Forecasting COVID-19 pandemic using an echo state neural network-based framework

J. H. K. Larcher[1], R. G. da Silva[2], M. H. D. M. Ribeiro[2, 3], V. C. Mariani[1,3], L. S. Coelho[2, 4]

gomes.ramon@pucpr.edu.br, mribeiro@utfpr.edu.br, jose.kleinubing@pucpr.edu.br, viviana.mariani@pucpr.br, leandro.coelho@pucpr.br

[1]Mechanical Engineering Graduate Program (PPGEM), Pontifical Catholic University of Parana (PUCPR). Curitiba, PR, Brazil.

[2]Industrial and Systems Engineering Graduate Program (PPGEPS), Pontifical Catholic University of Parana (PUCPR). Curitiba, PR, Brazil.

[3]Department of Mathematics, Federal Technological University of Parana (UTFPR). Pato Branco, PR, Brazil.

[4]Department of Electrical Engineering, Federal University of Parana (UFPR). Curitiba, PR, Brazil.

IJCNN 2021

Overview

1. Introduction
2. Objective
3. Dataset
4. Proposed forecasting framework
5. Findings
6. Conclusion

1. Introduction
2. Objective
3. Dataset
4. Proposed forecasting framework
5. Findings
6. Conclusion

Introduction

New coronavirus identified, currently named Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2), causing the coronavirus disease 2019 (COVID-19).

Brazil has more than 15 million cases and 440 thousand deaths.

Problems:

- Overcrowding of hospital beds;
- Lack of medical supplies to treat patients affected more severely.

Introduction

Short-term forecast allows to:

- Reallocate patients;
- Reapportion resources and staff;
- Alert residents of a determined region in case of a spike.

Models:

- ARIMA
- ESN

1. Introduction
2. Objective
3. Dataset
4. Proposed forecasting framework
5. Findings
6. Conclusion

Objective

- **Evaluate ESNs for epidemiological time series forecasting.**
- **Compare ESN and ARIMA for epidemiological time series forecasting.**
- **Present a ten days horizon forecast framework for COVID-19 cases and deaths in Brazilian states.**

1. Introduction
2. Objective
3. Dataset
4. Proposed forecasting framework
5. Findings
6. Conclusion

Dataset

Aggregated data from Brazilian State Health Offices from February 25 until December 31th, 2020.

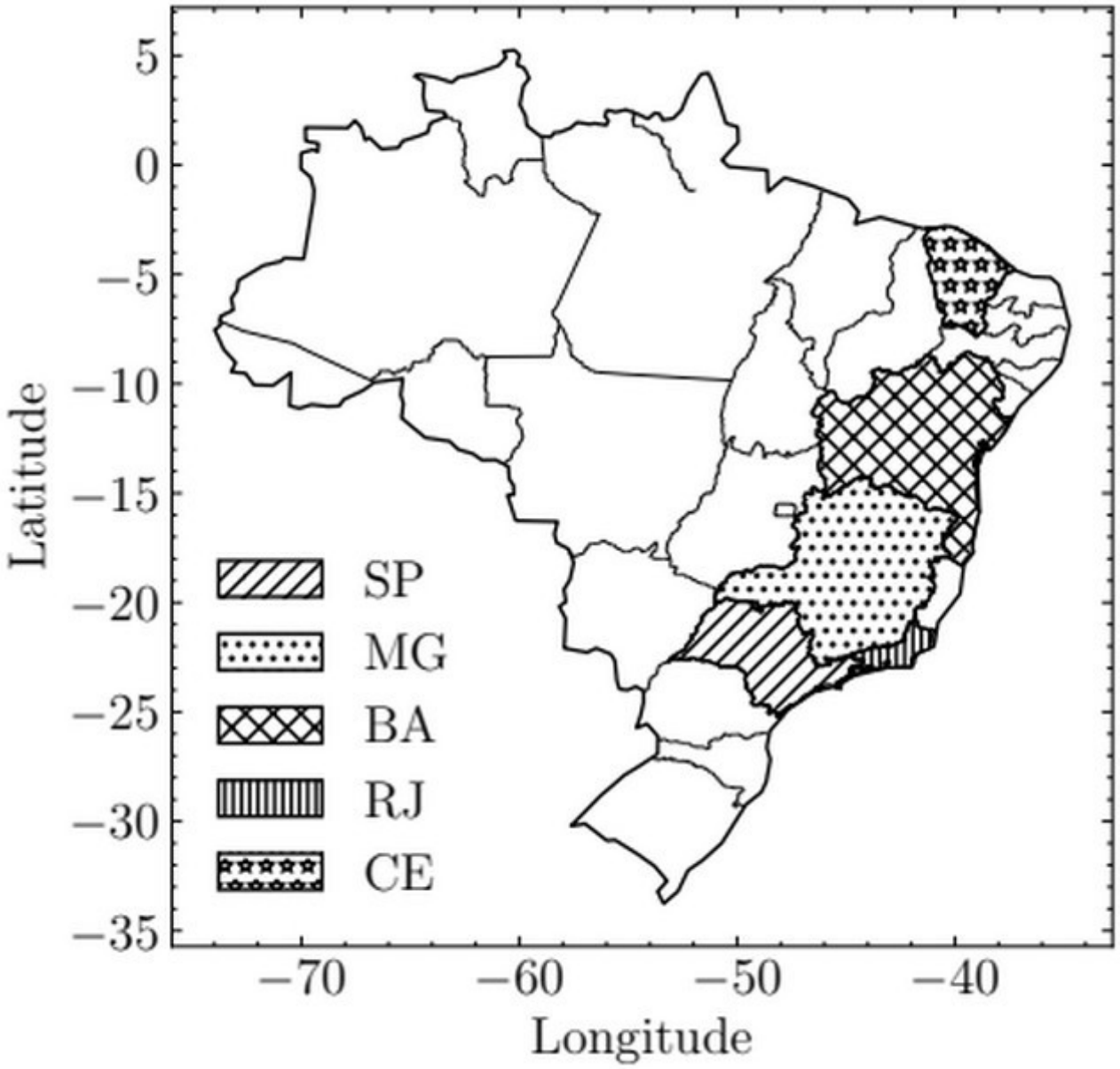
The data is gathered in a collaborative project with various contributors and made available with an API or direct download.

