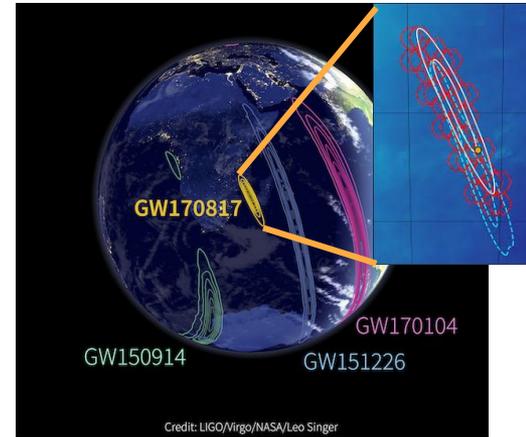
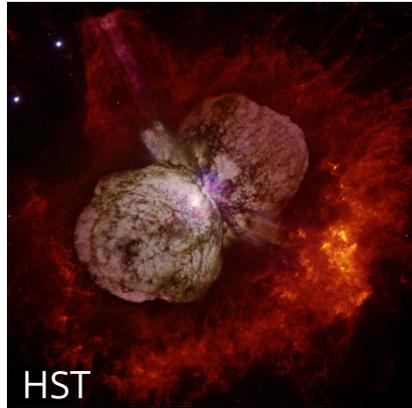

Recurrent VAE for anomaly detection in supernova time series

— V. Ashley Villar —
Simons Society Junior Fellow,
Columbia University/Flatiron Institute (CCA)

