Spatio-Temporal Time Series Modelling *20/12/2024*

With a Focus on Coral Reef Benthic Group Shifting in the Great Barrier Reef

By: Daud Waqas

I started off this project to tackle a fairly obscure and unexplored topic: *understanding and predicting shifts in* ***coral reef benthic cover*** *using data-driven approaches*. This was also my attempt to try and "handle" a much larger and unoptimised dataset. The data, initially provided at yearly intervals, represented the distribution of coral reef covers *(specifically* ***hard coral cover, soft coral cover*** *and* ***algae cover****)* across various reef systems throughout the Great Barrier Reef; due to the massive ecological size of the reef, I made an assumption that there would be distinct temporal patterns based on the spatial location that the data was sampled from. This was the basis of a study that was not just focused on fitting **time series**, but also on building a **spatio-temporal model** that would take those location-based dependencies into account.

The data was compiled from the **AIMS** and **eReefs** platform, both of which are managed by the Australian government; **AIMS** was used to source the primary data, and **eReefs** was used to source the auxiliary data. However, time constraints only allowed me to use the data from the **AIMS** platform *(ie. the primary data)*. The data was made up of spatio-temporal indexes *(latitude, longitude, date, etc)* alongside 4 cover values:

- **SOFT CORAL_COVER**
- **HARD CORAL_COVER**
- **ALGAE_COVER**
- **OTHER_COVER**

Most of the data came in an unusable format, and had to be *heavily augmented* to be used properly by the subsequent models. You should be able to find most of the context regarding these augmentations in the **`data_collection.ipynb`** notebook, which I wasn't able to knit due to the fact that it includes a HTML map (labelled as `folium_click_map.html`). This click map has also been included alongside this report, and you can observe how the details for each selected spatial location was given through the console *(within browser DevTools)* of that HTML file.

All of the modelling took place with the **primary data**, which *originally came sampled in the yearly format* from the **1990s** to **2004**. Since yearly data lacked the granularity needed for temporal analysis, I resampled it into weekly intervals using spline-based interpolation. This was a fairly safe augmentation, since biological metrics over square kilometres of area (like benthic group coverage) don't tend to have massive fluctuations or show volatility in that regard. To mimic **natural variability** and prevent models from fitting from the interpolated data, I added **Gaussian noise** and **scaled the data randomly** within a controlled range. This made sure that any models (especially neural networks) won't derive patterns from the interpolation.

To test temporal dependencies, I started with **univariate models** like **Auto ARIMA** and **TBATS**. These methods failed in capturing any trends at all; the average **nRMSE** result, which was close to **1** for both models *(exactly* ***0.98*** *and* ***0.94*** *respectively)* made it so that the model was completely off *(note that an acceptable nRMSE starts at around* ***0.2*** *or lower)*. The heat map showed how there were massive variations, with some values **above 1** and some closer to **0.05** *(which would have been considered a good fit if those results across spatial locations were consistent)*. Overall, the models were simply lacking and overly inconsistent.

