Hurricast

Hurricane Forecasting: A Novel Multimodal Machine Learning Framework

Léonard Boussioux, Cynthia Zeng
Joint work with Dimitris Bertsimas and Théo Guénais

OPERATIONS RESEARCH CENTER

I. Motivation

Why is Hurricane Forecasting worth our time?

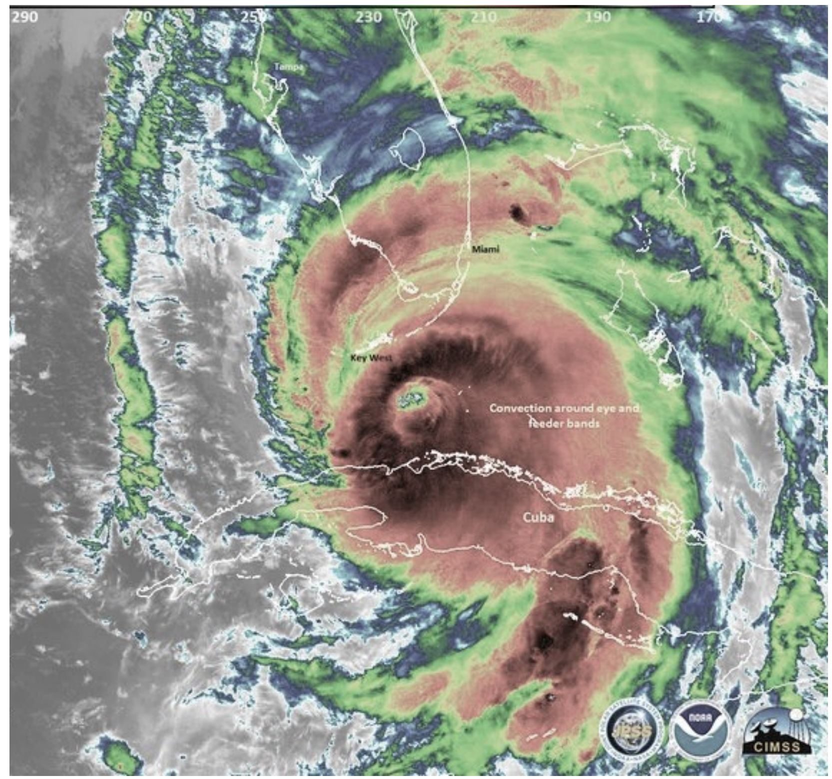

Katrina
1836 deaths
$250bn loss
Sources: [1] New York Times

The Problem of Hurricane Forecasting

Tropical Cyclones (TC)

Draw energy from the warm ocean waters.

Track and Intensity forecasting tasks.

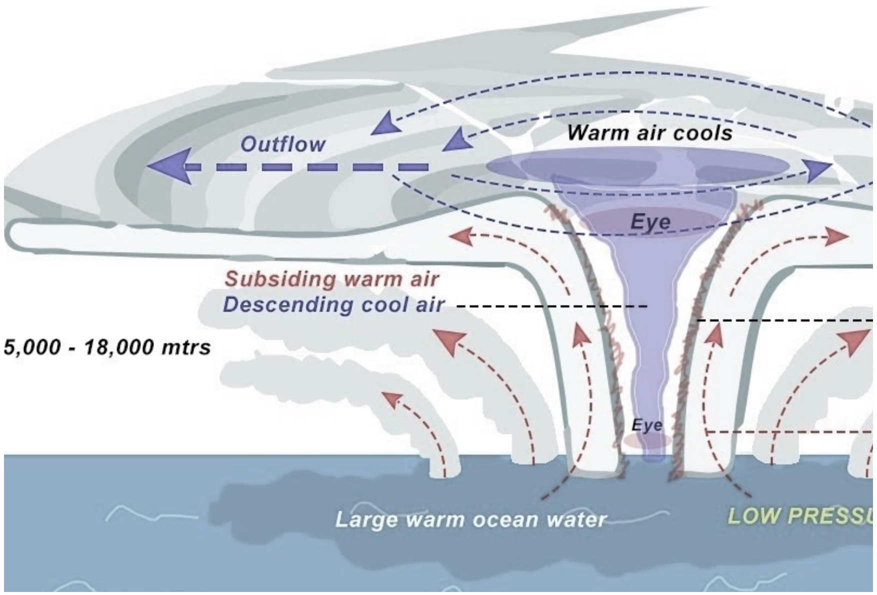


The Problem of Hurricane Forecasting

Tropical Cyclones (TC)

Draw energy from the warm ocean waters.

Track and Intensity forecasting tasks.

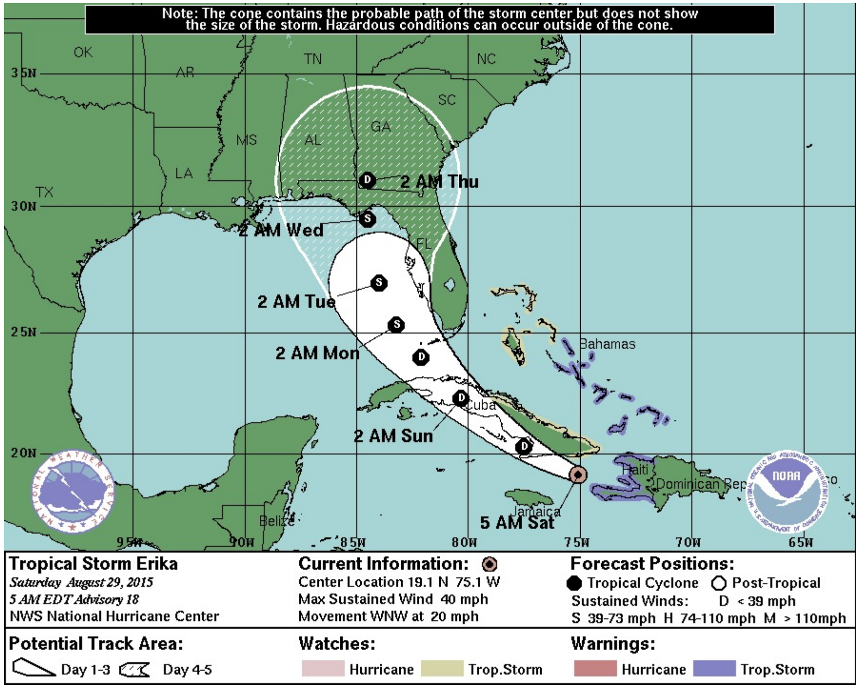


The Problem of Hurricane Forecasting

Tropical Cyclones (TC)

Draw energy from the warm ocean waters.

Track and Intensity forecasting tasks.



Current forecasting approaches

Dynamical

Fluid Mechanics, PDEs

Strong Modeling Power

Slow, computationally expensive

Highly sensitive to initialization

Statistical-Dynamical

Often regression-based

Uses outputs from dynamical models

Fast to compute

Limited predictive power

Hardly uses multiple data sources

Consensus

Often simple or weighted average of operational forecasts

Best performance

Relies on the availability of underlying models

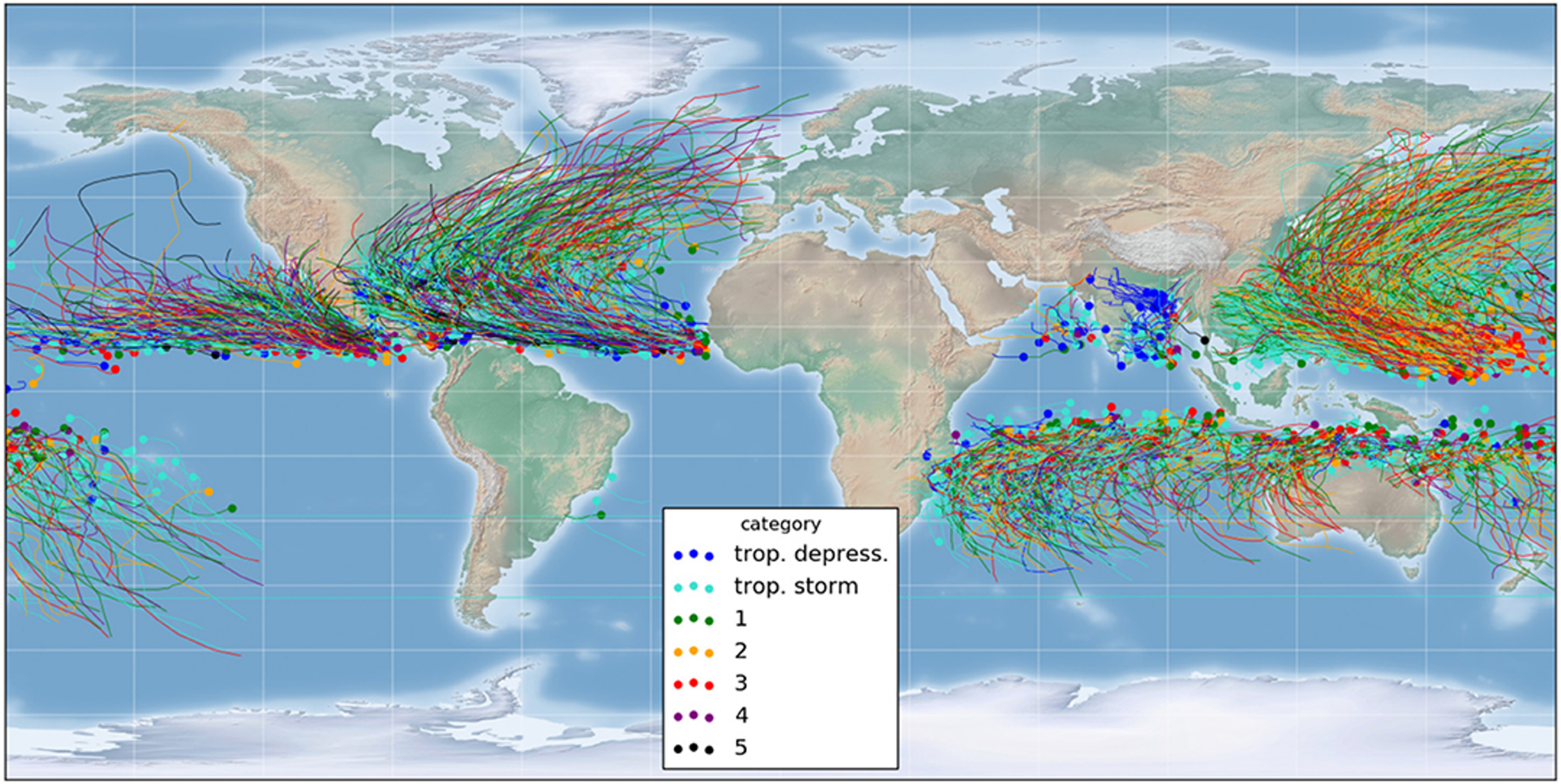
II. Formulate the problem and identify the challenges

Our Goal

Advance hurricane forecasting skills for both intensity and track

by utilizing distinct ML approaches and combining multiple data sources.