- Brasil.IO

States:

- São Paulo (SP)
- Minas Gerais (MG)
- Bahia (BA)
- Rio de Janeiro (RJ)
- Ceará (CE)

Dataset

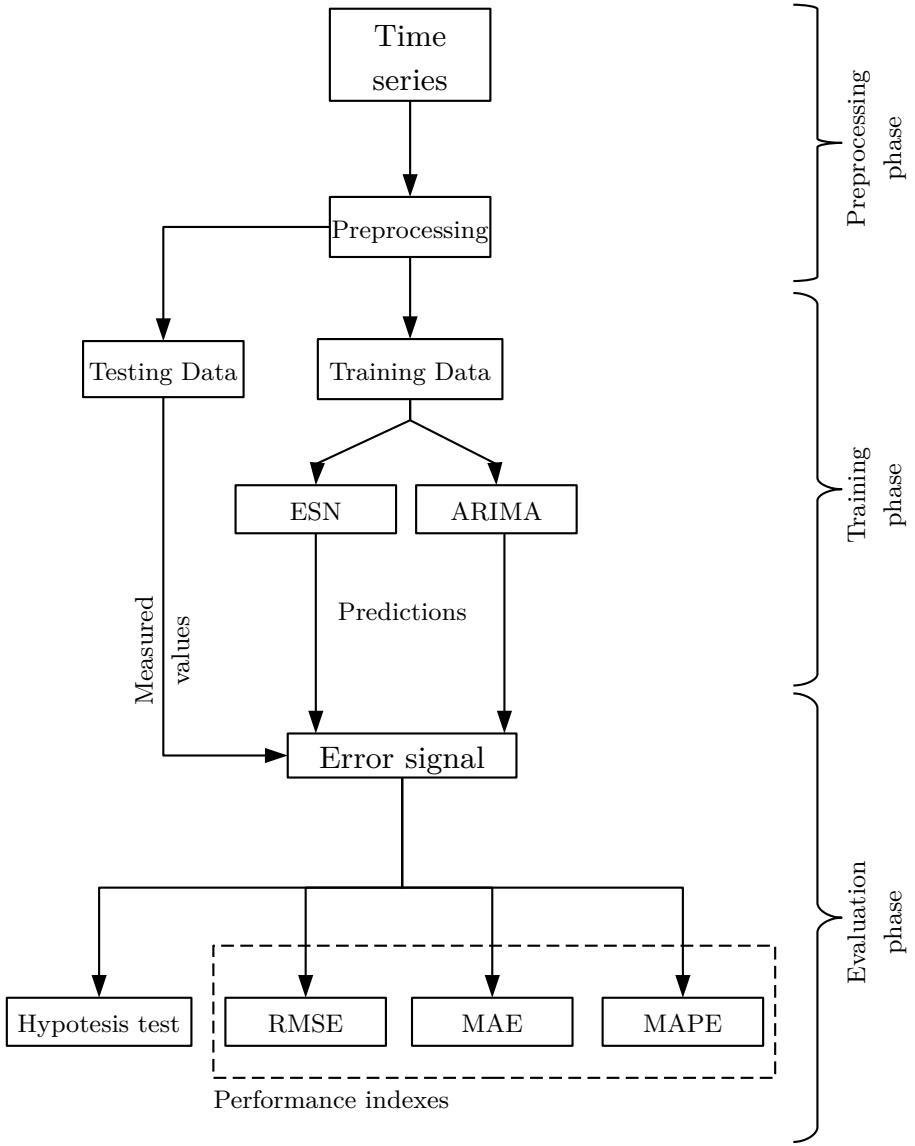

Latitude
5
0
−5
−10
−15
−20
−25
−30
−35
−70
−60
−50
−40
Longitude
SP
MG
BA
RJ
CE

Dataset

State	Number of Samples	First Report	Cases				Deaths			
			mean	min	max	std	mean	min	max	std
BA	301	2020-03-06	186326,99	1	493400	163785,85	3931,36	0	9129	3286,07
CE	291	2020-03-16	162827,01	9	335992	113064,66	6229,72	0	9993	3605,47
MG	299	2020-03-08	178409,63	1	542909	171296,27	4281,69	0	11902	4088,18
RJ	302	2020-03-05	173134,01	1	434648	135461,88	12169,91	0	25530	8428,85
SP	311	2020-02-25	575771,26	1	1462297	495462,73	21513,76	0	46717	16303,05

1. Introduction
2. Objective
3. Dataset
4. **Proposed forecasting framework**
5. Findings
6. Conclusion

Proposed forecasting framework



Proposed forecasting framework

1) **Train-test split.**
2) **Train ESN.**
    i. 7 lagged days are used.
    ii. Gridsearch for hyperparameter selection.
3) **ARIMA fitted.**
    i. Auto-arima.
4) **10-day-ahead forcast produced.**

Proposed forecasting framework

5) Evaluate performance metrics
   i. MAE
   ii. RMSE
   iii. MAPE
6) Wilcoxon Signed-Rank test
7) Visual inspection (plots)

Proposed forecasting framework

- **ESN Hyperparameters**

State	ESN	Size	α	ρ	λ	Density
SP	Cases	200	0.8	0.1	10	0.1
	Deaths	200	0.91	0.1	10	0.1
MG	Cases	200	0.58	0.52	0.24	0.1
	Deaths	200	1	0.1	1.25	0.1
BA	Cases	200	0.61	0.36	1.17	0.1
	Deaths	200	0.29	0.26	4.24	0.1
RJ	Cases	200	0.56	0.1	10	0.1
	Deaths	200	0.61	0.65	10	0.1
CE	Cases	200	0.84	0.1	0.7	0.1
	Deaths	200	0.73	1.22	10	0.1

Proposed forecasting framework

- **ARIMA parameters**

States	Cases							Deaths						
	p	d	q	P	D	Q	m	p	d	q	P	D	Q	m
SP	1	1	1	0	1	1	7	1	1	2	0	1	2	7
MG	1	1	1	2	1	0	7	1	1	1	0	1	1	7
BA	0	2	1	0	0	2	7	0	2	1	-	-	-	-
RJ	0	2	1	0	0	2	7	1	1	1	0	1	1	7
CE	0	2	1	1	0	1	7	1	2	2	1	0	1	7

