Hidden Markov Chain in Health and Well-being

explored using a sleep quality dataset

Natal and Maxence

Goal

- Use a Hidden Markov chain model to analyze health data and make predictions and to intervene to optimize the future outcome of some health parameters by manipulating others

What we did!

find some useful data in the area of health

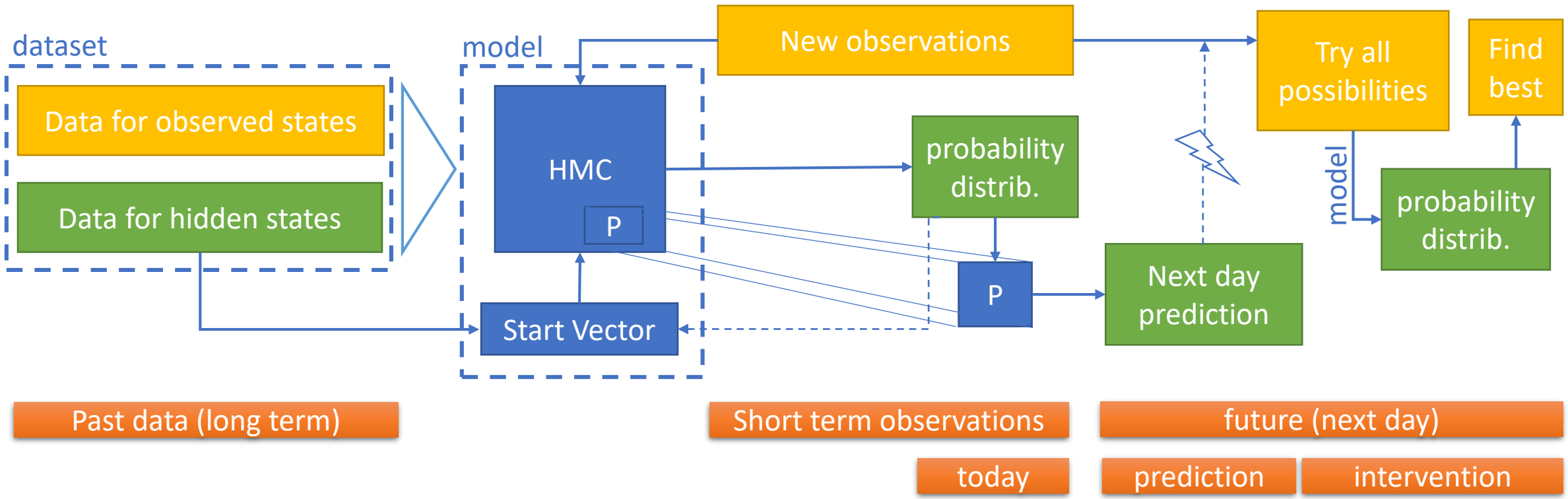
Difficult, as the datasets are often summarize and averaged and not sequential (anymore)

dataset in the area of sleep quality

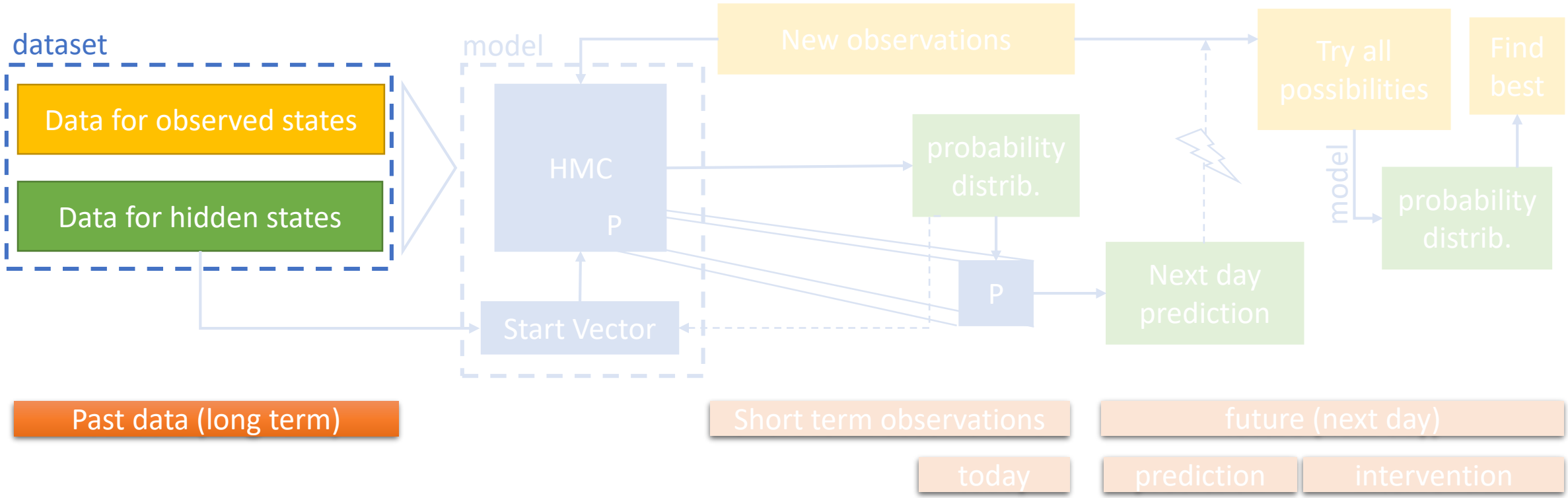
think about how to generate a HMC-model and what to do with it

