My Journey to an FFRDC and Machine Learning L-functions

Tamara Veenstra

IDA-CCR La Jolla

January 5, 2026, JMM

My Career Journey: a Roundabout Path to an FFRDC

1997

2001

SCAMPs$^+$
2007-2022

2023

[1] OK, really this is a photo of Torrey Pines State Beach and not CCRL.

What is IDA-CCR La Jolla and SCAMP?

Descriptions from IDA website

- "The Center for Communications and Computing (CCC) is a federally funded research and development center operated by the Institute for Defense Analyses (IDA). CCC consists of three distinct research centers: CCR La Jolla, CCR Princeton, CCS Bowie."
- "CCRL was founded in 1989 in La Jolla, California" (San Diego) and "now employs many PhD mathematicians, statisticians, and computer scientists working on problems in cryptography, cryptanalysis, machine learning, high-performance computing, and network security, as well as related areas of pure and applied mathematics."
- "SCAMP is an annual 10-week summer program (at each of the 3 centers) that brings together a diverse group of creative mathematicians and other scientists to provide practical solutions to critical real-world problems."

Interestingly, the SCAMP program dates back to 1952!

From the Friedman documents collection, declassified on 2-12-2014, link.

Report on SCAMP

to the

Director

of the

Armed Forces Security Agency

Submitted by

Stewart S. Cairns

8 September 1952

Report on SCAMP

1. Introduction

The summer symposium to which the name SCAMP was finally attached grew from discussions in SCAG. The initial proposals leading to SCAMP were made in 1951 by C. B. Tompkins in his role as a member of SCAG. Originally, a continuing project was intended. However, after various discussions, it was decided to conduct a summer symposium and, on the basis of experience therewith, to consider the desirability of later efforts. This decision was reached in March 1952, with the result that only about three months were available for the preliminaries to the symposium. Among the proposals by Tompkins was the use of the primarily mathematical part of SCAG, referred to as SCAM (Special Committee Advising in Mathematics), in an advisory capacity. The code name SCAMP was created in AFSA by adding the letter P to SCAM.

Why I SCAMPed many times...

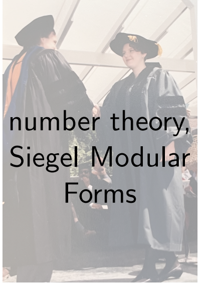
- Supportive (many advantages for doing research at SCAMP versus individually at small, private liberal arts school)
- Collaborative (great people to work with, collaboration encouraged; one of my favorite aspects!)
- Applied (interesting real-world problems)
- Mathematics (variety of different, exciting mathematical topics including both theoretical and computational; I enjoy learning new things!)
- Program (10-week, paid summer program, lodging provided, in San Diego; great opportunity to refresh and recharge in a very different setting.)

Balance

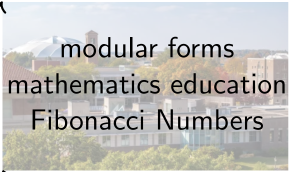
- As an academic; I found SCAMP a really great balance between focusing on teaching most of the year and getting to focus on a structured research area during the summer in a supportive, collaborative environment.
- Eventually balance didn't feel correct; I became more excited about the 10-week part of the year and less excited/more exhausted by the rest of the year.
- Very grateful to have been able to transition from one fantastic job to another.
- Leaving academia was a hard decision, (I love teaching and many aspects of academia!) but the right decision for me at that point in my career. (Overall, I was happier at CCRL.)

Returning to my Scholarly Interests

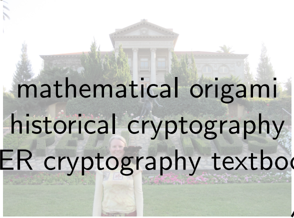
Most of my academic career was at a primarily undergraduate, teaching centered institution with limited time for research and included broad research interests.



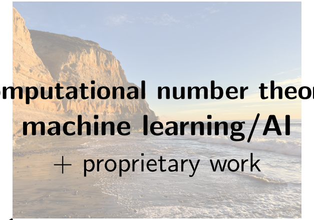
number theory,
Siegel Modular
Forms



modular forms
mathematics education
Fibonacci Numbers



mathematical origami
historical cryptography
OER cryptography textbook



computational number theory,
machine learning/AI
+ proprietary work

Machine Learning L-functions

- Even though I left academia, part of the mission at CCRL is to stay informed about current trends in mathematics; this includes attending conferences and collaborating with academic researchers.
- The research in this talk originated at the Mathematics and Machine Learning Program at Harvard's Center for Mathematical Sciences and Applications in Fall of 2024.[2]
- Big question: There are lots of new machine learning/AI tools. How can we use them to discover new mathematics?
- Similar to the spirit of the work at CCRL. How can we apply existing mathematical tools to help solve real-world problems? (Interested in all types of mathematical tools!)