1. Introduction
2. Objective
3. Dataset
4. Proposed forecasting framework
5. Findings
6. Conclusion

Findings - SP

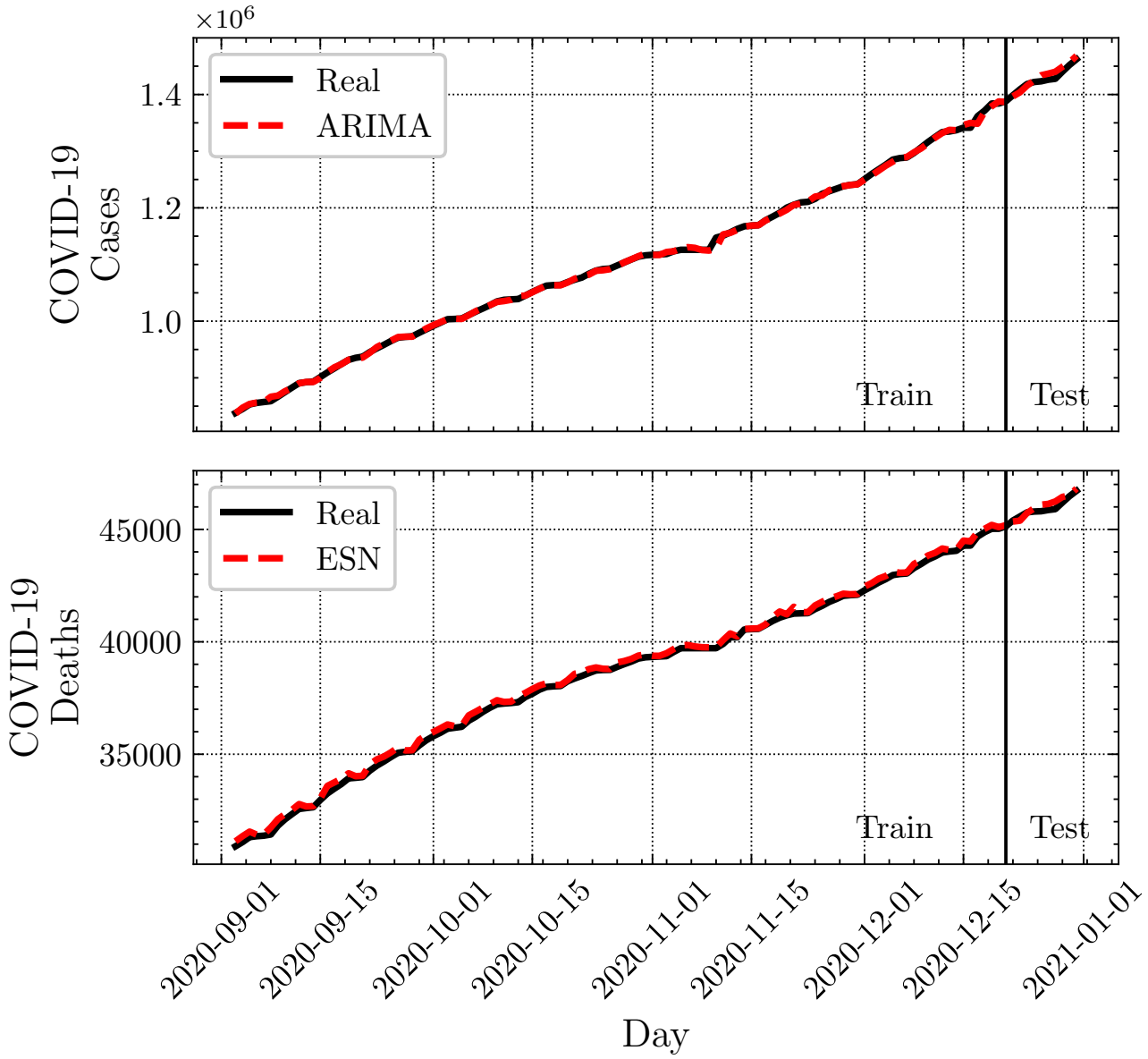
Cases

- ESN MAE: 7,538.23
- **ARIMA MAE: 6,562.83**
- Difference: 12.94%

Deaths

- **ESN MAE: 169.34**
- ARIMA MAE: 193.07
- Difference: 12.24%

Findings - SP


×10^6
1.4
1.2
1.0
COVID-19 Cases
Real
ARIMA
Train
Test
45000
40000
35000
COVID-19 Deaths
Real
ESN
Train
Test
2020-09-01
2020-09-15
2020-10-01
2020-10-15
2020-11-01
2020-11-15
2020-12-01
2020-12-15
2021-01-01
Day

