$
- Even when the Gaussian assumption of the data is not satisfied
  - The difference between the sample mean and actual mean function converges in distribution to a zero mean GP when N is large.

$$\sqrt{N}\{\bar{y}(t) - \mu(t)\} \xrightarrow{d} \mathrm{GP}(0, \gamma)$$

- From this, it can be shown that, the difference between group mean functions converges in distribution to a zero mean gaussian process under the null hypothesis

Why do we care?

How does F-ANOVA Work?

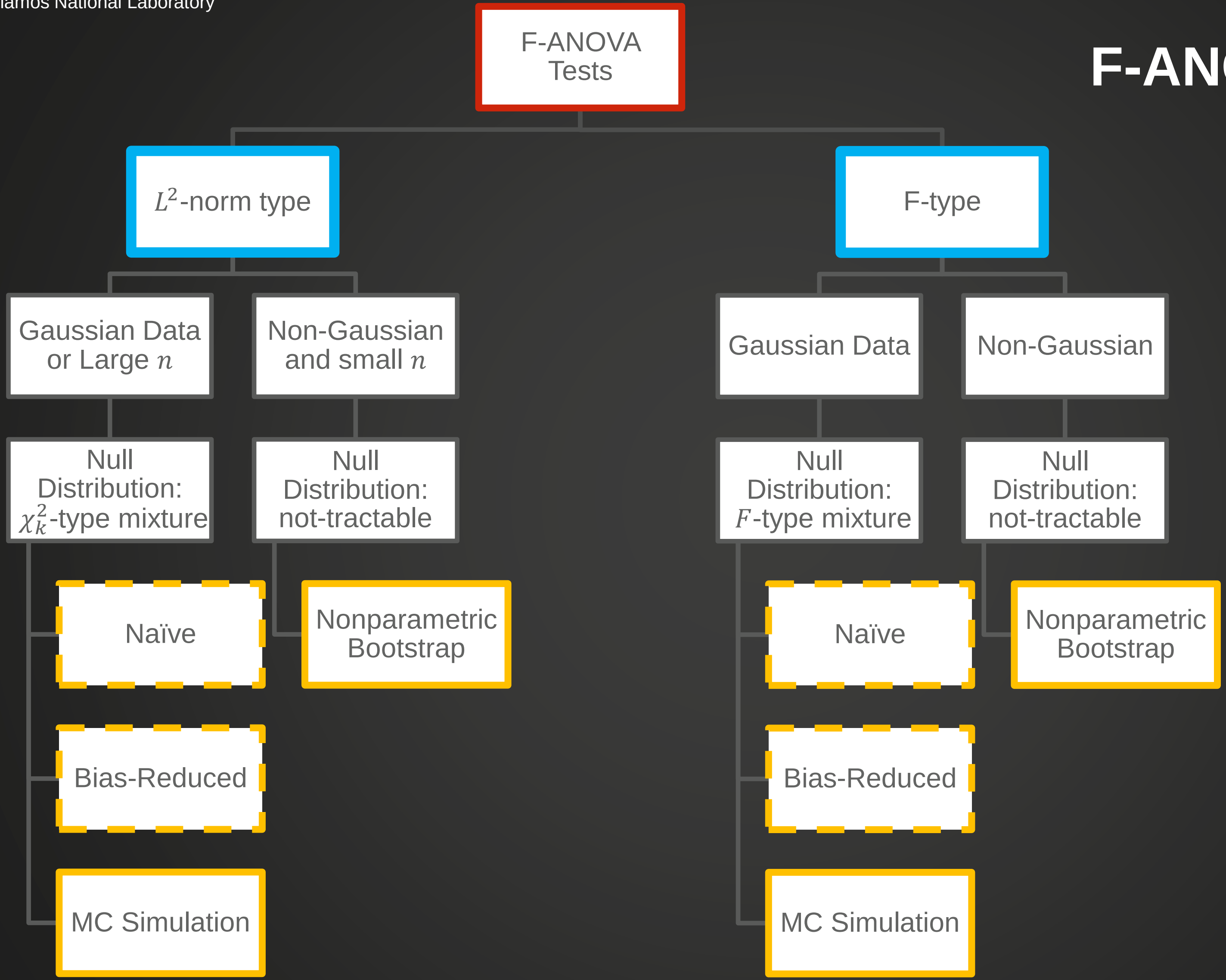
- The squared L^2-norm of a zero mean Gaussian process is a χ_k^2-type mixture:

$$\|y(t)\|^2 \overset{d}{=} \sum_{r=1}^{m} \lambda_r A_r$$

$$y(t) \sim \mathrm{GP}(0, \gamma) \text{ and } A_r \overset{i.i.d}{\sim} \chi_k^2$$

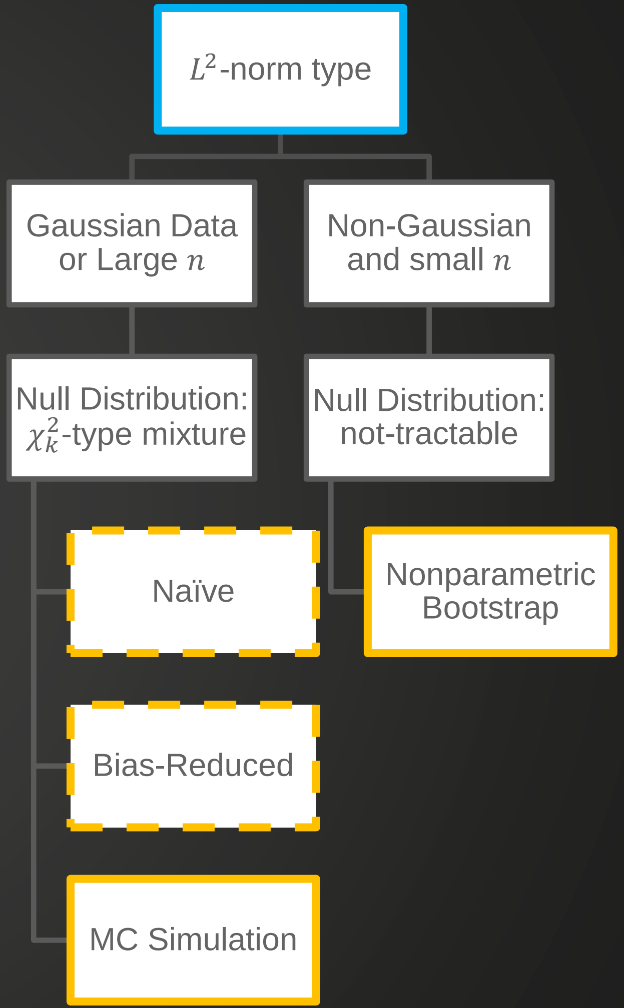
- All the F-ANOVA methods are based around this χ_k^2-type mixture distribution because its serves as our null distribution.
  - Approximations using naïve or biased-reduced cumulant matching
  - Direction simulation
- Compute a test statistic from the data and compare it to its null distribution to see how probable that statistic is under the null hypothesis.

