Predicting Resource Usage Trends with Southern California Petabyte Scale Cache

Caitlin Sim (Univ. of California, Berkeley)
Kesheng Wu, Alex Sim (Lawrence berkeley National Lab)
Inder Monga, Chin Guok (Energy Sciences Network)
Frank Wurthwein, Diego Davila (Univ. of California, San Diego)
Harvey Newman, Justas Balcas (Caltech)

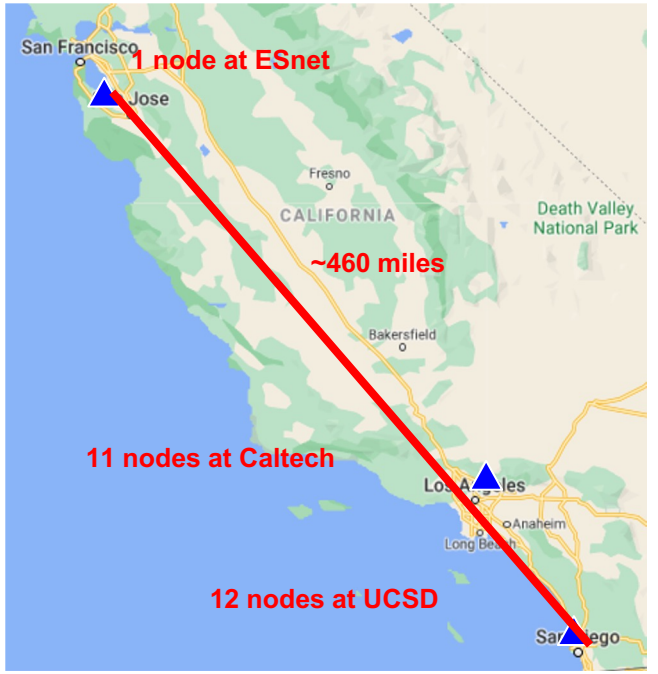
CHEP 2023, Track 1

Introduction

- **Observation**
  - Significant portion of the popular dataset is re-used in the user analysis
  - Storage cache allows data sharing among users in the same region
    - Reduce the repeated data transfers over the wide-area network
    - Decrease data access latency
    - Increase data access throughput
    - Improve overall application performance
  - Use case: Southern California Petabyte Scale Cache (SoCal Repo)
- **Goals**
  - Understand the resource characteristics
    - Explore measurements from SoCal Repo
    - Characterise the trends of network and cache utilization
    - Study the effectiveness of the SoCal Repo in reducing wide-area network traffic
  - Explore the predictability of the resource utilization
    - Help plan for additional deployments of caches in the science network infrastructure
  - Overall, study the effectiveness of the cache system for scientific applications

Southern California Petabyte Scale Cache

- SoCal Repo
  - Regional storage cache for US CMS user analysis
- SoCal Repo consists of 24 federated storage nodes
  - 12 nodes at UCSD: each with 24 TB, 10 Gbps network connection
  - 11 nodes at Caltech: each with storage sizes ranging from 96TB to 388TB, 40 Gbps network connections
  - 1 node at Silicon Valley (by ESnet): 44 TB storage, 40 Gbps network connection
  - Approximately 2.5PB of total storage capacity
  - ~100 miles between UCSD and Caltech nodes
    round trip time (RTT) < 3 ms
  - ~460 miles between Silicon Valley and UCSD nodes
    RTT ~10 ms



Previous work

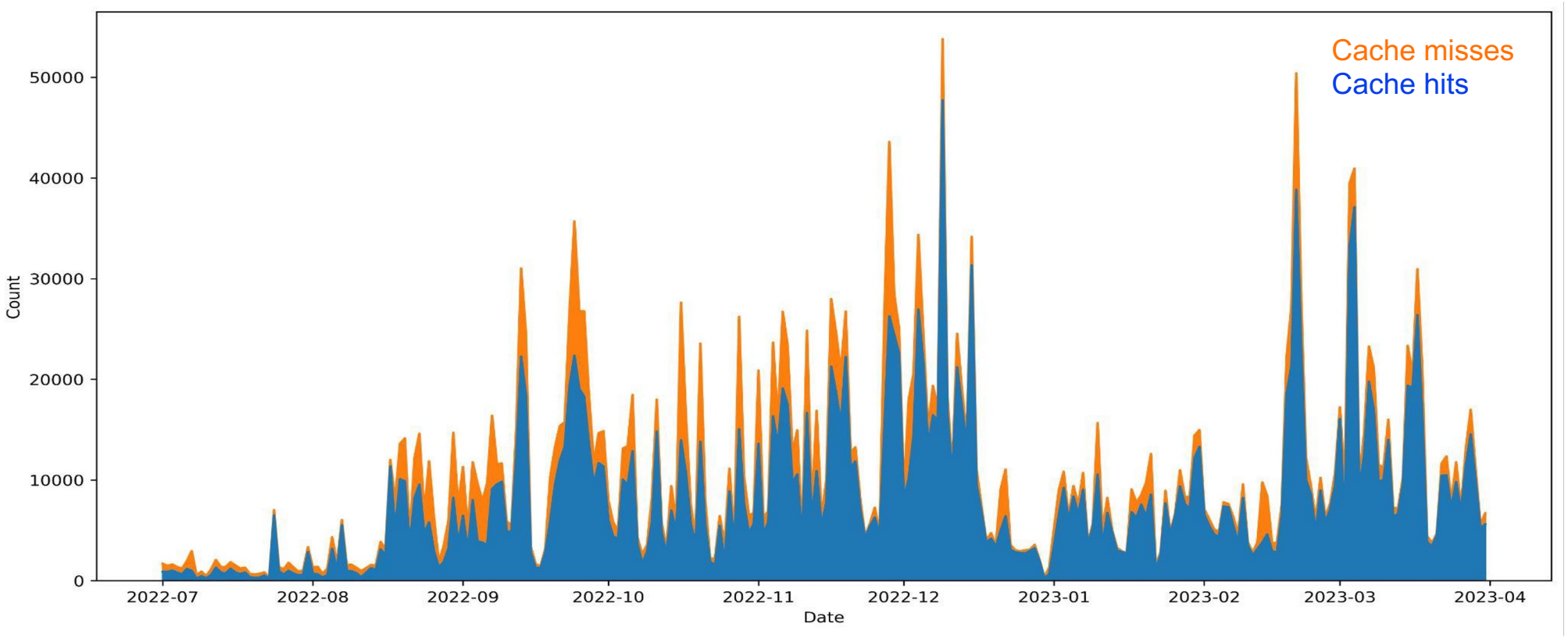
- "Effectiveness and predictability of in-network storage cache for Scientific Workflows", International Conference on Computing, Networking and Communication (ICNC 2023), 2023. https://sdm.lbl.gov/oapapers/icnc23-xcache-sim.pdf. Study duration from 7/2021 to 6/2022.
- "Access Trends of In-network Cache for Scientific Data", 5th ACM International Workshop on System and Network Telemetry and Analysis (SNTA), 2022, doi:10.1145/3526064.3534110. https://sdm.lbl.gov/oapapers/snta22-xcache.pdf. Study duration from 7/2021 to 1/2022.
- "Deploying in-network caches in support of distributed scientific data sharing", the US Community Study on the Future of Particle Physics (Snowmass 2021), 2022. doi:/10.48550/arXiv.2203.06843. https://arxiv.org/pdf/2203.06843.pdf. Study duration from 7/2021 to 12/2021.
- "Analyzing scientific data sharing patterns with in-network data caching", 4th ACM International Workshop on System and Network Telemetry and Analysis (SNTA 2021), 2021, doi:10.1145/3452411.3464441. https://sdm.lbl.gov/oapapers/snta21-xcache-esnet.pdf. Study duration from 5/2020 to 10/2020.