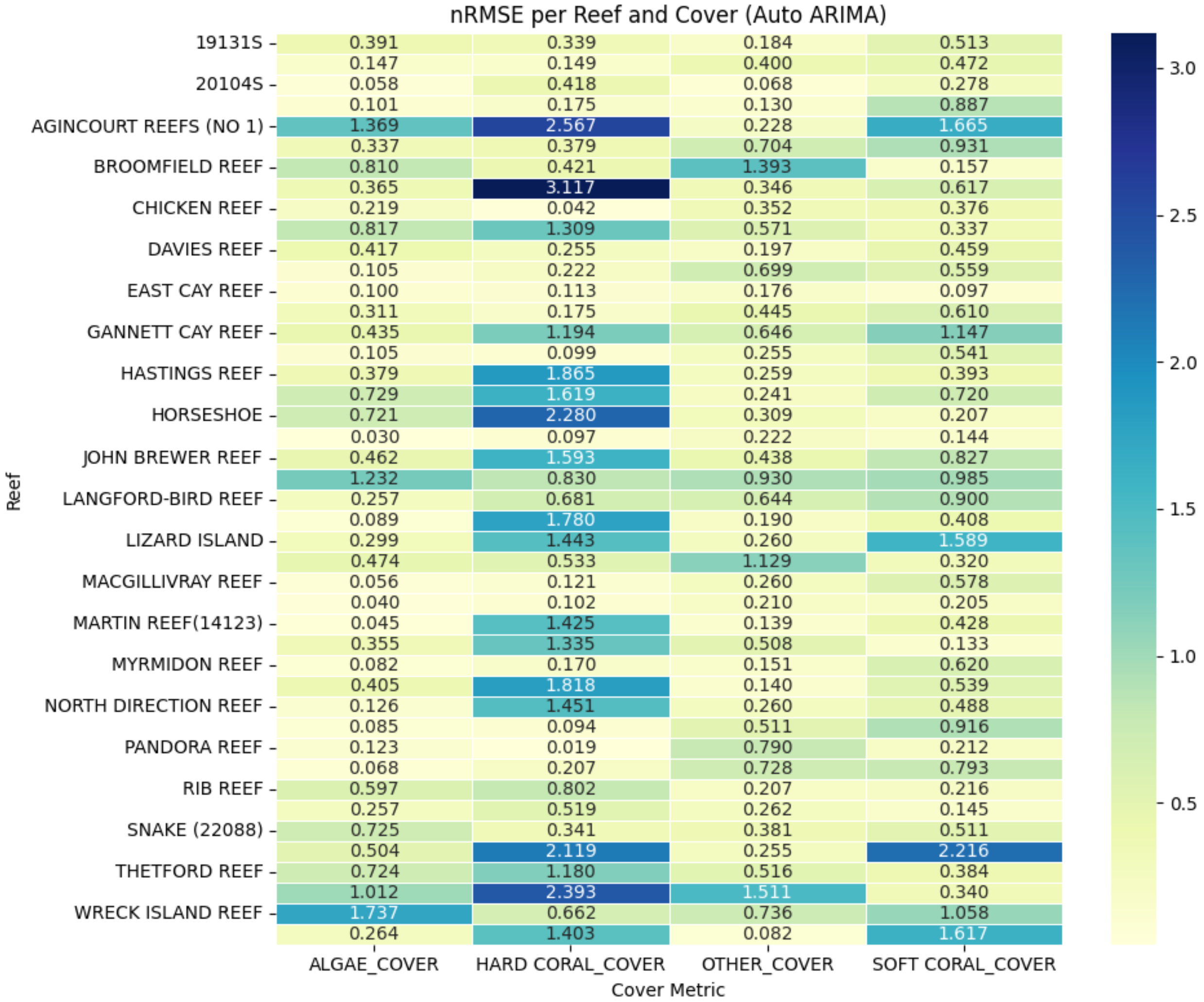
My choice for **multivariate time series model** was a **Temporal Convolutional Network (TCN)**. This neural network model leveraged the values of all 4 cover values to predict for the next time iteration. The **TCN** outperformed the **univariate models**, and quite aggressively so achieving an average **nRMSE** of **0.116**. This is a score that is considered a **"moderate fit"**, and goes to show that the granularity of the **neural network** within **TCN** did make up for some of the weaknesses of the data, such as its **multicollinearity**. Please note that the score was derived on an **80/20 split** on the data, where the **latter 20% of the temporal range is the test range**.

