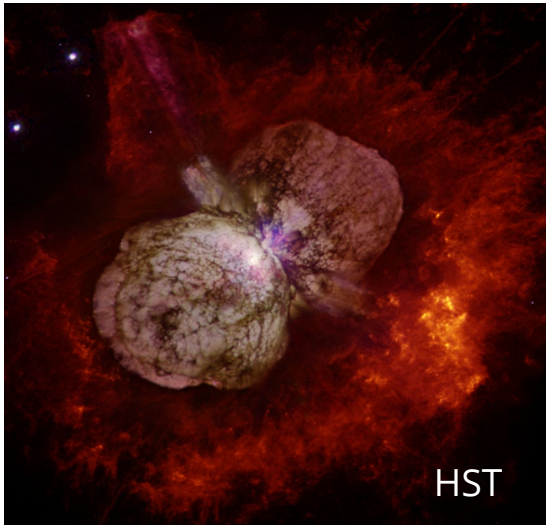
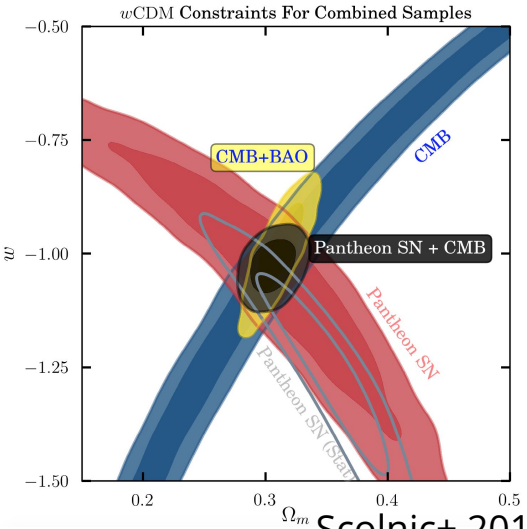
Time-domain Astrophysics in the Era of Big Data