Data Access Summary

	# of accesses	Data transfer size (TB)	Shared data size (TB)	# of cache misses	# of cache hits
Total	3,615,578	560.96	5,208.91	663,994	2,951,584
Daily average	13,147	2.04	18.94	2,414	10,733

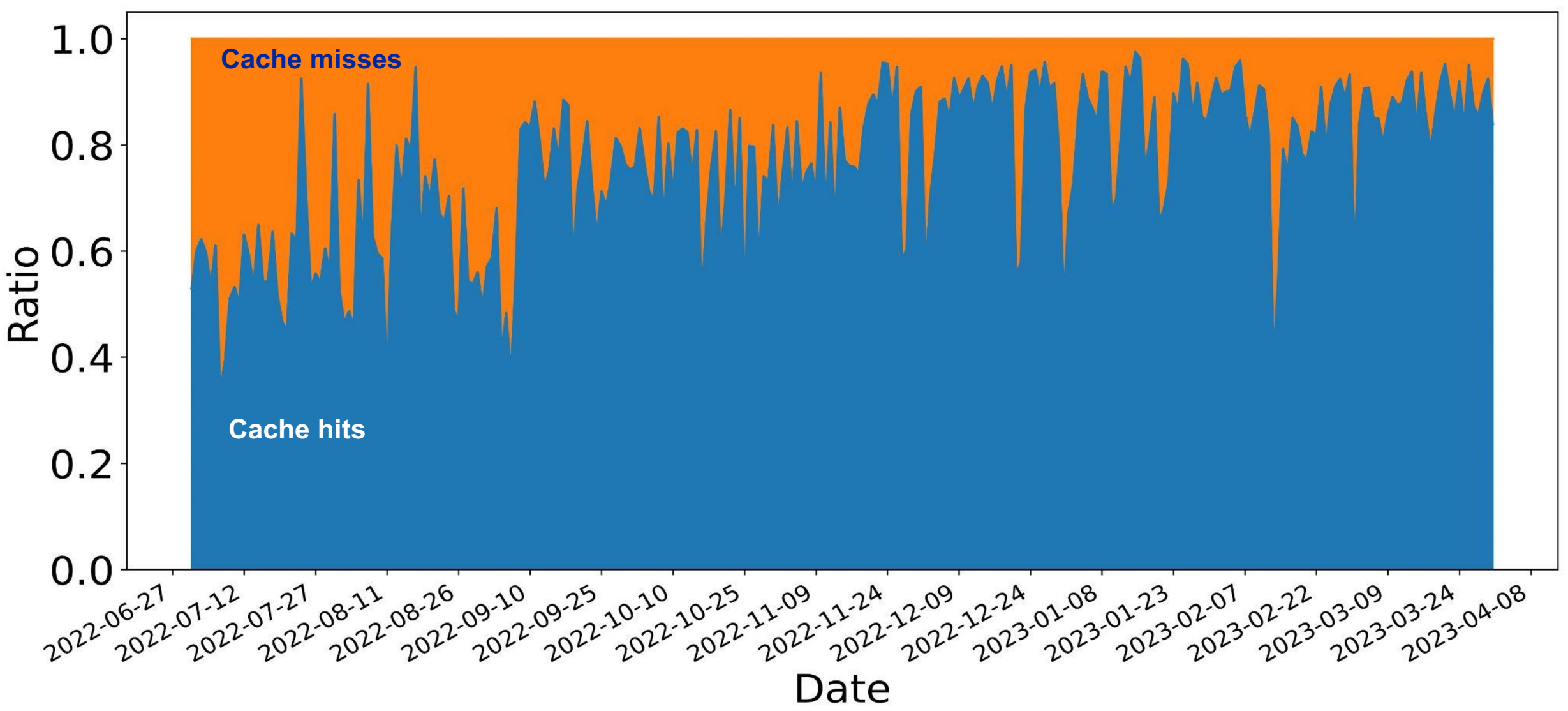
- **Logs**
  - **9 months operational logs from SoCal Repo from July 2022 to March 2023**
  - **5,838 log files, ~1.3GB**
- **Consisting of 3.6 million file requests between July 2022 and March 2023**
- **5.2PB (90.2%) of requested bytes (5.77PB) could be served from the cache**
- **81.6% of file requests are satisfied by the cache**