With thanks to CCA group: S. Ho, M. Cranmer, G. Contardo, J. Lin

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



Soares-Santos+2017





Extragalactic transients, like *supernovae*, are visible for days to months

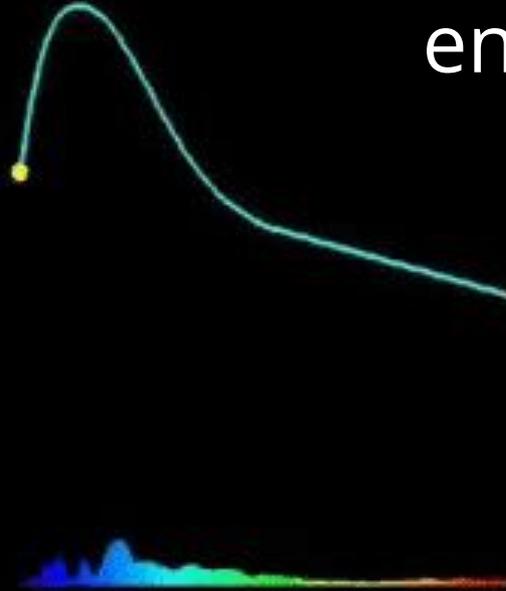


Bulk properties, like
the black body
temperature/radius

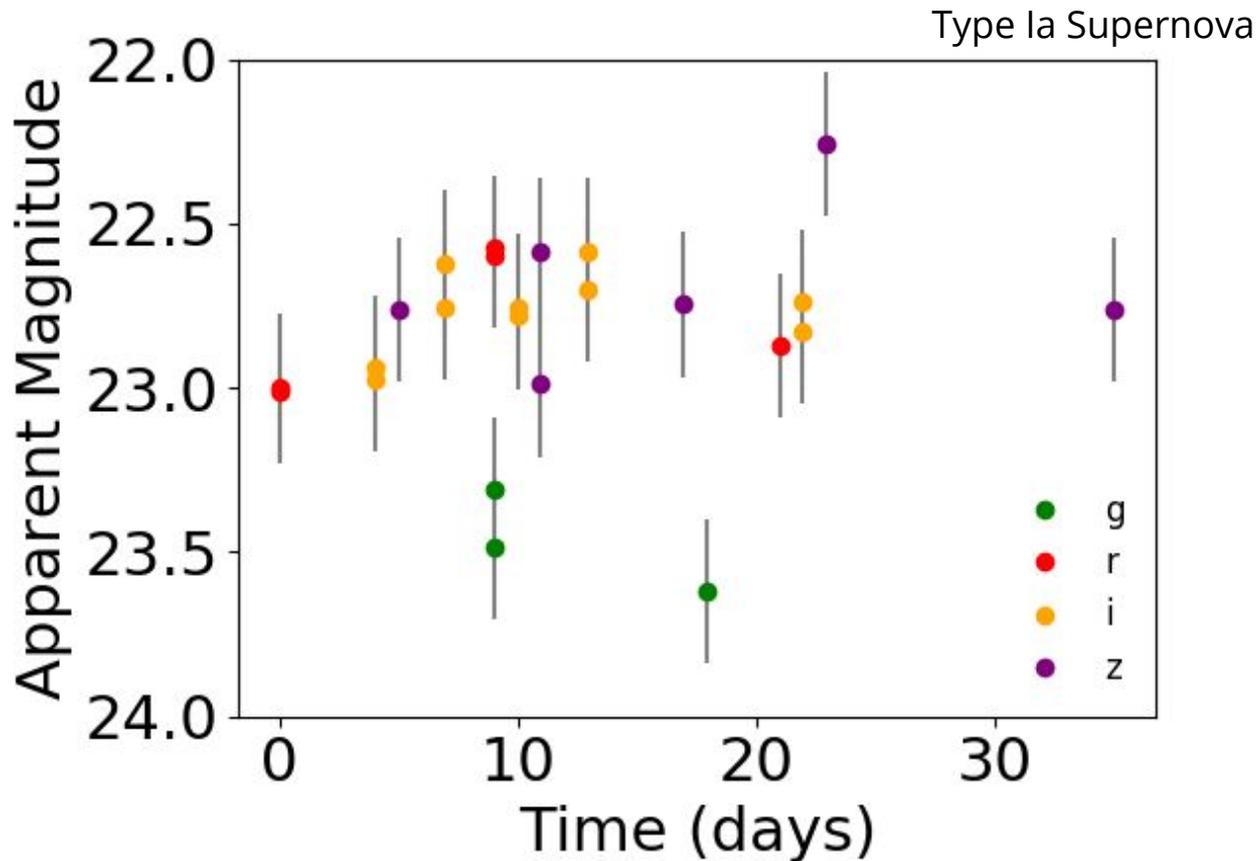
“Fingerprints” of
elements



Supernova
light curves
encode physics



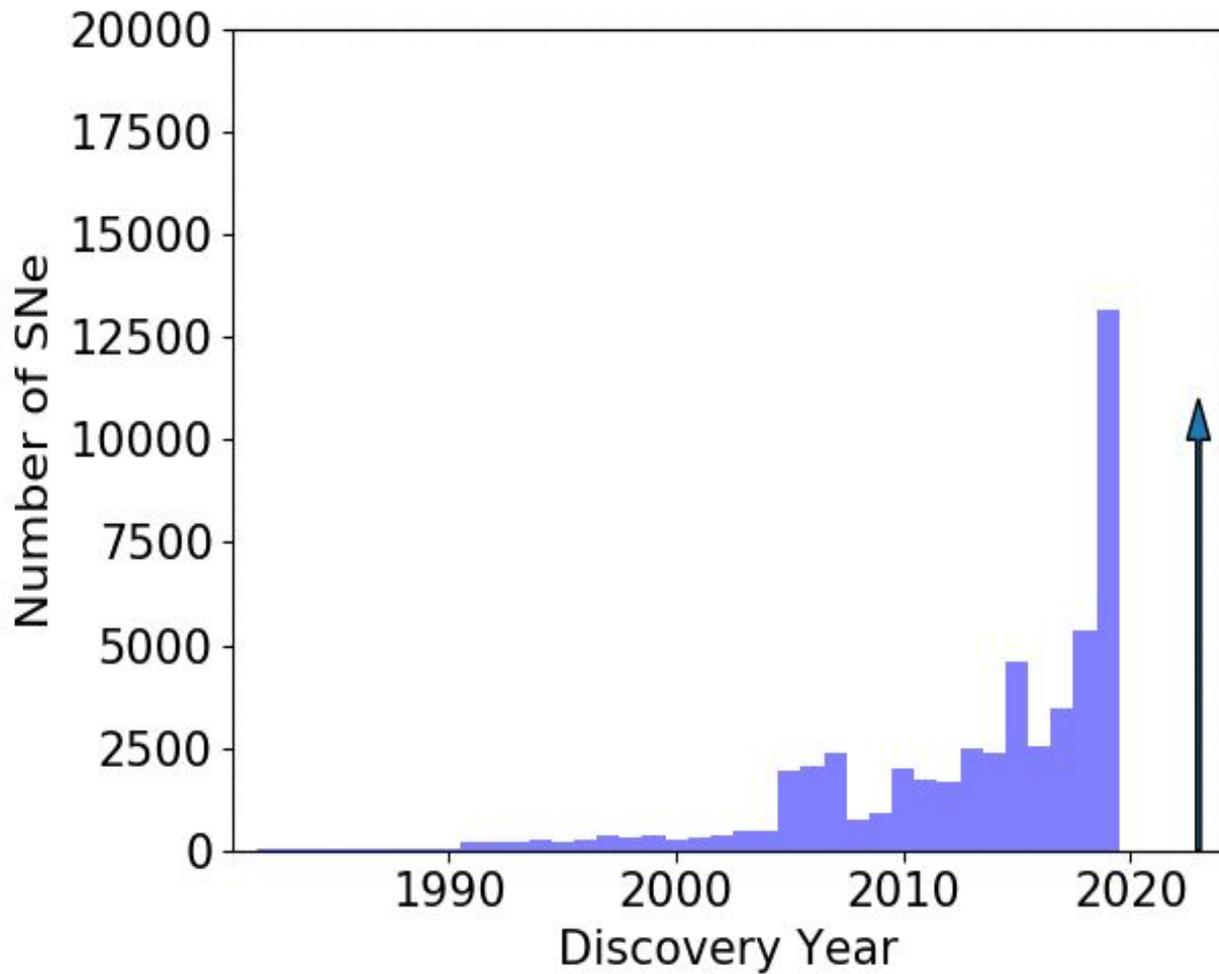