[2] Joint work with Joanna Bieri, Giorgi Butbaia, Edgar Costa, Alyson Deines, Kyu-Hwan Lee, David Lowry-Duda, Tom Oliver, and Yidi Qi.

What are L-functions?

L-functions are fundamental objects in number theory!

- Generalize the Riemann zeta function.
- Can study families of L-functions for various types of mathematical objects (elliptic curves, modular forms, Maass forms).
- Contain important arithmetic information (e.g., modularity theorem, class number formula, Birch-Swinnerton-Dyer conjecture)
- In particular, root numbers of L-functions[3] are amazing quantities that contain a huge amount of information.
- Goal: learn about root number from Dirichlet coefficients.

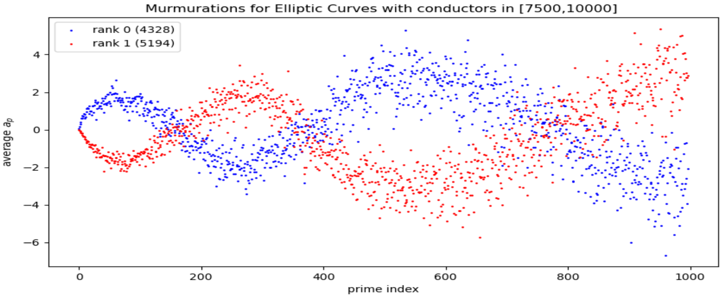
[3] signs occurring in functional equations, e.g. $\Lambda(s) = \epsilon\Lambda(1-s)$

What can we learn from? Dirichlet series of L-functions

- Dirichlet series
$$L(s) = \sum_{n \geq 1} a_n n^{-s}$$
- Dirichlet coefficients a_n. Also called Fourier coefficients (modular forms, Maass forms) or traces of Frobenius (elliptic curves).
- Multiplicative: $a_{nm} = a_n a_m$ if $gcd(n, m) = 1$.
- LMFDB (L-functions and Modular Forms Database: www.lmfdb.org) contains Dirichlet coefficients for many types of L-functions. What can we learn from them?

An unexpected result!

In 2022, "Murmurations of Elliptic Curves" paper, He, Lee, Oliver, and Pozdnyakov observed a curious fluctuation in average Dirichlet coefficients that gives information about order of vanishing. The root number $\epsilon = (-1)^v$ where v is the order of vanishing.



These statistical correlations between the root numbers of L-functions and Dirichlet coefficients seem to lead to ML's ability to predict accurately.

Maass Forms and Murmurations?

LMFDB recently added about 35K Maass forms. Maass forms are like modular forms but more complicated. In particular, Dirichlet/Fourier coefficients are transcendental.

- Murmurations?? $\Rightarrow$ predicting Fricke sign from Fourier coefficients??
- For Maass forms, the root number

$$\epsilon = (-1)^{\sigma(f)} w_N$$

  where w_N is the **Fricke sign** and $\sigma(f)$ is the parity of the **symmetry** (f even/odd.)
- Also, $f(z) = w_N f\left(\frac{-1}{Nz}\right)$ for Maass forms of level N (Fricke involution).

Incomplete data in LMFDB

Unfortunately, for about half of the Maass forms, the Fricke sign is missing. The Fricke sign is complicated to compute. In general, it requires:

- trace formula
- Hejhal's algorithm
- modularity properties

Best Case Scenario: Only need the trace formula and Hejhal's algorithm. These are the known Fricke signs in the LMFDB. (19,993 out of 35,416).

Worst Case Scenario: Zeros of the Bessel functions make the previous numerical calculations unstable, so need modularity conditions to produce an overdetermined system from which we can eventually deduce the Fricke sign. These are the unknown Fricke signs in the LMFDB. (15,423 out of 35,416)

We would like to predict w_N in these cases when it is hard to compute!