V. Ashley Villar
Center for Astrophysics | Harvard & Smithsonian
Ford Foundation Dissertation Fellow

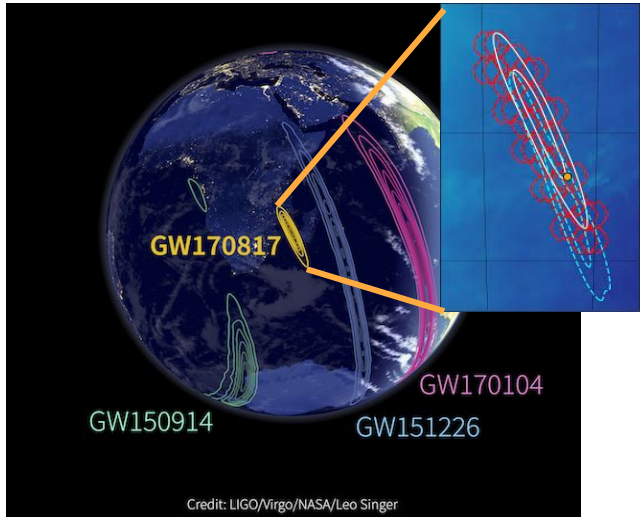
ML @ Ringberg 2019

Transients connect to all branches of astrophysics



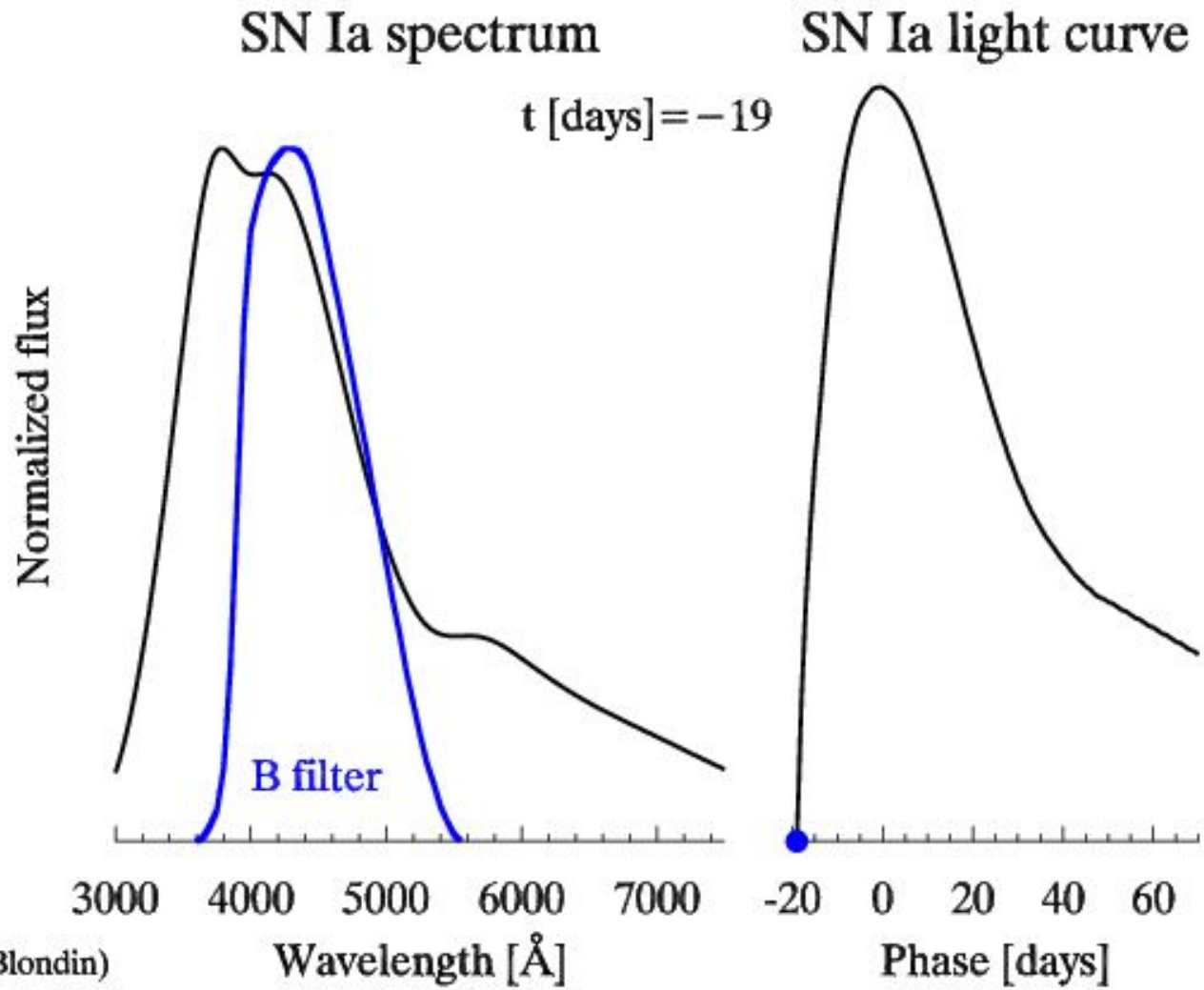


Scolnic+ 2018

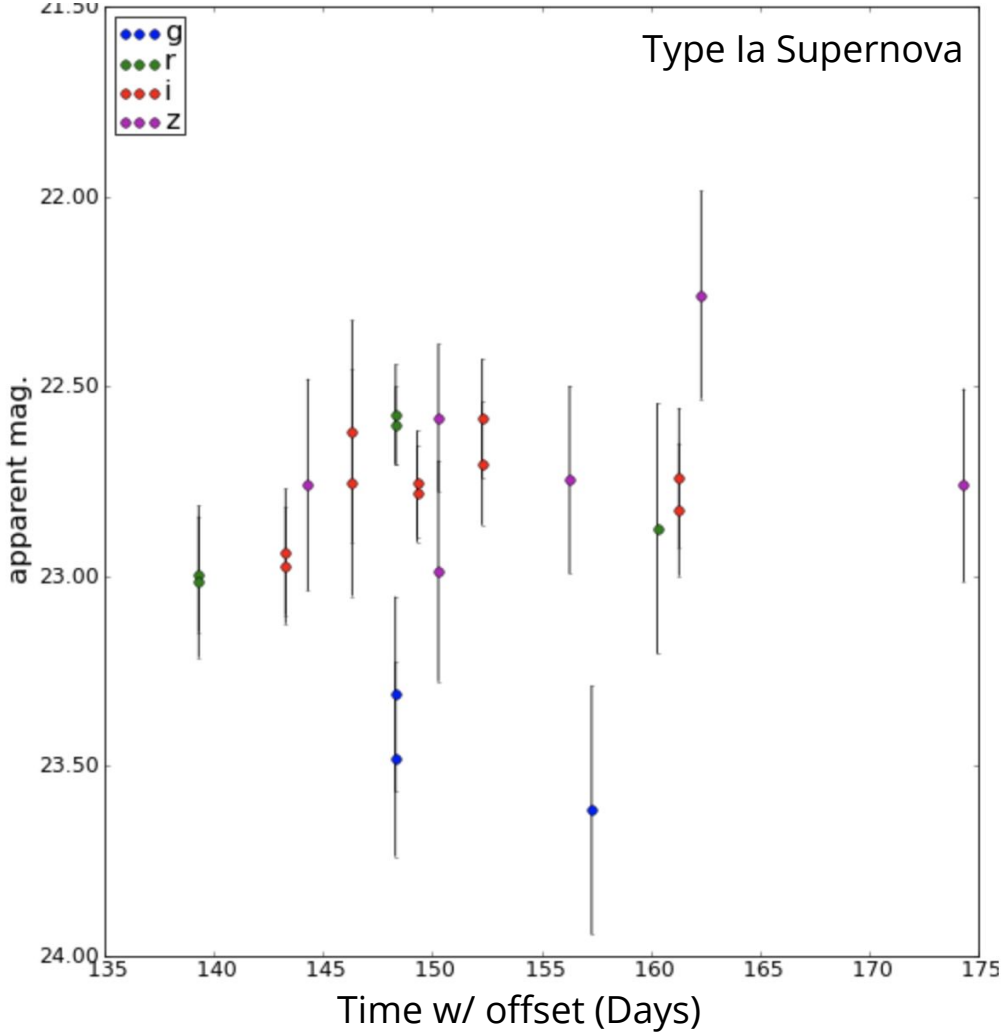


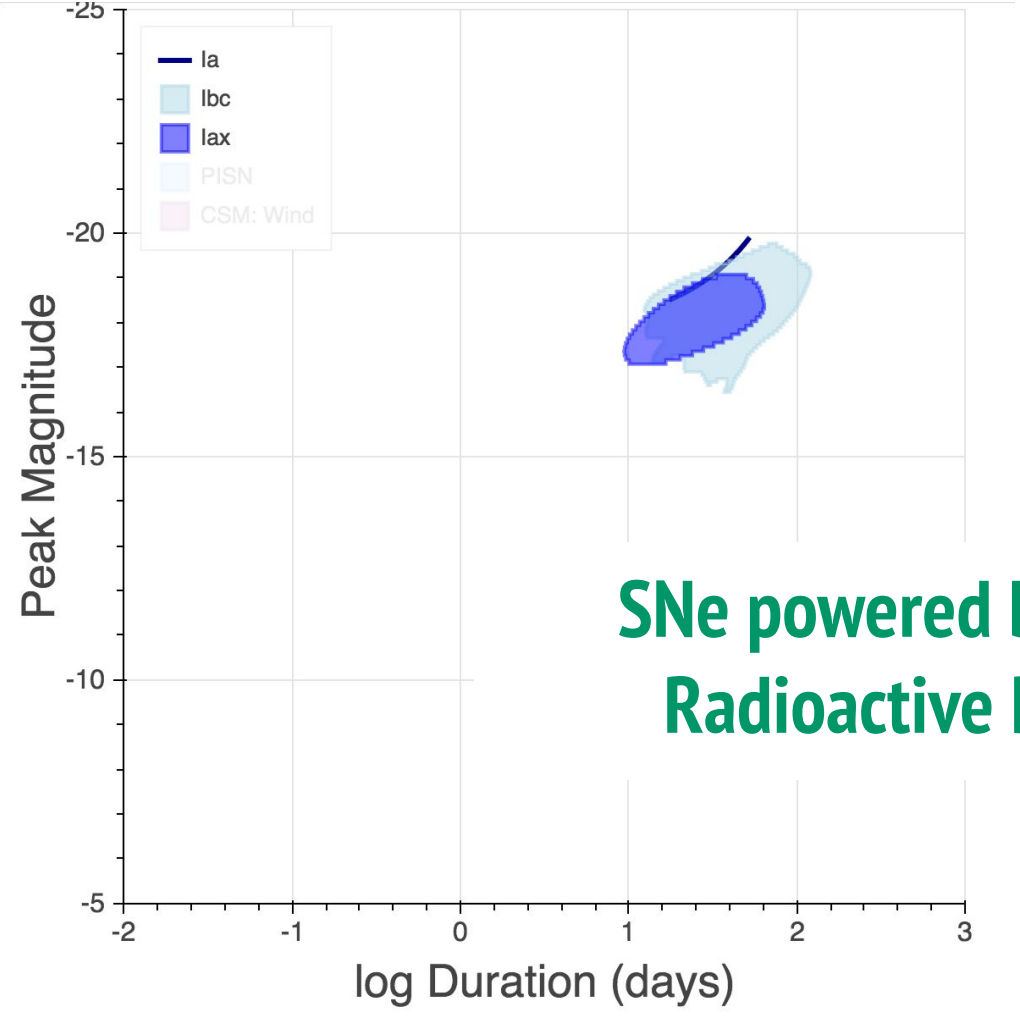
Soares-Santos+2017
L. Singer

How does the zoo of observed transients connect with the underlying (astro)physics?


SN Ia spectrum
SN Ia light curve
t [days] = −19
Normalized flux
B filter
3000
4000
5000
6000
7000
Wavelength [Å]
-20
0
20
40
60
Phase [days]

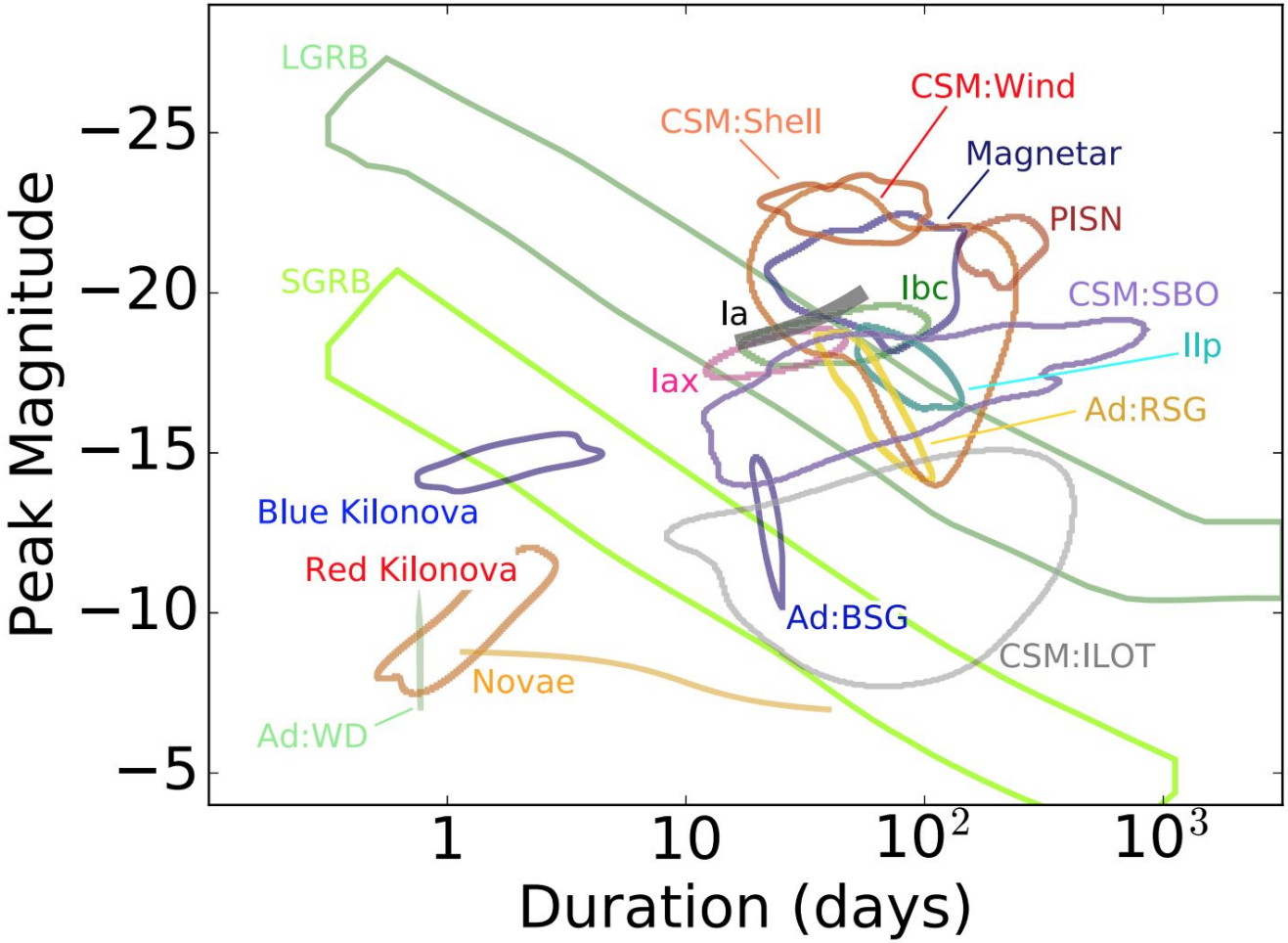
(S. Blondin)


Type Ia Supernova
g
r
i
z
apparent mag.
21.50
22.00
22.50
23.00
23.50
24.00
135
140
145
150
155
160
165
170
175
Time w/ offset (Days)