F-ANOVA Tests



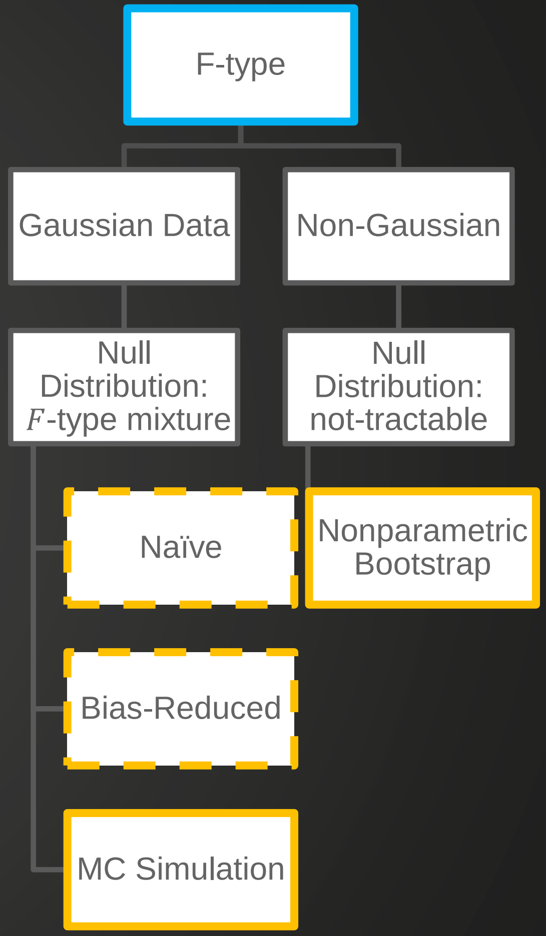
F-ANOVA Tests: L^2-norm type test

- Gaussian or large sample sizes:
  - Null distribution: χ^2_k-type mixture
    - Naïve and biased-reduced are approximations to the null distribution
    - Monte Carlo (MC) simulation of null distribution $(b = 10000)$
      - Estimate the PDF using a kernel density estimate (KDE)
- Non-Gaussian and small n
  - Bootstrap the critical values, resulting in a L^2-norm bootstrap test
  - Bootstrapping the data and calculating a bootstrap test statistic from each bootstrap resample
  - Estimate the $100(1-\alpha)$-percentile for the test statistic for the given data



F-ANOVA Tests: *F*-type test

- Gaussian Assumption
  - Partially takes the variations of $\hat{\gamma}(t,s)$ into account
  - Normalization factor to the L^2-norm type tests using the $\mathrm{trace}\left(\hat{\gamma}(t,s)\right)$
  - Scale invariant, multiply the data by a non-zero constant, doesn't change the test-statistic
  - Expected to outperform L^2-norm type test
  - Null distribution: F-type mixture
    - Either approximate or MC simulate this distribution
- Non-Gaussian
  - Bootstrap the critical values, resulting in a F-type bootstrap



What test should I use?