Daily File Requests

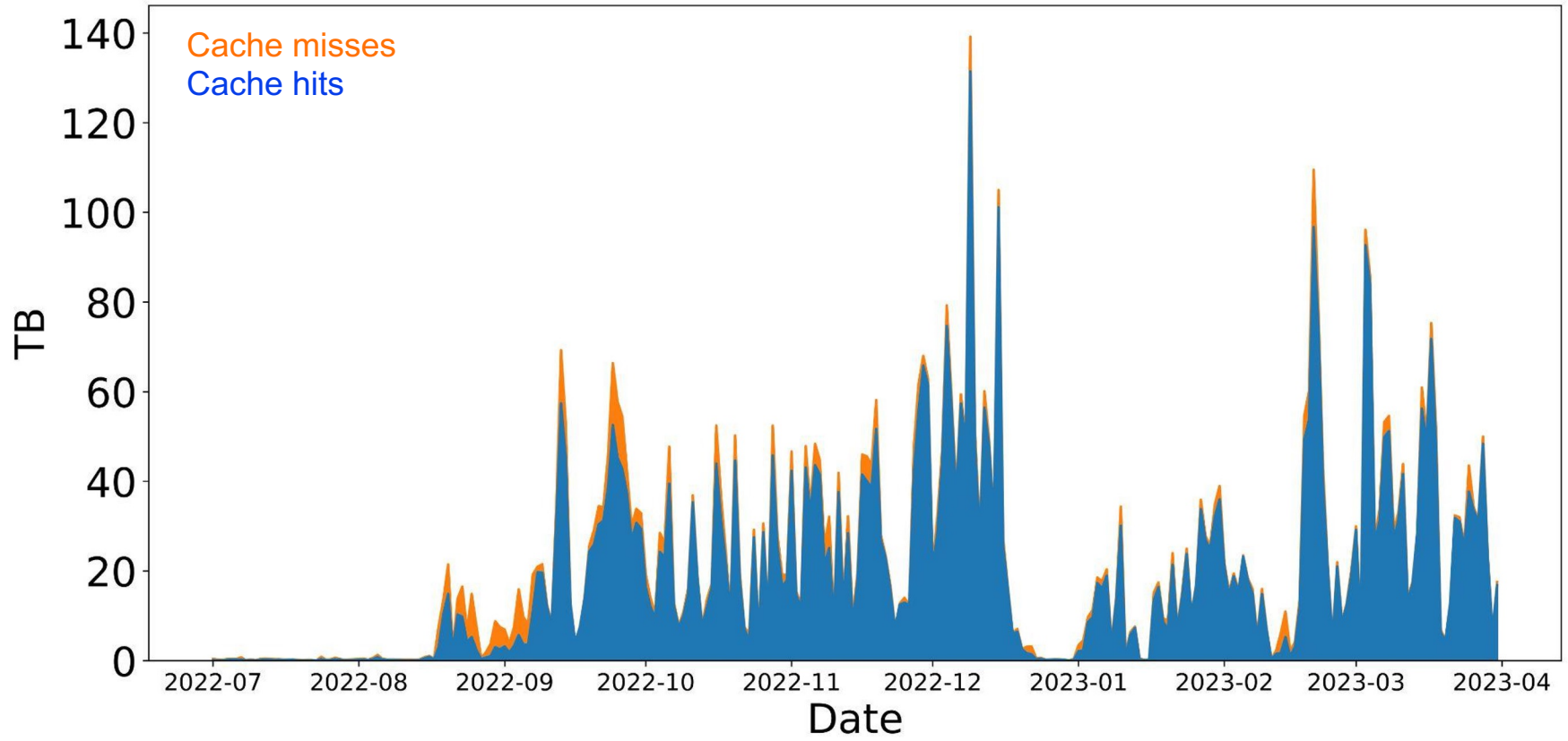


- On average, ~13,000 file requests per day are served from the storage cache nodes, while 10,000 requests are cache hits
- Only file requests that miss the cache trigger remote file transfers