Hurricanes since 1980



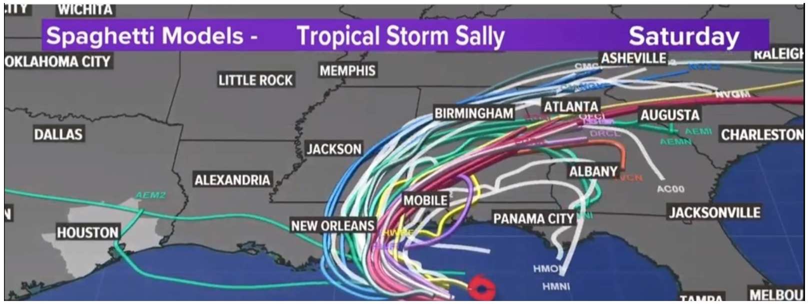
Picture source: www.frontiersin.org/articles/10.3389/fdata.2020.00001/full

Multimodality: Three distinct data sources

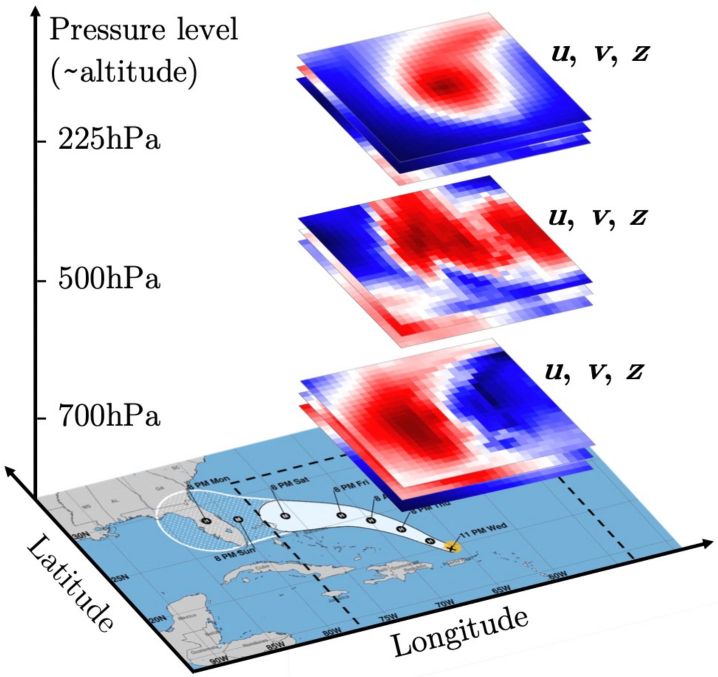
Historical data

BASIN	ISO_TIME	LAT	LON	STORM_SPEED	STORM_DIR
		degrees_north	degrees_east	kts	degrees
EP	2016-01-05 06:00:00	2.00000	-173.500	3	73
EP	2016-01-05 09:00:00	2.04500	-173.353	3	71
EP	2016-01-05 12:00:00	2.10000	-173.200	3	67
EP	2016-01-05 15:00:00	2.17750	-173.042	4	56

Forecast data



Vision data: reanalysis maps



Key Results

Our framework demonstrates a successful approach to combine of multiple data sources.

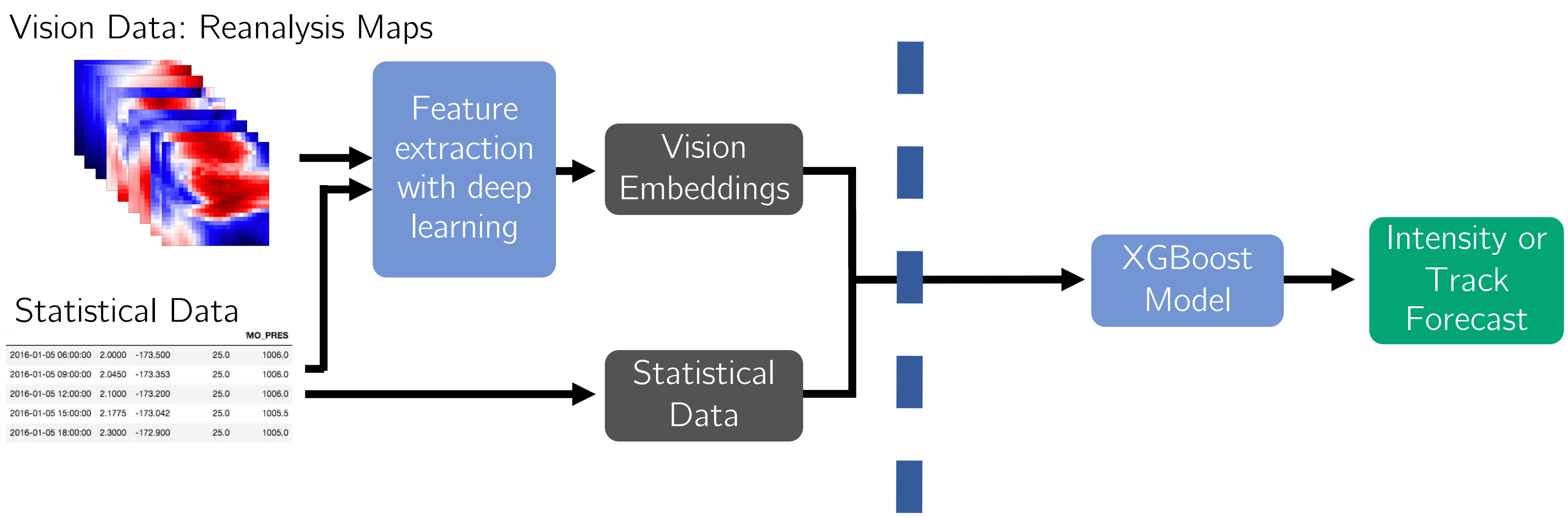
ML models outperform statistical models, and competes with dynamical models.

Inclusion of Hurricast into an operational consensus model leads to a significant improvement of 5% - 15% over NHC's official forecast.

III. The Hurricast Methodology

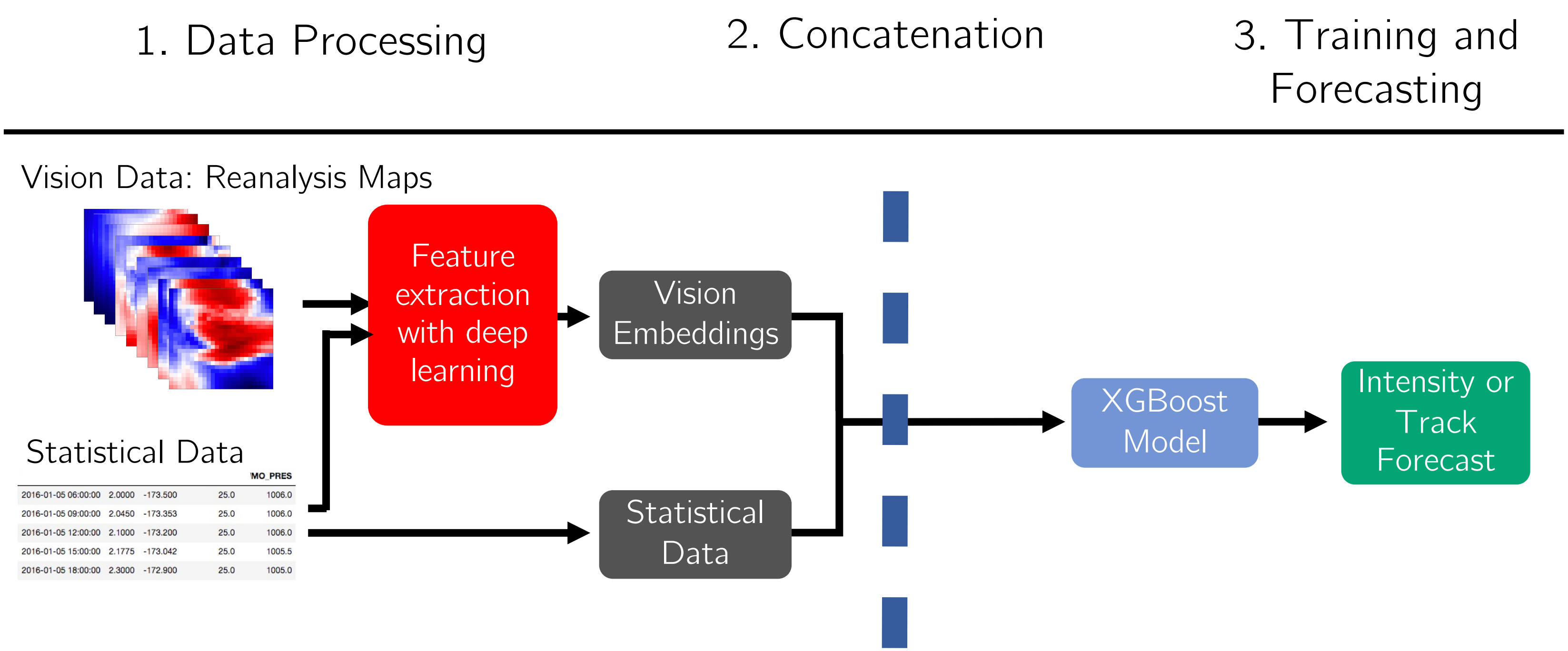
General Framework

1. Data Processing
2. Concatenation
3. Training and Forecasting



				'MO_PRES
2016-01-05 06:00:00	2.0000	-173.500	25.0	1006.0
2016-01-05 09:00:00	2.0450	-173.353	25.0	1006.0
2016-01-05 12:00:00	2.1000	-173.200	25.0	1006.0
2016-01-05 15:00:00	2.1775	-173.042	25.0	1005.5
2016-01-05 18:00:00	2.3000	-172.900	25.0	1005.0

General Framework



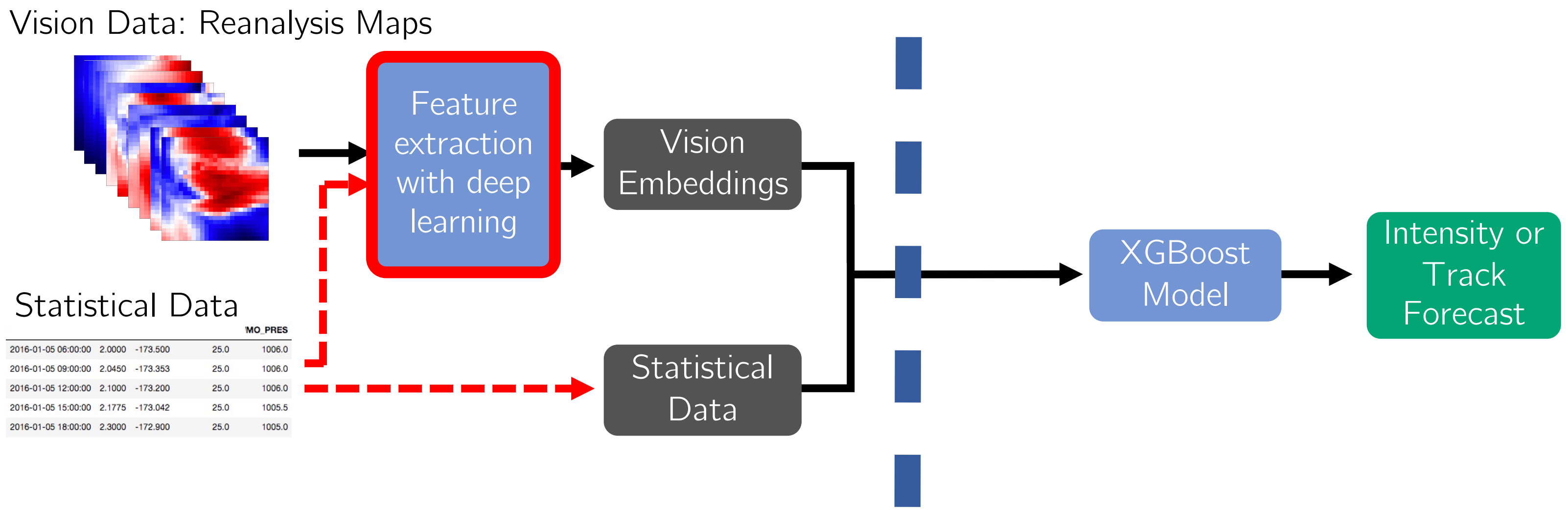
Feature extraction is challenging.

General Framework

1. Data Processing

2. Concatenation

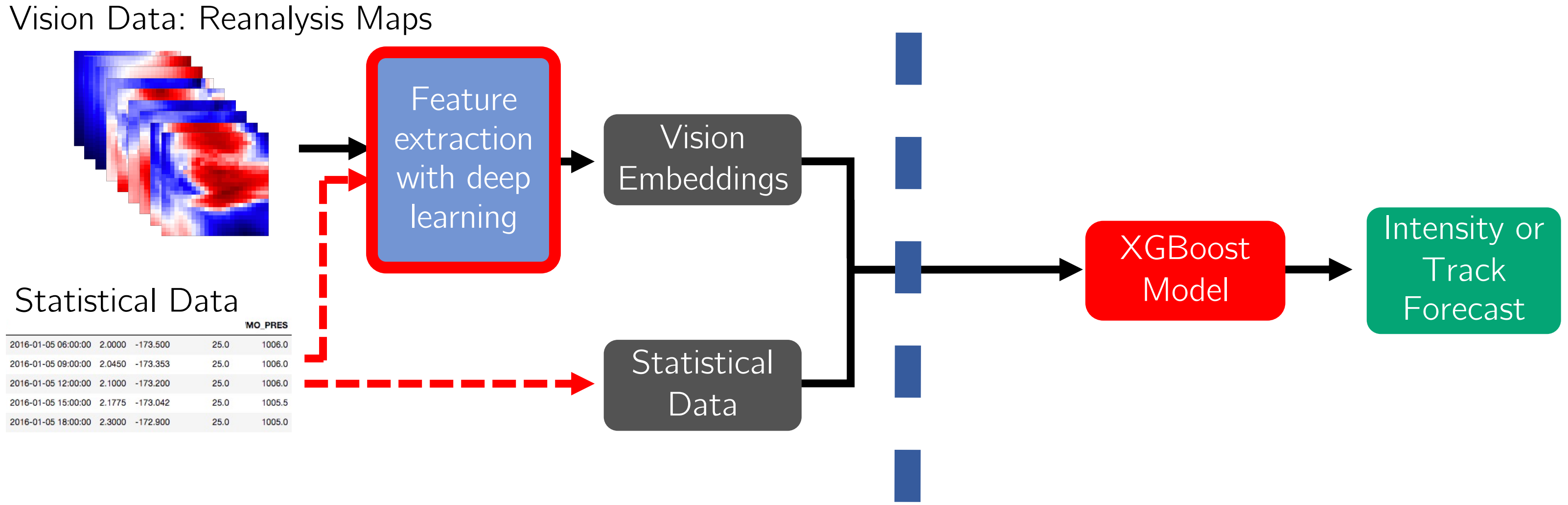
3. Training and Forecasting



Statistical data is used twice.