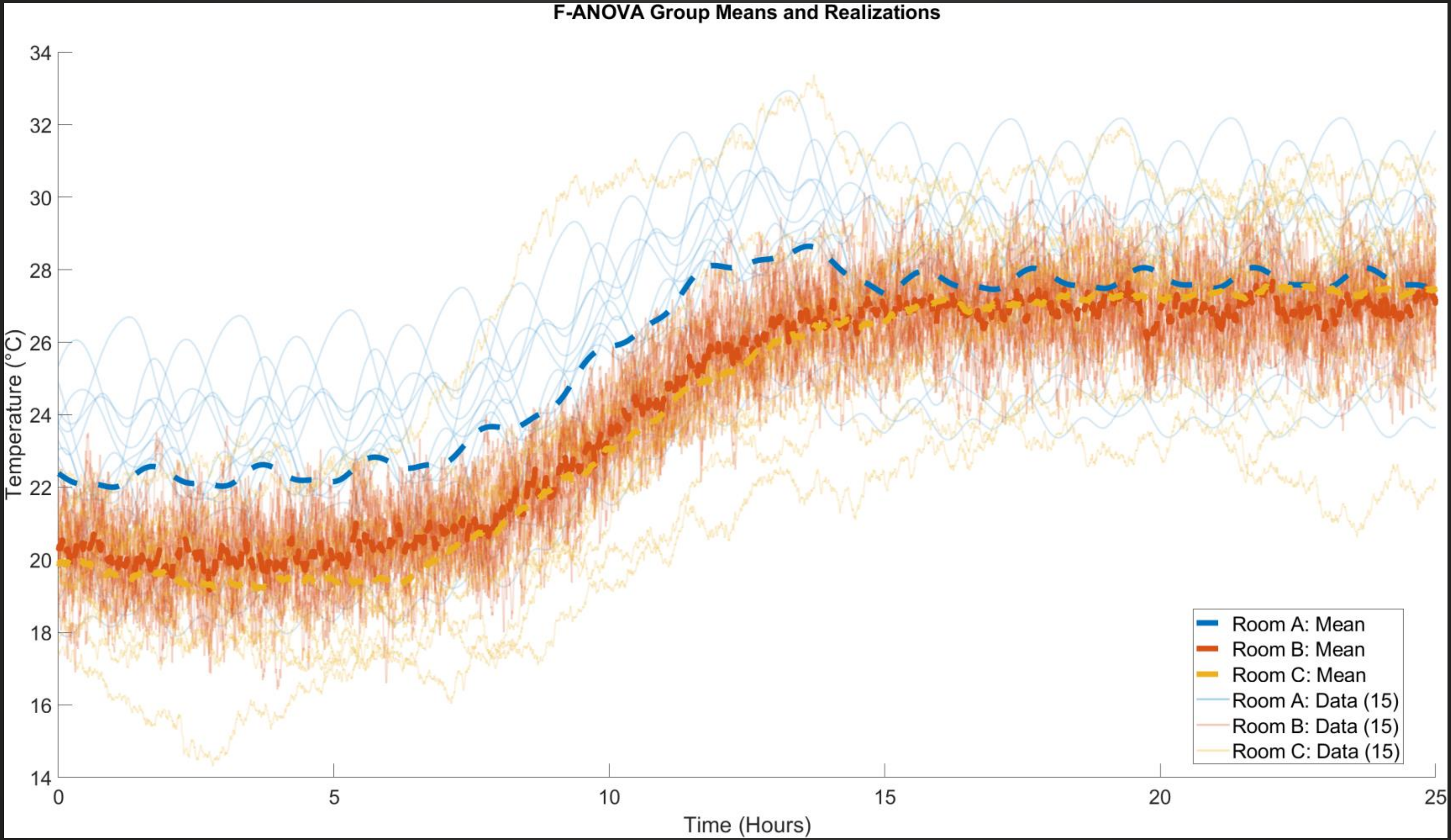
- The nonparametric $L^2\mathrm{-norm}$ and F-type bootstrap tests are preferred
  - Robust and require the least number of assumptions (such as being Gaussian)
  - They can still struggle with very low sample sizes, $n \lesssim 7$
- Cumulant matching and direction MC simulations are fast
  - Doesn't hurt to include them
  - Comparing p-values from various tests allows one to see what assumptions hold and which do not

n	Unique Bootstrap Resamples
4	35
5	126
6	462
7	1,716
8	6,435
9	24,310
10	92,378

Variants of F-ANOVA: Heteroscedastic

- The heteroscedastic F-ANOVA variant doesn’t pool the covariances
- Requires modifications to the $L^2 - \mathrm{norm}$ and F-type test null distributions or test statistics compared the ones used in the homoscedastic case
- Check equality of covariances to pick which F-ANOVA variant to use
  - Or, run both heteroscedastic and homoscedastic F-ANOVAs
- If homoscedastic:
  - Both variants will provide similar p-values.
  - The homoscedastic F-ANOVA variant is expected to have greater statistical power