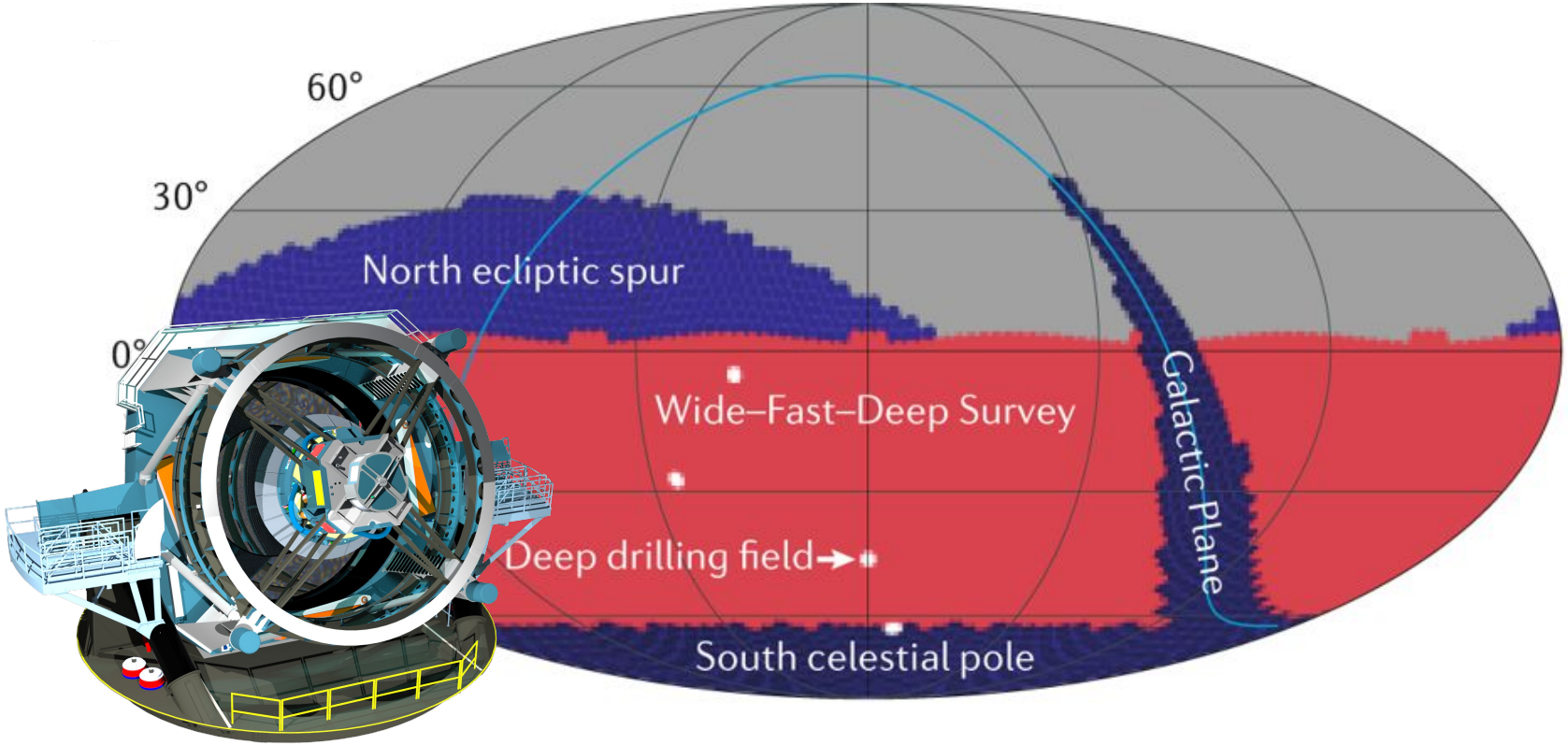
SNe powered by 56Ni Radioactive Decay

ashleyvillar.com/dlps


LGRB
CSM:Wind
CSM:Shell
Magnetar
PISN
SGRB
Ibc
CSM:SBO
Ia
IIp
Iax
Ad:RSG
Blue Kilonova
Red Kilonova
Ad:BSG
CSM:ILOT
Novae
Ad:WD
Peak Magnitude
−25
−20
−15
−10
−5
1
10
10^2
10^3
Duration (days)