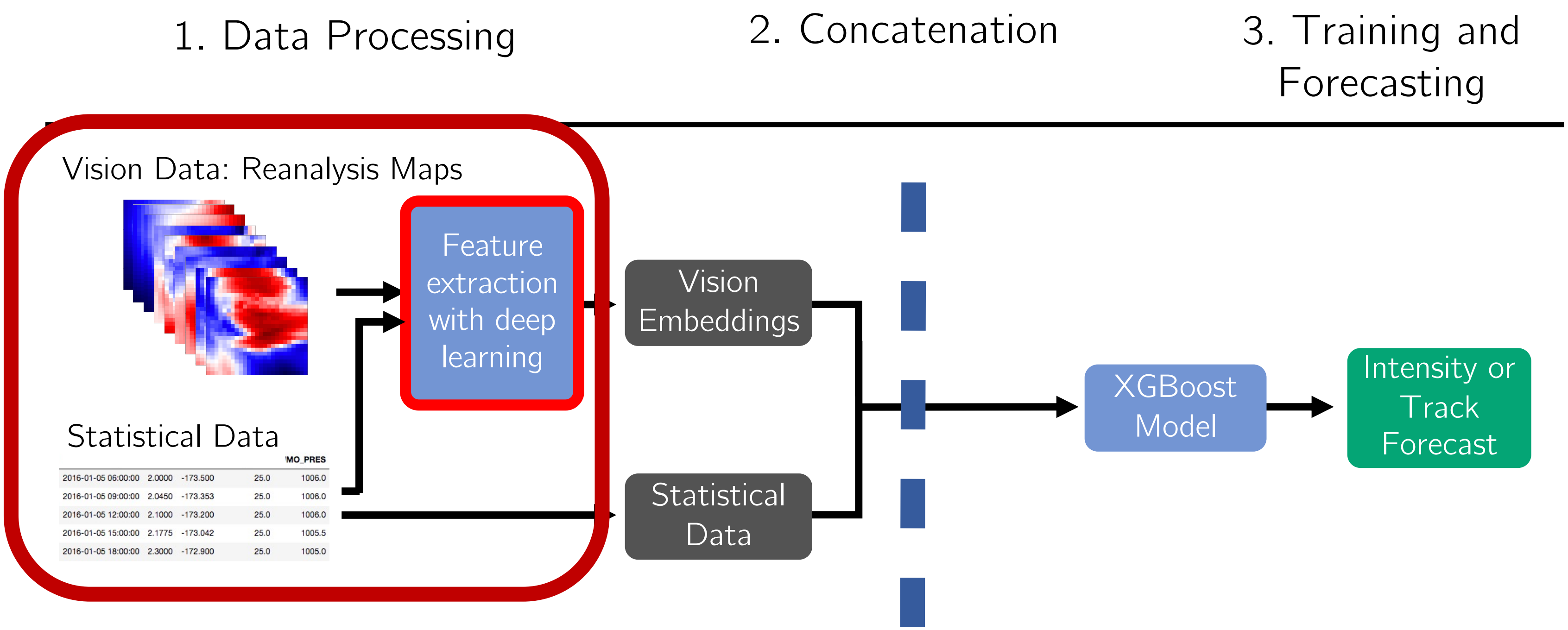
General Framework

1. Data Processing | 2. Concatenation | 3. Training and Forecasting



Tree-based models are powerful.

General Framework

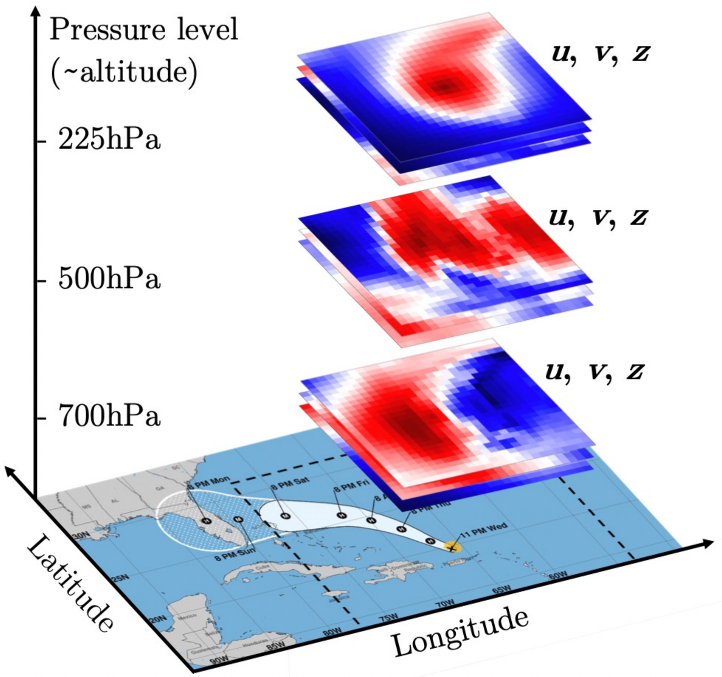


Multimodality

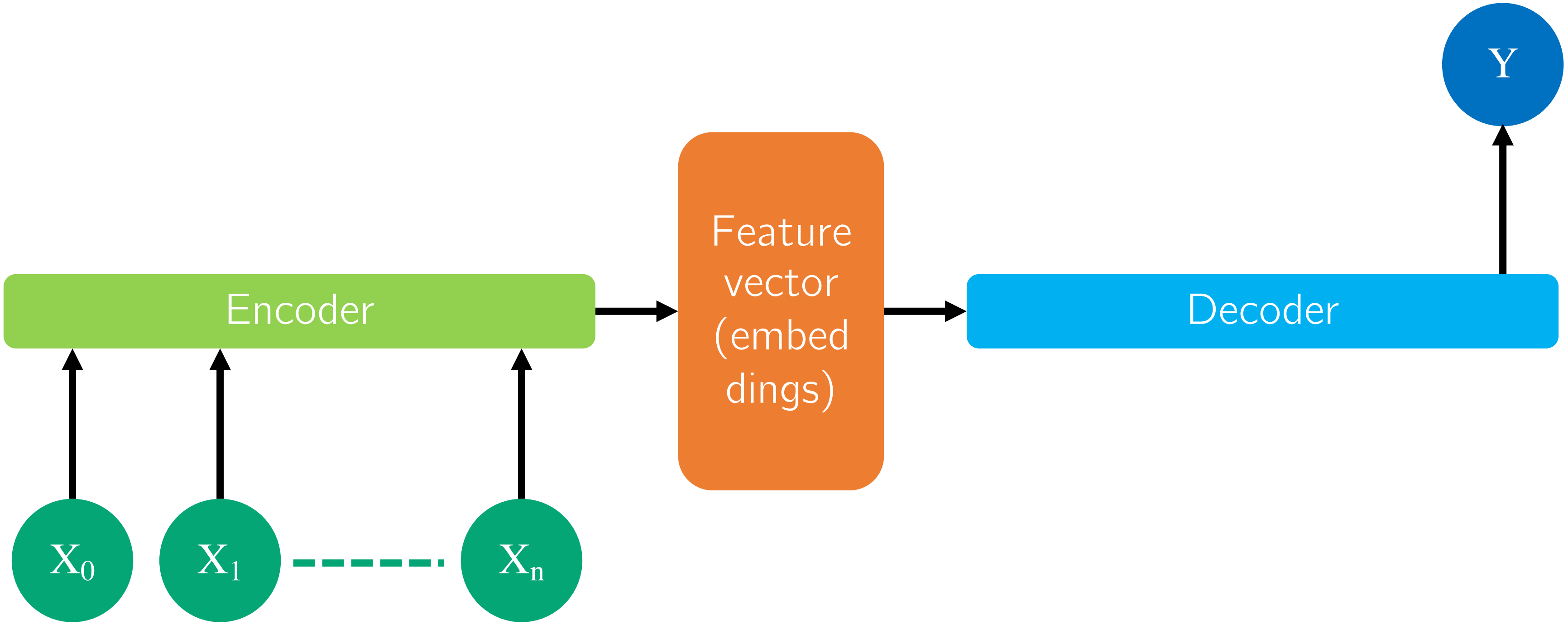
Tabular data

BASIN	ISO_TIME	LAT	LON	STORM_SPEED	STORM_DIR
		degrees_north	degrees_east	kts	degrees
EP	2016-01-05 06:00:00	2.00000	-173.500	3	73
EP	2016-01-05 09:00:00	2.04500	-173.353	3	71
EP	2016-01-05 12:00:00	2.10000	-173.200	3	67
EP	2016-01-05 15:00:00	2.17750	-173.042	4	56