generalization and further applications

Our pipeline

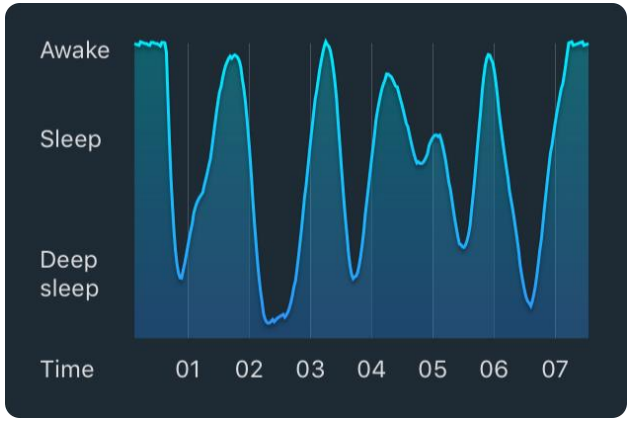


Our dataset

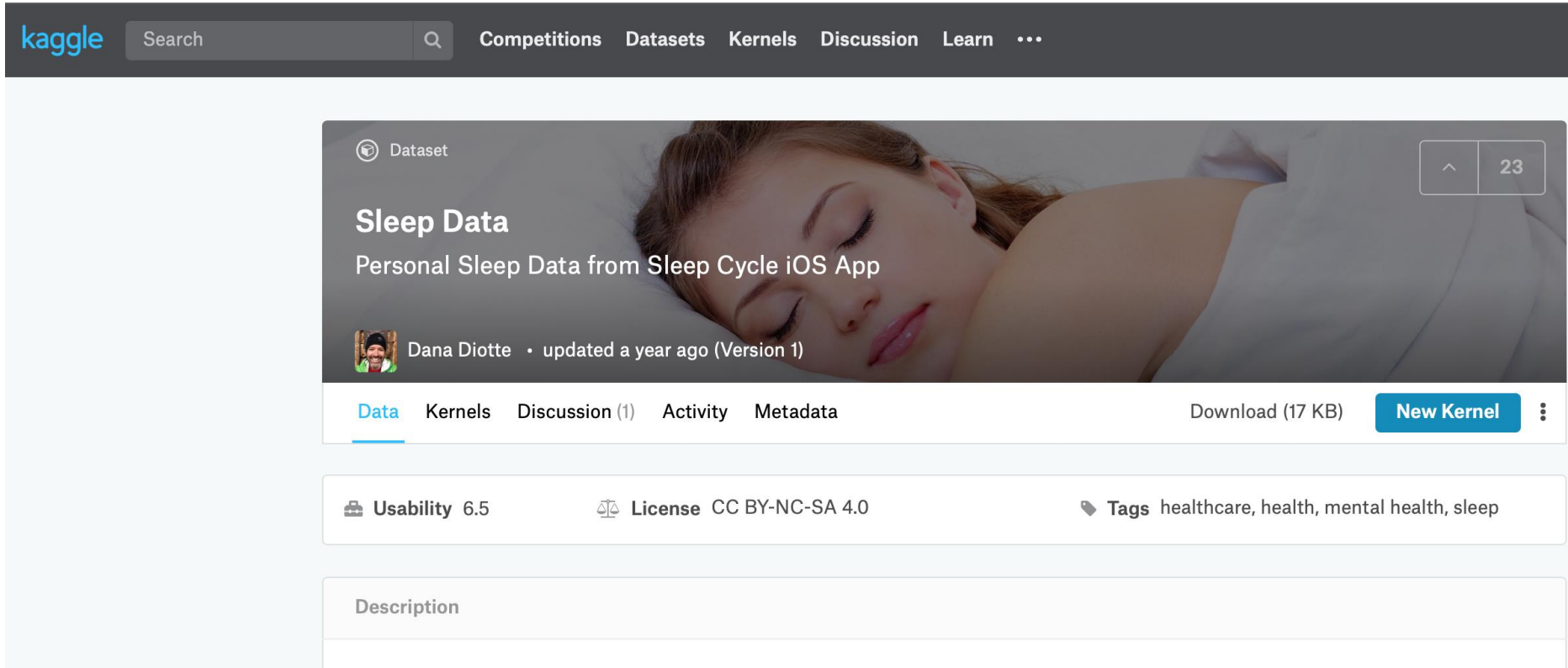

dataset
Data for observed states
Data for hidden states
model
HMC
P
Start Vector
New observations
probability distrib.
P
Next day prediction
Try all possibilities
Find best
model
probability distrib.
Past data (long term)
Short term observations
future (next day)
today
prediction
intervention

Our dataset

Data from Sleep Cycle



App estimates your quality of sleep



Our dataset

sleepdata.csv (66.11 KB) — 8 of 8 columns — Views

	Start	End	Sleep quality	Time in bed	Wake up	Sleep Notes	Heart rate	Activity (steps)
	29Dec14 – 16Feb18	29Dec14 – 16Feb18	81% 5% 83% 4% Other (77) 91%	22Mar18 – 23Mar18	:) 24% :| 3% Other (1) 72%	Drank coffee:Dra... 18% Drank coffee:Dra... 14% Other (17) 68%	58 2% 57 2% Other (28) 96%	0 – 21.9k
1	2014-12-29 22:57:49	2014-12-30 07:30:13	100%	8:32	:)		59	0
2	2014-12-30 21:17:50	2014-12-30 21:33:54	3%	0:16	:|	Stressful day	72	0
3	2014-12-30 22:42:49	2014-12-31 07:13:31	98%	8:30	:|		57	0
4	2014-12-31 22:31:01	2015-01-01 06:03:01	65%	7:32				0
5	2015-01-01 22:12:10	2015-01-02 04:56:35	72%	6:44	:)	Drank coffee:Drank tea	68	0
6	2015-01-03 00:34:57	2015-01-03 07:47:23	83%	7:12	:)	Drank coffee:Drank tea	60	0