Findings – MG

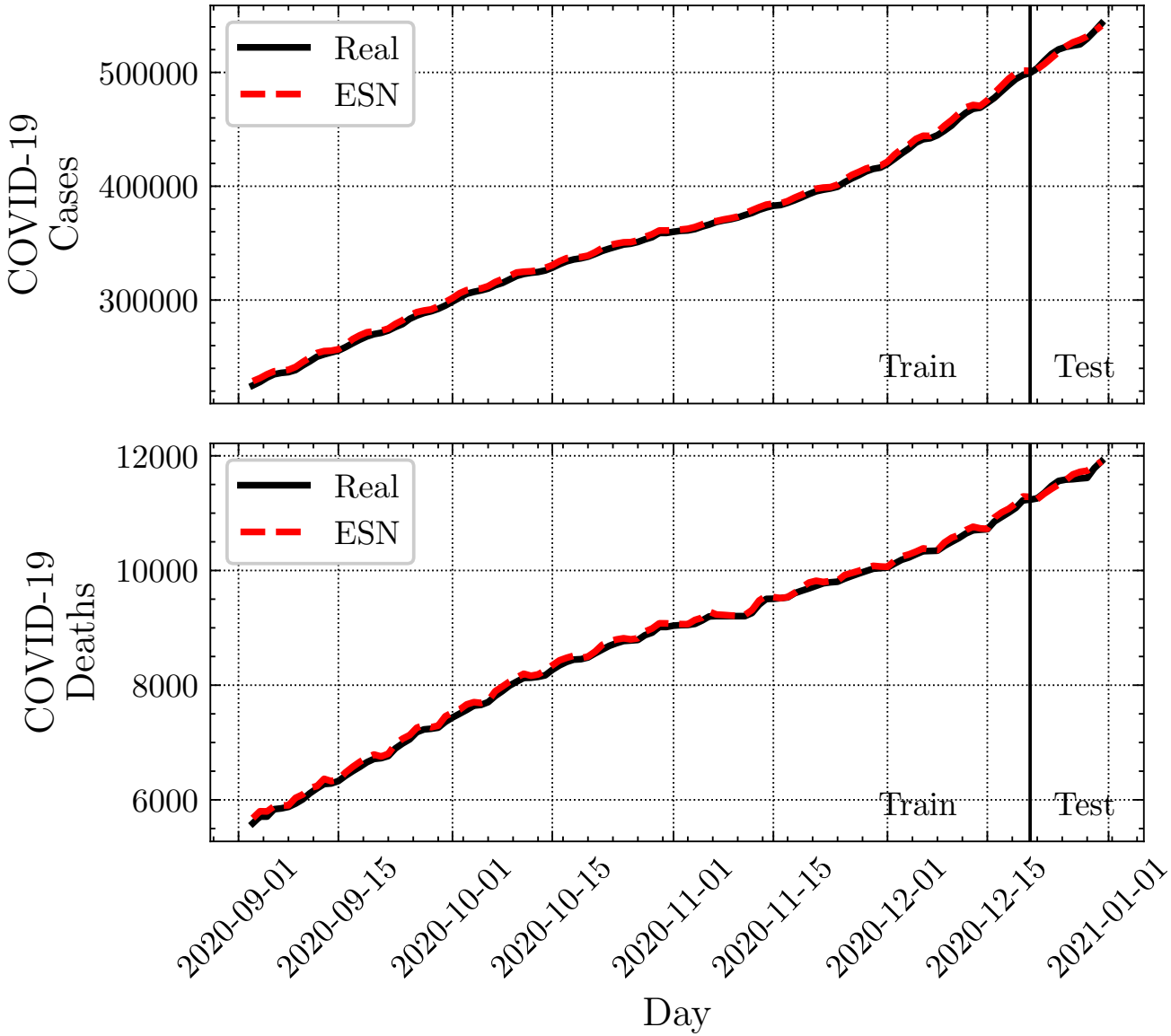
Cases

- **ESN MAE: 2088.39**
- ARIMA MAE: 4122.39
- Difference: 49.34%

Deaths

- **ESN MAE: 51.13**
- ARIMA MAE: 60.78
- Difference: 15.88%

Findings - MG


Real
ESN
COVID-19 Cases
500000
400000
300000
Train
Test
Real
ESN
COVID-19 Deaths
12000
10000
8000
6000
Train
Test
2020-09-01
2020-09-15
2020-10-01
2020-10-15
2020-11-01
2020-11-15
2020-12-01
2020-12-15
2021-01-01
Day

