A Template Problem in Spatio-temporal Modeling and Data Mining

Adway Mitra

4 January 2021

Notations

- Consider S locations in a region
- A geo-physical variable, say X may be measured at every location
- Readings are taken at regular time intervals, say hourly/daily
- Denote the readings by X^s_{dh} (s: location, d: day, h: hour)
- Or maybe, X^s_{ymd} (y: year, m: month, d: day)
- Observations are available at only a subset of the locations!!

Template Setting

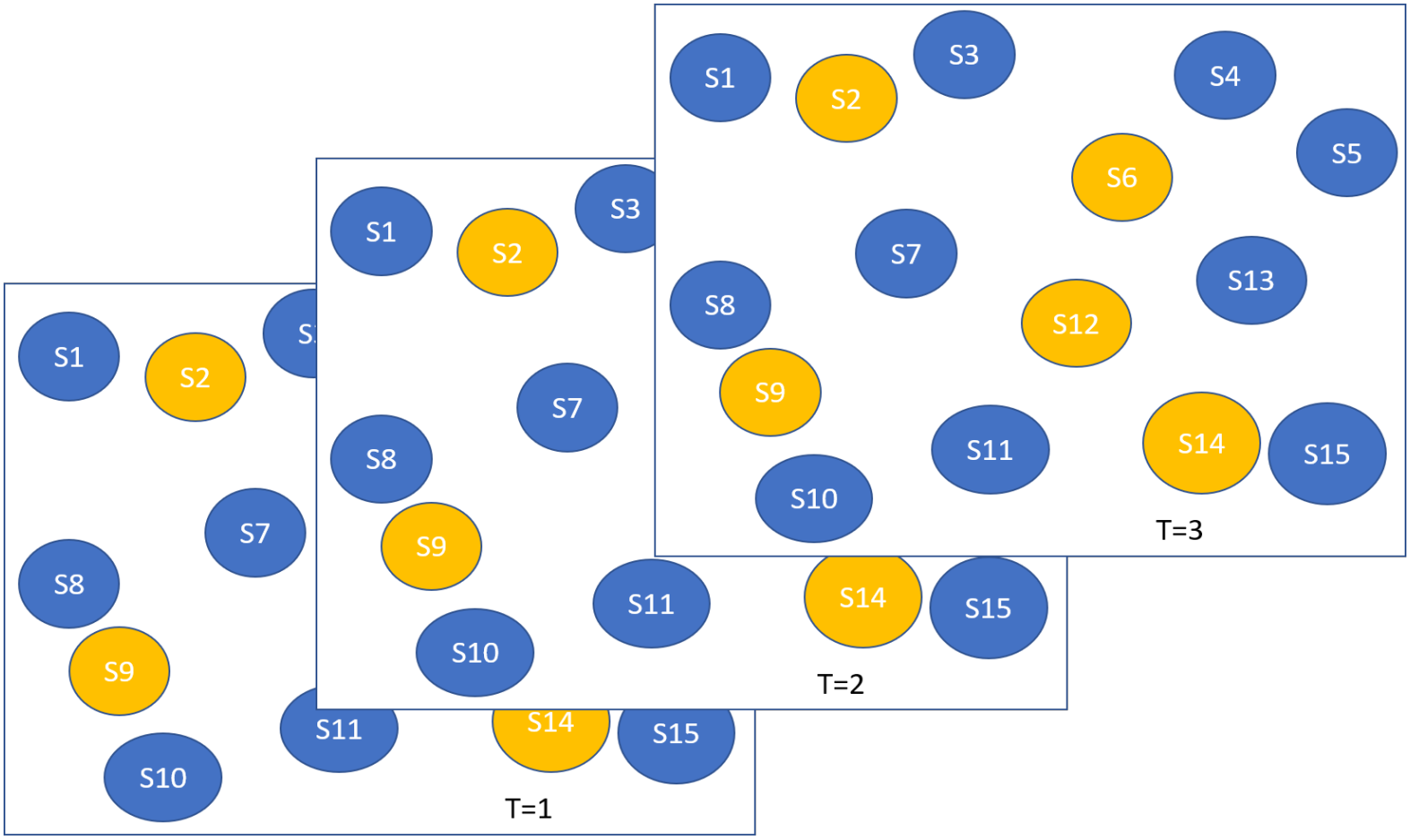

Figure: 15 locations: observations available in blue locations, not in orange locations

Template Problems

- Estimate the values of X at the locations which have no observations
- Predict future values at all locations
- Identify spatial relationships between locations
- Identify trends and periodic/seasonal behavior
- Identify "anomalies" or unusual events

Probabilistic Modeling

- Consider $\{X\}$ as random variables, whose values are sometimes known
- Each of them can be considered as separate R.V. (not useful)
- Each of them can be considered as a realization of the same R.V. (may not make sense physically)
- We can divide them into groups - all values in same group are realizations of one R.V.?
- How to define such groups?

Probabilistic Modeling

- Let us consider X_h^s as a R.V. (value at location s at hour h) *(Total* $24S$ *variables)*
- Its realizations are available for each day: $\{x_{dh}^s\}$
- We utilize the property of periodicity (values at same location, same hour likely to be similar on different days)
- Similarly, we can define X_m^s as a R.V. (value at location s for month m) *(Total* $12S$ *variables)*
- Its realizations are available for each year and each day: $\{x_{ymd}^s\}$

Probabilistic Modeling

- What sort of R.V. is X_h^s or X_m^s?
- Continuous or discrete?
  - *Decide based on the nature of the data*
- Follows which distribution?
  - *Decide based on the histogram of the data*
- Parameters of the distribution?
  - *Parameter Estimation techniques!*