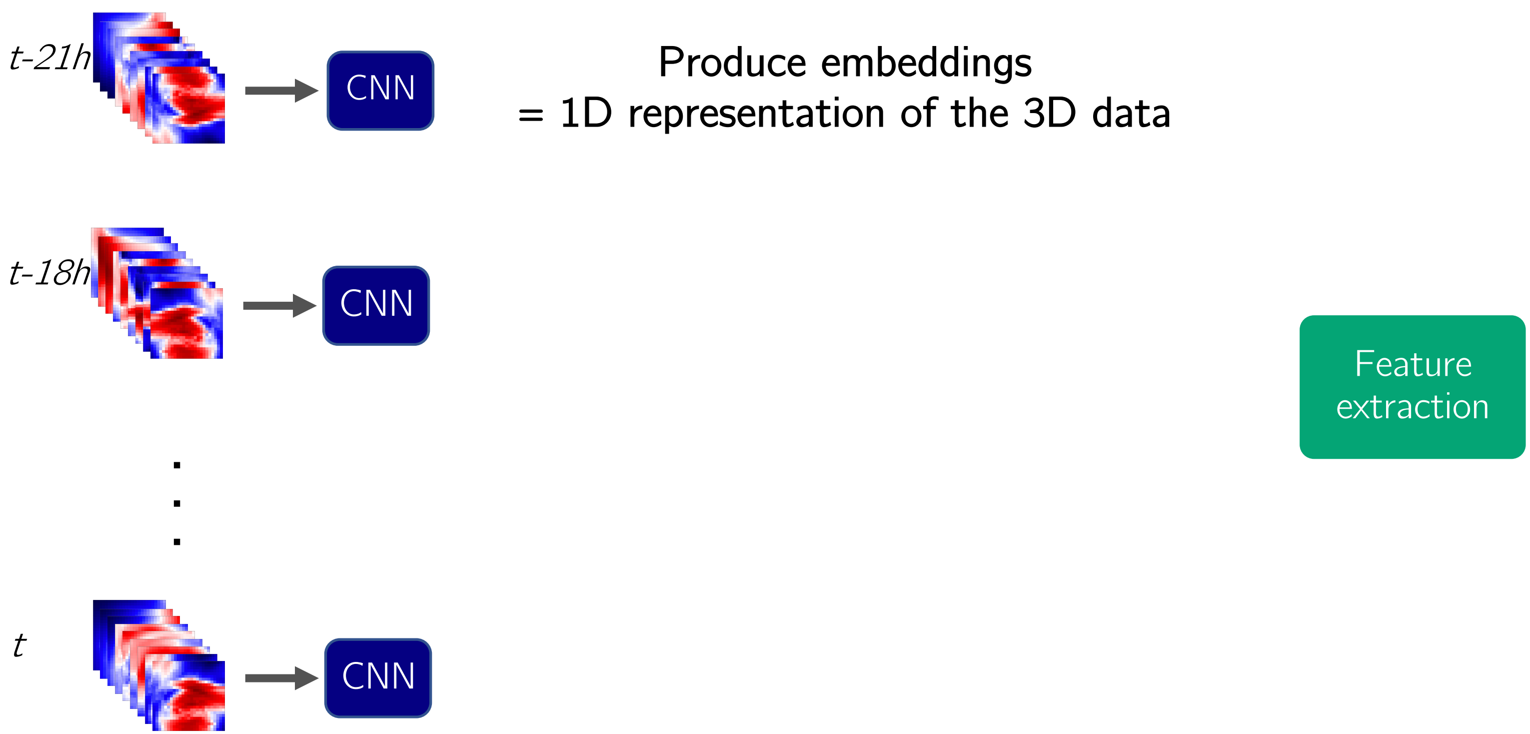
Vision data: reanalysis maps



Encoder-Decoder Architecture

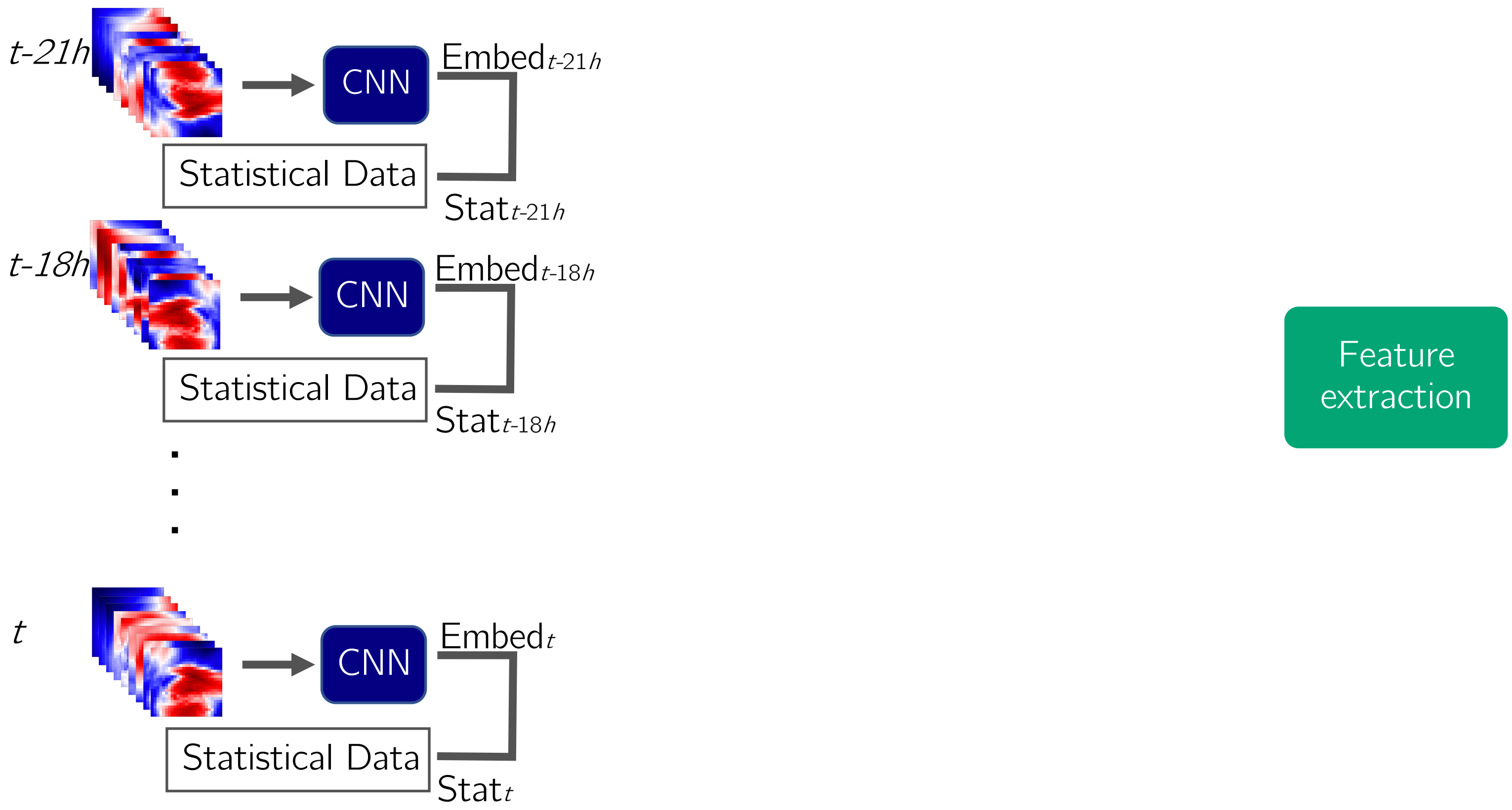


Encoder-Decoder Architecture



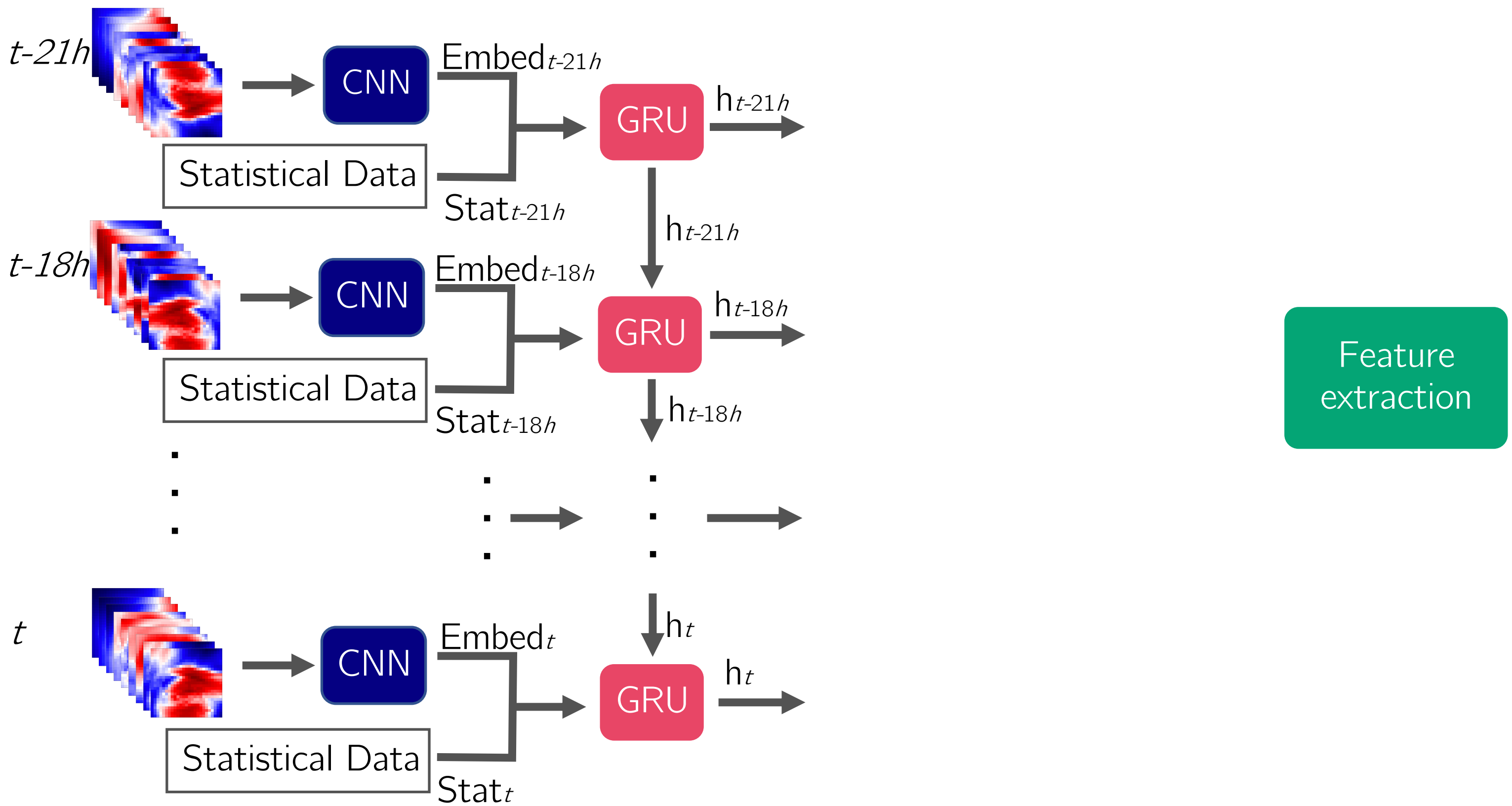
Encoder-Decoder Architecture

1. Encoding

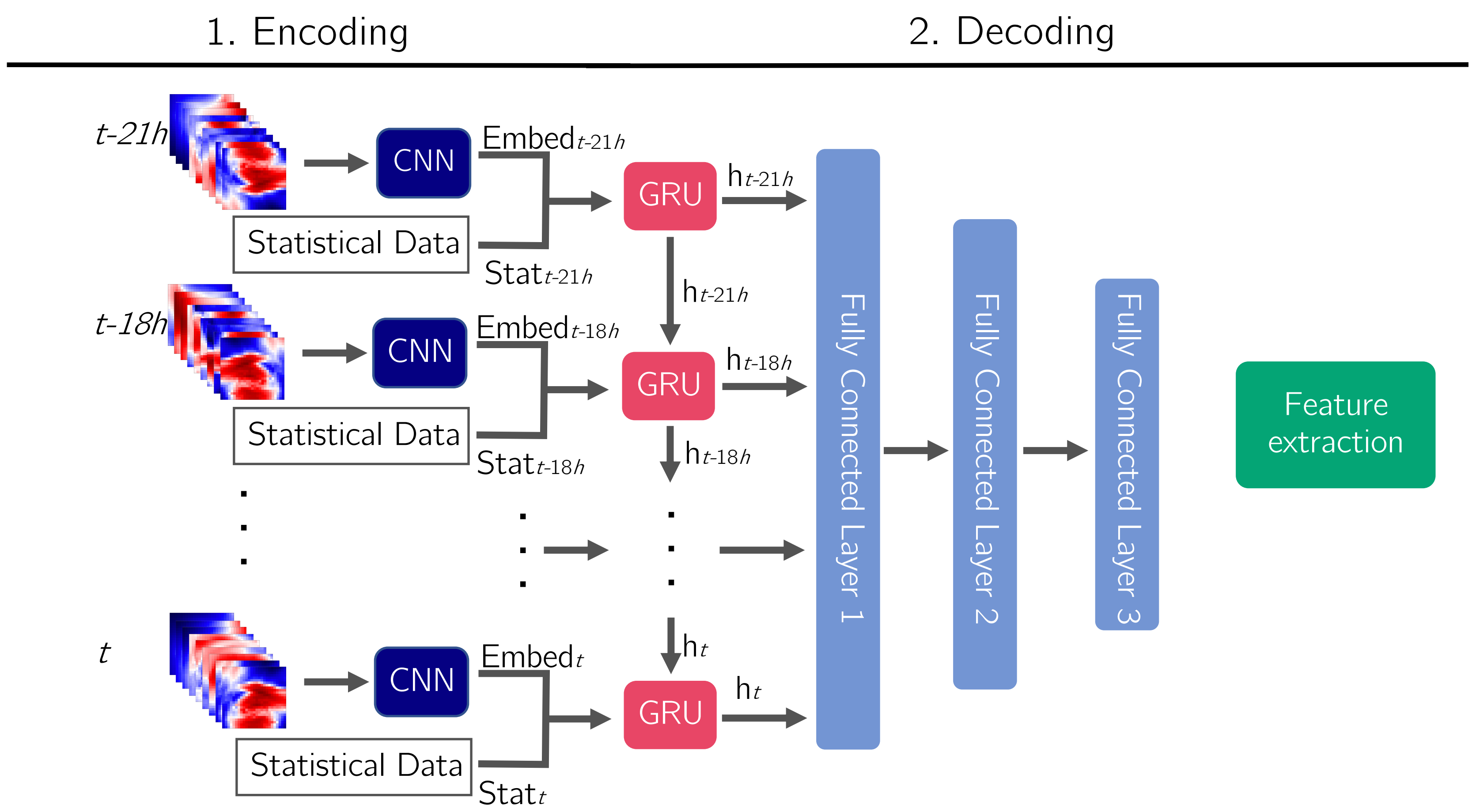


Encoder-Decoder Architecture