Findings – BA

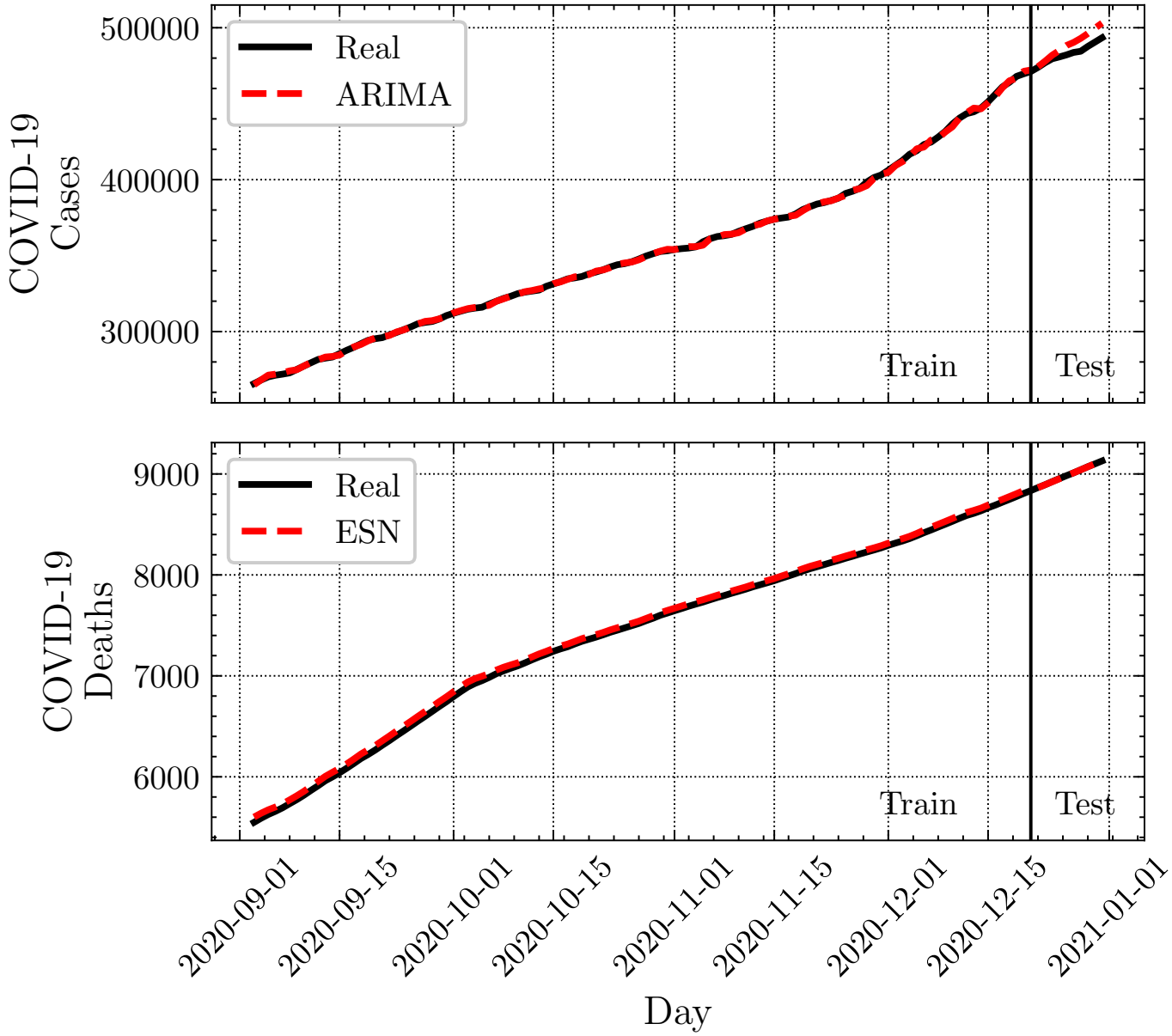
Cases

- ESN MAE: 5827.98
- **ARIMA MAE: 5662.76**
- Difference: 28.34%

Deaths

- **ESN MAE: 1.14**
- ARIMA MAE: 5.08
- Difference: 77.56%

Findings - BA


Real
ARIMA
COVID-19 Cases
500000
400000
300000
Train
Test
Real
ESN
COVID-19 Deaths
9000
8000
7000
6000
Train
Test
2020-09-01
2020-09-15
2020-10-01
2020-10-15
2020-11-01
2020-11-15
2020-12-01
2020-12-15
2021-01-01
Day