A real light curve



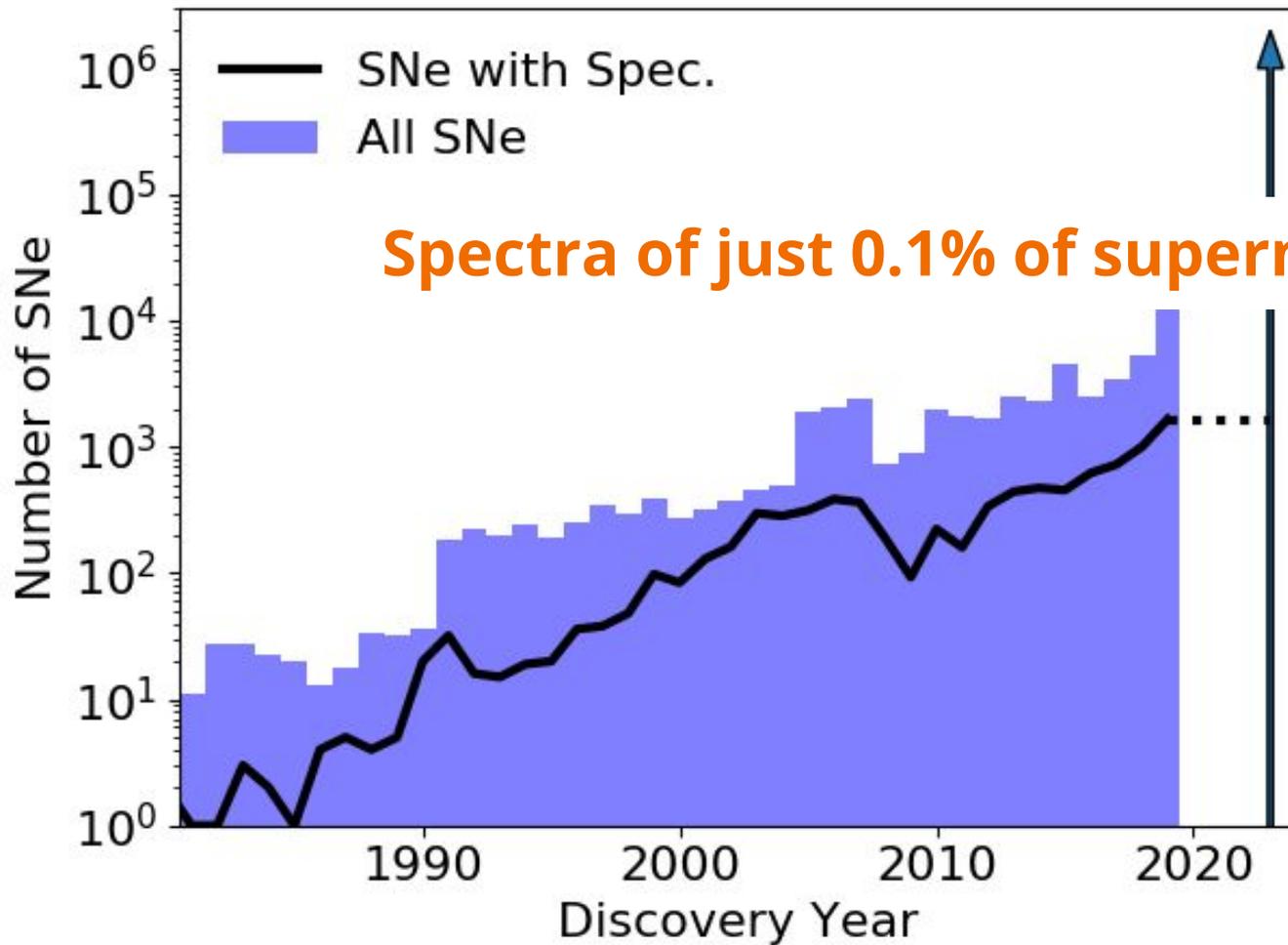
sparse,

noisy,

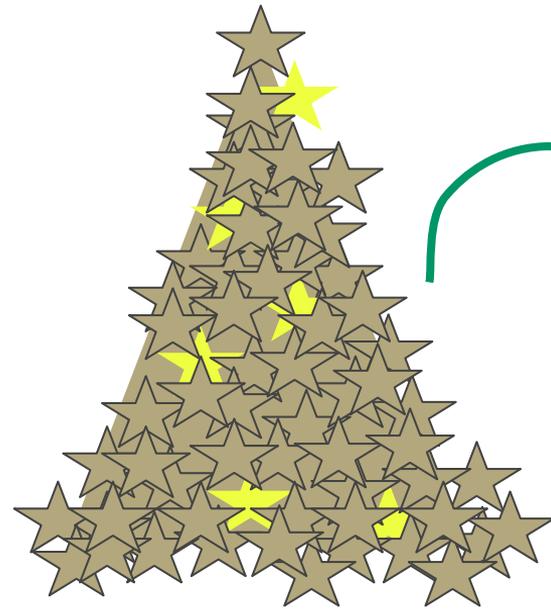
irregularly-sampled



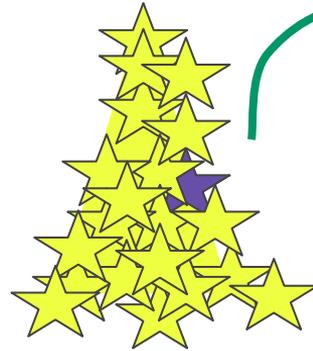
**Rubin
Observatory**



The Rubin Needles & the Haystack



~Million supernovae /
year

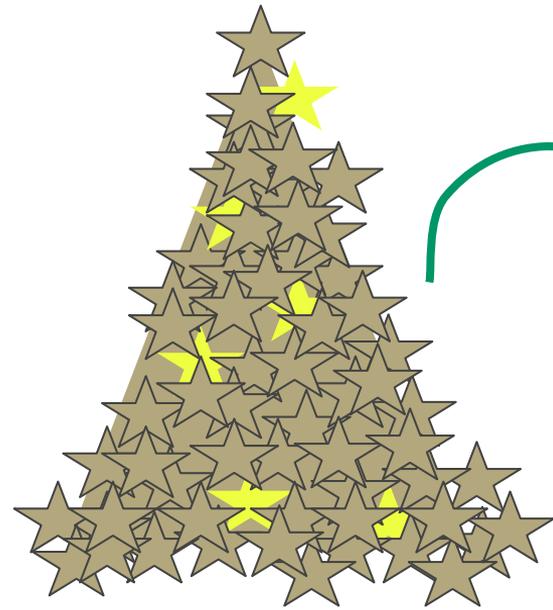


~1000s / year
with traditional
(spectroscopic)
classification

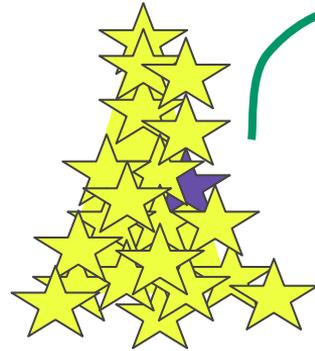


~100 supernovae we actively