7	2015-01-04 00:23:06	2015-01-04 07:37:09	78%	7:14		Drank tea		0
8	2015-01-04 21:34:44	2015-01-05 04:53:34	78%	7:18	:)	Ate late:Drank coffee	57	0
9	2015-01-05 21:32:25	2015-01-06 05:00:03	69%	7:27	:)	Drank coffee:Drank tea:Worked out	56	0
10	2015-01-06 21:24:56	2015-01-07 05:00:02	74%	7:35	:|	Drank tea:Worked out	64	0

Our dataset

sleepdata.csv (66.11 KB) — 8 of 8 columns — Views

	Start	End	Sleep quality	Time in bed	Wake up	Sleep Notes	Heart rate	Activity (steps)
	29Dec14 – 16Feb18	29Dec14 – 16Feb18	81% 5% 83% 4% Other (77) 91%	22Mar18 – 23Mar18	:) 24% :| 3% Other (1) 72%	Drank coffee:Dra... 18% Drank coffee:Dra... 14% Other (17) 68%	58 2% 57 2% Other (28) 96%	0 – 21.9k
1	2014-12-29 22:57:49	2014-12-30 07:30:13	100%	8:32	:)		59	0
3	2014-12-30 22:42:49	2014-12-31 07:13:31	98%	8:30	:|		57	0
4	2014-12-31 22:31:01	2015-01-01 06:03:01	65%	7:32				0
5	2015-01-01 22:12:10	2015-01-02 04:56:35	72%	6:44	:)	Drank coffee:Drank tea	68	0
6	2015-01-03 00:34:57	2015-01-03 07:47:23	83%	7:12	:)	Drank coffee:Drank tea	60	0
7	2015-01-04 00:23:06	2015-01-04 07:37:09	78%	7:14		Drank tea		0
8	2015-01-04 21:34:44	2015-01-05 04:53:34	78%	7:18	:)	Ate late:Drank coffee	57	0
9	2015-01-05 21:32:25	2015-01-06 05:00:03	69%	7:27	:)	Drank coffee:Drank tea:Worked out	56	0
10	2015-01-06 21:24:56	2015-01-07 05:00:02	74%	7:35	:|	Drank tea:Worked out	64	0