Findings – RJ

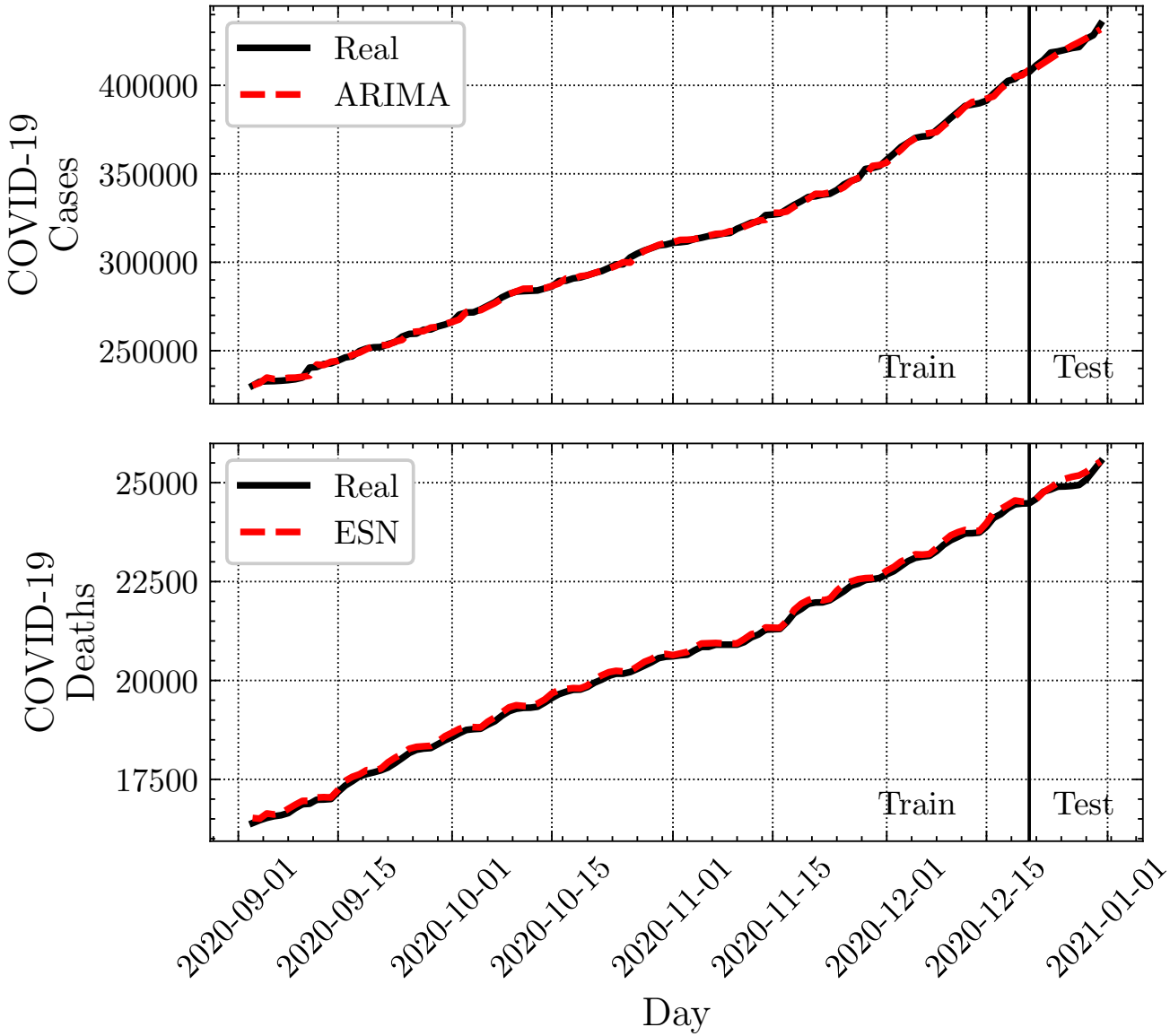
Cases

- ESN MAE: 2739.91
- **ARIMA MAE: 1620.13**
- Difference: 40.84%

Deaths

- **ESN MAE: 113.92**
- ARIMA MAE: 118.57
- Difference: 3.92%

Findings - RJ


Real
ARIMA
COVID-19 Cases
400000
350000
300000
250000
Train
Test
Real
ESN
COVID-19 Deaths
25000
22500
20000
17500
Train
Test
2020-09-01
2020-09-15
2020-10-01
2020-10-15
2020-11-01
2020-11-15
2020-12-01
2020-12-15
2021-01-01
Day