follow with other resources

The Rubin Needles & the Haystack



~Million supernovae /
year



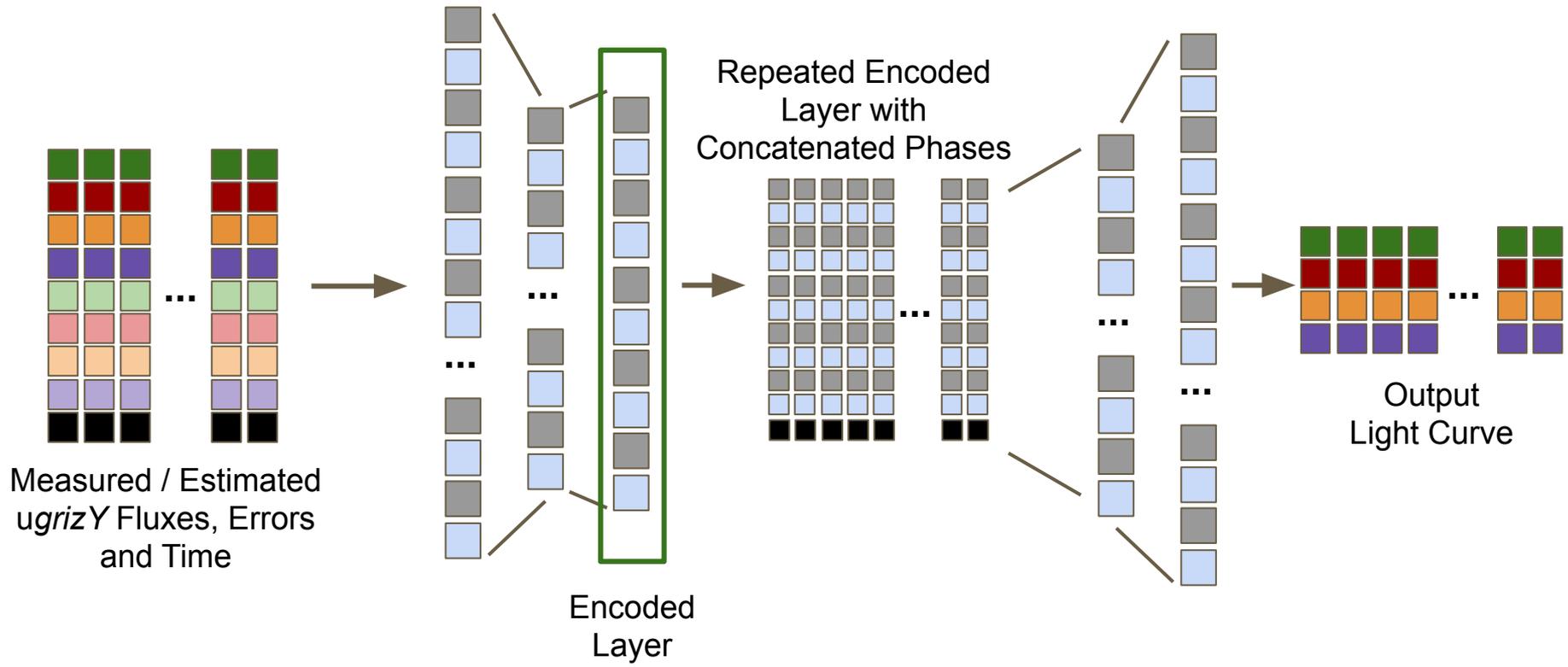
~1000s / year
with traditional
(spectroscopic)
classification



~100 supernovae we actively
follow with other resources

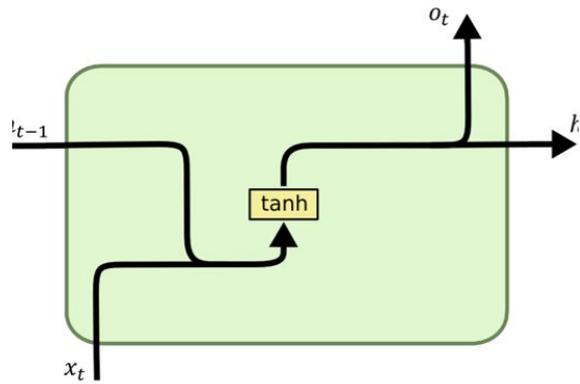
**The rarest needles: the
unknown unknowns**

A data-driven, **unsupervised**
anomaly detection method
using a **variational recurrent**
autoencoder