1. Encoding | 2. Decoding



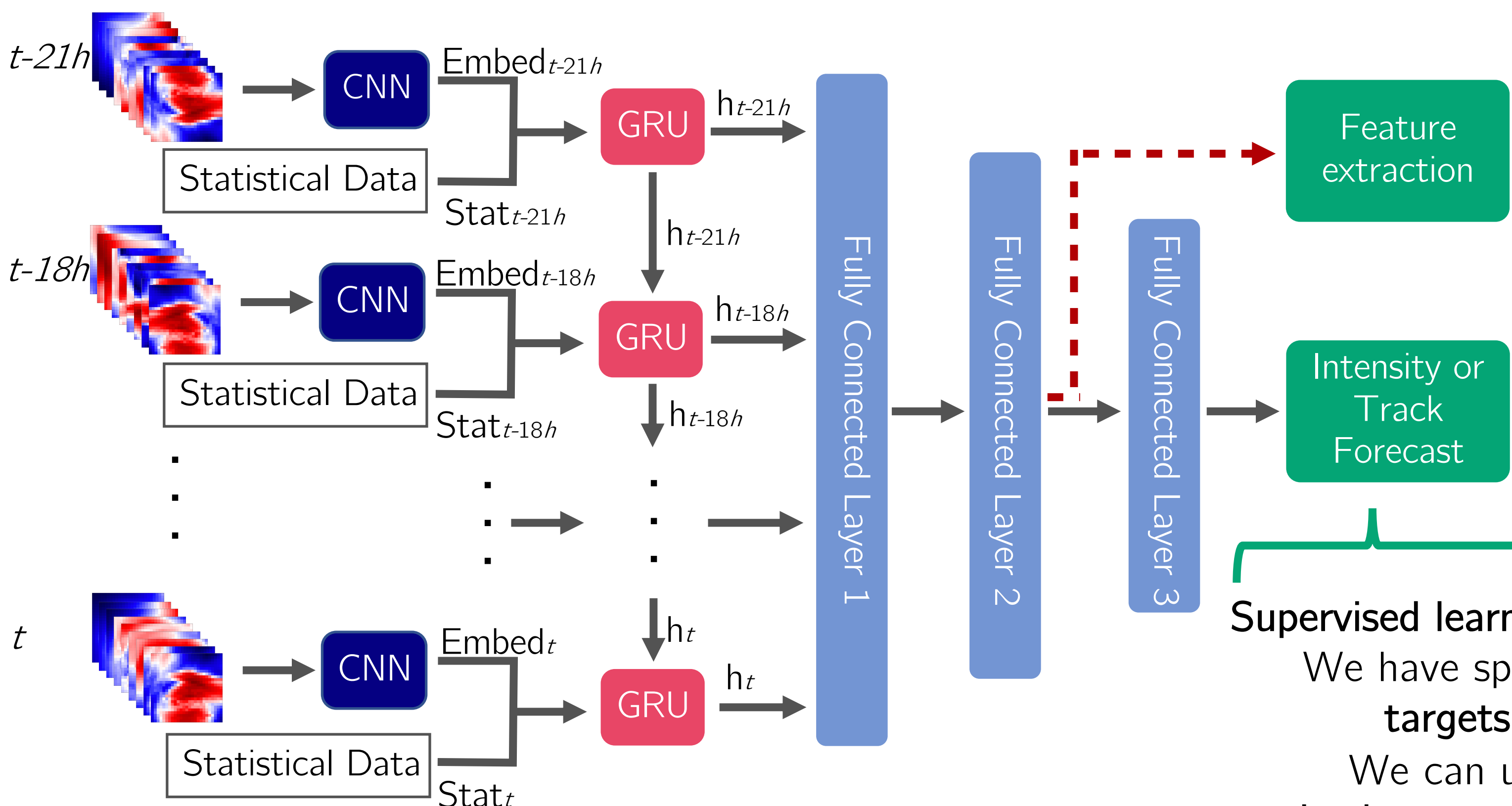
Encoder-Decoder Architecture



Encoder-Decoder Architecture

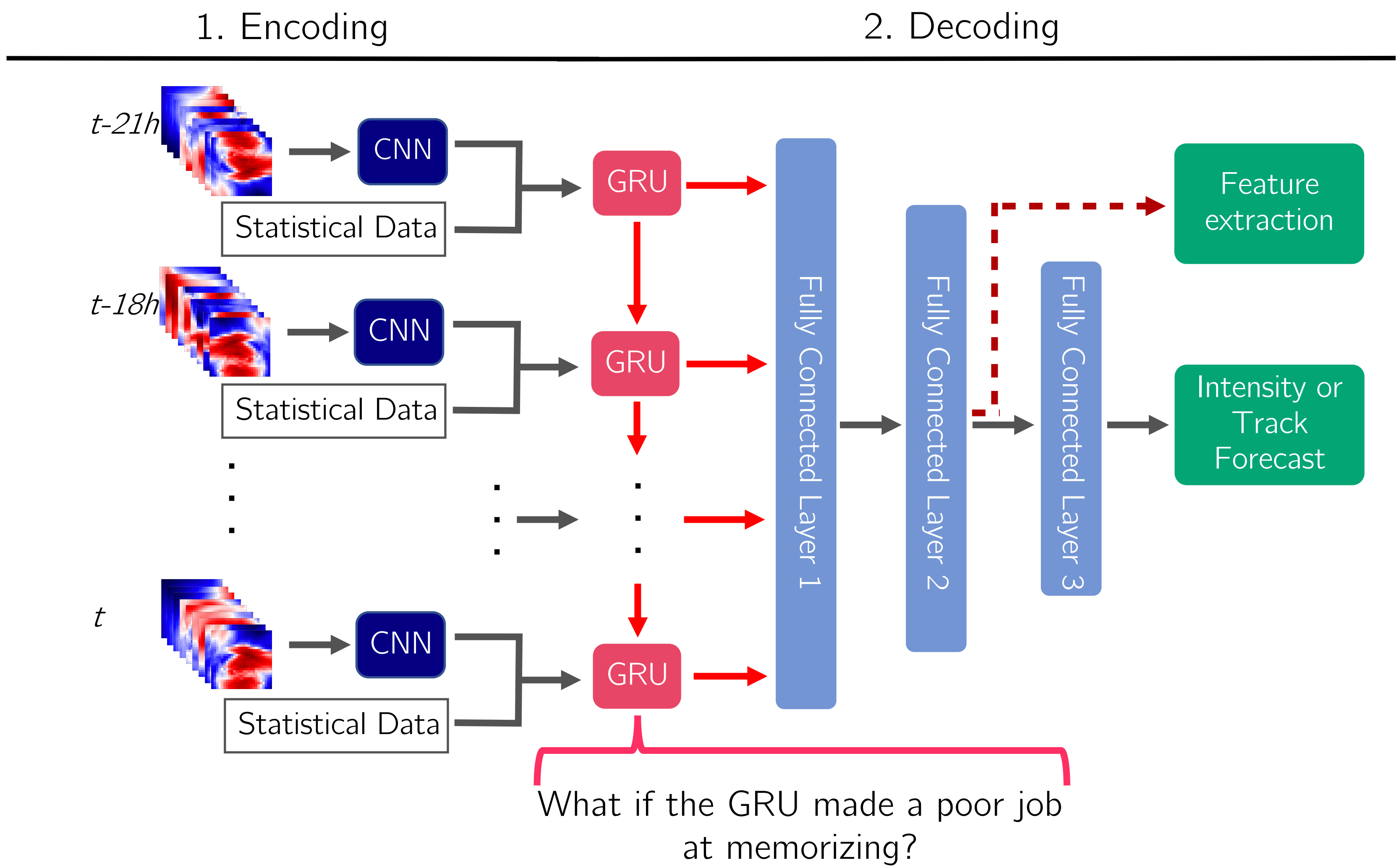
1. Encoding

2. Decoding

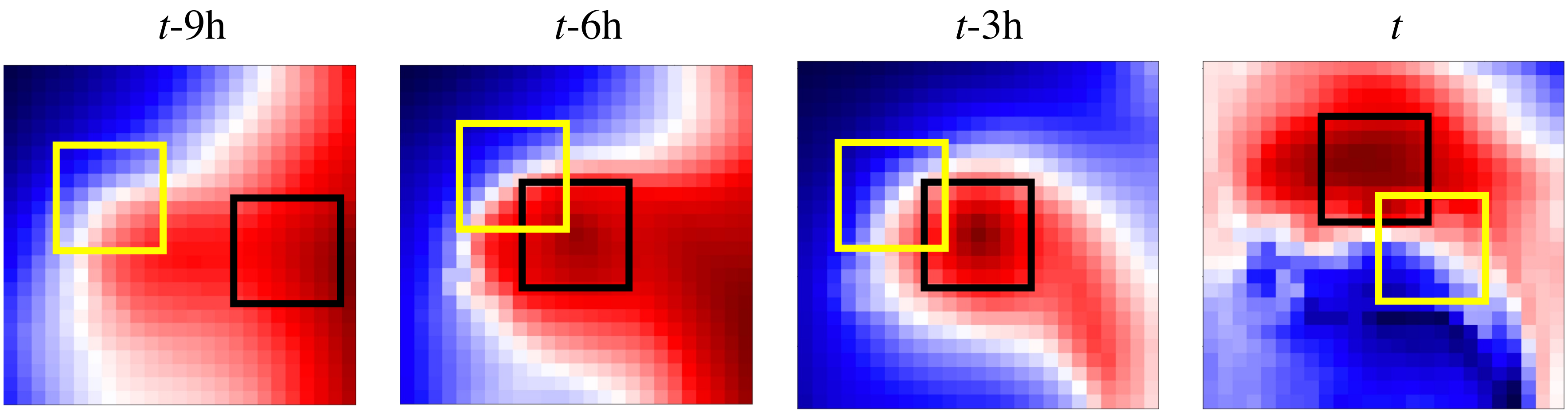


Supervised learning task!
We have specific **targets**.
We can use **backpropagation**.