Our Dataset

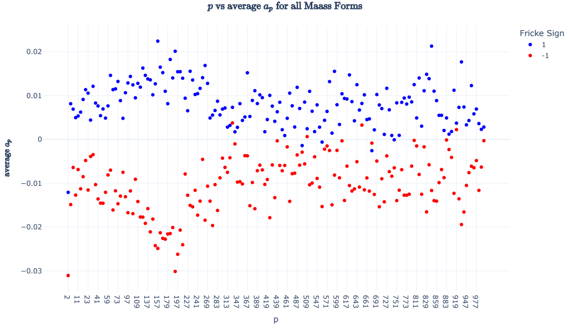
35,416 rigorously computed Maass forms from the LMFDB, all with weight 0, trivial character, and level N in the range from 1 to 105.

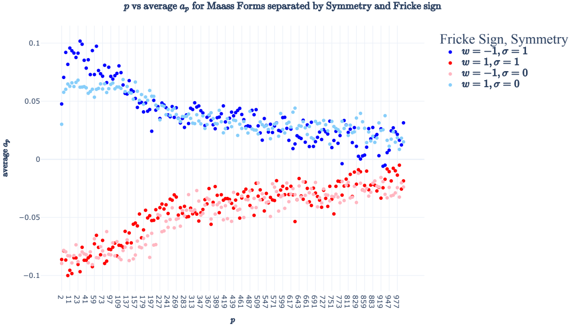
The dataset contains the first 1000 Fourier coefficients a_n for each Maass form.

Separated by symmetry (σ) and Fricke sign (w):

	$w = -1$	$w = 1$	$w = 0$ (unknown)	Total
$\sigma = 0$ (even)	5,009	7,171	6,173	18,353
$\sigma = 1$ (odd)	3,724	4,089	9,250	17,063
Total	8,733	11,260	15,423	35,461

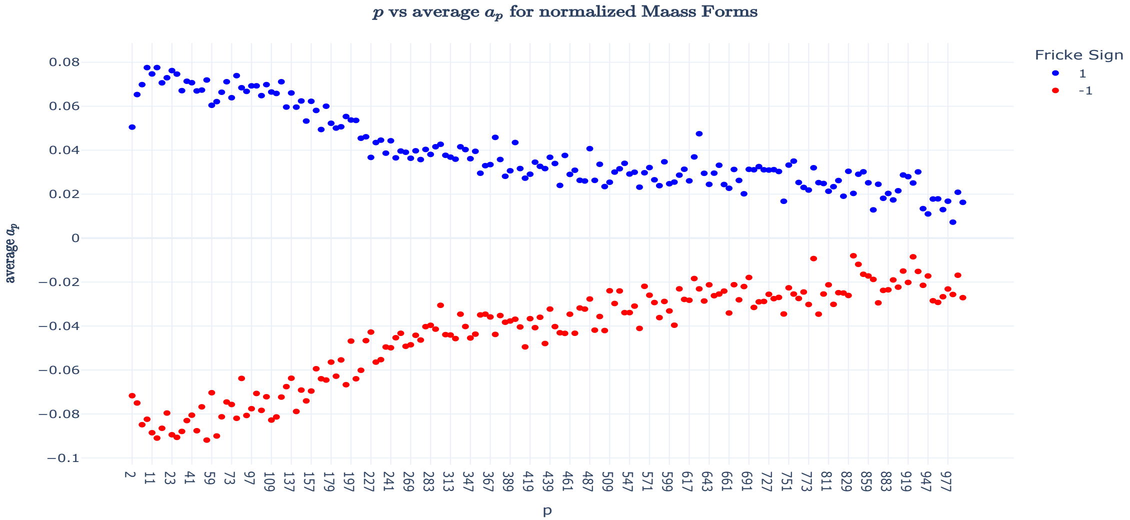
Are there murmuration type patterns by Fricke sign?





Better! Let's normalize for the symmetry of the Maass form!

Normalize the Fourier coefficients by $(-1)^\sigma a_p$ (σ, symmetry). I.e, separate by root number.



Predicting Fricke signs

- Clear separation in Fricke sign by average $(-1)^\sigma a_p$
- Should be possible to predict Fricke sign, maybe even with a simple technique.
- Used Linear Discriminant Analysis (LDA)
- LDA is a supervised dimensionality reduction to maximize separability between classes
- Learns linear decision boundary based on maximizing distance between the means of the two classes and minimizing variance within class
- Separate the normalized data for which we have known Fricke sign into testing, training, and validation sets, stratifying the sample sets based on the Fricke sign. 80/20 split.