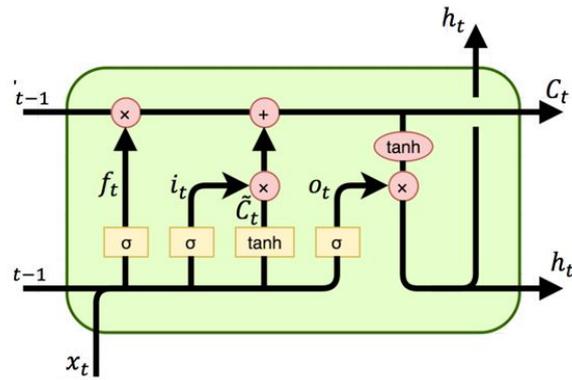


Recurrent neurons

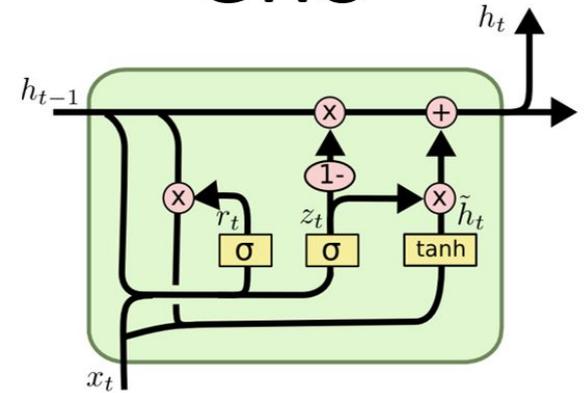
RNN



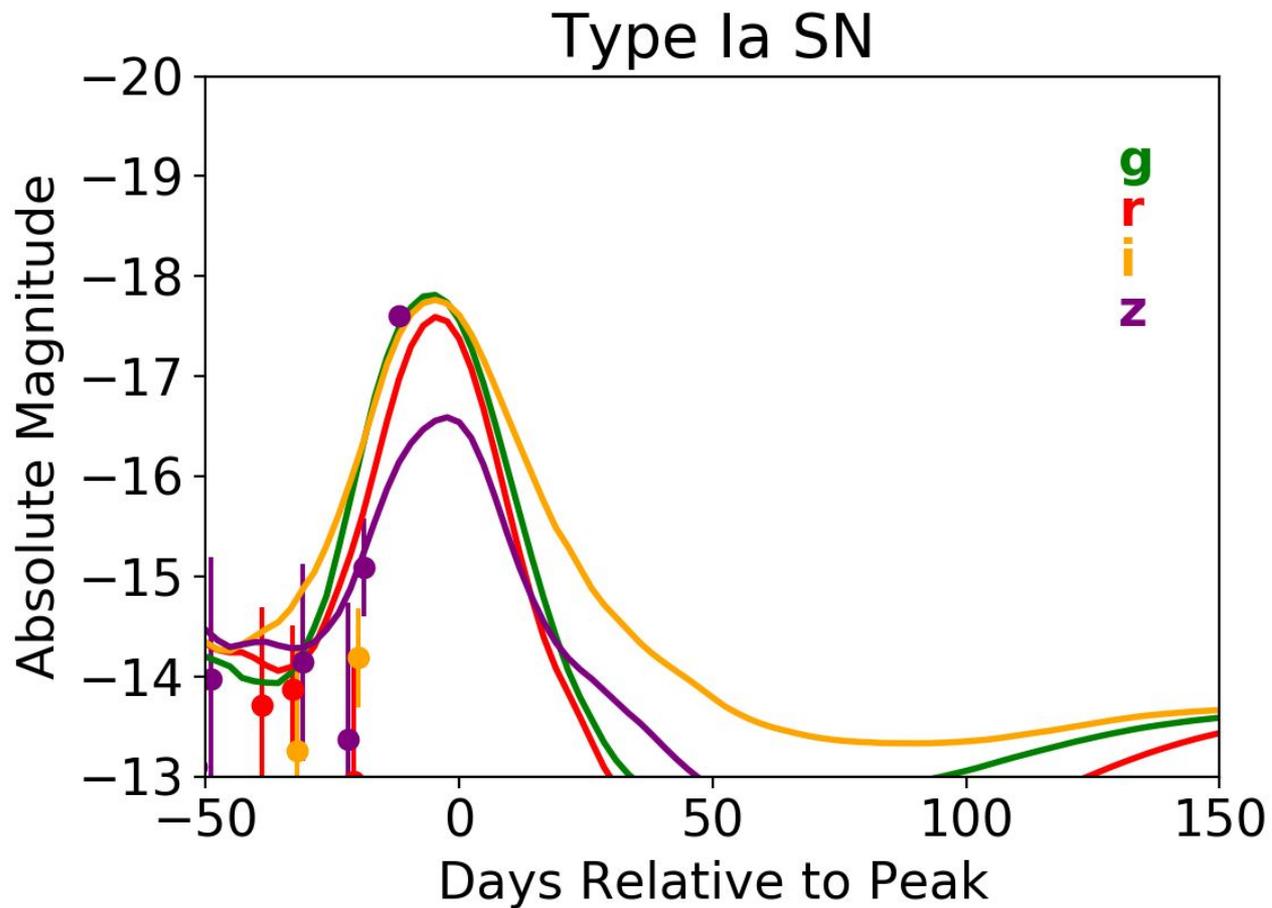
LSTM



GRU