Encoder-Decoder Architecture



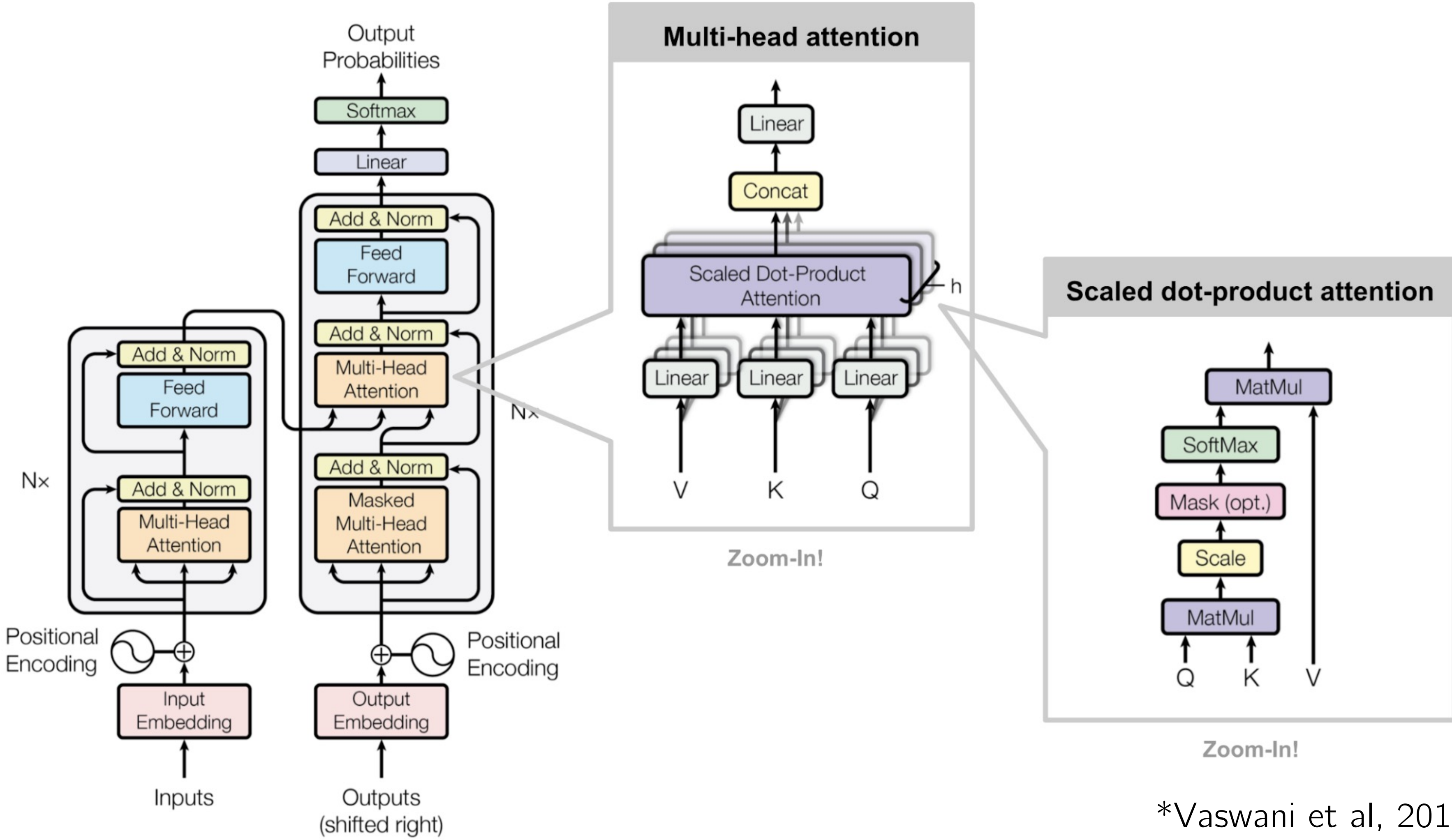
Attention Mechanisms



- Features A
- Features B

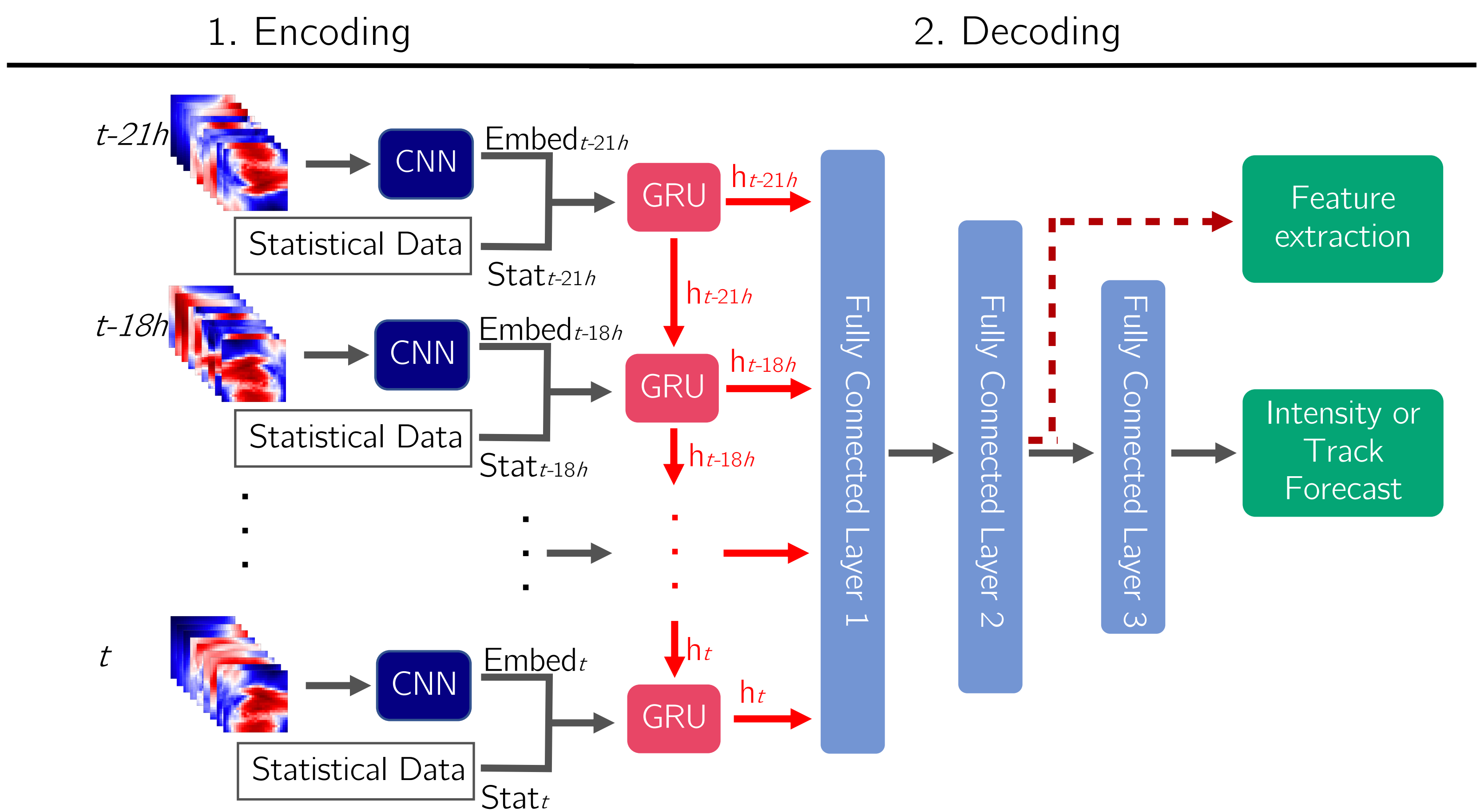
Disclaimer: this is a very schematic representation of what happens in reality