81.6% of File Requests - Cache Hits

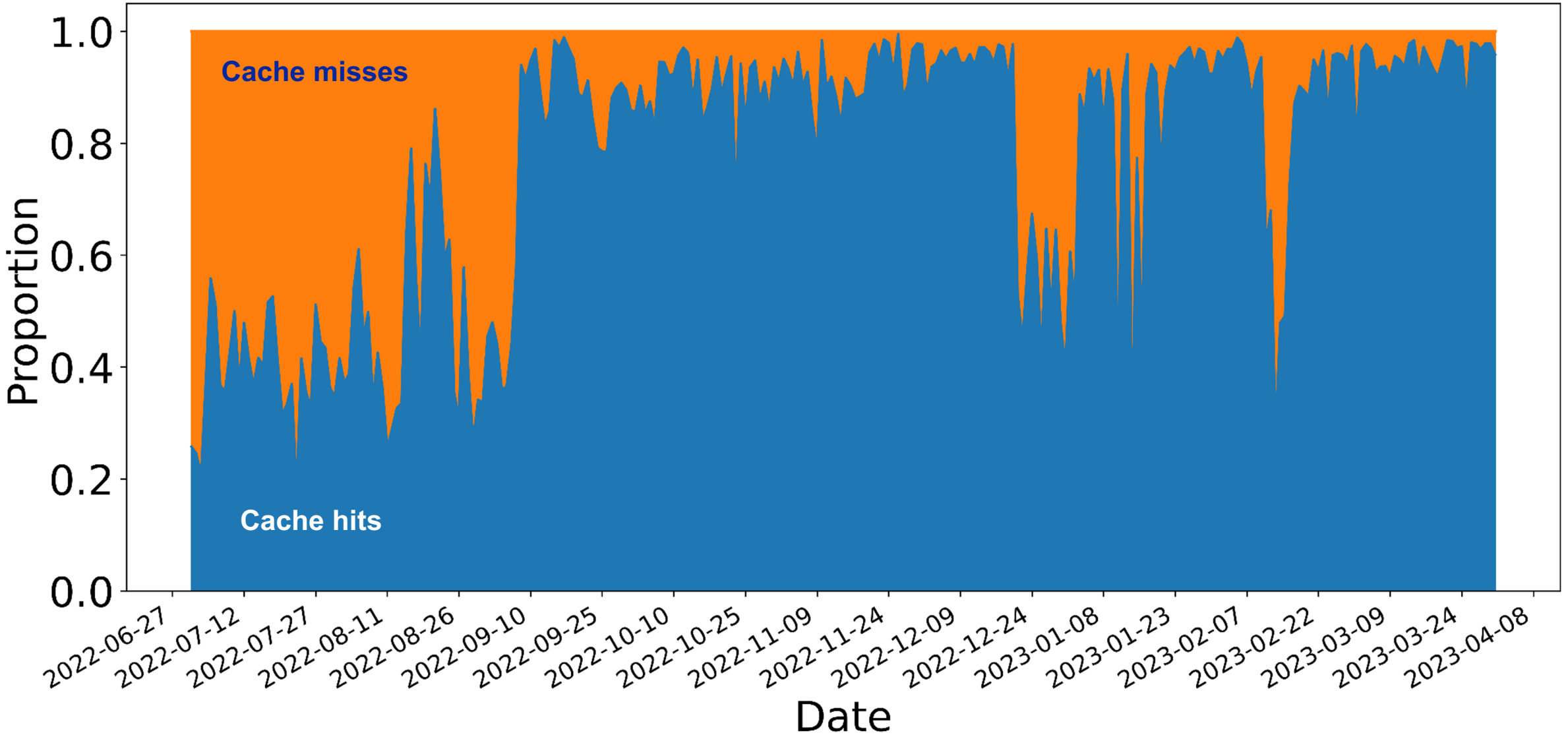


Daily Cache Hits: Some days have over 100TB

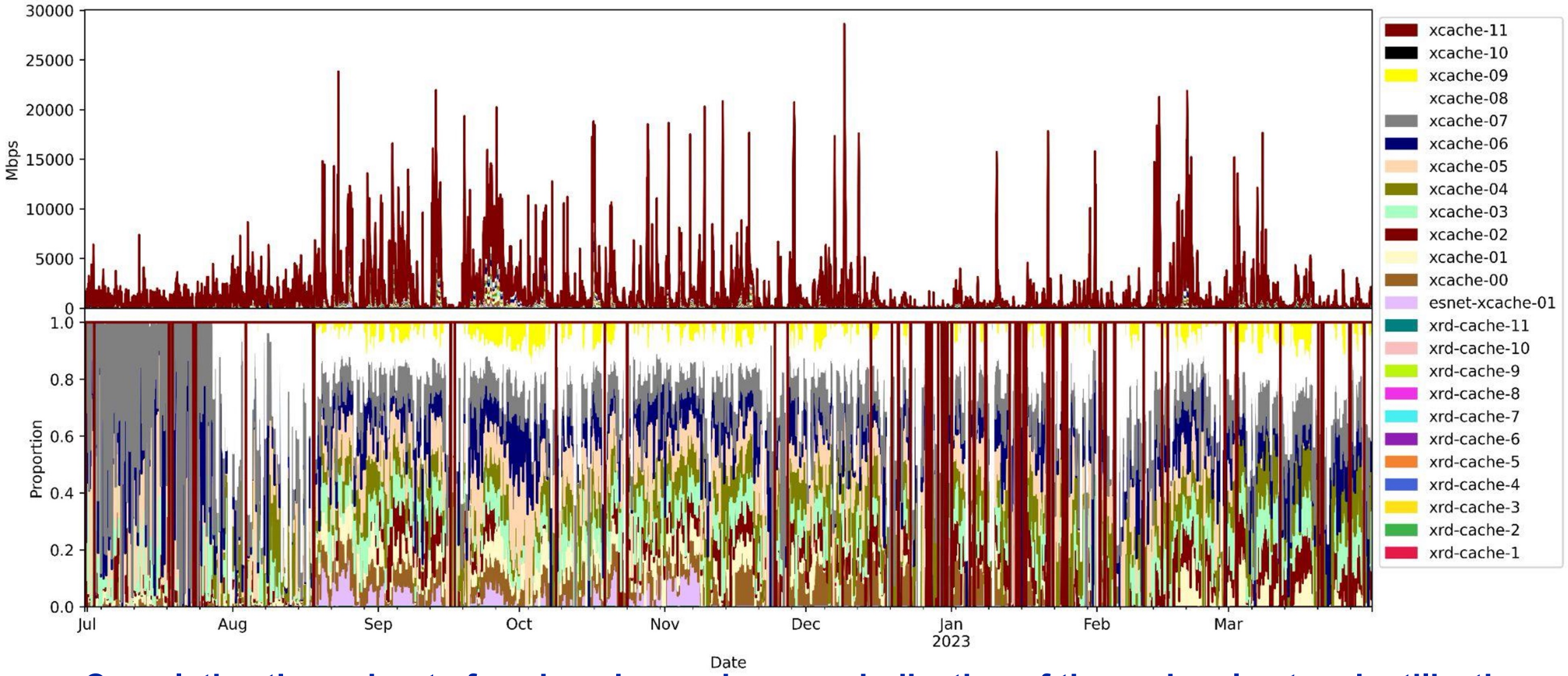


- On average, ~21 TB per day are served out of the storage cache
- On average, ~19 TB are cache hits (90.2%)

90.2% of Requested Bytes - Cache Hits



Hourly Cumulative Throughput of Cache Misses



- Cumulative throughput of cache misses shows an indication of the regional network utilization

Modeling Resource Usage Trends

- Mainly for cache and network utilization
- Use Long Short-Term Memory (LSTM) architecture to model the hourly and daily average utilization
- Training data: the first 80% of the monitoring period (7/22-3/23, 274 days)
- Testing data: the last 20%
- Hyper-parameters

	# of LSTM units	activation function	dropout rate	# of training epoch