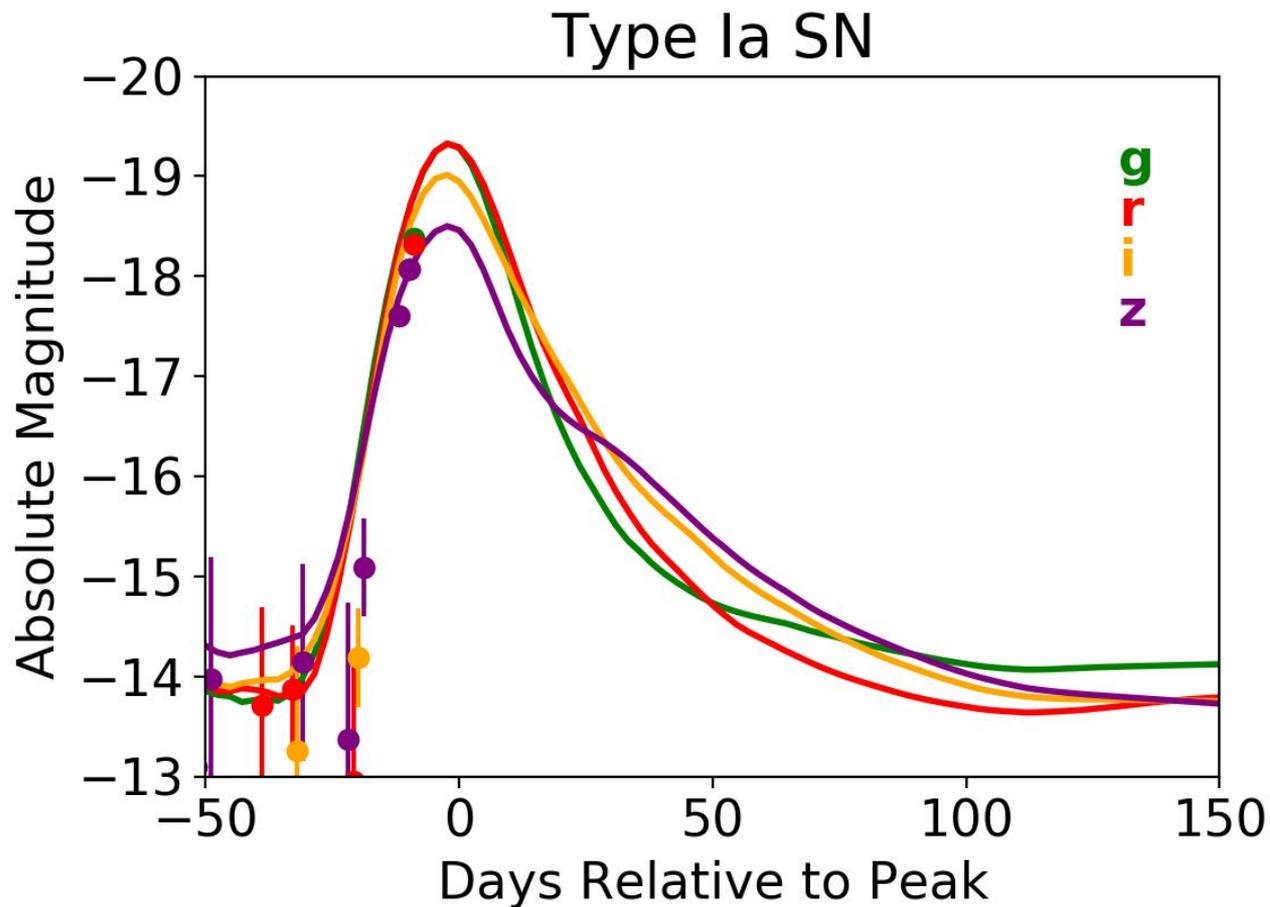


Decoded light curve updated with new data



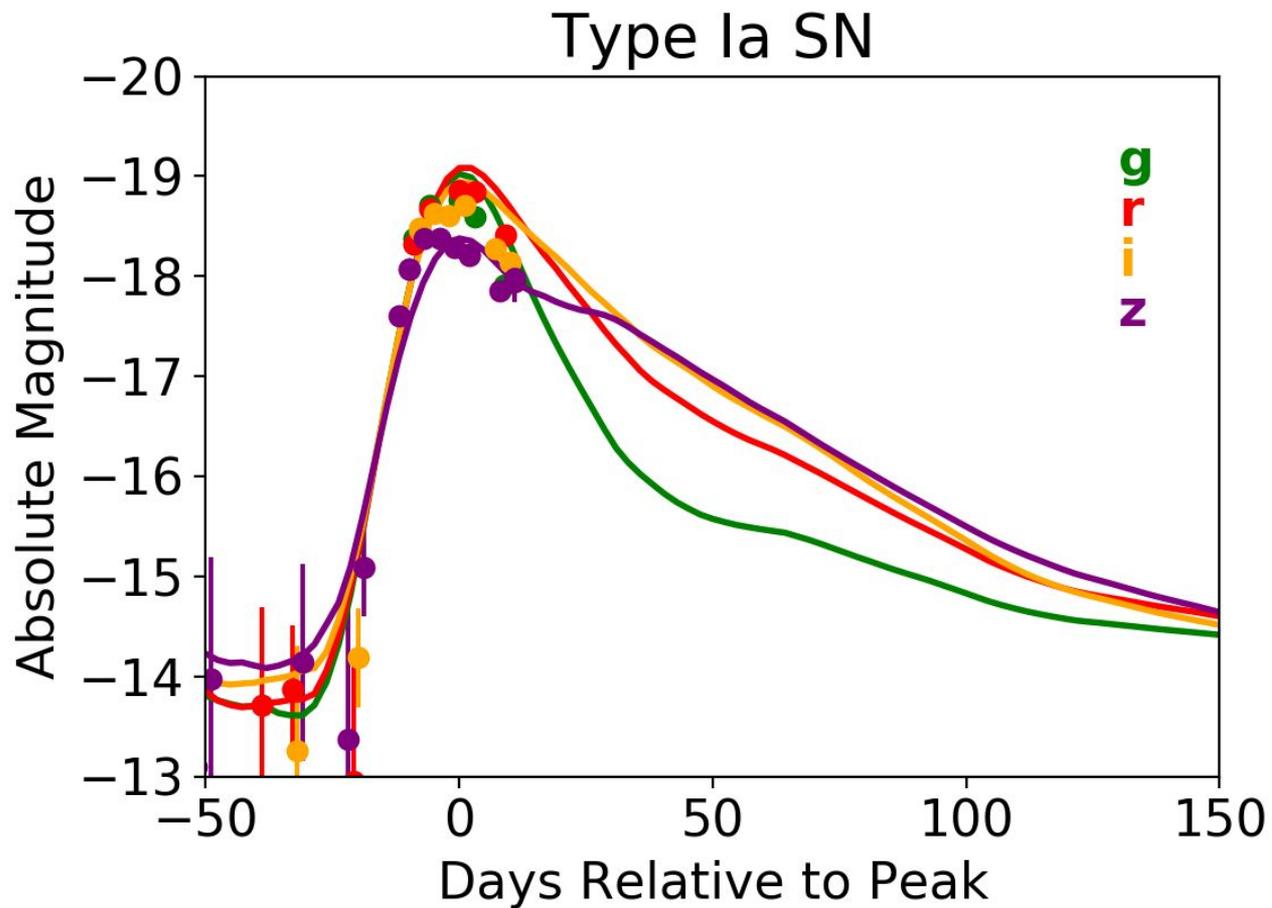
VAE estimate is a little odd,
thinks it is short and dim.