Transformers*



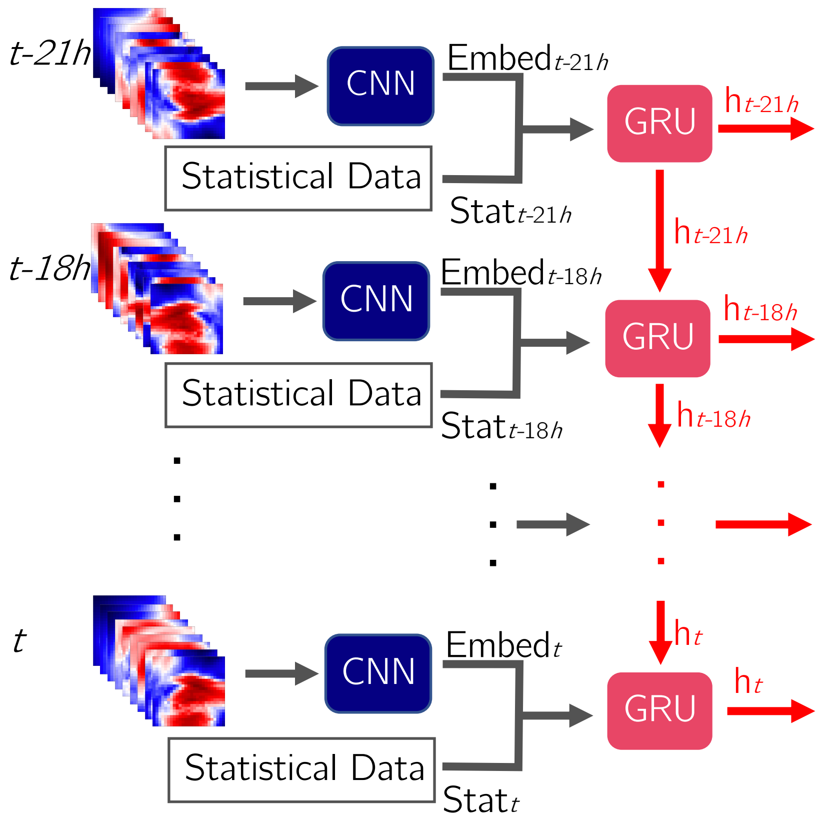
*Vaswani et al, 2017

Encoder-Decoder Architecture



Encoder-Decoder Architecture

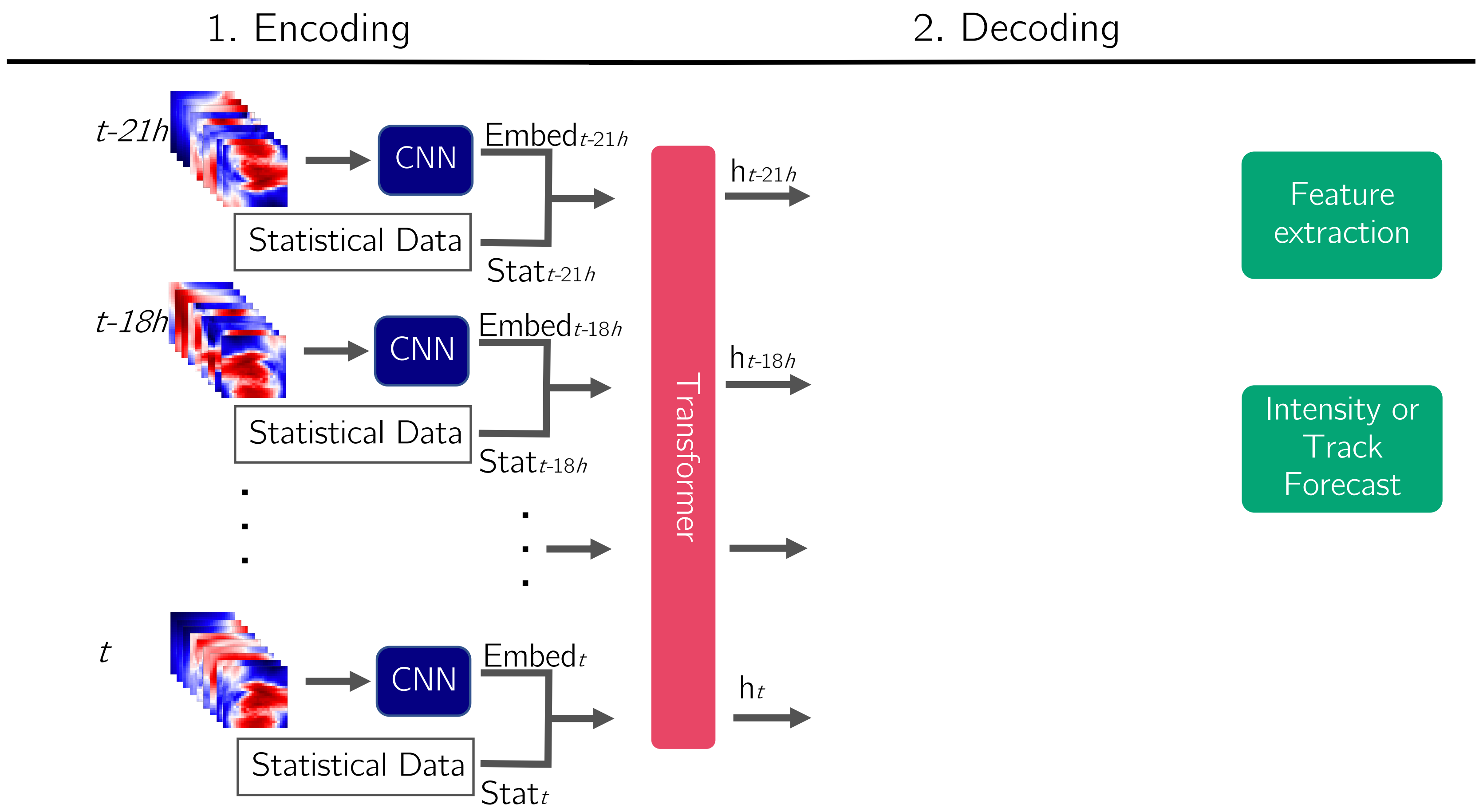
1. Encoding | 2. Decoding



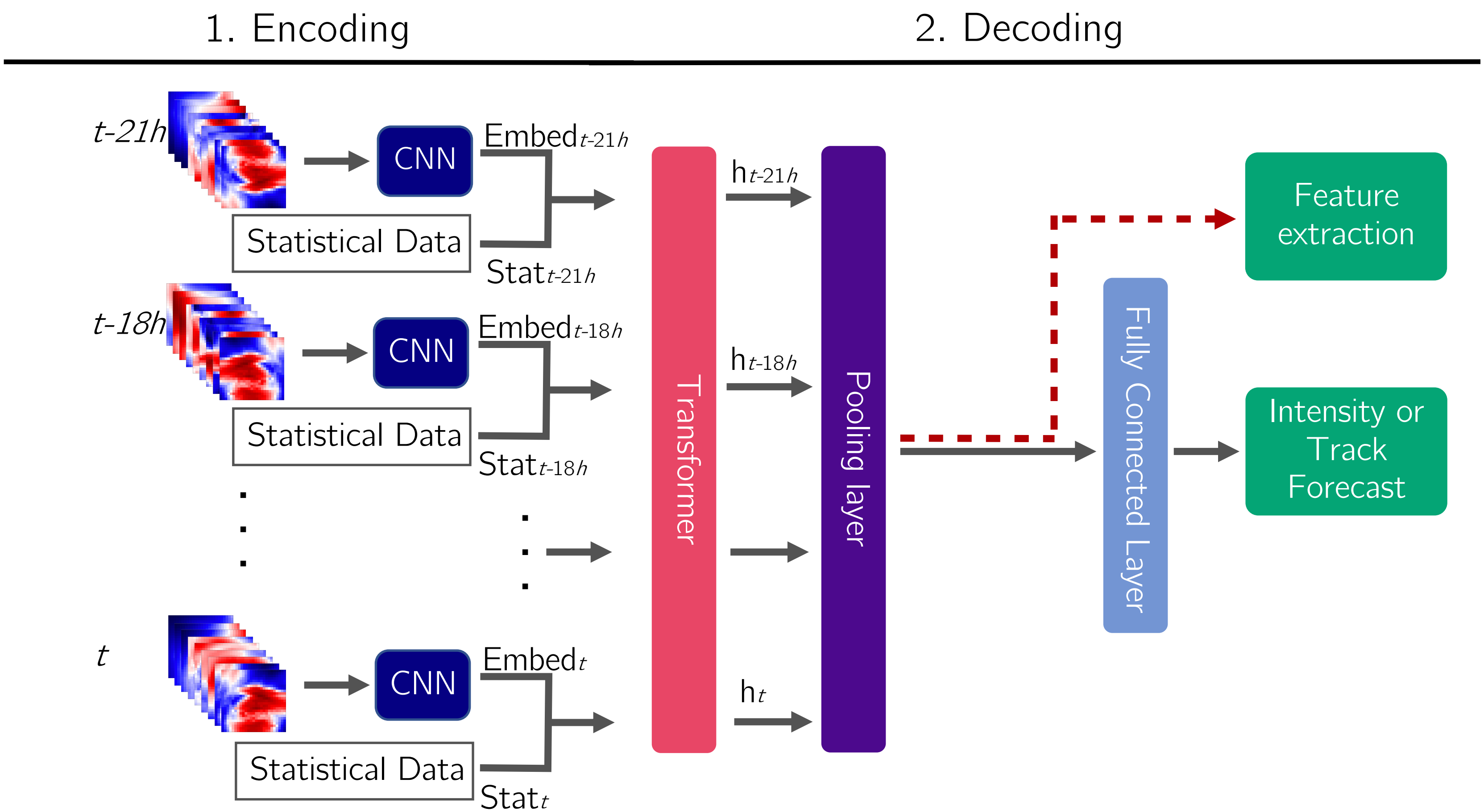
Feature extraction

Intensity or Track Forecast

Encoder-Decoder Architecture



Encoder-Decoder Architecture

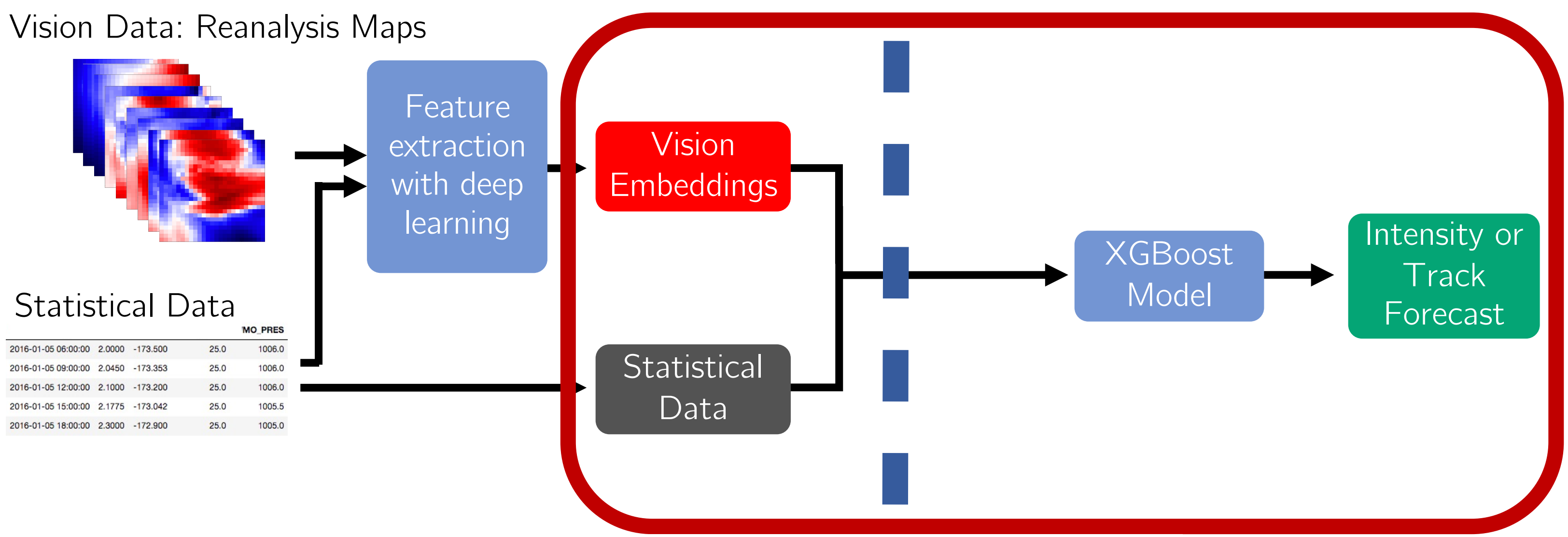


General Framework

1. Data Processing

2. Concatenation

3. Training and Forecasting



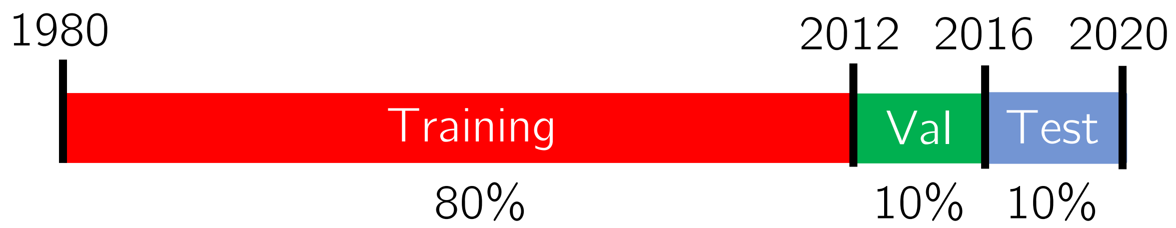
One-dimensional vector

IV. Results

Training, Validation, Testing

Data ranges from 1980 to 2020

We use a validation strategy.



Intensity results

Model Type	Model Name	Eastern Pacific Basin Comparison on 36 TC		North Atlantic Basin Comparison on 45 TC	
		MAE (kn)	**Error sd (kn)**	**MAE (kn)**	**Error sd (kn)**
Hurricast (HURR) Methods	HURR-(viz, cnn/gru)	**10.7**	10.1	**11.4**	9.6
	HURR-(viz, cnn/transfo)	**10.5**	**10.0**	**11.4**	**9.5**
	HURR-(stat, xgb)	10.5	10.4	10.8	9.3
	HURR-(stat/viz, xgb/cnn/gru)	**10.3**	10.1	10.8	9.3
	HURR-(stat/viz, xgb/cnn/transfo)	**10.3**	**9.8**	**10.4**	**8.8**

- Combining data sources has a significant edge.
- Using XGBoost on top of Deep Learning-extracted features has a clear edge.
- Transformer slightly better than GRU.

Intensity results

Model Type	Model Name	Eastern Pacific Basin Comparison on 36 TC		North Atlantic Basin Comparison on 45 TC	
		MAE (kn)	**Error sd (kn)**	**MAE (kn)**	**Error sd (kn)**
Hurricast (HURR) Methods	HURR-(viz, cnn/gru)	**10.7**	10.1	**11.4**	9.6
	HURR-(viz, cnn/transfo)	**10.5**	**10.0**	**11.4**	**9.5**
	HURR-(stat, xgb)	10.5	10.4	10.8	9.3
	HURR-(stat/viz, xgb/cnn/gru)	**10.3**	10.1	10.8	9.3
	HURR-(stat/viz, xgb/cnn/transfo)	**10.3**	**9.8**	**10.4**	**8.8**
Standalone Operational Forecasts	Decay-SHIPS	11.7	**10.4**	10.2	9.3
	HWRF	**10.6**	11.0	**9.7**	**9.0**
	GFSO	15.7	14.7	14.2	14.1

Very competitive or better performance than the top statistical-dynamical and best dynamical models!

Track results

Model Type	Model Name	Eastern Pacific Basin Comparison on 36 TC		North Atlantic Basin Comparison on 45 TC	
		MAE (km)	**Error sd (km)**	**MAE (km)**	**Error sd (km)**
Hurricast (HURR) Methods	HURR-(viz, cnn/gru)	**73**	**43**	114	83
	HURR-(viz, cnn/transfo)	**73**	44	**110**	**71**
	HURR-(stat, xgb)	81	47	144	109
	HURR-(stat/viz, xgb/cnn/gru)	**71**	**43**	**110**	79
	HURR-(stat/viz, xgb/cnn/transfo)	72	43	**110**	**72**

- Combining data sources has a significant edge.
- Using XGBoost on top of Deep Learning-extracted features is useful.
- GRU and Transformer approaches perform similarly.

Track results

Model Type	Model Name	Eastern Pacific Basin Comparison on 36 TC		North Atlantic Basin Comparison on 45 TC	
		MAE (km)	**Error sd (km)**	**MAE (km)**	**Error sd (km)**
Hurricast (HURR) Methods	HURR-(viz, cnn/gru)	**73**	**43**	114	83
	HURR-(viz, cnn/transfo)	**73**	44	**110**	**71**
	HURR-(stat, xgb)	81	47	144	109
	HURR-(stat/viz, xgb/cnn/gru)	**71**	**43**	**110**	79
	HURR-(stat/viz, xgb/cnn/transfo)	72	43	**110**	**72**
Standalone Operational Forecasts	CLP5	121	67	201	149
	HWRF	67	**42**	75	**49**
	GFSO	**65**	45	**71**	54

Performance getting close to the top operational forecast models.