values	128	tanh	0.04	50

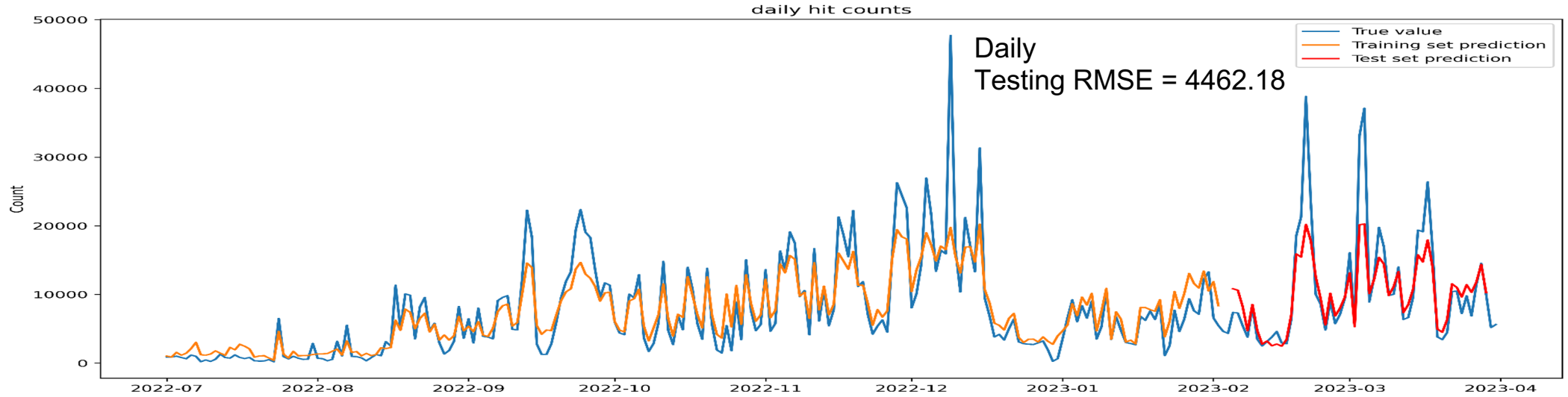
- Overall performance measure: root-mean-square error (RMSE)

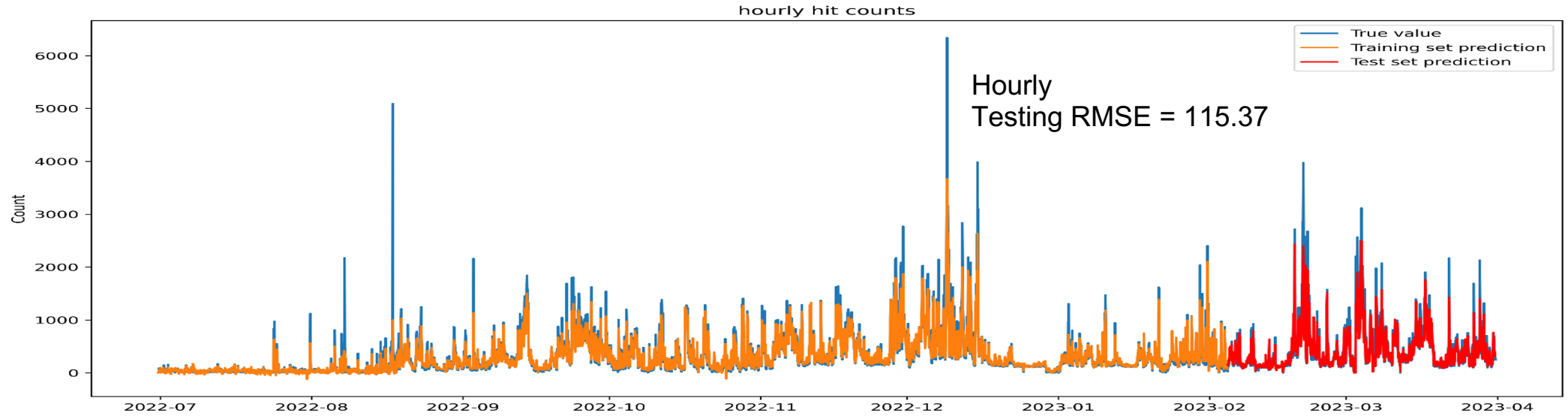
RMSE of LSTM Model Results

	Training RMSE	Testing RMSE	Standard deviation
Daily cache hits counts	3078.01	4462.18 (0.61)	7312.03
Daily cache misses counts	1737.65	1668.72 (0.63)	2639.63
Hourly cache hits counts	100.02	115.37 (0.28)	407.09
Hourly cache misses counts	36.93	50.02 (0.27)	182.88
Daily cache hits volume	8.65	12.44 (0.59)	20.91
Daily cache misses volume	1.30	1.42 (0.57)	2.50
Hourly cache hits volume	0.16	0.23 (0.25)	0.93
Hourly cache misses volume	0.09	0.10 (0.50)	0.20
Daily average throughput for cache miss	5.43	3.37 (0.23)	14.71
Hourly average throughput for cache miss	16.82	4.59 (0.19)	23.92