Decoded light curve updated with new data



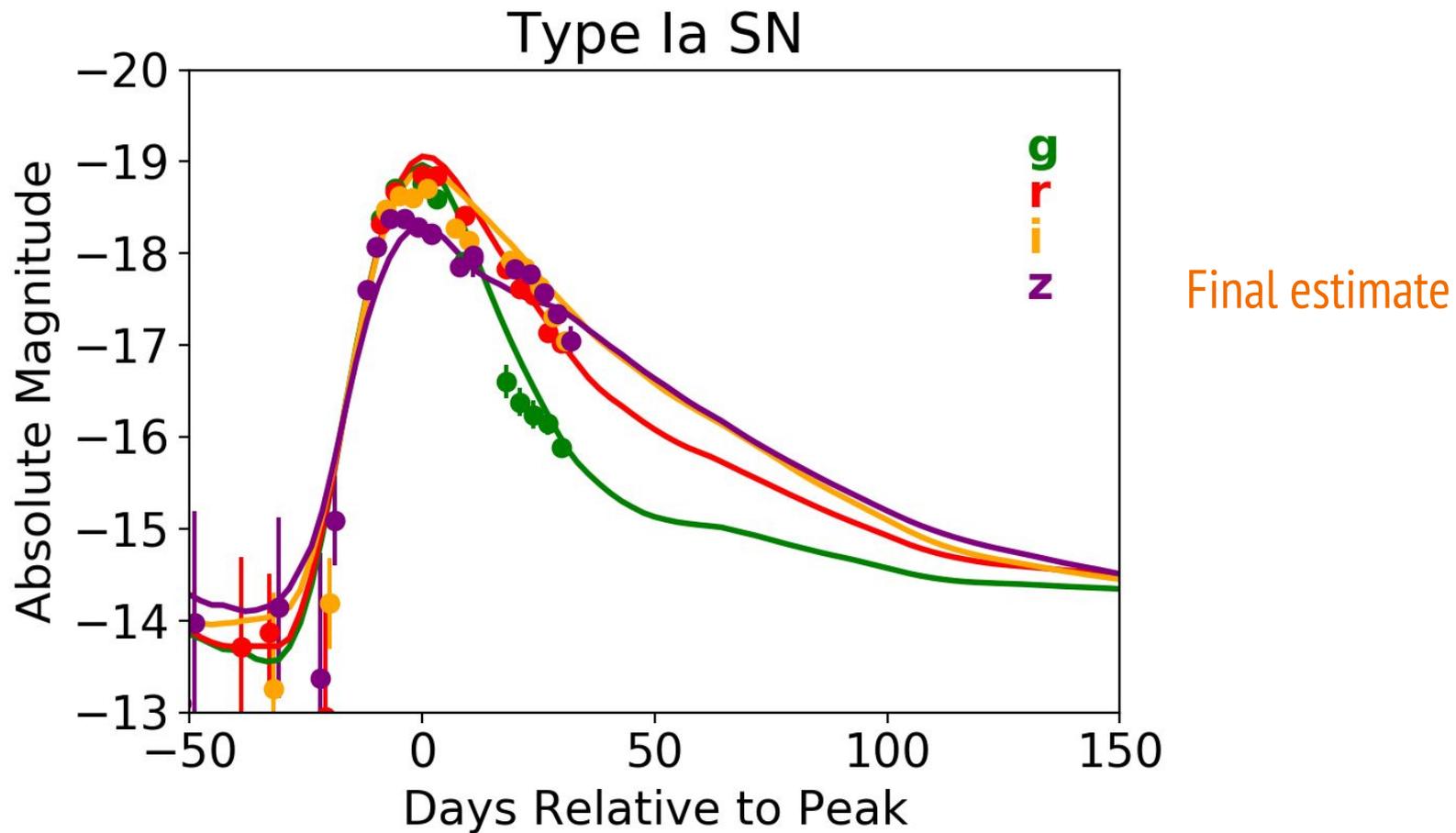
VAE estimate hits the “correct” peak flux for this type of supernova

Decoded light curve updated with new data

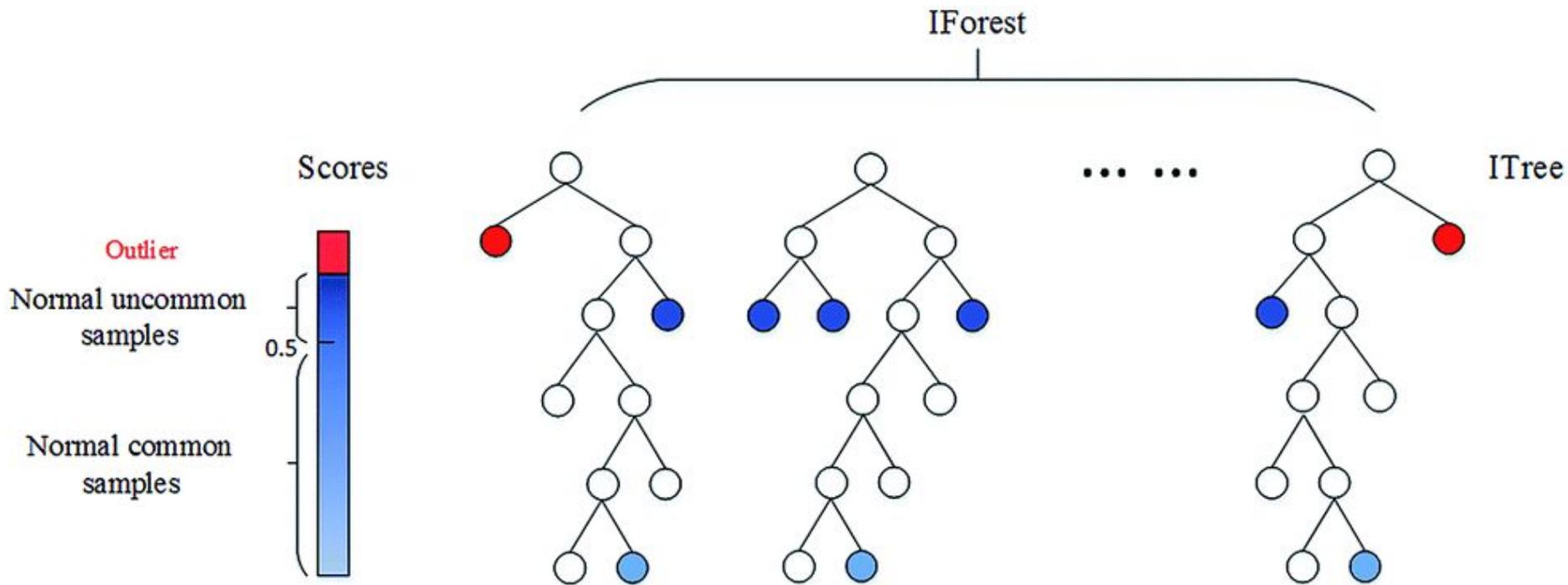


VAE estimate correctly predicts the 'bump' in z-band (again a distinct feature for this supernova type)