remove naps

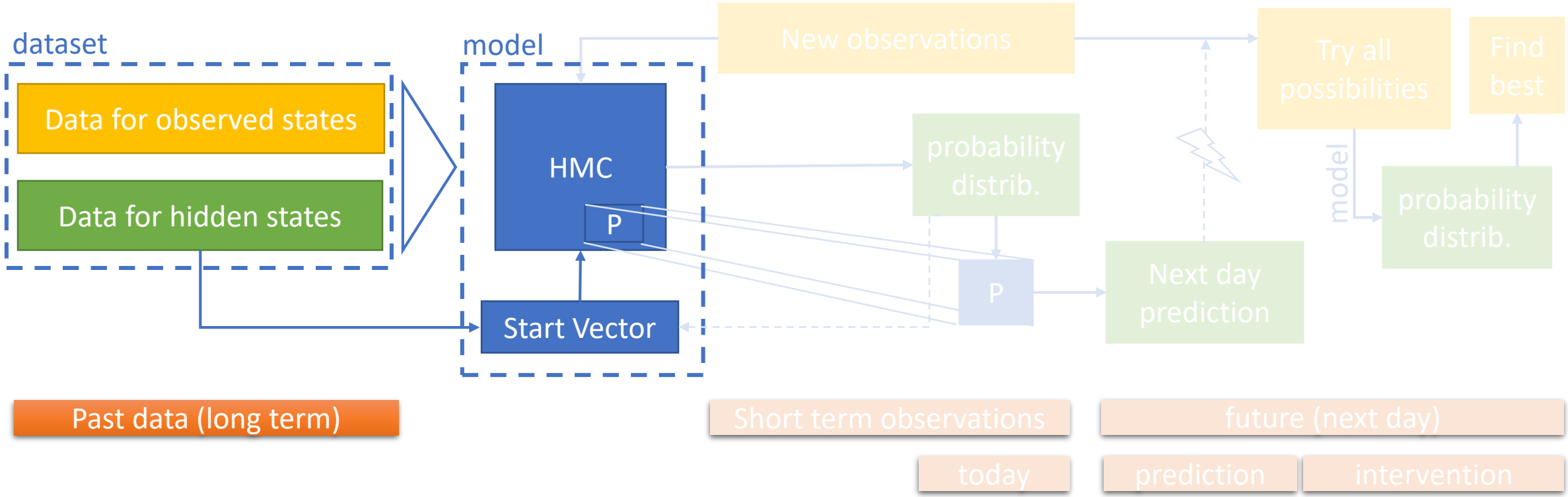
Our dataset

- Sleep data over 4 years

- Minimum quality of sleep 30% and maximum 100%
- Average quality of sleep 76.5%

- Minimum length of sleep 5:01h and maximum 10:46h
- Average length of sleep 7:48h

Creation of our model



Creation of our model

sleepdata.csv (66.11 KB)

8 of 8 columns · Views

	Start	End	Sleep quality	Time in bed	Wake up	Sleep Notes	Heart rate	Activity (steps)
	29Dec14 – 16Feb18	29Dec14 – 16Feb18	81% 5% 83% 4% Other (77) 91%	22Mar18 – 23Mar18	:) 24% :| 3% Other (1) 72%	Drank coffee:Dra... 18% Drank coffee:Dra... 14% Other (17) 68%	58 2% 57 2% Other (28) 96%	0 – 21.9k
1	2014-12-29 22:57:49	2014-12-30 07:30:13	100%	8:32	:)		59	0
3	2014-12-30 22:42:49	2014-12-31 07:13:31	98%	8:30	:|		57	0
4	2014-12-31 22:31:01	2015-01-01 06:03:01	65%	7:32				0
5	2015-01-01 22:12:10	2015-01-02 04:56:35	72%	6:44	:)	Drank coffee:Drank tea	68	0
6	2015-01-03 00:34:57	2015-01-03 07:47:23	83%	7:12	:)	Drank coffee:Drank tea	60	0
7	2015-01-04 00:23:06	2015-01-04 07:37:09	78%	7:14		Drank tea		0
8	2015-01-04 21:34:44	2015-01-05 04:53:34	78%	7:18	:)	Ate late:Drank coffee	57	0
9	2015-01-05 21:32:25	2015-01-06 05:00:03	69%	7:27	:)	Drank coffee:Drank tea:Worked out	56	0
10	2015-01-06 21:24:56	2015-01-07 05:00:02	74%	7:35	:|	Drank tea:Worked out	64	0