Experiments on training and validation data

Performed LDA on all the training data (12,795 observations) normalized by symmetry. Tested accuracy on validation data (3199 observations).

LDA method	all symmetry (normalized)	even symmetry	odd symmetry
LDA using all a_n	0.9612	0.9488	0.9633
LDA using only a_p	0.8625	0.8261	0.8800

Surprising results:

- a_n is better than a_p.
- unsupervised clustering algorithms such as K-means and PCA were not successful at distinguishing the Fricke sign.

Why are a_n so much better than a_p?

Conjecture: small prime effects, but . . .

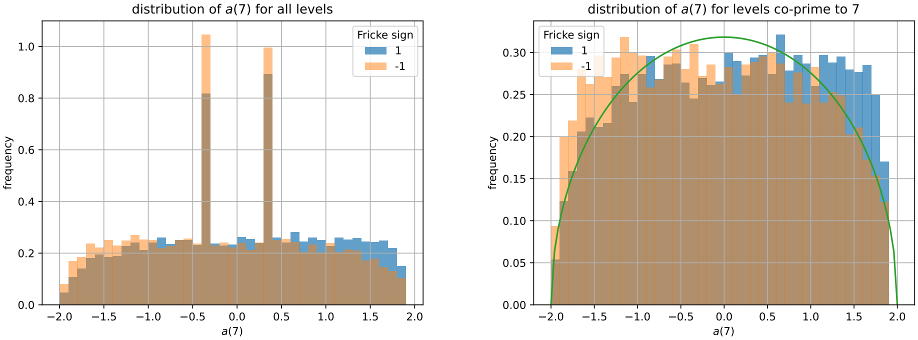
Feature indices[4]	LDA accuracy	number of a_p
all n	96.1%	1000
prime n	86.2%	168
n is divisible by $2, 3$ or 5	78.5%	733
n is divisible by $2, 3$ or 5 or n is prime	89.8%	898
n is divisible by 1 or 2 prime factors	95.3%	702
n even	70.6%	500
n odd	93.4%	500
n 45-smooth	75.3%	497
$\{n \in \mathbb{Z} : 1 \leq n \leq 500\}$	93.3%	500

[4] Exploring subsets: $1 \leq n \leq 1000$ unless otherwise specified.

Is LDA cheating?

Fourier coefficients contain information about Fricke sign when $p|N$.

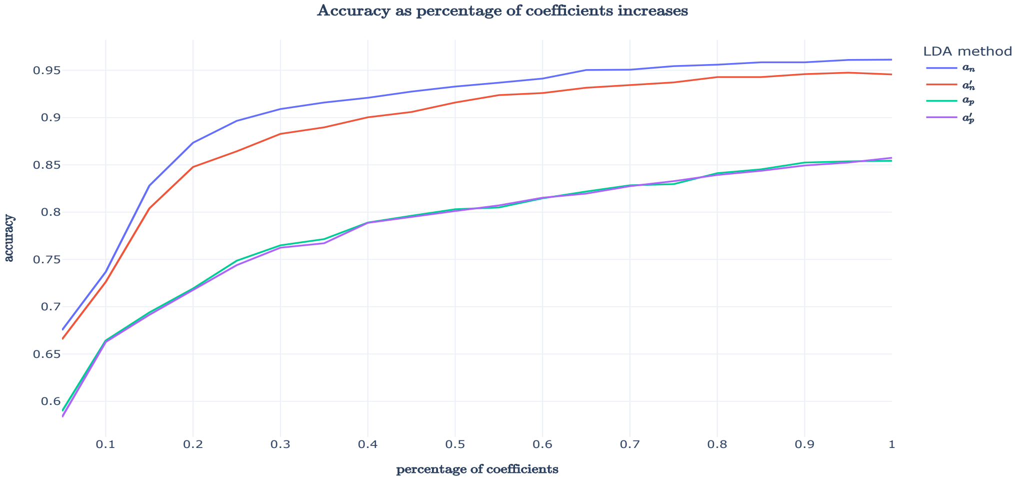
- If Fricke sign is known, $a_p = -w_p/\sqrt{p}$ and $w_N = \pm 1 = \prod_{p|N} w_p$.
- If Fricke sign is unknown, $a_n := 0$ for $\gcd(n, N) > 1$.