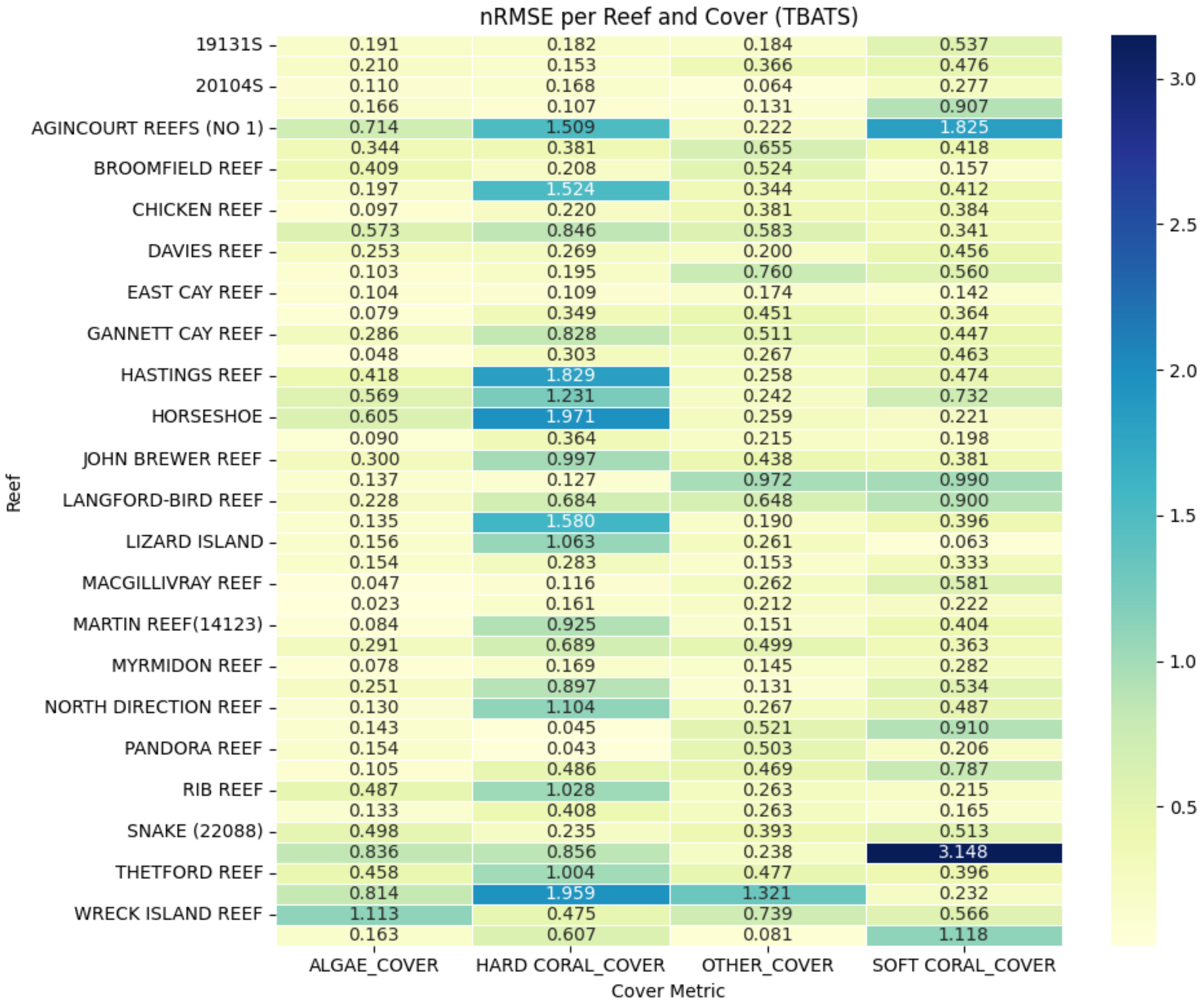
The main part of this project was the implementation of a **Spatio-Temporal Graph Neural Network (ST-GNN)**. By *building a spatial graph where reef locations were connected based on proximity*, I would have been (in theory) able to capture the temporal relationships between the cover values **relative to different reef locations** along the spatial map. This part of the project was quite a struggle, since I couldn't find any proper documentation on how to best fit ST-GNNs. There were many papers out there for the fitting of ST-GNNs for traffic monitoring, but not that many for sensitive environmental-based modelling *(admittedly, I spent too long trying to straighten out this part of the notebook, and had to redesign multiple times to get the results I wanted. I had fairly inadequate results, but this the afternoon before the submission of this assignment I managed to figure out how to implement this properly)*.

