Tom de Groot

Chemistry - Mathematics
Programming - AI

Born: 30-03-2004

Grijpskerk, NL

Herestraat 45, 9843 AJ

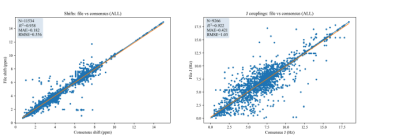
+31 6 314 90 416

tomthegreatest04@gmail.com

github.com/Tomodovodoo

x.com/Tomodovodoo

Research Showcase



(a) Shifts: file vs. consensus. (b) J couplings: file vs. consensus.

Figure 4.1: Scatter plots showing the agreement of file values with the per-atom consensus. See Table 4.2 for summary statistics.

Agreement across files (labels vs. consensus)

To quantify how consistent the raw labels are, we compared each file's values to a per-atom *consensus* (as defined in Methods). Across all matched pairs, the agreement is strong for both shifts and couplings (Fig. 4.1 and Table 4.2), while the multiplicity analysis shows where disagreements accumulate as sequences get longer (Fig. 4.3).

Table 4.2: Agreement of per-file labels with per-atom consensus across all matched pairs.

Quantity	N	R^2	MAE	RMSE
Shift (ppm)	11,534	0.938	0.182	0.556
J coupling (Hz)	9,266	0.922	0.421	1.030

Cross-file variability (per atom). We computed per-atom variability across files for molecules with sufficient replicate spectra. In total this covered **3,792 atoms** across **350 molecules** (for J statistics, **2,206** atoms had usable couplings; **58.2%** J coverage among those atoms). Distributions are summarized in Table 4.3 and visualized in Fig. 4.2.

Table 4.3: Per-atom cross-file variability (distributions across atoms).

Quantity	Count	Mean	Median	p90	Max	SD
Shift std. dev. (ppm)	3,792	0.1406	0.0700	0.3200	2.9081	0.2485
Mean $\lvert\Delta\text{shift}\rvert$ (ppm)	3,792	0.1401	0.0650	0.3200	3.5133	0.2636
Mean $\lvert\Delta J\rvert$ (Hz)	2,206	0.3307	0.1280	0.9338	7.4667	0.5491
Mean $\lvert\Delta\rvert$ (mult. length)	3,791	0.2110	0.0000	0.6667	2.4000	0.3490

31

Bachelor Thesis: *Evaluating and Optimizing NMR Prediction with GNNs (Pollice Research Group).*

EDUCATION

BSc Chemistry
University of Groningen
Sep 2021 - Aug 2025
Track: Chemistry of Life

AWARDS

Chess.com Member of the Year 2021

Harmonic Rising Mathematician (confidential)

ABOUT ME

Hi, I'm Tom! I graduated from the University of Groningen with a BSc in Chemistry. During this, I chose the Life Sciences track. I wrote my thesis in the Pollice Group, evaluating and optimizing NMR prediction with Graph Neural Networks.

After graduation, I got more into math and programming. I wrote a program to automatically search jobs agentically, to no avail yet. So, during my free time, I started to work on Project Euler (complex math challenges), which eventually led me to **PNT+**, an effort to create a human-readable proof of the Prime Number Theorem in Lean.

Since my time at uni, I've developed myself to be a competent student mathematician and programmer. So, without working experience in the field, I'm trying to make do with contributing to Open Source, and perform my hobbies in math, AI, and in the sciences.

SKILLS

Math: Lean (Beginner), Formal Verification

Code: Python (Medior), AI/ML (Medior)

Science: Analytical (Bio)Chemistry

Tools: Git, GitHub, VS Code, LaTeX

Native: Dutch

Fluent: English

FEATURED PROJECTS

PNT+ (Formal Verification)

Formalizing complex number theory in Lean. Solving impactful theorems and lemmas. View Contributions

Automated Job Search Agent

Developed a Python-based agentic tool to automate and optimize job searching workflows.

Collection of Projects

From Audio to Text from scratch, plotting immigration, vaccine, and water data, to manipulating single player savegame data, these are my University hobby projects. View on Github

WORK & COMMUNITY EXPERIENCE

Transcom (for Odido)

May 2024 - Nov 2024

Customer Service Representative. Consistent top performer in customer satisfaction.

Teamspot

Feb 2022 - July 2024

Regional Director (Northern Netherlands). Coordinated school projects and performed testing/data entry.

Chess.com

Jan 2022 - June 2022

Event Proctor & Moderator. Assisted in moderation of international online chess events and the "Leagues" project.