Findings – CE

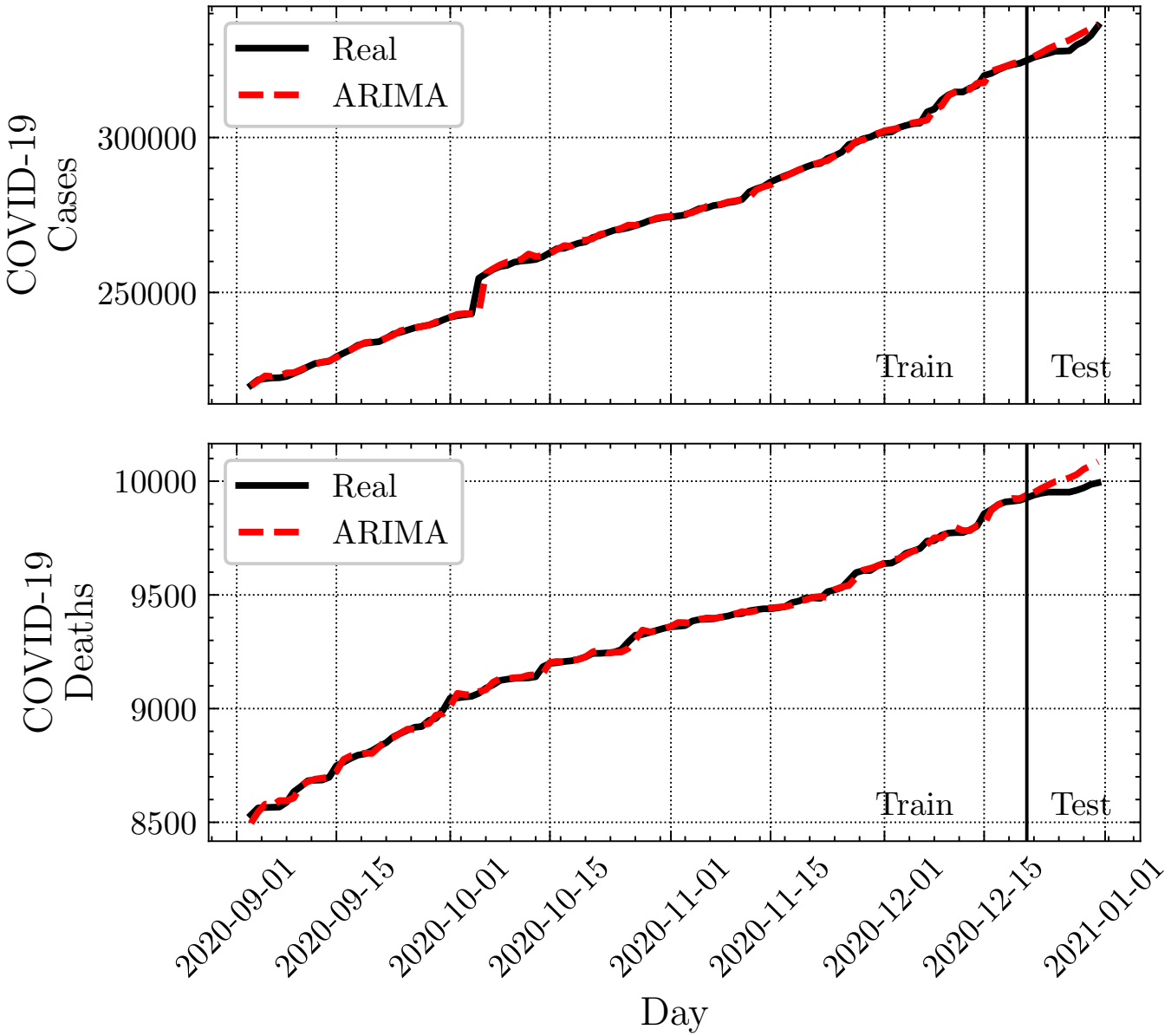
Cases

- ESN MAE: 2882.43
- **ARIMA MAE: 1924.41**
- Difference: 49.34%

Deaths

- ESN MAE: 84.76
- **ARIMA MAE: 55.17**
- Difference: 3.92%

Findings - CE


Real
ARIMA
COVID-19 Cases
300000
250000
Train
Test
Real
ARIMA
COVID-19 Deaths
10000
9500
9000
8500
Train
Test
2020-09-01
2020-09-15
2020-10-01
2020-10-15
2020-11-01
2020-11-15
2020-12-01
2020-12-15
2021-01-01
Day

Findings - Metrics

State	Metric	Cases		Deaths	
		ESN	ARIMA	ESN	ARIMA
SP	RMSE	8442.37	**7477.63**	**201.57**	227.07
	MAE	7538.23	**6562.83**	**169.34**	193.07
	MAPE	0.53%	**0.46%**	**0.37%**	0.42%
MG	RMSE	**2425.43**	4994.78	**66.48**	77.09
	MAE	**2088.39**	4122.39	**51.13**	60.78
	MAPE	**0.4%**	0.78%	**0.44%**	0.52%
BA	RMSE	6614.05	**6534.67**	**1.41**	5.68
	MAE	5827.98	**5662.76**	**1.14**	5.08
	MAPE	1.2%	**1.16%**	**0.01%**	0.06%
RJ	RMSE	3279.99	**1932.4**	145.37	**143.51**
	MAE	2739.91	**1620.13**	**113.92**	118.57
	MAPE	0.65%	**0.38%**	**0.46%**	0.47%
CE	RMSE	3121.06	**2199.06**	92.99	**61.11**
	MAE	2882.43	**1924.41**	84.76	**55.17**
	MAPE	0.87%	**0.58%**	0.85%	**0.55%**

Findings – Wilcoxon test

State	Target	V-value	p-value	H_0
SP	Cases	55	0,001953125	Rejected
	Deaths	55	0,001953125	Rejected
MG	Cases	55	0,001953125	Rejected
	Deaths	55	0,001953125	Rejected
BA	Cases	15	0,232421875	Accepted
	Deaths	0	0,001953125	Rejected
RJ	Cases	55	0,001953125	Rejected
	Deaths	40	0,232421875	Accepted
CE	Cases	0	0,001953125	Rejected
	Deaths	0	0,001953125	Rejected

1. Introduction
2. Objective
3. Dataset
4. Proposed forecasting framework