- **Relative prediction errors (inside parentheses) are less than 0.65.**
  - **Measured against the standard deviations**
- **More accurate with the hourly time series than with the daily**
  - **Most likely reason might be there are more training data records for the hourly time series**

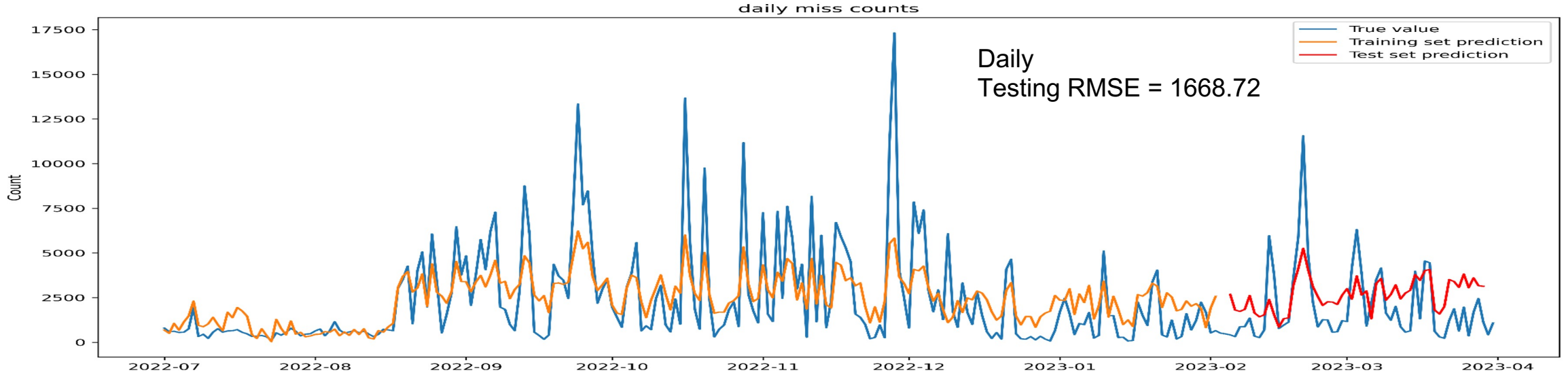
Predicted Number of Cache hits

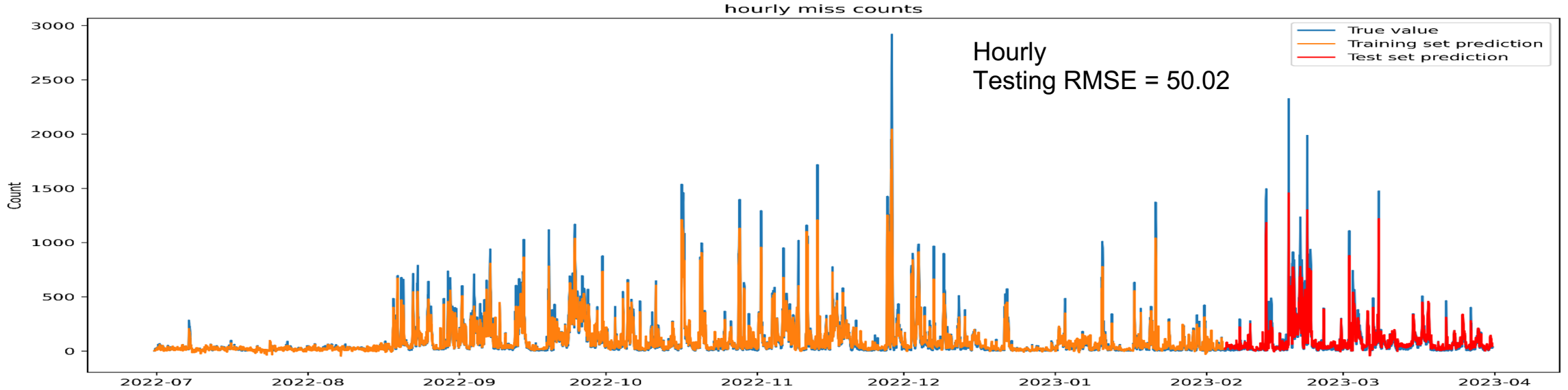




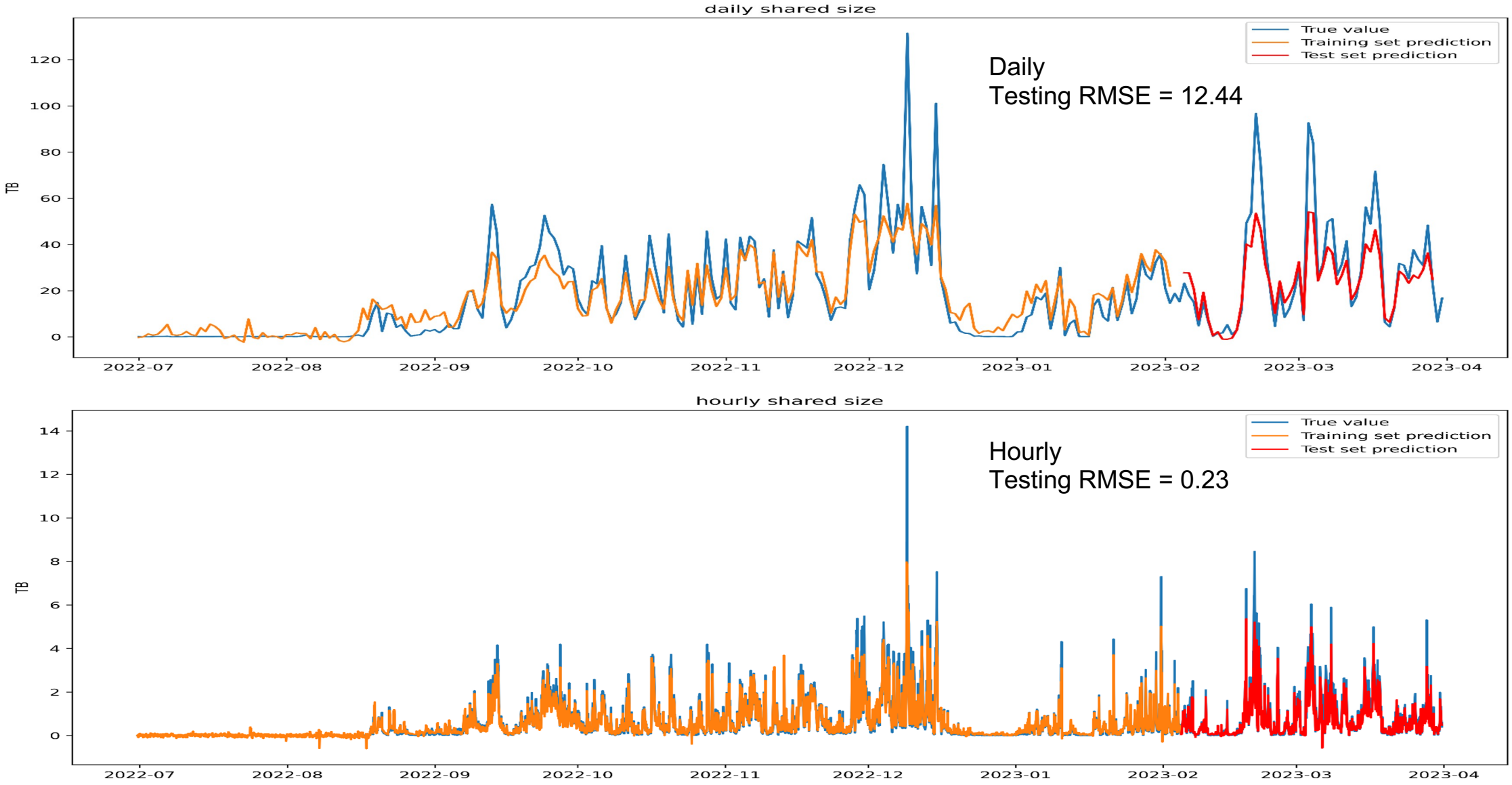
Predictions on the hourly time series are closer to the actual values than with the daily