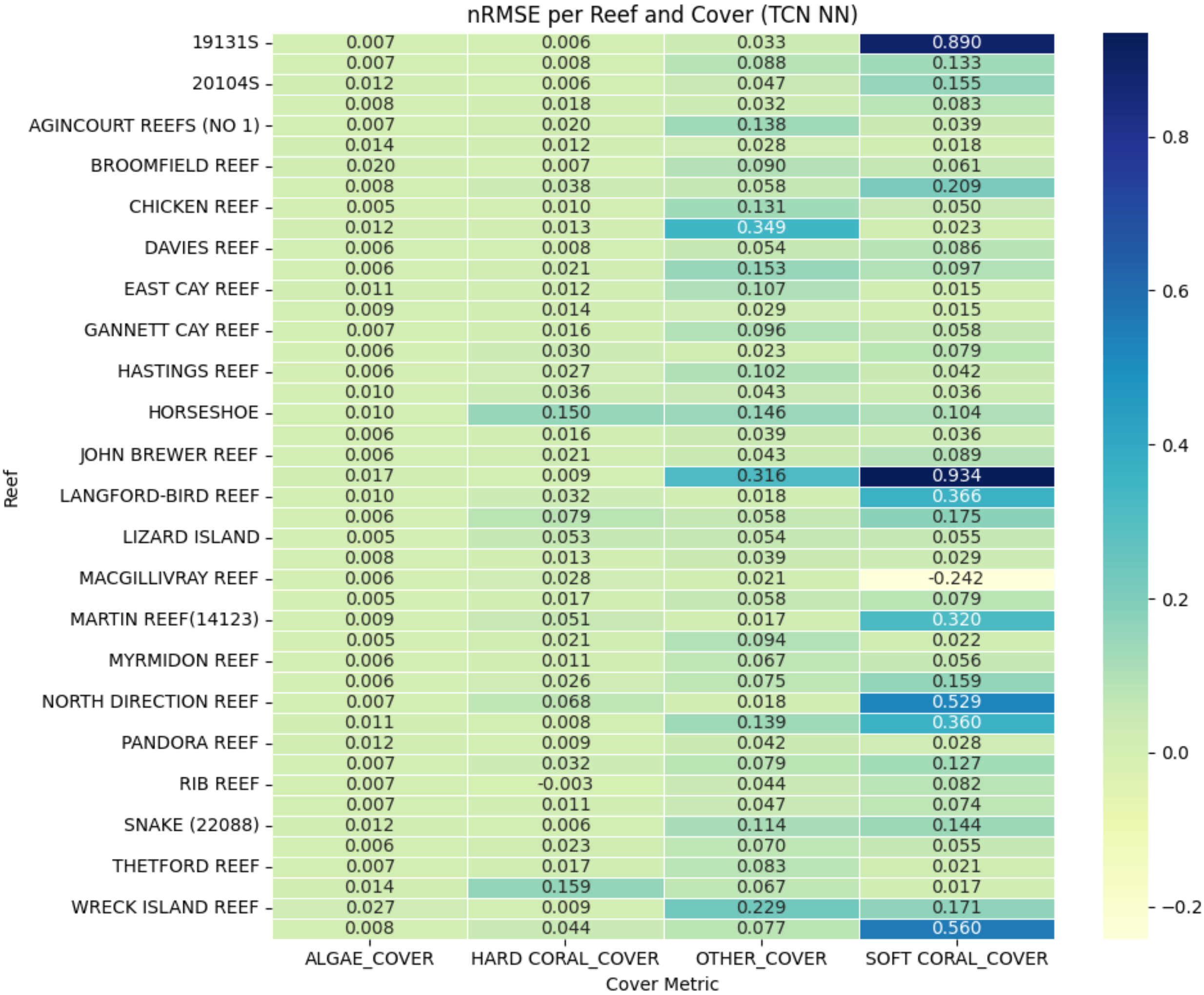
The **temporal component** of the ST-GNN was handled by a **TCN**, keeping consistent to the multivariate model done beforehand. This **ST-GNN** combined spatial correlations between reefs with temporal patterns provided by the **TCN** to provide a holistic view of the system. Once I did manage to figure out how to implement all this properly, the **nRMSE** values **improved significantly** for several cover types, and managed to not just beat **TCN** but also reduce the sensitivity of the more hard-to-predict cover types in particular reefs. The average **nRMSE** ended up being **0.033**, which is considered a good fit according to nRMSE standards *(and far better than the 0.116 of the TCN)*.