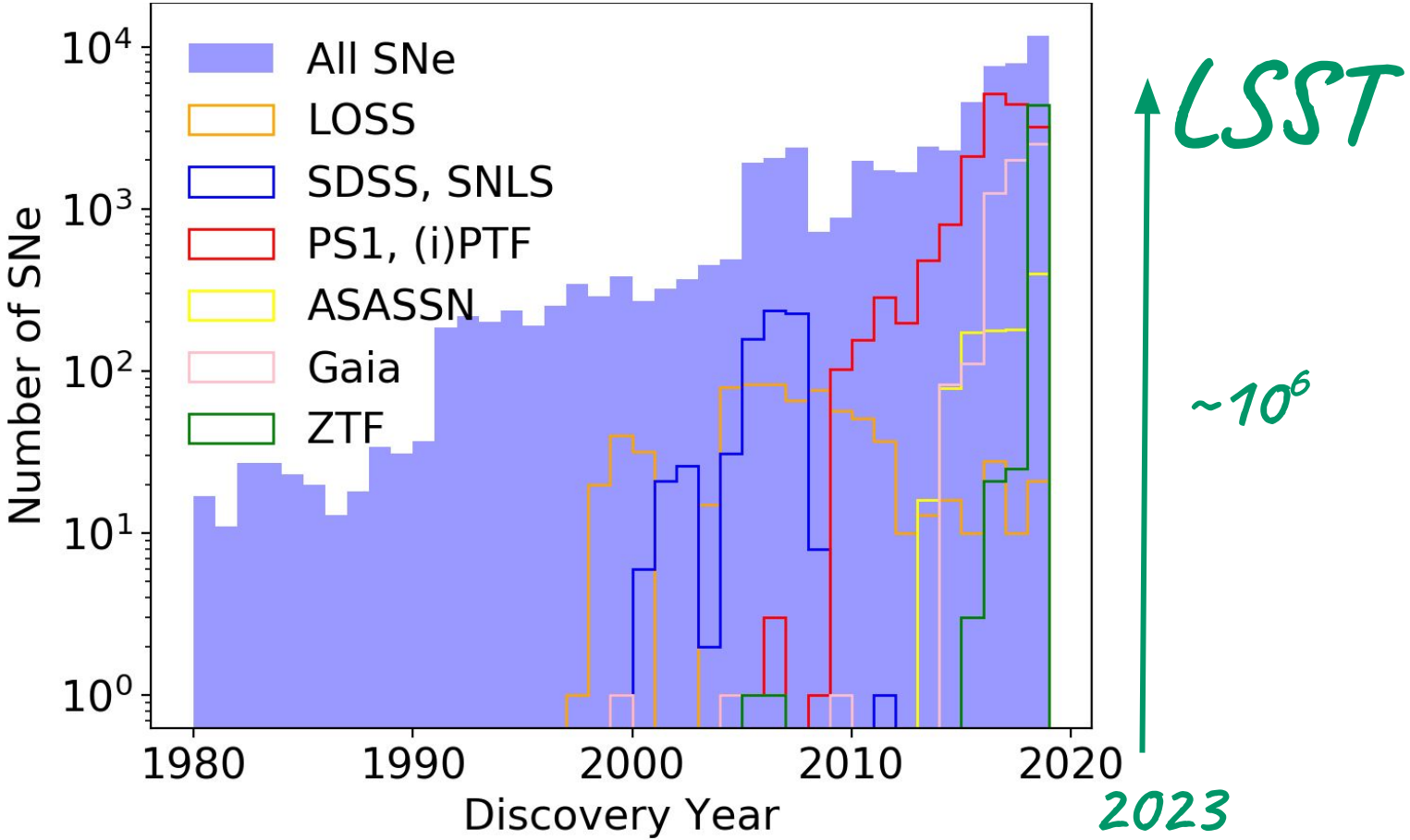
VAV+ 2017a

The Large Synoptic Survey Telescope



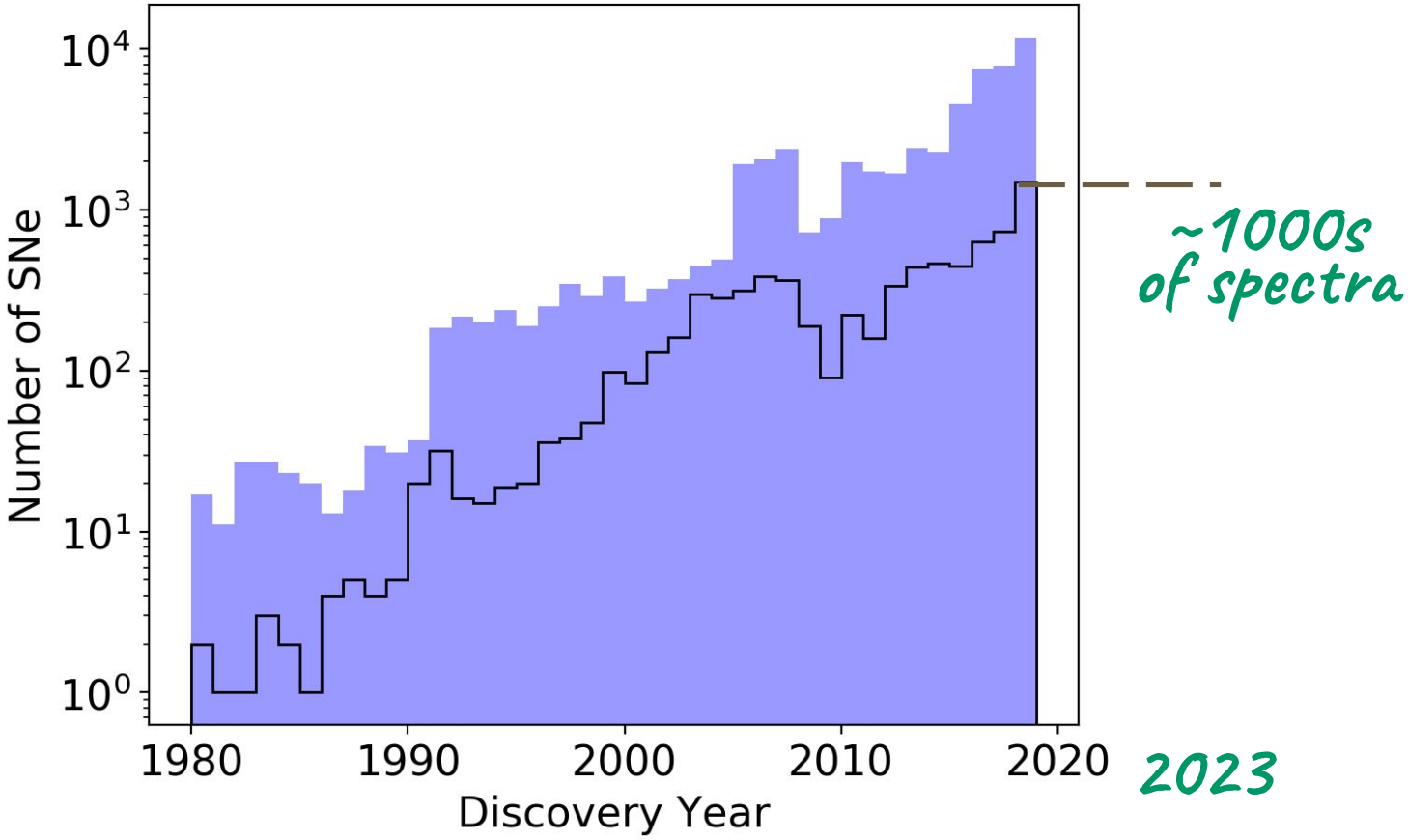
LSST

LSST will discover >1 million supernovae annually!



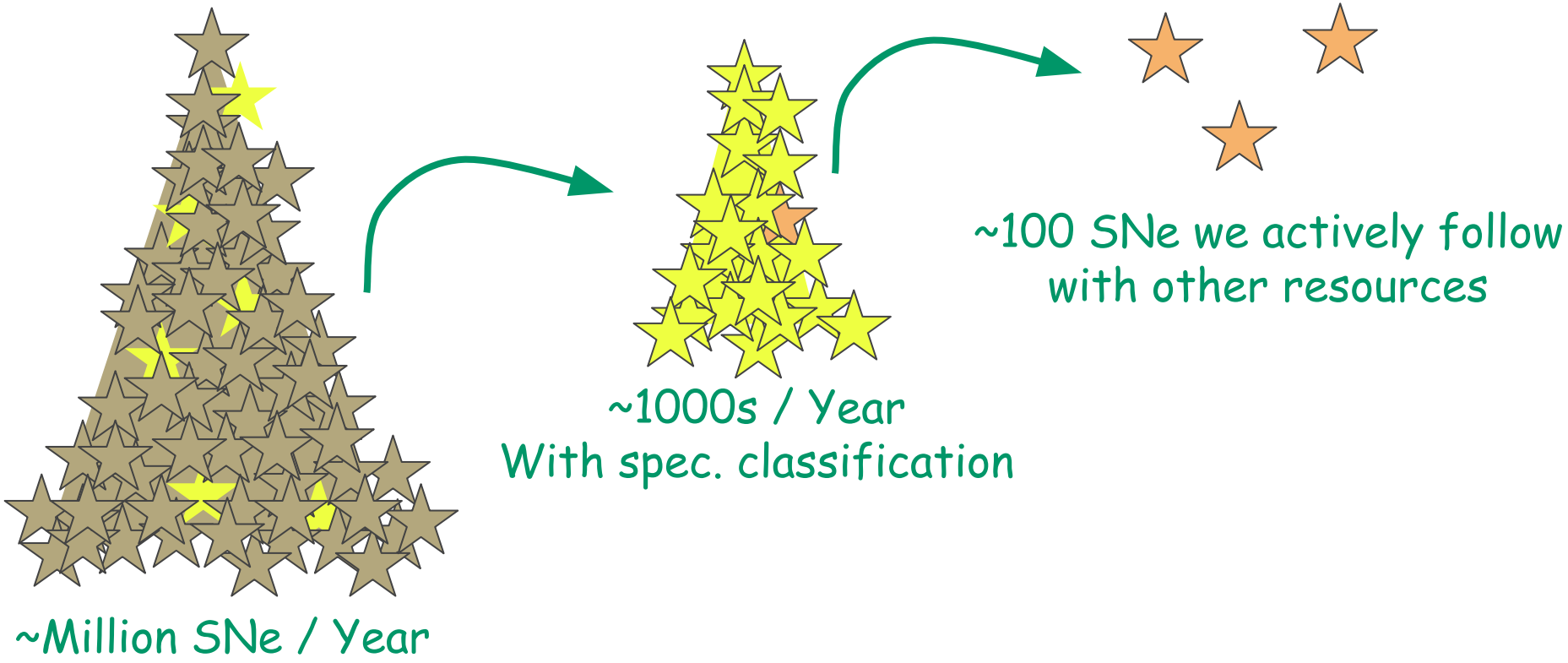
Data from OSC; Guillochon+ 2018

We will follow up *some* of these supernovae



Data from OSC; Guillochon+ 2018

The LSST Needles & the Haystack



A Christmas list for SN classification:

1. Meaningful feature extraction which can handle noisy, sparse data
2. Feature extraction which can utilize unclassified data
3. Classification which can work on incoming data
4. *A method which can search for needles in real time*

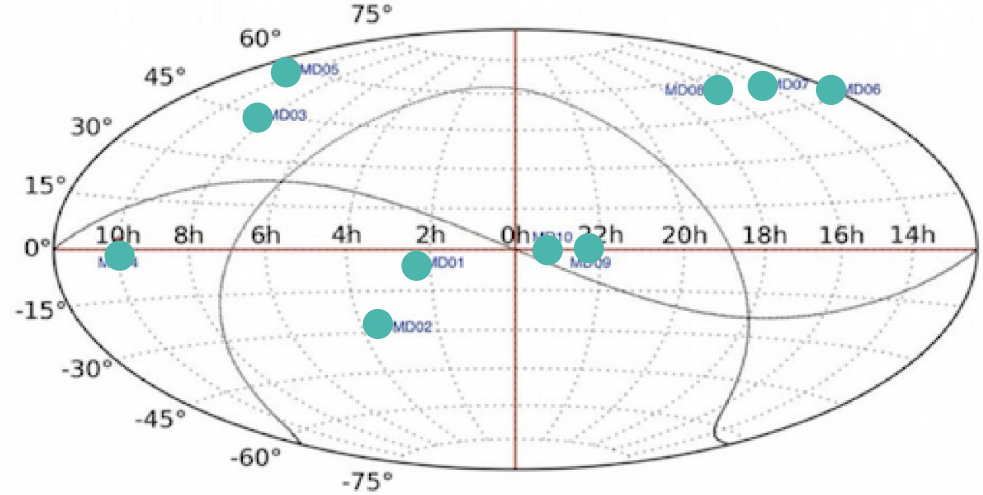
A Christmas list for SN classification:

1. Meaningful feature extraction which can handle noisy, sparse data
2. Feature extraction which can utilize unclassified data
3. Classification which can work on incoming data
4. *A method which can search for needles in real time*

Recurrent neuron-based autoencoder

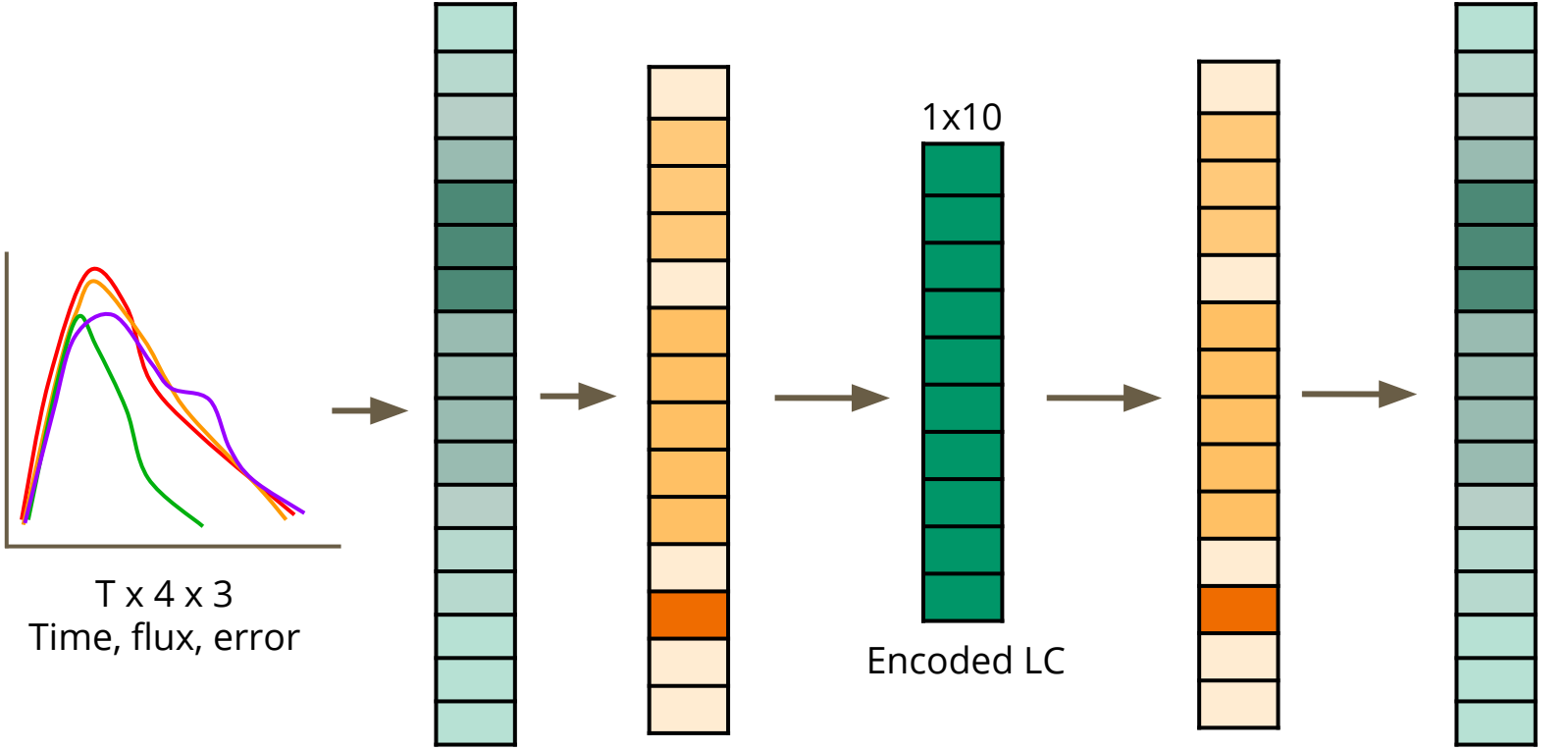
Pan-STARRS Medium Deep Survey is a milliLSST