Creation of our model

sleepdata.csv (66.11 KB)

8 of 8 columns · Views

	Start	End	Sleep quality	Time in bed	Wake up	Sleep Notes	Heart rate	Activity (steps)
	29Dec14 – 16Feb18	29Dec14 – 16Feb18	81% 5% 83% 4% Other (77) 91%	22Mar18 – 23Mar18	:) 24% :| 3% Other (1) 72%	Drank coffee:Dra... 18% Drank coffee:Dra... 14% Other (17) 68%	58 2% 57 2% Other (28) 96%	0 – 21.9k
1	2014-12-29 22:57:49	2014-12-30 07:30:13	100%	8:32	:)		59	0
3	2014-12-30 22:42:49	2014-12-31 07:13:31	98%	8:30	:|		57	0
4	2014-12-31 22:31:01	2015-01-01 06:03:01	65%	7:32				0
5	2015-01-01 22:12:10	2015-01-02 04:56:35	72%	6:44	:)	Drank coffee:Drank tea	68	0
6	2015-01-03 00:34:57	2015-01-03 07:47:23	83%	7:12	:)	Drank coffee:Drank tea	60	0
7	2015-01-04 00:23:06	2015-01-04 07:37:09	78%	7:14		Drank tea		0
8	2015-01-04 21:34:44	2015-01-05 04:53:34	78%	7:18	:)	Ate late:Drank coffee	57	0
9	2015-01-05 21:32:25	2015-01-06 05:00:03	69%	7:27	:)	Drank coffee:Drank tea:Worked out	56	0
10	2015-01-06 21:24:56	2015-01-07 05:00:02	74%	7:35	:|	Drank tea:Worked out	64	0

hidden observations

Creation of our model

Hidden states:

- each state representing a 10% interval of sleep quality (starting at 30%)

```
qualityInterval = ['21-30%', '31-40%', '41-50%', '51-60%', '61-70%', '71-80%', '81-90%', '91-100%']
```

Observations:

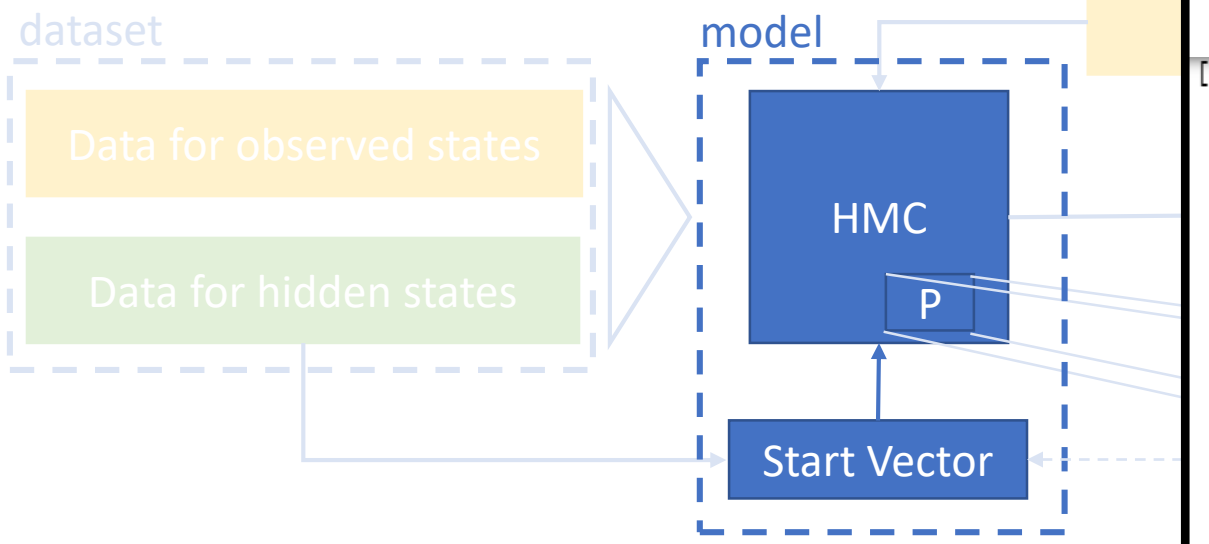
- Each state represents a interval of 30min
- Except first and last state

```
timeInterval = ['1-360', '361-390', '391-420', '421-450', '451-480', '481-510','511-540', '541-570', '571-600', '601-1000']
```