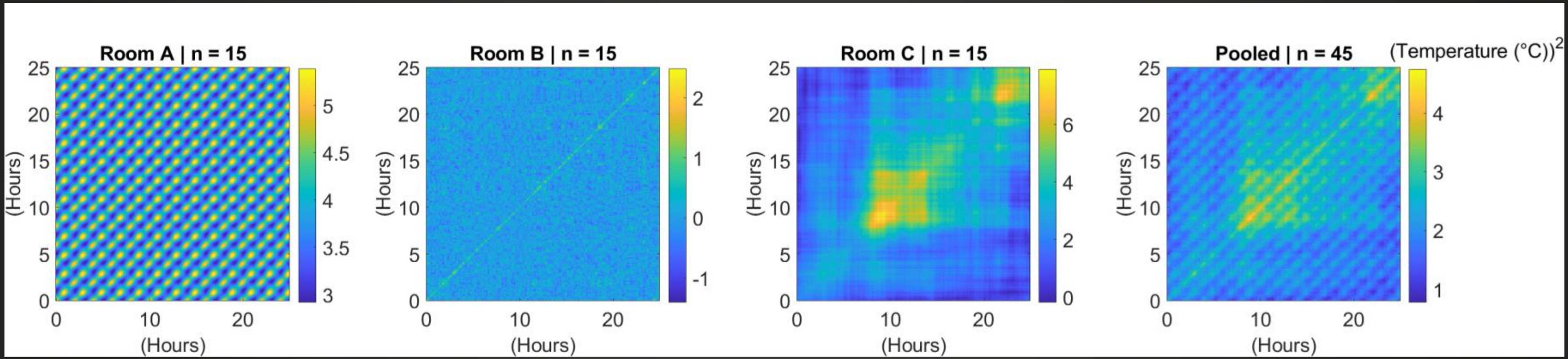
Example Data Again



Ideal Workflow

1. Data Preprocessing (if needed)
   - Sampled over a common domain
   - ~ 1000 elements per sample Plot Data
2. Pick Statistical Significance $(\alpha = 0.05)$
3. Check Assumptions (Heteroscedastic or Homoscedastic)
   - Visual Check
   - Statistical Test for the Equality of Covariances
4. Run Appropriate F-ANOVA variations
5. Post-Hoc test

Check Assumptions (Heteroscedastic or Homoscedastic)



Method	P-Value	Verdict
L2-Simul	0.02844	Reject Null Hypothesis for Alternative Hypothesis
L2-Naive	0.00086	Reject Null Hypothesis for Alternative Hypothesis
L2-Bias Reduced	0.00059	Reject Null Hypothesis for Alternative Hypothesis
Permutation-Test	0.01560	Reject Null Hypothesis for Alternative Hypothesis
Bootstrap-Test	0.02670	Reject Null Hypothesis for Alternative Hypothesis

Run F-ANOVA (Heteroscedastic)

Family-Wise Method	Test-Statistic	P-Value	Verdict
L2-Simul	116667.5466	0.00018	Reject Null Hypothesis for Alternative Hypothesis
L2-Naive	116667.5466	0.00032	Reject Null Hypothesis for Alternative Hypothesis
L2-BiasReduced	116667.5466	0.00005	Reject Null Hypothesis for Alternative Hypothesis
L2-Bootstrap	116667.5466	0.00060	Reject Null Hypothesis for Alternative Hypothesis
F-Simul	5.6473724	0.00083	Reject Null Hypothesis for Alternative Hypothesis
F-Naive	5.6473724	0.00032	Reject Null Hypothesis for Alternative Hypothesis
F-Bias Reduced	5.6473724	0.00007	Reject Null Hypothesis for Alternative Hypothesis
F-Bootstrap	5.6473724	0.00250	Reject Null Hypothesis for Alternative Hypothesis

*For this example, the homoscedastic and heteroscedastic F-ANOVA variants provided very similar P-Values