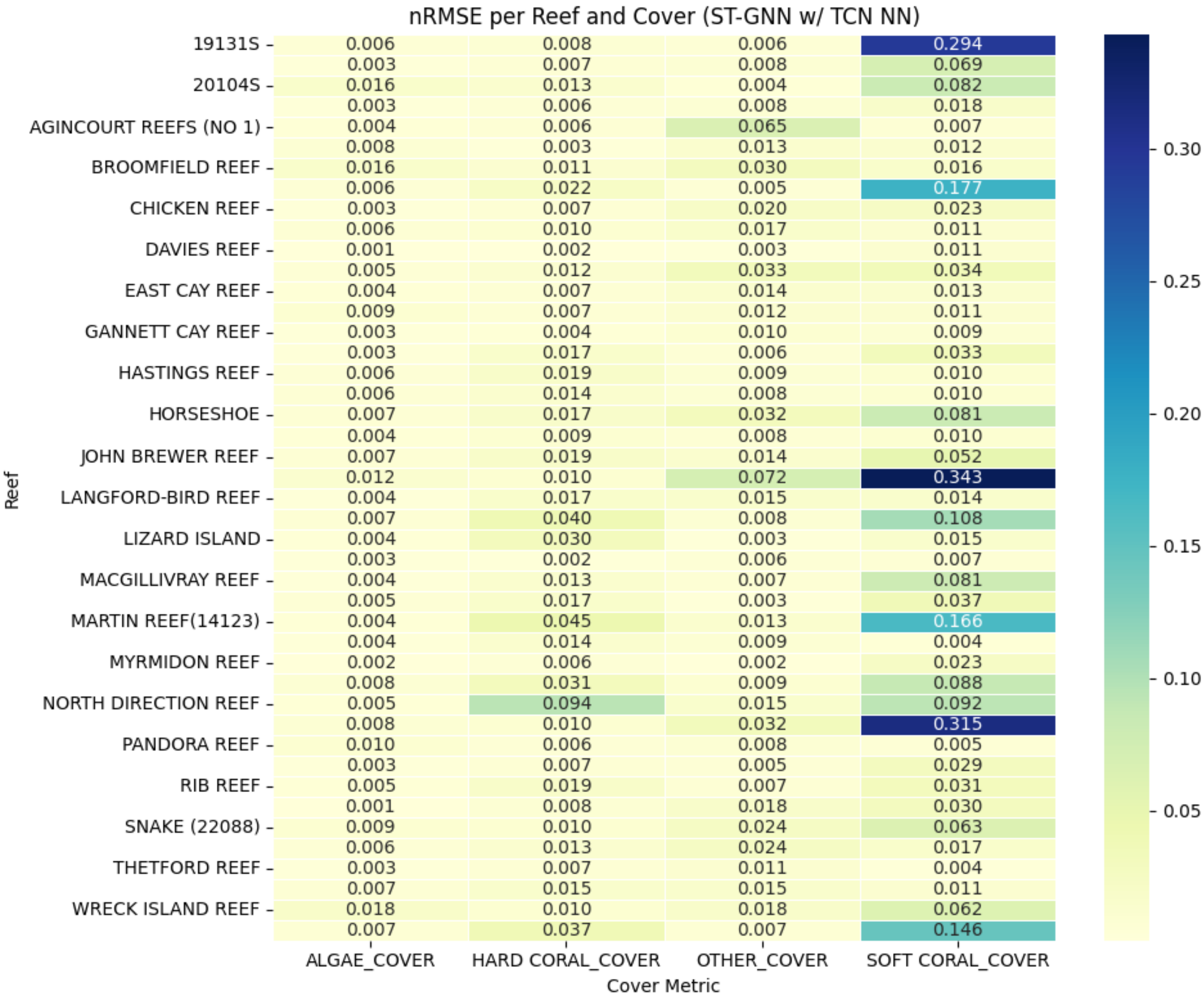
While the deep learning models showed promising results, there are still some areas of interest left around this particular topic. A main issue was the implementation of the **auxiliary data**; it ends up that implementing varying feature sets into **neural networks** is a large ongoing subject with not that much knowledge surrounding it, ***especially for ST-GNNs where documentation was already sparse***. The paper which did initially fuel my interest, **Spatial-temporal graph neural networks for groundwater data**, uses a quite a complicated setup for their auxiliary data; and for the most part, it involves the combination of **multiple different ST-GNNs knitted together for a custom-purposed solution**. After learning how to interpret and build models at a more mathematical and case-specific level, I may come back to this project in the future.