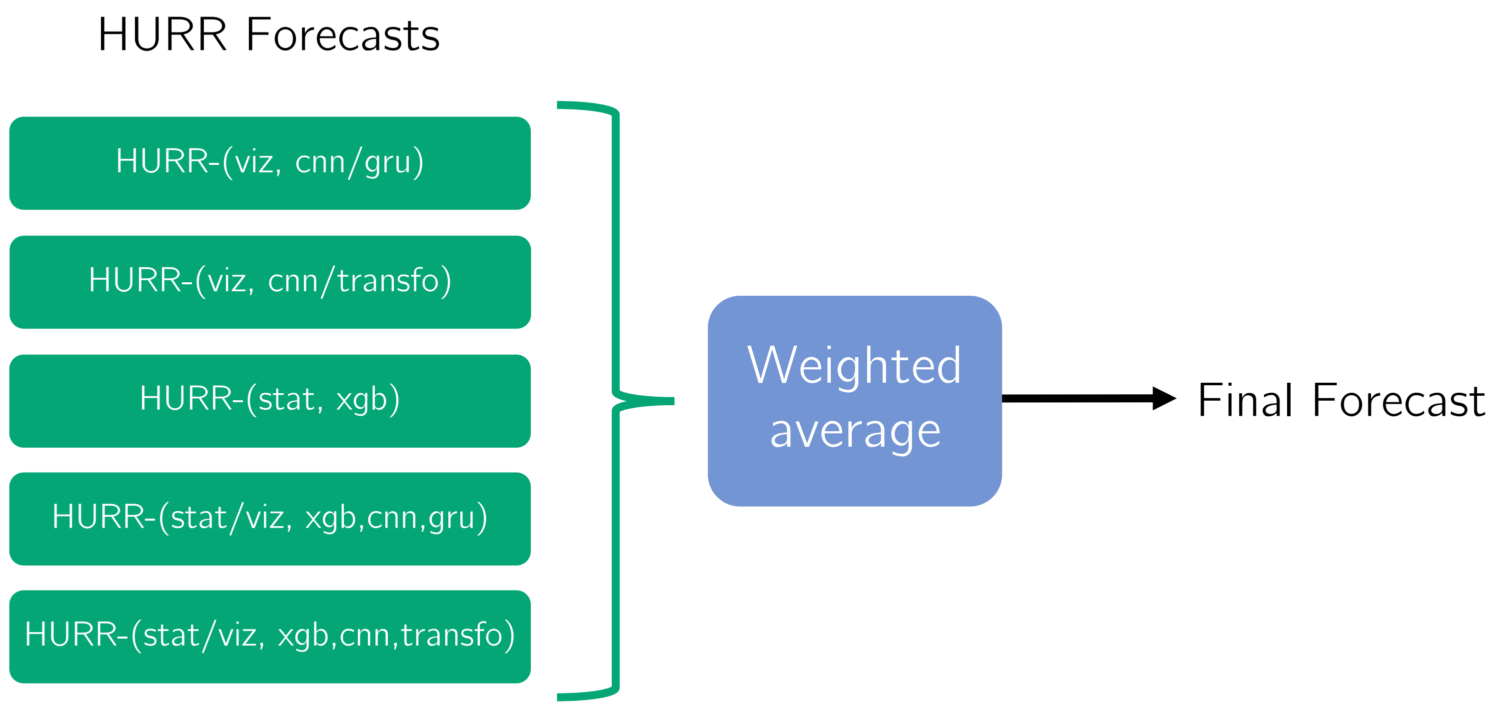
Track results

Model Type	Model Name	Eastern Pacific Basin Comparison on 36 TC		North Atlantic Basin Comparison on 45 TC	
		MAE (km)	**Error sd (km)**	**MAE (km)**	**Error sd (km)**
Hurricast (HURR) Methods	HURR-(viz, cnn/gru)	**73**	**43**	114	83
	HURR-(viz, cnn/transfo)	**73**	44	**110**	**71**
	HURR-(stat, xgb)	81	47	144	109
	HURR-(stat/viz, xgb/cnn/gru)	**71**	**43**	**110**	79
	HURR-(stat/viz, xgb/cnn/transfo)	72	43	**110**	**72**
Standalone Operational Forecasts	CLP5	121	67	201	149
	HWRF	67	**42**	75	**49**
	GFSO	**65**	45	**71**	54

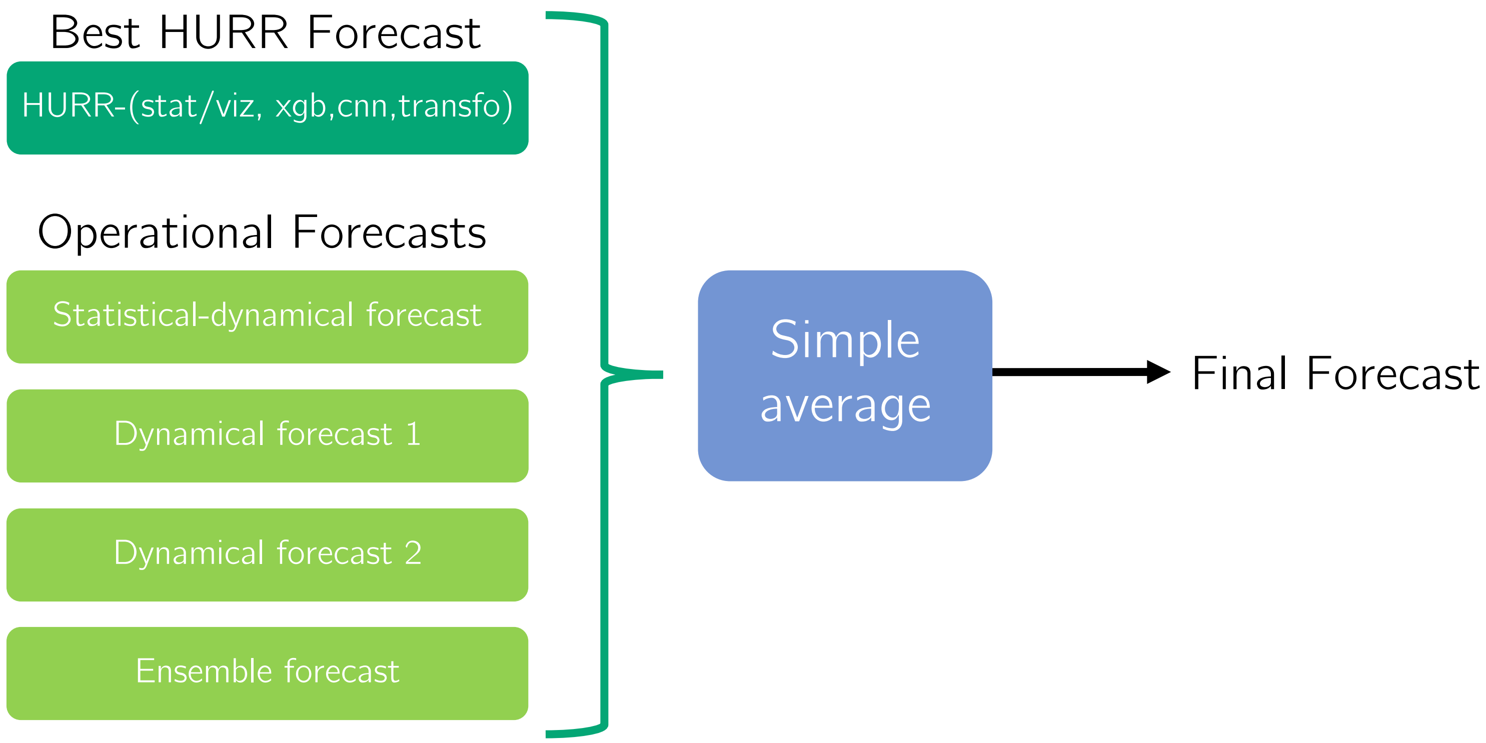
We have many models, let's ensemble them!

V. Ensemble models

Hurricast ensemble



Hurricast + Operational forecasts average



Intensity Results

Model Type	Model Name	Eastern Pacific Basin Comparison on 36 TC		North Atlantic Basin Comparison on 45 TC	
		MAE (kn)	**Error sd (kn)**	**MAE (kn)**	**Error sd (kn)**
Hurricast (HURR) Methods	HURR-(stat/viz, xgb/cnn/transfo)	10.3	**9.8**	10.4	**8.8**
	HURR-consensus	**10.2**	9.9	**10.2**	8.9
Operational Forecasts	FSSE	**9.7**	**9.5**	**8.5**	**7.8**
	OFCL	10.0	10.1	**8.5**	8.1
Consensus Models	Average consensus op. forecast	9.6	9.7	8.5	7.9
	HURR/OP-average consensus	**9.2**	**9.0**	**8.3**	**7.6**

- Ensembling our models improves performance.
- Including HURR into a simple operational forecast consensus is beneficial.

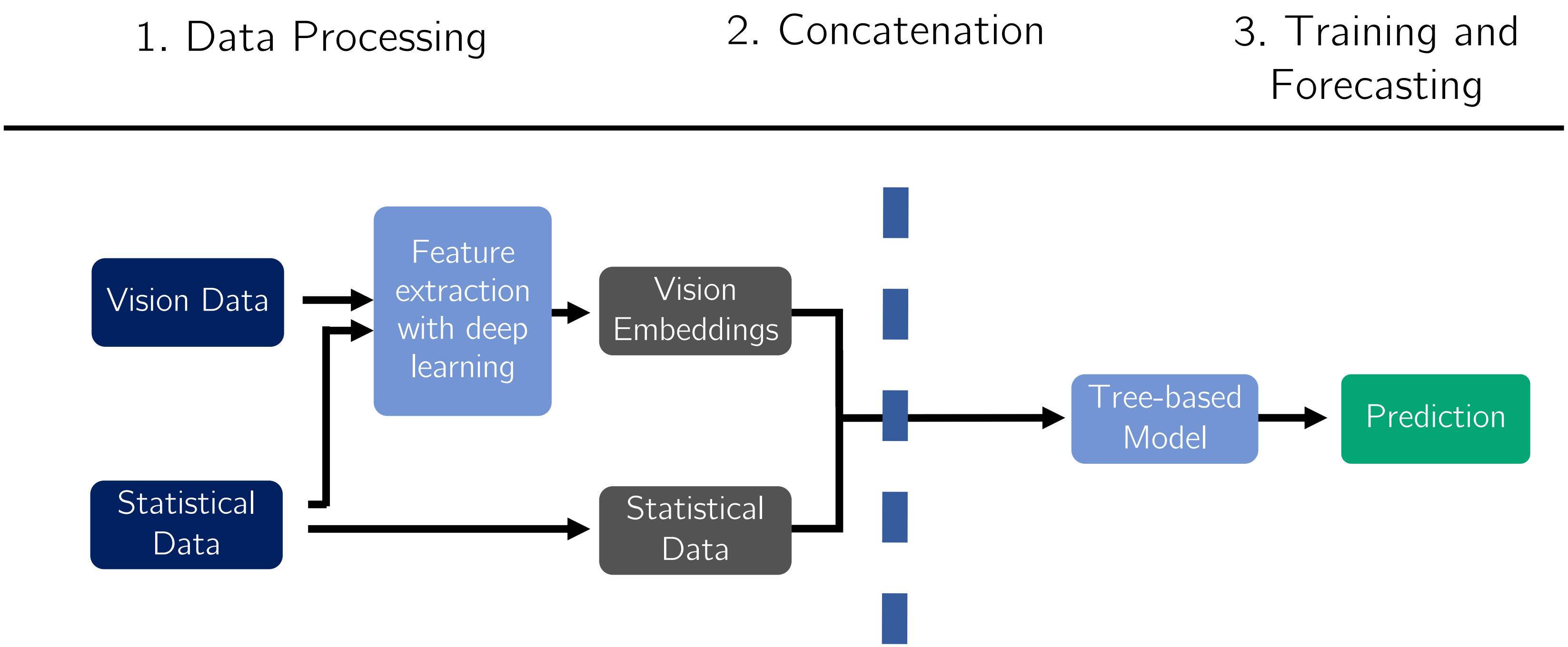
Track Results

Model Type	Model Name	Eastern Pacific Basin Comparison on 36 TC		North Atlantic Basin Comparison on 45 TC	
		MAE (km)	**Error sd (km)**	**MAE (km)**	**Error sd (km)**
Hurricast (HURR) methods	HURR-(stat/viz, xgb/cnn/transfo)	72	43	110	**72**
	HURR-consensus	**68**	**41**	**107**	77
Operational Forecasts	AEMN	60	37	73	55
	FSSE	56	47	**69**	**53**
	OFCL	**54**	**33**	71	56
Consensus Models	Average consensus op. forecast	55	37	64	48
	HURR/OP-average consensus	**50**	**32**	**61**	**43**

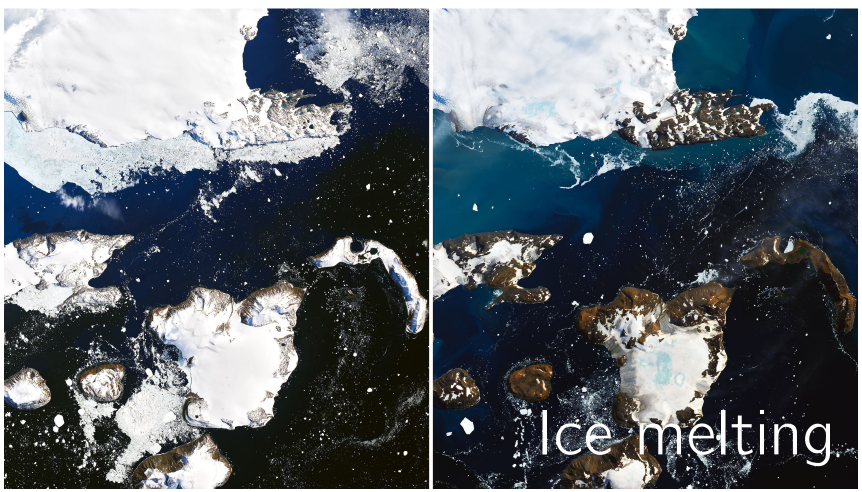
- Ensembling our models improves performance.
- Including HURR into a simple operational forecast consensus is beneficial.

VI. Exciting applications

Other applications of the framework



Satellite data holds a lot of potential



Conclusion

- Multimodality and ensemble models are powerful!
- Machine Learning can advance hurricane forecasting.
- Significant potential of feature extraction techniques combined with tree-based models.

Thank you for your attention!

Hurricane Forecasting: A Novel Multimodal Machine Learning Framework

Léonard Boussioux and Cynthia Zeng
Joint work with Dimitris Bertsimas and Théo Guénais

https://arxiv.org/abs/2011.06125

OPERATIONS RESEARCH CENTER