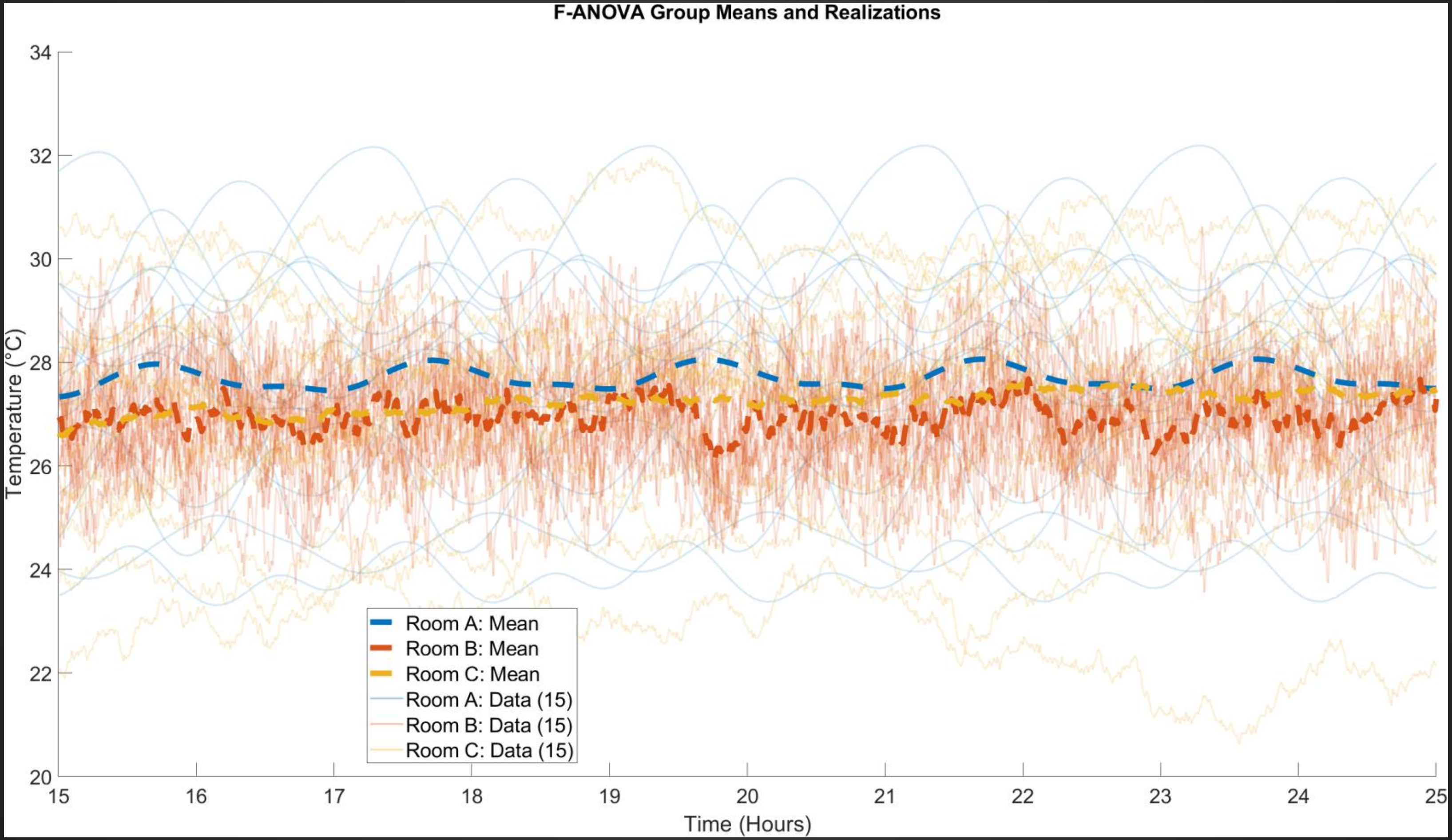
Post-Hoc Test: Pair-Wise Heteroscedastic F-ANOVA

Which Room(s) are statistically different from each other?