All the **heat maps** of the **nRMSE results** sorted by **reef** and **cover type** may be found in the below section, including the **graph network** of the **ST-GNN**.


nRMSE per Reef and Cover (Auto ARIMA)
Reef
19131S
20104S
AGINCOURT REEFS (NO 1)
BROOMFIELD REEF
CHICKEN REEF
DAVIES REEF
EAST CAY REEF
GANNETT CAY REEF
HASTINGS REEF
HORSESHOE
JOHN BREWER REEF
LANGFORD-BIRD REEF
LIZARD ISLAND
MACGILLIVRAY REEF
MARTIN REEF(14123)
MYRMIDON REEF
NORTH DIRECTION REEF
PANDORA REEF
RIB REEF
SNAKE (22088)
THETFORD REEF
WRECK ISLAND REEF
0.391 0.339 0.184 0.513
0.147 0.149 0.400 0.472
0.058 0.418 0.068 0.278
0.101 0.175 0.130 0.887
1.369 2.567 0.228 1.665
0.337 0.379 0.704 0.931
0.810 0.421 1.393 0.157
0.365 3.117 0.346 0.617
0.219 0.042 0.352 0.376
0.817 1.309 0.571 0.337
0.417 0.255 0.197 0.459
0.105 0.222 0.699 0.559
0.100 0.113 0.176 0.097
0.311 0.175 0.445 0.610
0.435 1.194 0.646 1.147
0.105 0.099 0.255 0.541
0.379 1.865 0.259 0.393
0.729 1.619 0.241 0.720
0.721 2.280 0.309 0.207
0.030 0.097 0.222 0.144
0.462 1.593 0.438 0.827
1.232 0.830 0.930 0.985
0.257 0.681 0.644 0.900
0.089 1.780 0.190 0.408
0.299 1.443 0.260 1.589
0.474 0.533 1.129 0.320
0.056 0.121 0.260 0.578
0.040 0.102 0.210 0.205
0.045 1.425 0.139 0.428
0.355 1.335 0.508 0.133
0.082 0.170 0.151 0.620
0.405 1.818 0.140 0.539
0.126 1.451 0.260 0.488
0.085 0.094 0.511 0.916
0.123 0.019 0.790 0.212
0.068 0.207 0.728 0.793
0.597 0.802 0.207 0.216
0.257 0.519 0.262 0.145
0.725 0.341 0.381 0.511
0.504 2.119 0.255 2.216
0.724 1.180 0.516 0.384
1.012 2.393 1.511 0.340
1.737 0.662 0.736 1.058
0.264 1.403 0.082 1.617
ALGAE_COVER
HARD CORAL_COVER
OTHER_COVER
SOFT CORAL_COVER
Cover Metric
3.0
2.5
2.0
1.5
1.0
0.5