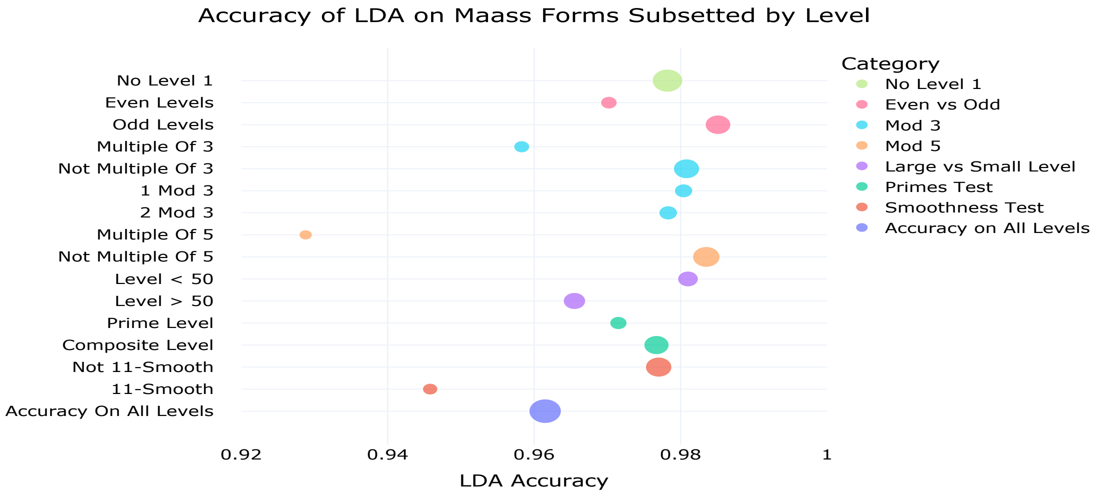
Spikes at $\pm 1/\sqrt{p}$ versus Sato-Tate circular distribution.

Eliminating Fricke sign info in a_p and how many a_n do we need?

Accuracy relative to the number of coefficients, for a_p, a_n and a'_p, a'_n where $a'_n := 0$ if $\gcd(n, N) > 1$ (as when Fricke sign is unknown).

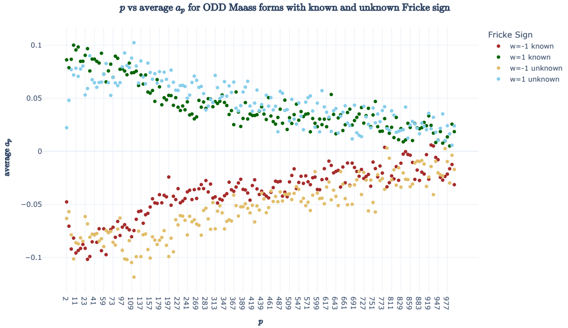


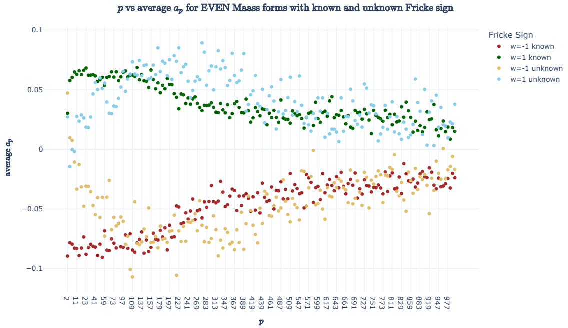
Level Analysis: Are some levels easier to predict?



Let's use the predictions!

- We have unknown Fricke signs in the dataset.
- Using LDA to make predictions on these. How do we assess?
- Compare to murmurations.





- Also good accuracy with heuristic methods of computing Fricke sign.

Conclusion for Maass forms

- Really good accuracy!
- Predicting with LDA is very easy!
- Predictions with Neural Networks had similar accuracy.
- Even the heuristic calculations were surprisingly hard.
- Hope to use the predicted values to help determine the unknown signs precisely!

Thank you!

- Maass Forms paper available: arxiv.org/abs/2501.02105.
- Code and data available on Zenodo: doi.org/10.5281/zenodo.15716014, doi.org/10.5281/zenodo.15490636
- Slides and references to other papers available on my website, tamarabveenstra.com