5. Findings
6. Conclusion

Conclusion

- ARIMA and ESN show comparable results.
- For the number of cases, ARIMA performed better for most states.
- For the number of deaths, ESN showed better results for most states.

Conclusion

- The state of MG was the only one where ESNs showed a better result than ARIMA regarding the number of cases.
- Regarding the number of deaths, ARIMA performed better just for CE, while ESN achieved better results for every other state when MAE is considered.

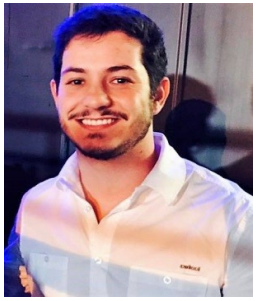
Conclusion

- The selected parameters via grid-search might not be optimal
  - Finding another tuning framework using metaheuristics might improve these results.
- The use of cumulative cases and deaths introduces a trend in the development of the series.
  - The comparison with MAE and RMSE would be the same using new cases and new deaths but models could perform differently fitting in thease.
- The smaller values of spectral radius of the ESNs fitted for the number of cases can show dependency on the recent values, performing worst when not taking into account the older values as much as new ones.

Team

J.H.K. Larcher
Master Student

Pontifical Catholic University of Parana (PUC-PR).



M. H. D. M. Ribeiro
Phd Candidate

Pontifical Catholic University of Parana (PUC-PR).

R.G. da Silva
Phd Candidate

Pontifical Catholic University of Parana (PUC-PR).

Dr[a]. V. C. Mariani
Full Professor

Pontifical Catholic University of Parana (PUC-PR).

Dr. L.S. Coelho
Full Professor

Pontifical Catholic University of Parana (PUC-PR).

Obrigado!