nRMSE per Reef and Cover (TBATS)
Reef
19131S
20104S
AGINCOURT REEFS (NO 1)
BROOMFIELD REEF
CHICKEN REEF
DAVIES REEF
EAST CAY REEF
GANNETT CAY REEF
HASTINGS REEF
HORSESHOE
JOHN BREWER REEF
LANGFORD-BIRD REEF
LIZARD ISLAND
MACGILLIVRAY REEF
MARTIN REEF(14123)
MYRMIDON REEF
NORTH DIRECTION REEF
PANDORA REEF
RIB REEF
SNAKE (22088)
THETFORD REEF
WRECK ISLAND REEF
0.191 0.182 0.184 0.537
0.210 0.153 0.366 0.476
0.110 0.168 0.064 0.277
0.166 0.107 0.131 0.907
0.714 1.509 0.222 1.825
0.344 0.381 0.655 0.418
0.409 0.208 0.524 0.157
0.197 1.524 0.344 0.412
0.097 0.220 0.381 0.384
0.573 0.846 0.583 0.341
0.253 0.269 0.200 0.456
0.103 0.195 0.760 0.560
0.104 0.109 0.174 0.142
0.079 0.349 0.451 0.364
0.286 0.828 0.511 0.447
0.048 0.303 0.267 0.463
0.418 1.829 0.258 0.474
0.569 1.231 0.242 0.732
0.605 1.971 0.259 0.221
0.090 0.364 0.215 0.198
0.300 0.997 0.438 0.381
0.137 0.127 0.972 0.990
0.228 0.684 0.648 0.900
0.135 1.580 0.190 0.396
0.156 1.063 0.261 0.063
0.154 0.283 0.153 0.333
0.047 0.116 0.262 0.581
0.023 0.161 0.212 0.222
0.084 0.925 0.151 0.404
0.291 0.689 0.499 0.363
0.078 0.169 0.145 0.282
0.251 0.897 0.131 0.534
0.130 1.104 0.267 0.487
0.143 0.045 0.521 0.910
0.154 0.043 0.503 0.206
0.105 0.486 0.469 0.787
0.487 1.028 0.263 0.215
0.133 0.408 0.263 0.165
0.498 0.235 0.393 0.513
0.836 0.856 0.238 3.148
0.458 1.004 0.477 0.396
0.814 1.959 1.321 0.232
1.113 0.475 0.739 0.566
0.163 0.607 0.081 1.118
ALGAE_COVER
HARD CORAL_COVER
OTHER_COVER
SOFT CORAL_COVER
Cover Metric
3.0
2.5
2.0
1.5
1.0
0.5