Predicted Number of Cache Misses

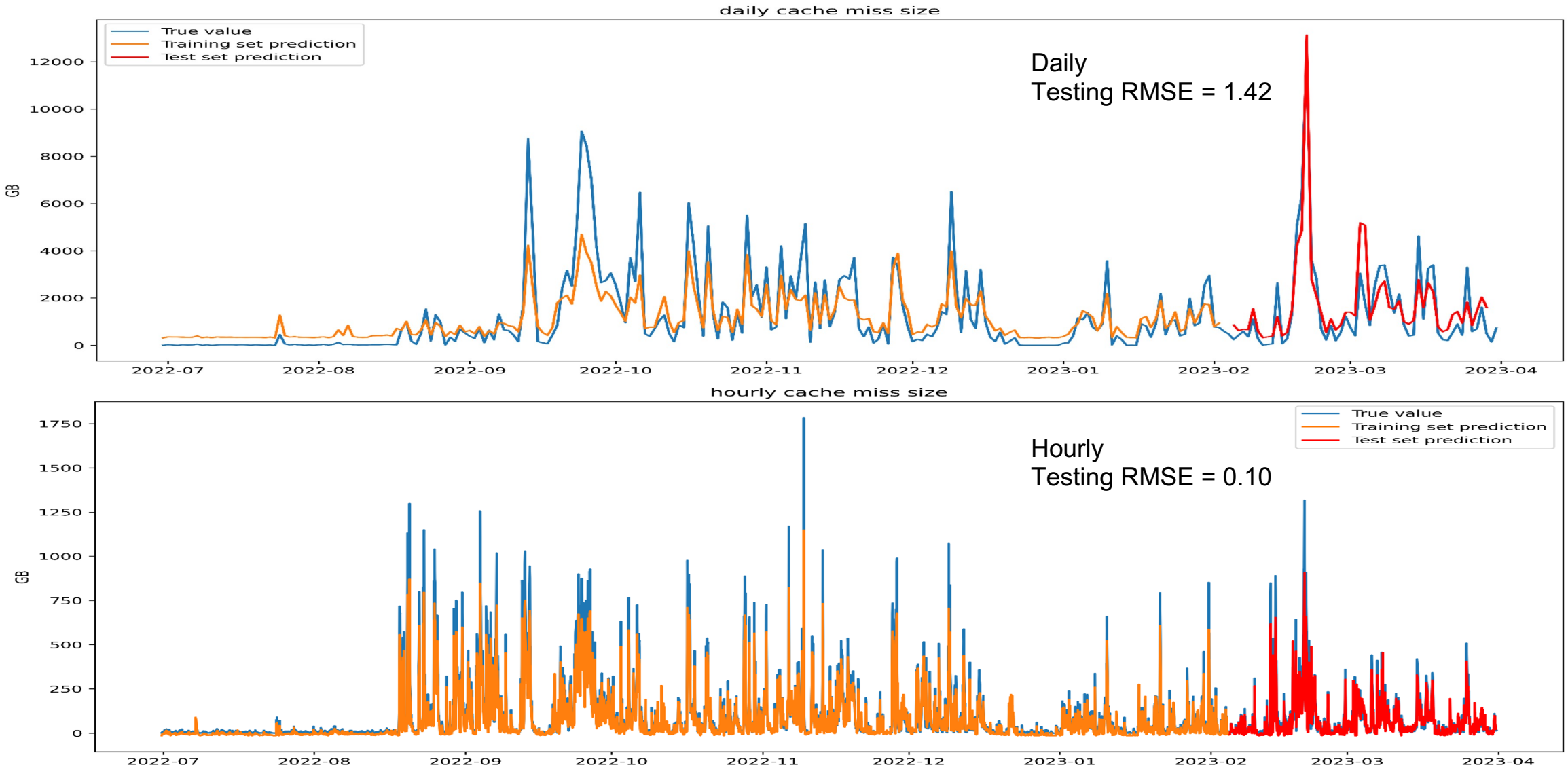




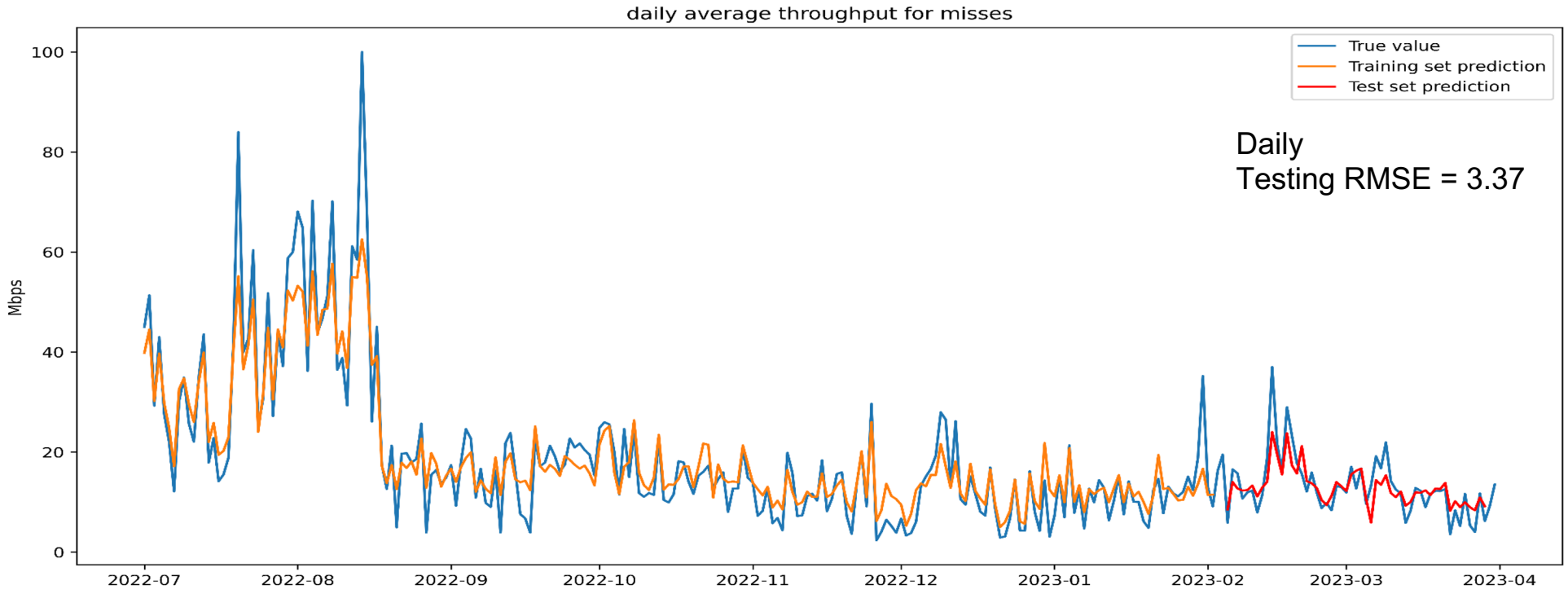
Predicted Volume of Cache Hits



Predicted Volume of Cache Misses

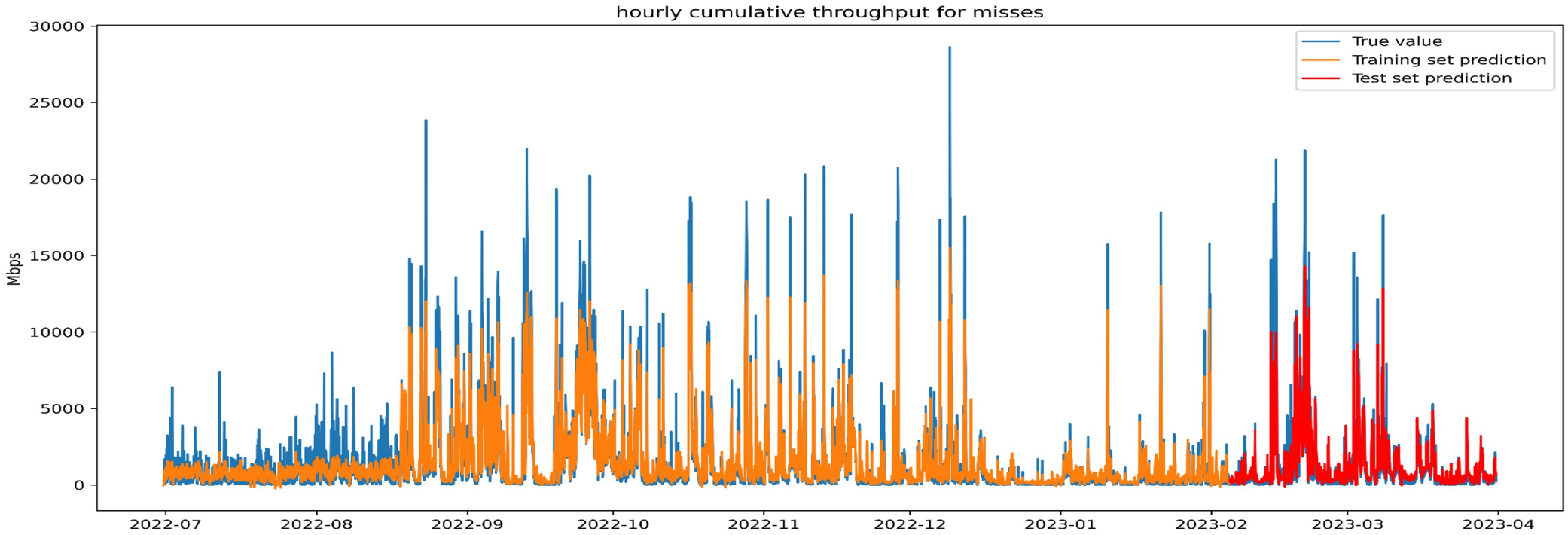


Daily Throughput of Cache Misses from LSTM



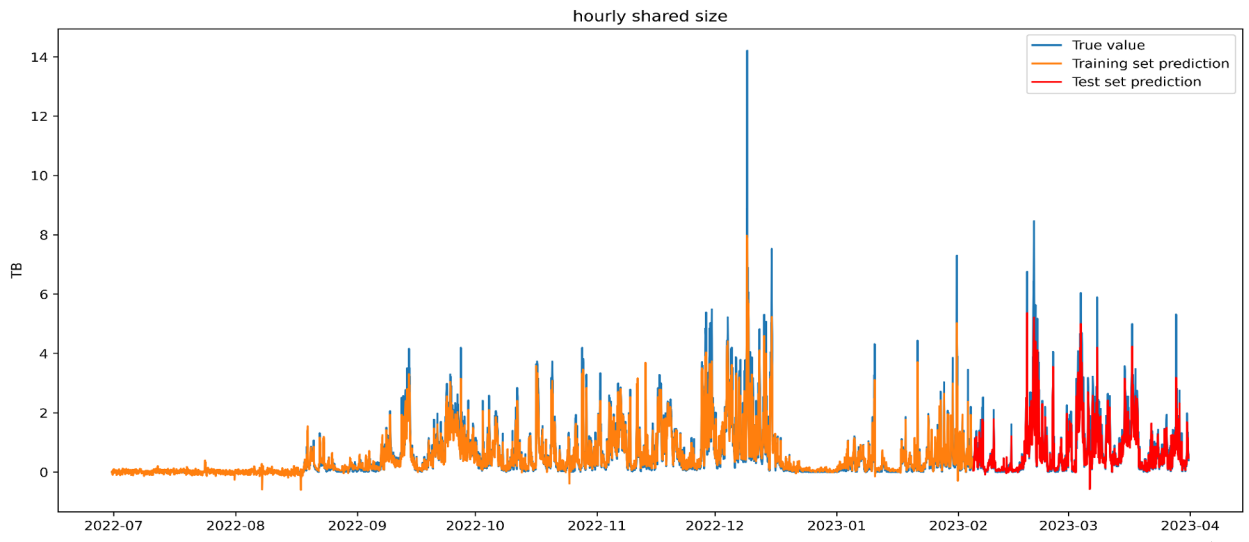
- **Prediction follows actual observations well despite the considerable variation**
- **The average throughput is relatively low due to small files being transferred**