- ~5200 SNe-like transients in PS1 MDS (Jones+2017)
- **~3200 SNe** have host redshift measurements
- **~520 SNe** are spectroscopically classified with host redshift measurements



Chambers+ 2016

A semi-supervised method to encode/classify SNe

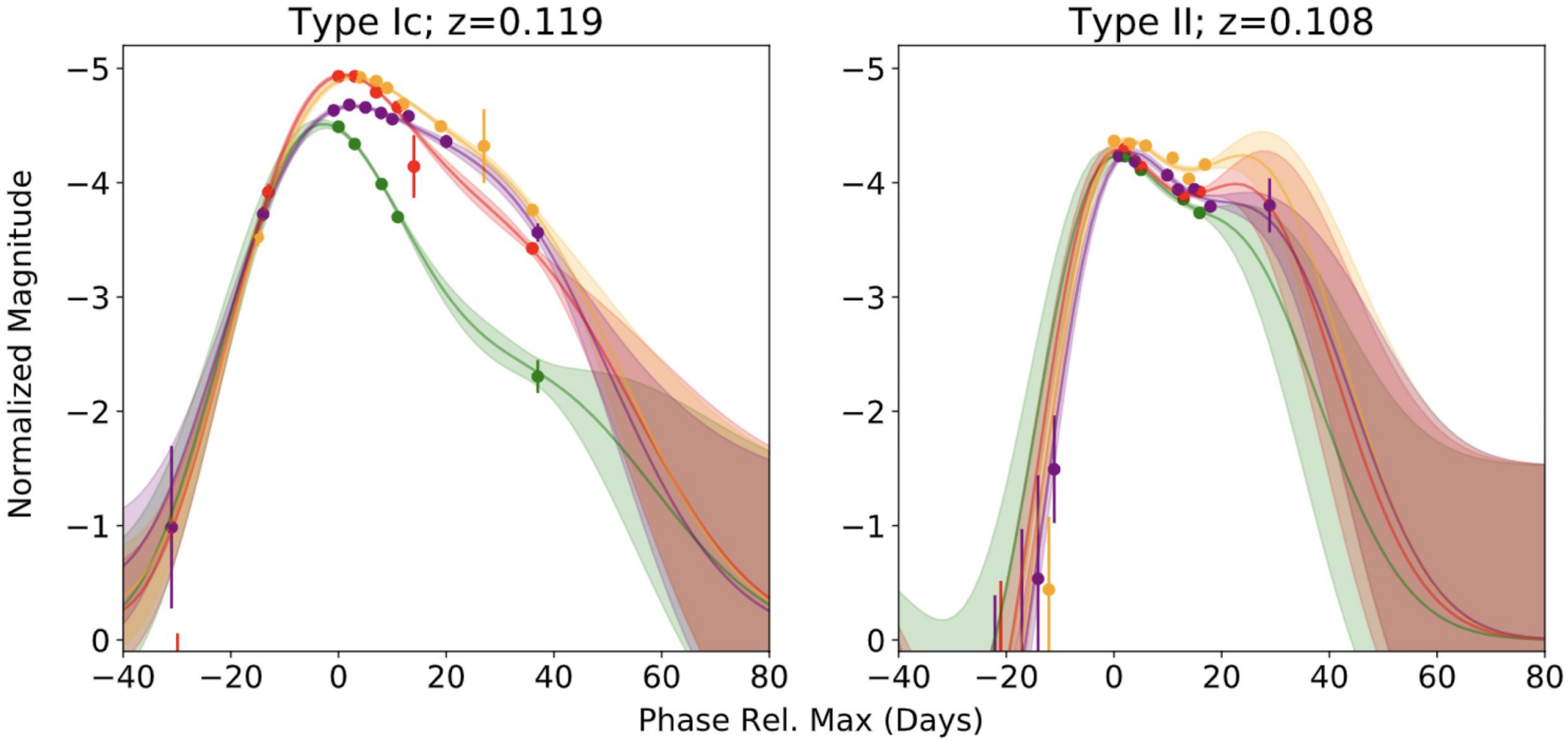


Encoder

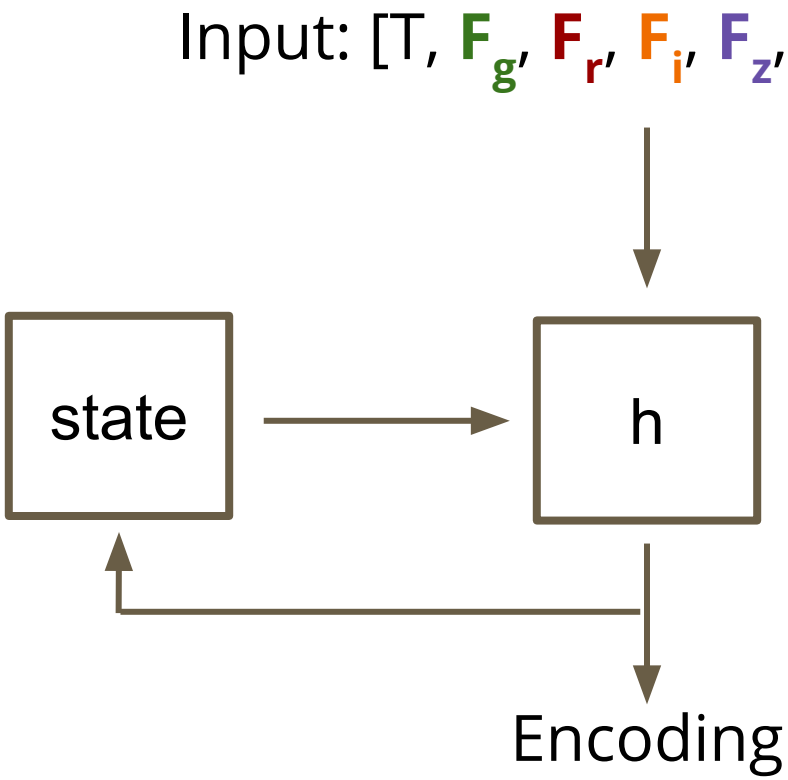
Decoder

VAV+ in prep.