nRMSE per Reef and Cover (TCN NN)
Reef
Cover Metric
ALGAE_COVER HARD CORAL_COVER OTHER_COVER SOFT CORAL_COVER
19131S 0.007 0.006 0.033 0.890
0.007 0.008 0.088 0.133
20104S 0.012 0.006 0.047 0.155
0.008 0.018 0.032 0.083
AGINCOURT REEFS (NO 1) 0.007 0.020 0.138 0.039
0.014 0.012 0.028 0.018
BROOMFIELD REEF 0.020 0.007 0.090 0.061
0.008 0.038 0.058 0.209
CHICKEN REEF 0.005 0.010 0.131 0.050
0.012 0.013 0.349 0.023
DAVIES REEF 0.006 0.008 0.054 0.086
0.006 0.021 0.153 0.097
EAST CAY REEF 0.011 0.012 0.107 0.015
0.009 0.014 0.029 0.015
GANNETT CAY REEF 0.007 0.016 0.096 0.058
0.006 0.030 0.023 0.079
HASTINGS REEF 0.006 0.027 0.102 0.042
0.010 0.036 0.043 0.036
HORSESHOE 0.010 0.150 0.146 0.104
0.006 0.016 0.039 0.036
JOHN BREWER REEF 0.006 0.021 0.043 0.089
0.017 0.009 0.316 0.934
LANGFORD-BIRD REEF 0.010 0.032 0.018 0.366
0.006 0.079 0.058 0.175
LIZARD ISLAND 0.005 0.053 0.054 0.055
0.008 0.013 0.039 0.029
MACGILLIVRAY REEF 0.006 0.028 0.021 -0.242
0.005 0.017 0.058 0.079
MARTIN REEF(14123) 0.009 0.051 0.017 0.320
0.005 0.021 0.094 0.022
MYRMIDON REEF 0.006 0.011 0.067 0.056
0.006 0.026 0.075 0.159
NORTH DIRECTION REEF 0.007 0.068 0.018 0.529
0.011 0.008 0.139 0.360
PANDORA REEF 0.012 0.009 0.042 0.028
0.007 0.032 0.079 0.127
RIB REEF 0.007 -0.003 0.044 0.082
0.007 0.011 0.047 0.074
SNAKE (22088) 0.012 0.006 0.114 0.144
0.006 0.023 0.070 0.055
THETFORD REEF 0.007 0.017 0.083 0.021
0.014 0.159 0.067 0.017
WRECK ISLAND REEF 0.027 0.009 0.229 0.171
0.008 0.044 0.077 0.560
0.8
0.6
0.4
0.2
0.0
−0.2


nRMSE per Reef and Cover (ST-GNN w/ TCN NN)
Reef
Cover Metric
ALGAE_COVER HARD CORAL_COVER OTHER_COVER SOFT CORAL_COVER
19131S 0.006 0.008 0.006 0.294
0.003 0.007 0.008 0.069
20104S 0.016 0.013 0.004 0.082
0.003 0.006 0.008 0.018
AGINCOURT REEFS (NO 1) 0.004 0.006 0.065 0.007
0.008 0.003 0.013 0.012
BROOMFIELD REEF 0.016 0.011 0.030 0.016
0.006 0.022 0.005 0.177
CHICKEN REEF 0.003 0.007 0.020 0.023
0.006 0.010 0.017 0.011
DAVIES REEF 0.001 0.002 0.003 0.011
0.005 0.012 0.033 0.034
EAST CAY REEF 0.004 0.007 0.014 0.013
0.009 0.007 0.012 0.011
GANNETT CAY REEF 0.003 0.004 0.010 0.009
0.003 0.017 0.006 0.033
HASTINGS REEF 0.006 0.019 0.009 0.010
0.006 0.014 0.008 0.010
HORSESHOE 0.007 0.017 0.032 0.081
0.004 0.009 0.008 0.010
JOHN BREWER REEF 0.007 0.019 0.014 0.052
0.012 0.010 0.072 0.343
LANGFORD-BIRD REEF 0.004 0.017 0.015 0.014
0.007 0.040 0.008 0.108
LIZARD ISLAND 0.004 0.030 0.003 0.015
0.003 0.002 0.006 0.007
MACGILLIVRAY REEF 0.004 0.013 0.007 0.081
0.005 0.017 0.003 0.037
MARTIN REEF(14123) 0.004 0.045 0.013 0.166
0.004 0.014 0.009 0.004
MYRMIDON REEF 0.002 0.006 0.002 0.023
0.008 0.031 0.009 0.088
NORTH DIRECTION REEF 0.005 0.094 0.015 0.092
0.008 0.010 0.032 0.315
PANDORA REEF 0.010 0.006 0.008 0.005
0.003 0.007 0.005 0.029
RIB REEF 0.005 0.019 0.007 0.031
0.001 0.008 0.018 0.030
SNAKE (22088) 0.009 0.010 0.024 0.063
0.006 0.013 0.024 0.017
THETFORD REEF 0.003 0.007 0.011 0.004
0.007 0.015 0.015 0.011
WRECK ISLAND REEF 0.018 0.010 0.018 0.062
0.007 0.037 0.007 0.146
0.30
0.25
0.20
0.15
0.10
0.05

Reef Graph Visualization