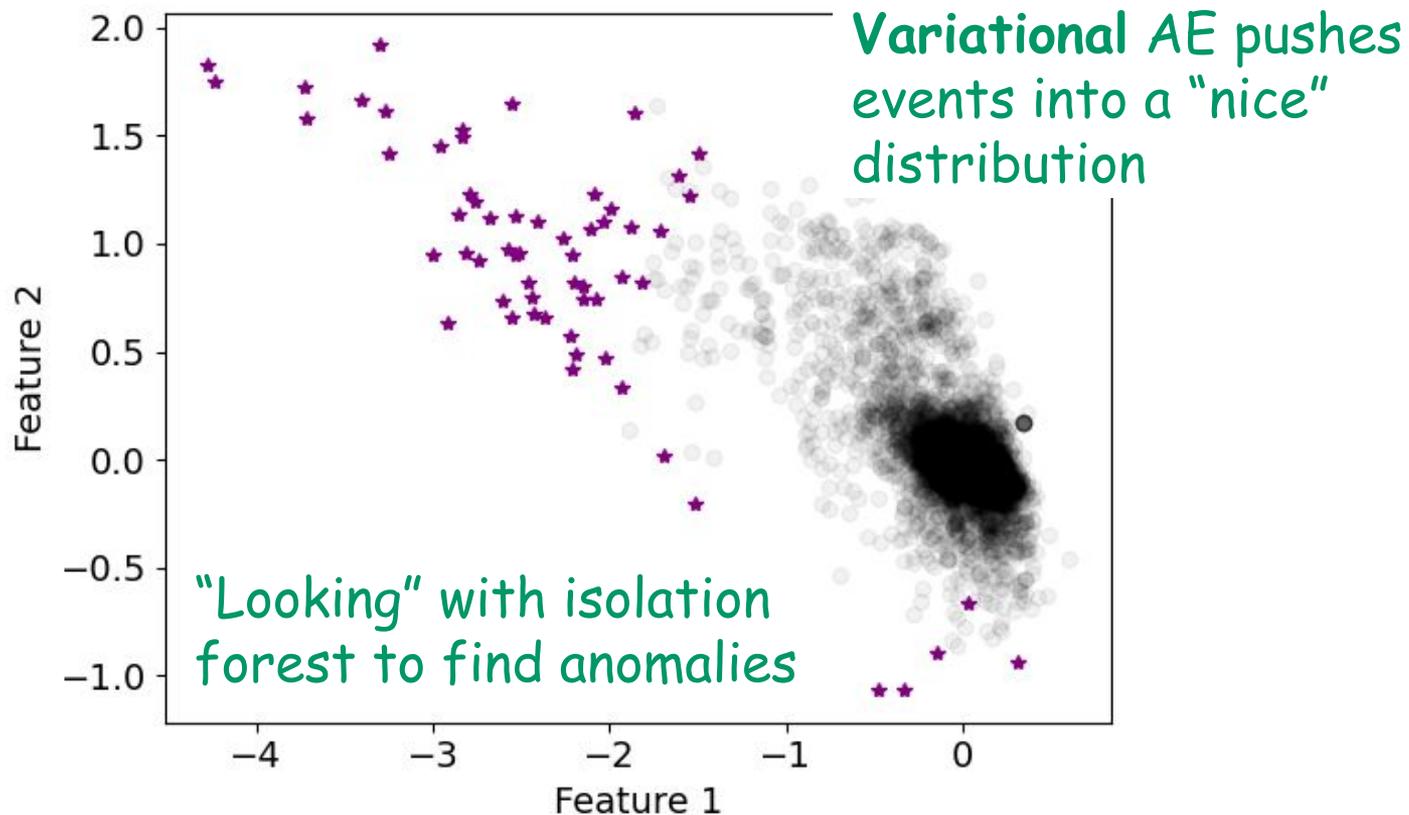
Decoded light curve updated with new data



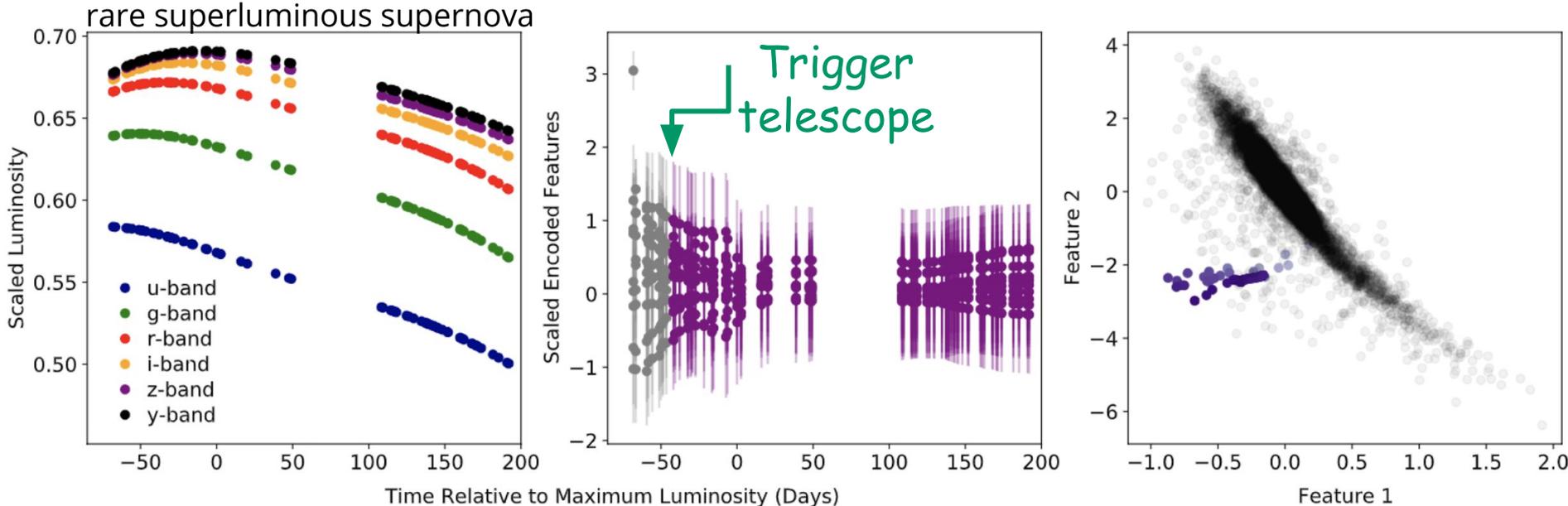
Step 3: Isolation forest to look for outliers



Look at the encoded space for “needles”



Look at encoded space as the event evolves!



In progress: what is a useful metric for anomaly detection vs. time?

RVAE for anomaly detection in SN time series

- Our problem: Need to identify unknown 'needles' (anomalies) in a 'haystack' of transient events in real time
- Solution:
 - Step 1: Pre-process (interpolate) time series with a GP
 - Step 2: Train RVAE on the data to learn an encoded space
 - Step 3: Search for anomalies in the encoding space using an isolation forest

Thank you!