
1 Pseudo-code for hierarchical selection of data sources for influenza forecasting

S1 Algorithm. Hierarchical data source selection.

- 1: **Given** : Set of predictor data sources \mathbf{D} and forecasting target G
 - 2: $\mathbf{D}_{chosen} : \{D_1, D_2, D_3, \dots, D_{n-1}\}$
 - 3: $\mathbf{D}_{remaining} : \mathbf{D} \setminus \mathbf{D}_{chosen}$
 - 4: To choose D_n ,
 - 5: **for** $q \in \mathbf{D}_{remaining}$ **do**
 - 6: Fit $\Omega_q : G \sim D_1 + D_2 + \dots + D_{n-1} + q$ on the full historical time series.
 - 7: \mathbf{S} : historical seasons of G
 - 8: **for** $s_i \in \mathbf{S}$ **do**
 - 9: **for** $d \in \mathbf{D}_{chosen} \cup q$ **do**
 - 10: Make Bayes' forecast of s_i for d from week w , with $(\mathbf{S} \setminus s_i)$ as priors.
 - 11: Apply Ω_q to forecast G for season s_i from forecasts generated in Step 10.
 - 12: Score forecast from Step 11 by comparison to actual G (call this score $\sigma(q, i)$).
 - 13: Calculate the grand score $\sum_{s_i} \sigma(q, i)$.
 - 14: Choose q that maximizes the grand score.
 - 15: Set $D_n = q$.
-