Hypothesis	L2-Simul	L2-Naive	L2-Bias Reduced	L2-Bootstrap	F-Simul	F-Naive	F-Bias Reduced	F-Bootstrap
Room A & Room B	0.00065	0.00027	0.00003	0.00030	0.00198	0.00027	0.00006	0.00360
Room A & Room C	0.00217	0.00106	0.00044	0.00100	0.00362	0.00106	0.00058	0.00410
Room B & Room C	0.50303	0.53021	0.55530	0.45900	0.51839	0.53024	0.57365	0.52880

- Room A statistically differs compared to Room B and Room C
- However, we fail to do so for Room B and Room C

What if we subset from 15 hours to 25 hours?



Run F-ANOVA (Heteroscedastic)

Family-Wise Method	Test-Statistic	P-Value	Verdict
L2-Simul	6851.3198	0.4799	Fail to Reject Null Hypothesis
L2-Naive	6851.3198	0.4801	Fail to Reject Null Hypothesis
L2-BiasReduced	6851.3198	0.4852	Fail to Reject Null Hypothesis
L2-Bootstrap	6851.3198	0.4167	Fail to Reject Null Hypothesis
F-Simul	0.8171	0.4872	Fail to Reject Null Hypothesis
F-Naive	0.8171	0.4801	Fail to Reject Null Hypothesis
F-BiasReduced	0.8171	0.5020	Fail to Reject Null Hypothesis
F-Bootstrap	0.8171	0.4589	Fail to Reject Null Hypothesis

- We don’t have enough evidence to state that the room mean temperatures are statistically different from each other

Variants of F-ANOVA: Two-Way

F-ANOVA, just like traditional ANOVA, can be performed One-Way or Two-Way

- One-Way: whether the mean functional response differs across multiple *levels or groups* of a single factor.
  - Factor = Temperature
  - Levels = {30, 60, 100}°F

- Two-way: whether the mean functional response differs *jointly* across levels of *two* factors, and can also test for **interaction** between those factors
  - Factors = {Temperature, Humidity}
  - Levels = {30, 60, 100}°F and {10, 90}%
  - Interaction: temperature 100°F and 10% humidity

Differences between Existing F-ANOVA Libraries

Features	Library Name	
	fdANOVA	LANL's f-ANOVA
Programming Language	R	MATLAB

F-ANOVA Methods	▪ **Basis-function Representation**
	▪ Random Projections
	▪ L^2-norm and F-test
	  ▪ Naïve
	  ▪ Bias-reduced
	  ▪ Nonparametric Bootstrap
		▪ L^2-norm and F-test
		  ▪ Naïve
		  ▪ Bias-reduced
		  ▪ Nonparametric Bootstrap
		  ▪ **MC Simulation**

Data Types	Univariate and **Multivariate**	Univariate
Heteroscedastic Support	Just one (L^2-norm parametric bootstrap)	**All methods**
Design	One-Way	One-Way and **Two-Way**
Post-hoc / Pairwise Tests	Not Included	Easy **pairwise** comparisons between factor levels
Equality of Covariance Testing	Not Included	**Visual** and **statistical tests**

Key Take Aways

- F-ANOVA works over a variety of scientific fields.
  - If your data can be sampled over a common domain, you can utilize F-ANOVA methods
- Provide sufficient resolution for each sample
  - ~ 1000 elements per sample of your functional data.
- F-ANOVA can support:
  - Heteroscedastic or homoscedastic data (pick the right F-ANOVA based on your data)
  - Data that does not originate from a Gaussian process.
  - One-way and Two-way analyses
  - Low sample sizes, bootstrap options exist.

Thank You

Convergence of test statistic

- Convergence of the test statistic is not trivial

- Depends on
  - Covariance function
  - Sample size
  - Smoothing of Data (Post-Processing)

- Safe assumption is to discretize data to at least 1000 elements

Benefits of reconstruction

- Evaluate at any desired resolution over the domain, $\mathcal{D}$ [1]
- Remove measurement errors/noise so the variance arises from the between-subject variation
- Convergence of test statistic requires sufficiently large enough resolution $\sim$1000 elements per realization