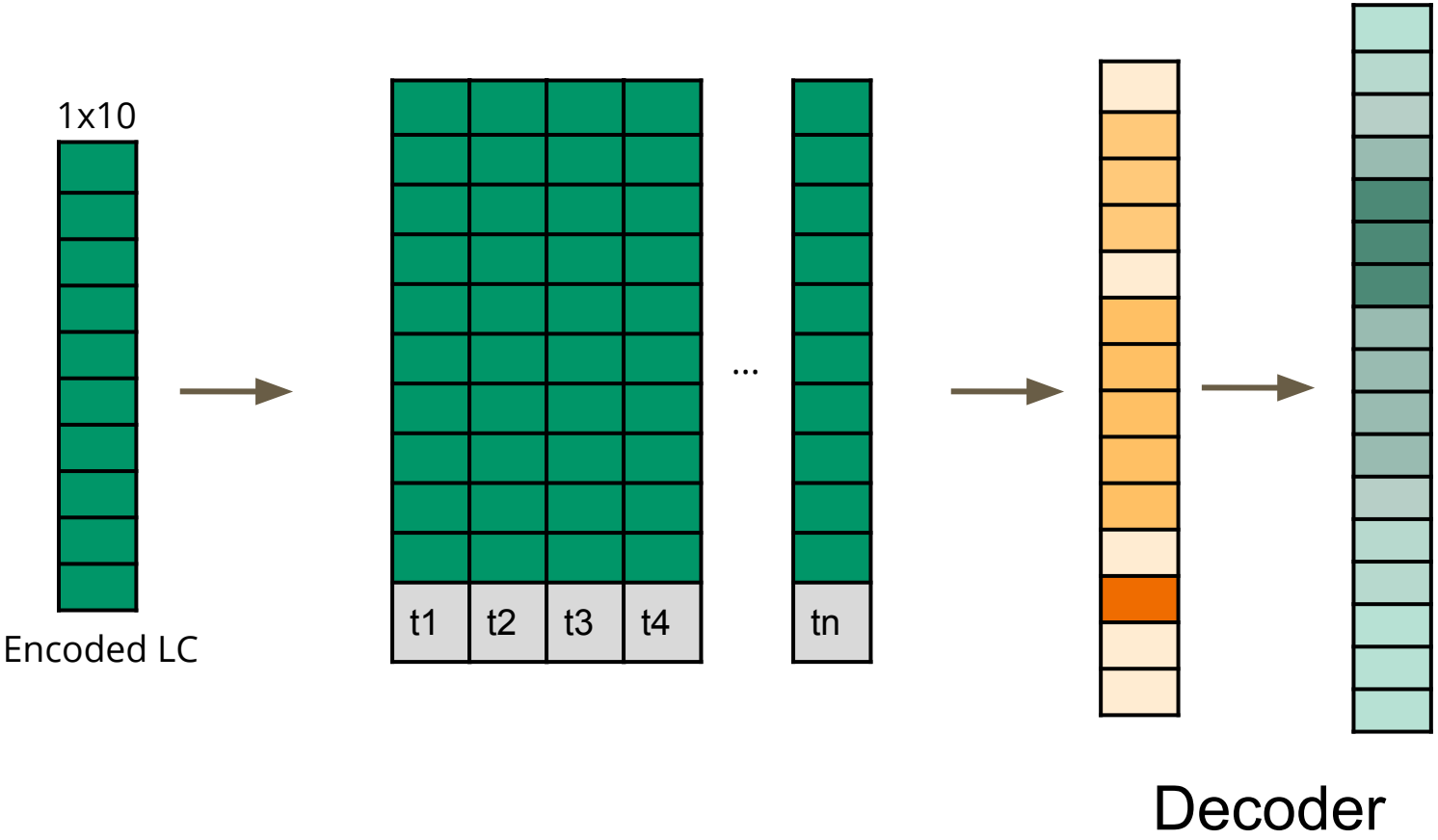
Use a GP to deal with uneven sampling in filters



Recurrent neurons update the encoded light curve

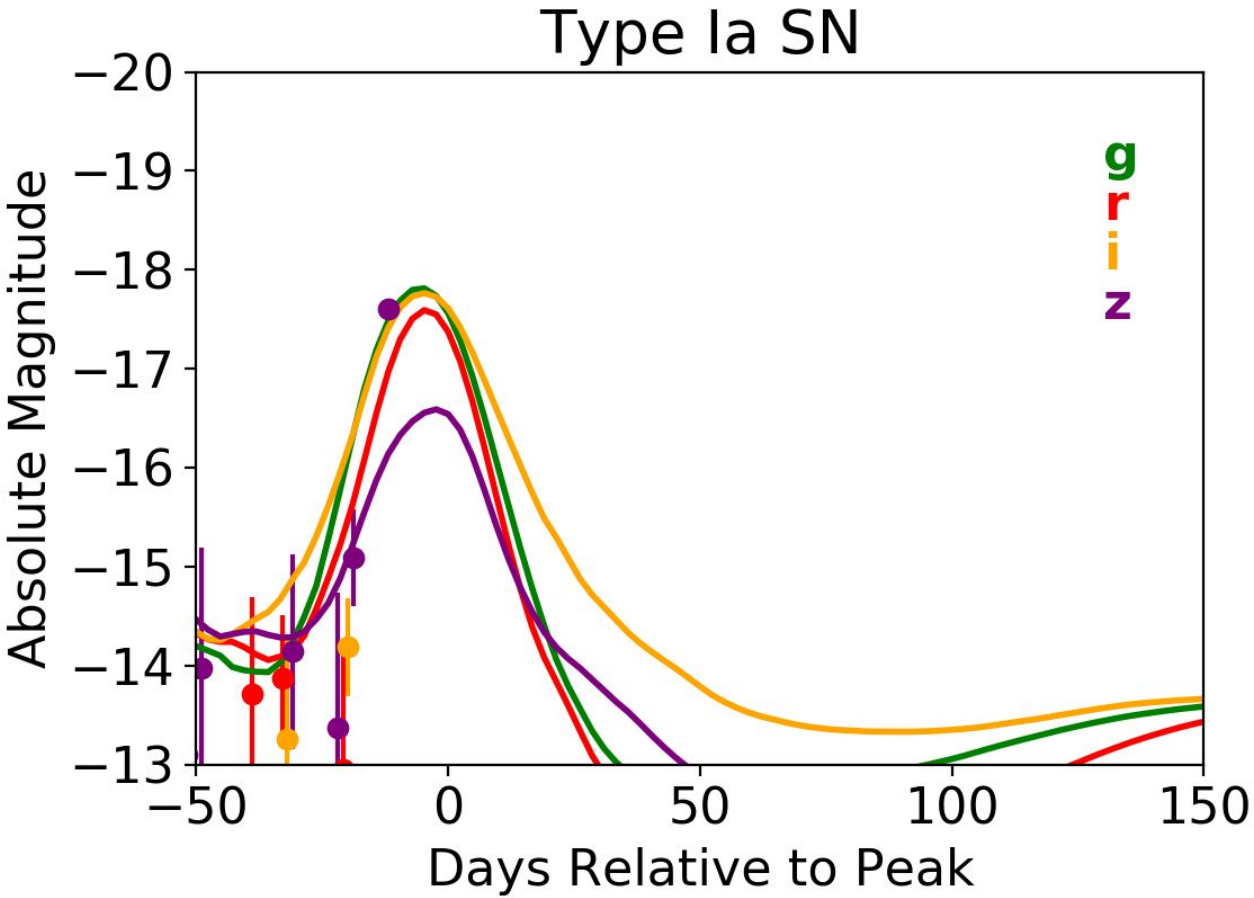


Repeat encoded LC with a new set of times



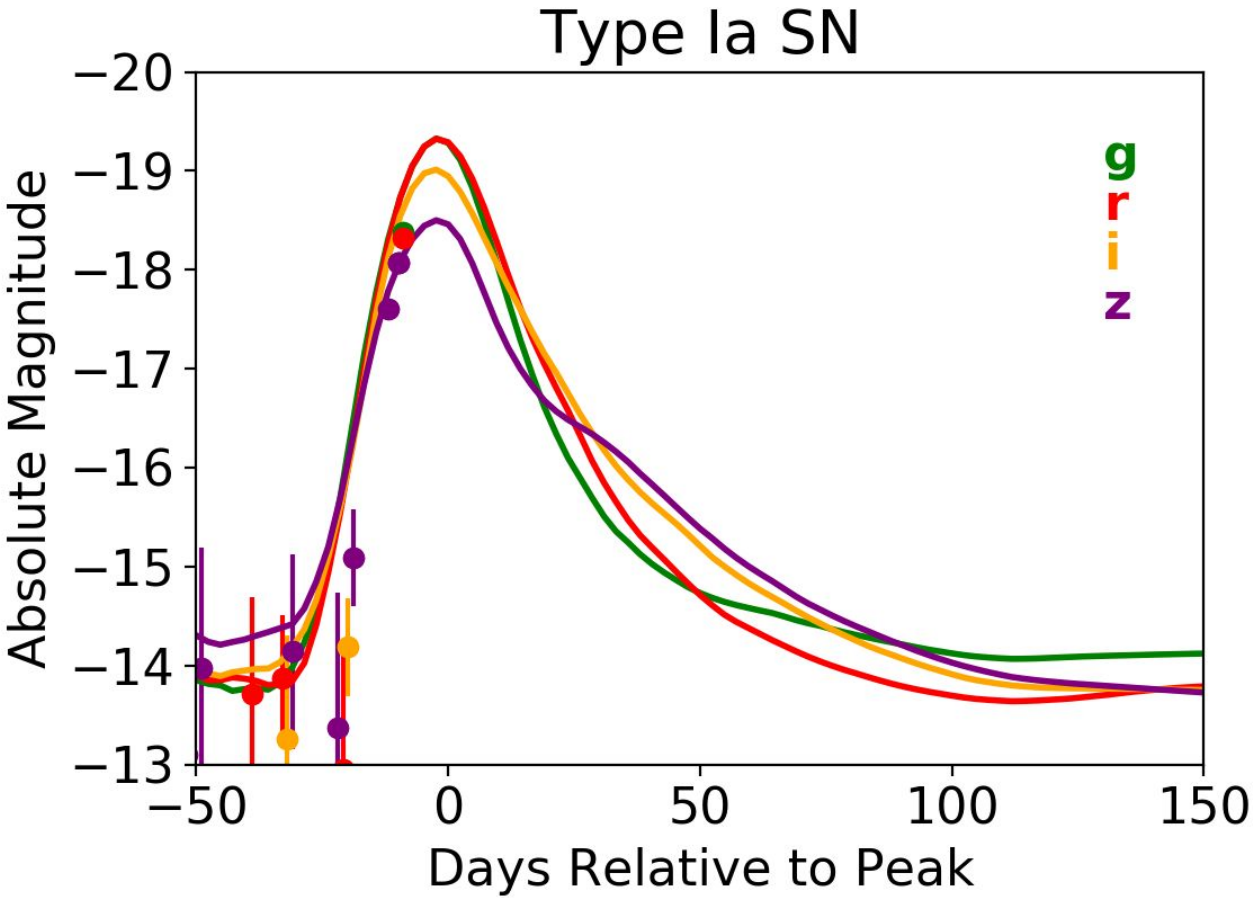
VAV+ in prep.