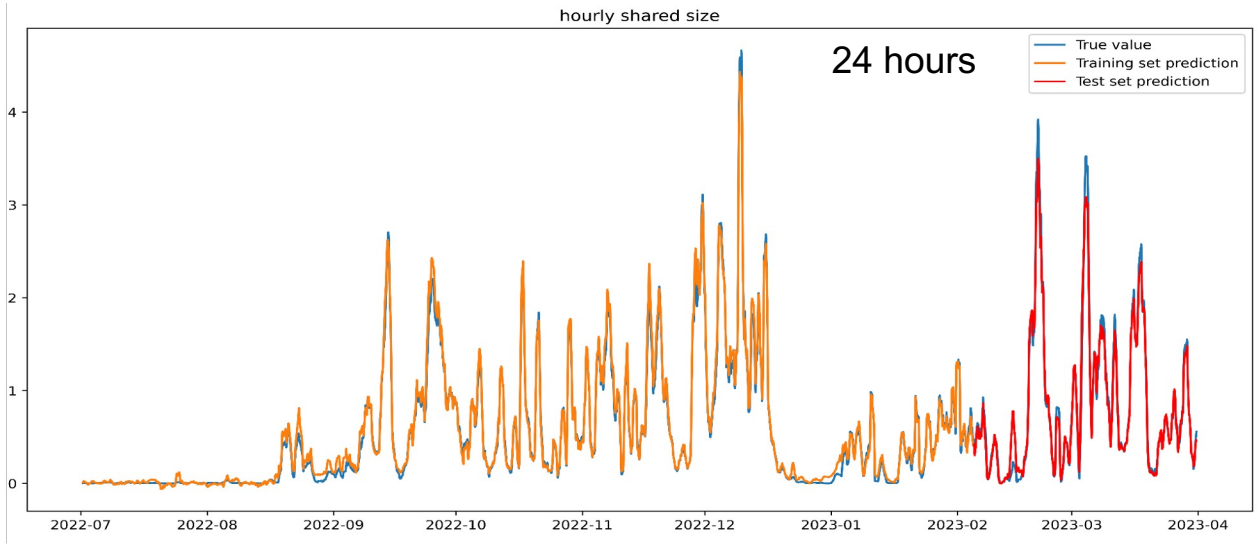
Hourly Cumulative Throughput of Cache Misses



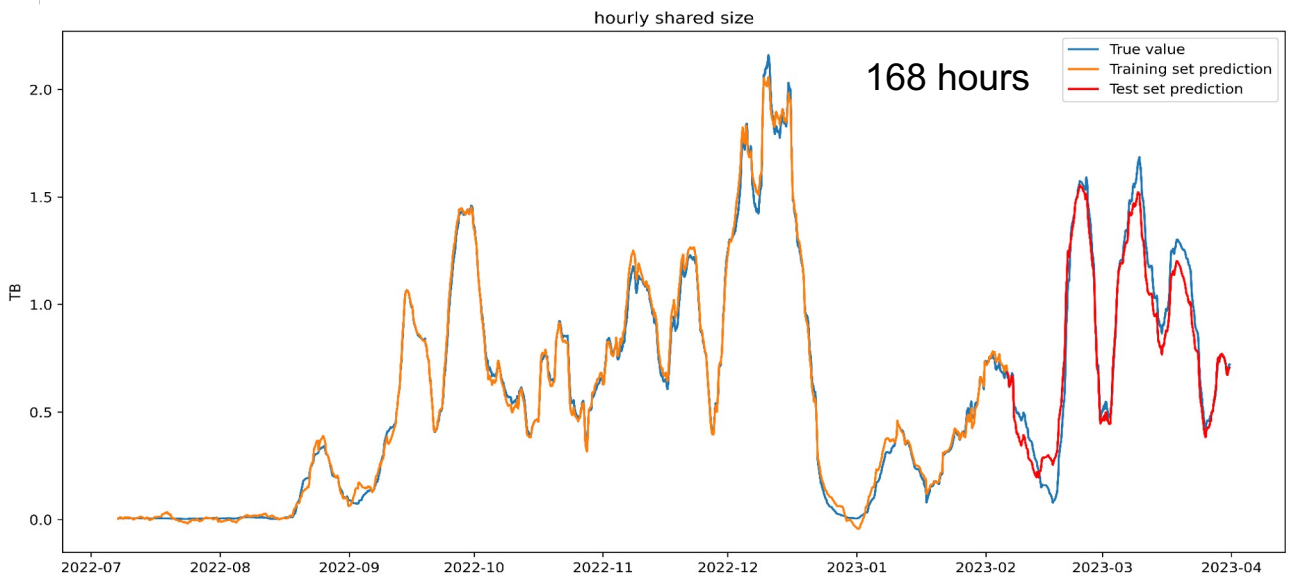
- Cumulative throughput shows an indication of the regional network utilization
- Prediction follows actual observations well despite the considerable variation

Moving Average of Hourly Cache Hits Volume Could Be Modelled More Accurately





- Even though the 24-hour moving averages look like the daily throughput time series, the LSTM predictions match the 24-hour moving averages much better, based on visual inspection.
  - RMSE on the hourly data is 0.1
  - RMSE on the 24-hour moving-averages is 0.021
  - RMSE on the 168-hour moving-averages is 0.013
- LSTM model based on the moving averages is likely to work better.



Summary

- SoCal Repo could serve on average about 81.6% of file requests from its storage cache without remote transfers, and on average 90.2% of bytes requested could be shaed from the cache
  - The average volume of cache hits is about 18.9TB per day
- Explored modeling of the cache and network utilization with the LSTM neural network architecture
  - Tests show that the prediction errors (measured as RMSE) are relatively small
- Our study on understanding the resource characteristics
  - Characterize the trends of network and cache utilization
  - Study the effectiveness of the SoCal Repo in reducing wide-area network traffic
  - Explore the predictability of the resource utilization
  - Study the effectiveness of the cache system for scientific applications
- We plan to study other storage caches currently under deployment to gain better understanding of in-network storage caches.
  - Chicago and Boston caching nodes in collaboration with US CMS
  - Additional caching nodes in London and Amsterdam, working with OSG/OSDF, mainly targeting LIGO and DUNE for transatlantic traffic from the US to Europe
  - Collaboration with ATLAS, DUNE, and LHCb

Acknowledgements

- **Alex Sim, K. John Wu** (Lawrence Berkeley National Lab)
- **Chin Guok, Damian Hazen, Ezra Kissel, Inder Monga, Engineering team, Infrastructure team** (ESnet)
- **Diego Davila, Dmitry Mishin, Fabio Adnrijauskas, Frank Wuerthwein** (Univ. of California at San Diego)
- **Justas Balcas, Harvey Newman** (Caltech)
- **Caitlin Sim, Jack Ruize Han** (Univ. of California at Berkeley)
- **Ellie Copps** (Middlebury College)

- **US Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Contract No. DE-AC02-05CH11231**
- **National Science Foundation through the grants OAC-2030508, OAC-1836650, MPS-1148698, PHY-2121686 and OAC-1541349**
- **US Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internship (SULI) program**
- **Used resources of the National Energy Research Scientific Computing Center (NERSC)**