Decoded light curve updated with new data



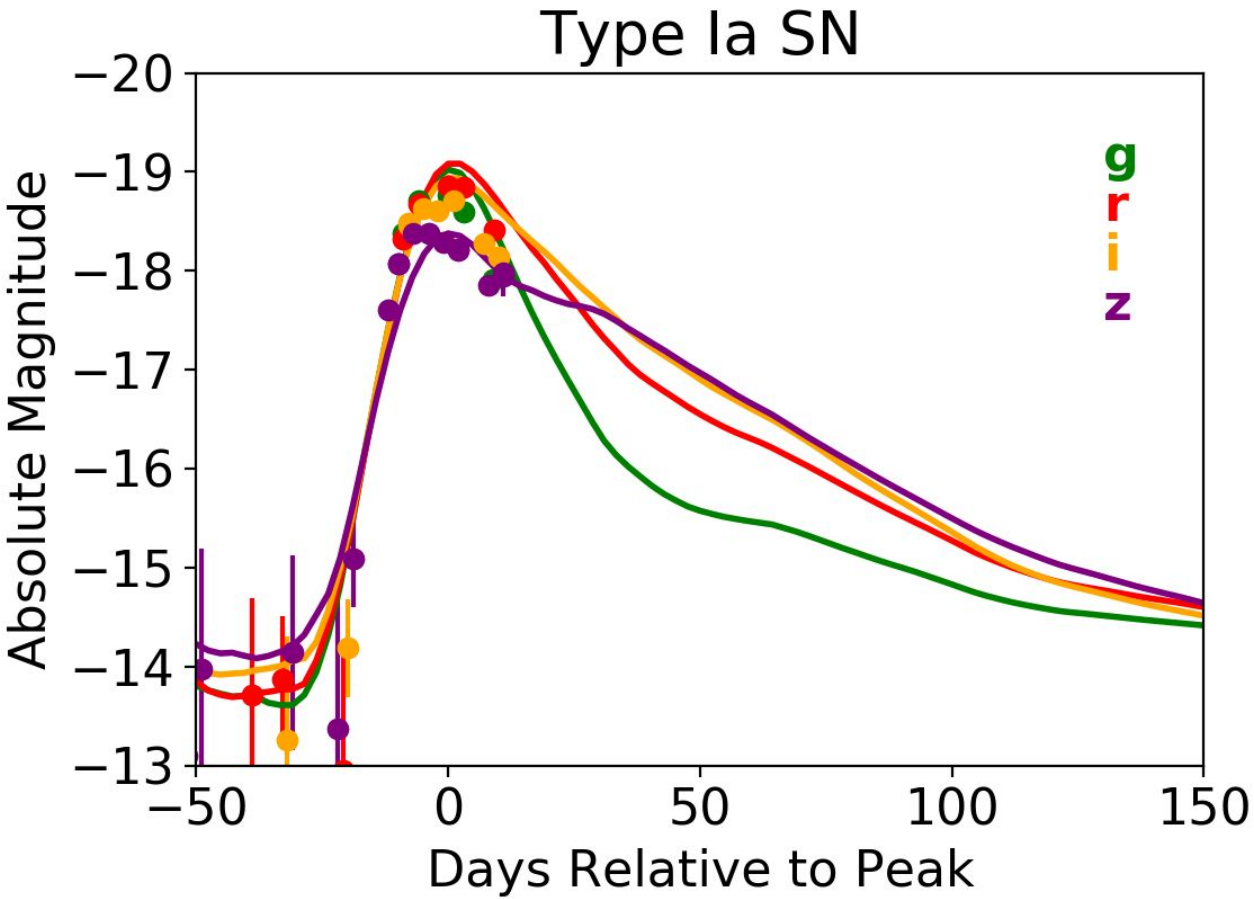
VAV+ in prep.

Decoded light curve updated with new data



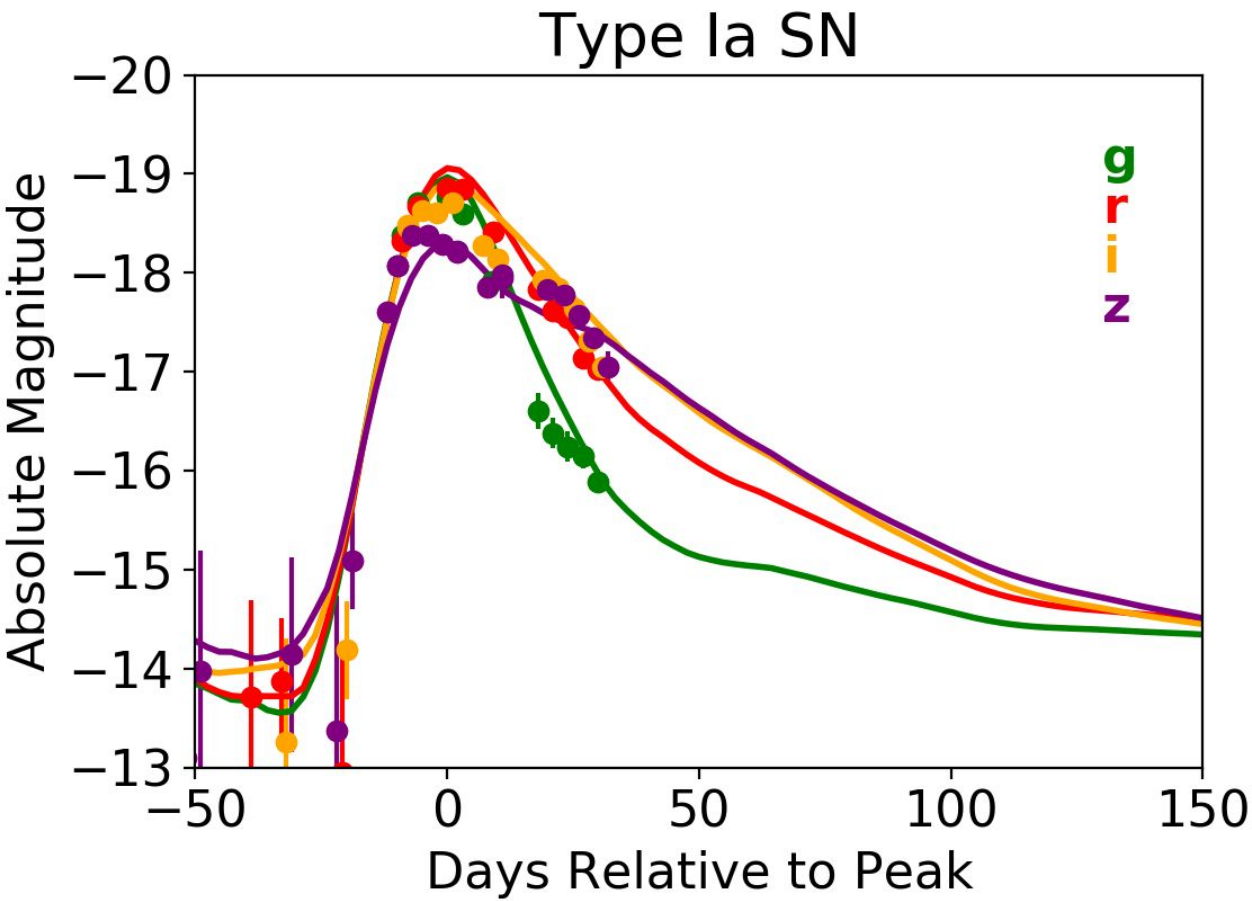
VAV+ in prep.

Decoded light curve updated with new data



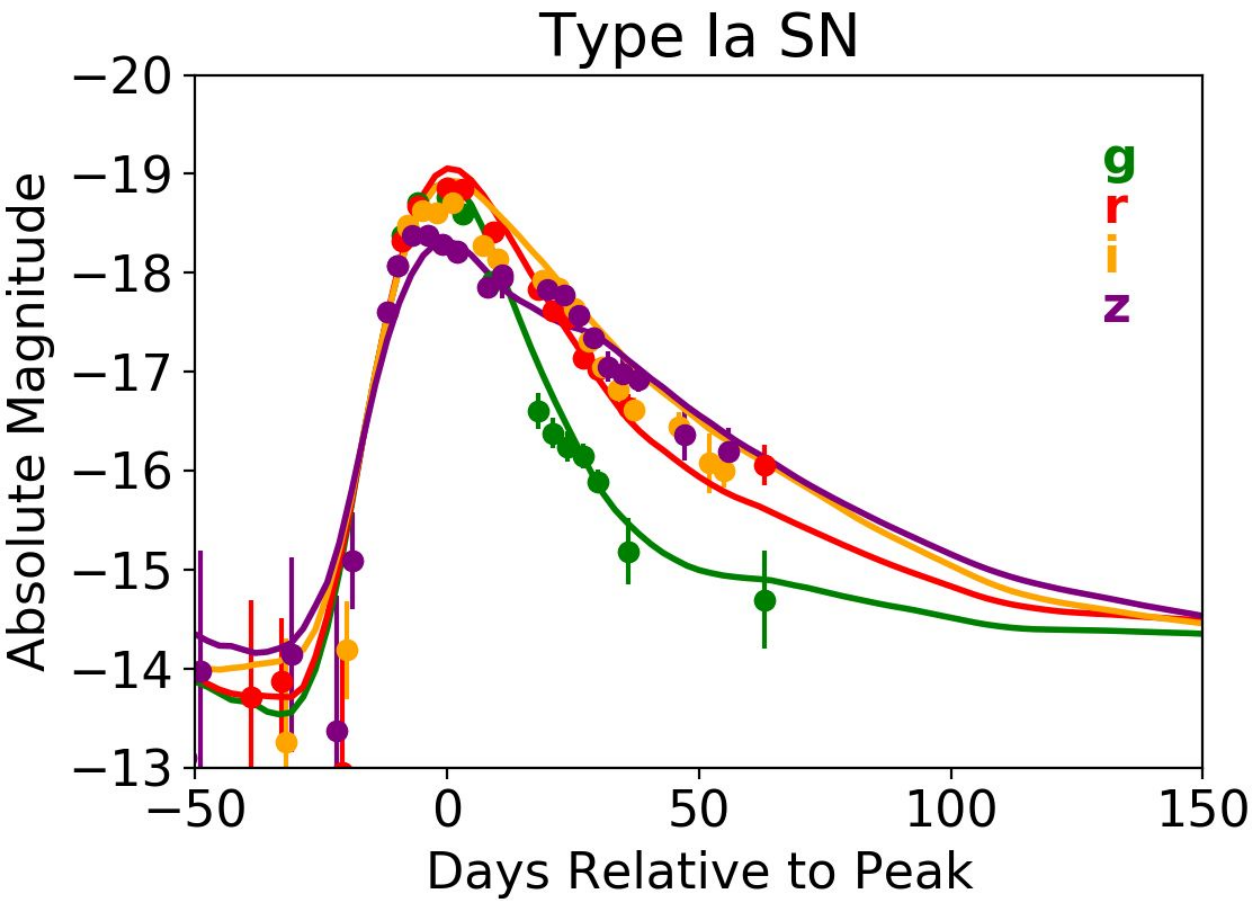
VAV+ in prep.

Decoded light curve updated with new data



VAV+ in prep.

Decoded light curve updated with new data

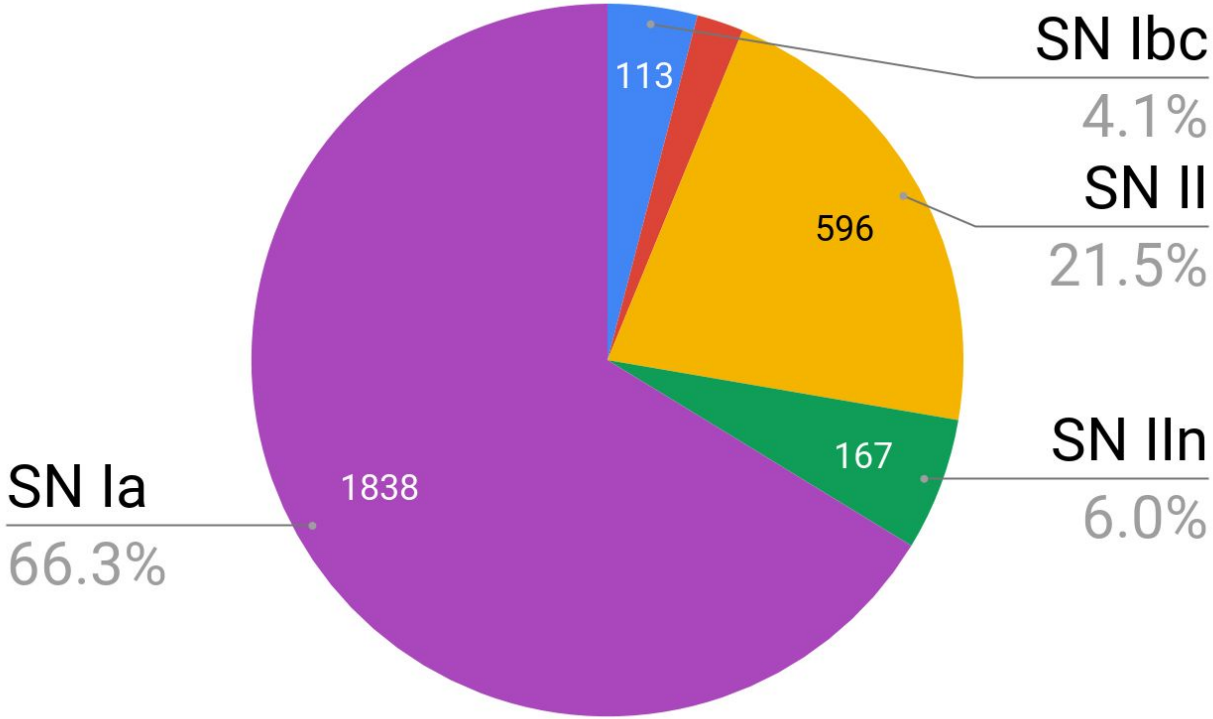


VAV+ in prep.

Why use a RNN autoencoder?

- Semi-supervised methods allow us to use information from *the full dataset*
- We can extract *unique, nonlinear* features directly from the light curves
- Actively makes *forecasting* predictions, which may be used to hunt for anomalies aka **the needles**

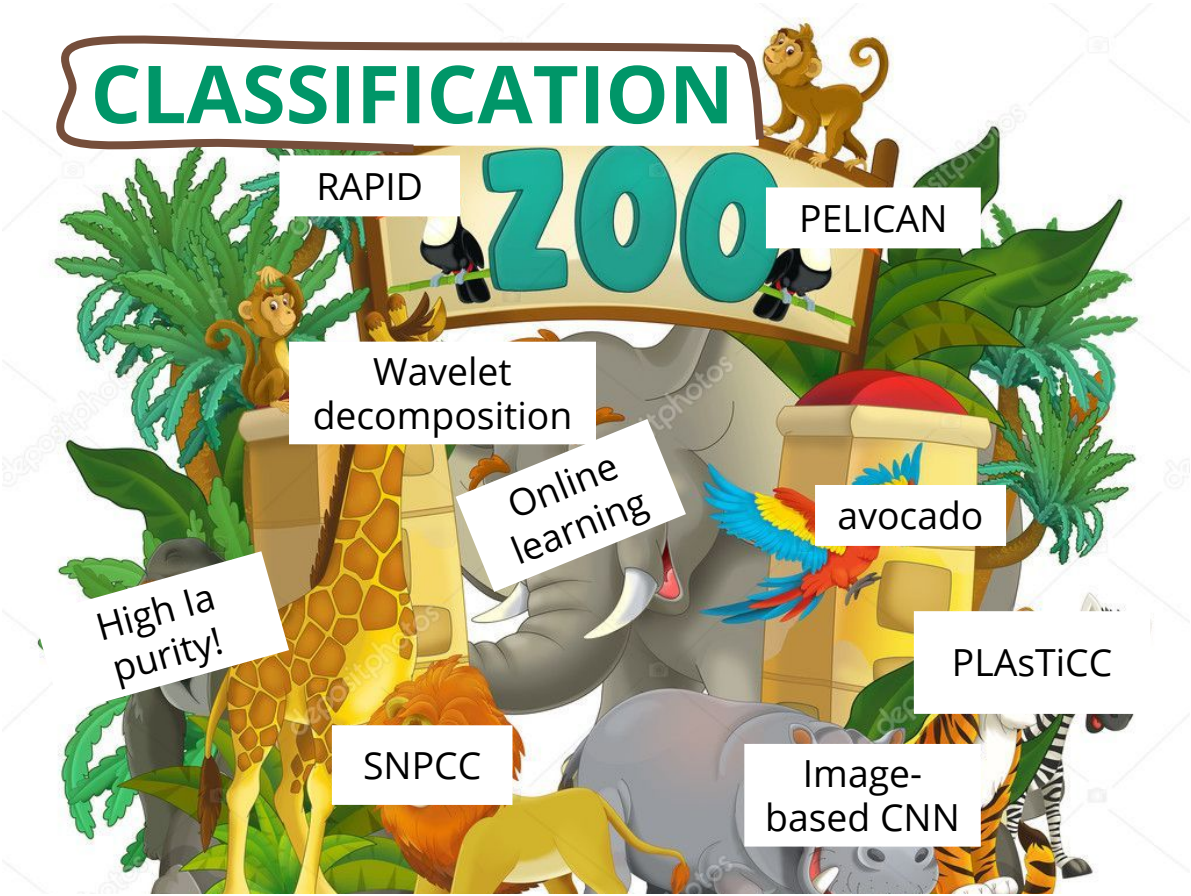
Using a random forest classifier, we classify the full sample of 3200 SNe



VAV+ in prep.

Time-domain Astrophysics in the Era of Big Data

- LSST will bring TDA into a new era of big data, thanks to both a deep and wide survey strategy
- LSST light curves will be noisy and sparse, but simple features correlate with underlying physics
- RNN-based AEs are a promising strategy to classify SNe in real time
- RNN-based AEs *may* be a promising strategy for real time anomaly detection



https://tinyurl.com/transienttable

Do we have a suitable training set for classification?

PLAsTiCC

Real datastream!
gr(i) filters
Depth ~21 mag

Simulated dataset
LSST filters/cadence

see e.g., Bellm+ 2019; Kessler+ 2019