Creation of the model

- We used the Pomegranate package (Python)

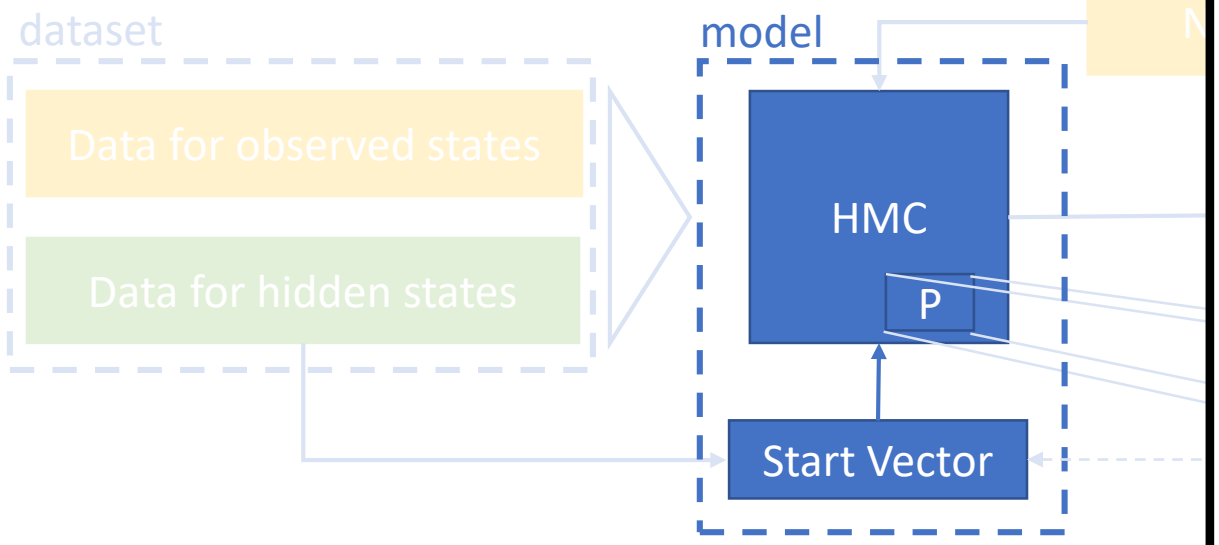
Our model



The hidden matrix P, incl. the starting distribution

```
[[0.         0.         0.         0.         0.         0.
  0.         1.         0.         0.        ]
 [0.11111111 0.11111111 0.11111111 0.33333333 0.         0.11111111
  0.11111111 0.11111111 0.         0.        ]
 [0.         0.09090909 0.18181818 0.09090909 0.27272727 0.09090909
  0.27272727 0.         0.         0.        ]
 [0.         0.05084746 0.03389831 0.23728814 0.20338983 0.28813559
  0.10169492 0.08474576 0.         0.        ]
 [0.         0.00632911 0.01265823 0.08227848 0.28481013 0.34810127
  0.21518987 0.05063291 0.         0.        ]
 [0.         0.00342466 0.00684932 0.05479452 0.17465753 0.32876712
  0.30136986 0.13013699 0.         0.        ]
 [0.         0.00413223 0.00826446 0.04132231 0.12396694 0.38016529
  0.32644628 0.11570248 0.         0.        ]
 [0.         0.01098901 0.         0.02197802 0.18681319 0.32967033
  0.34065934 0.10989011 0.         0.        ]
 [0.01041667 0.01041667 0.06365741 0.15856481 0.33449074 0.29398148
  0.11342593 0.0150463  0.         0.        ]
 [0.         0.         0.         0.         0.         0.
  0.         0.         0.         0.        ]]
```

Our model



The emission probabilities

```
[[0.         0.         1.         0.         0.         0.
  0.         0.         0.         0.        ]
 [0.22222222 0.         0.11111111 0.11111111 0.11111111 0.44444444
  0.         0.         0.         0.        ]
 [0.09090909 0.         0.09090909 0.09090909 0.54545455 0.
  0.09090909 0.09090909 0.         0.        ]
 [0.05084746 0.10169492 0.18644068 0.25423729 0.20338983 0.15254237
  0.01694915 0.01694915 0.         0.01694915]
 [0.05063291 0.05063291 0.10126582 0.25949367 0.30379747 0.17088608
  0.02531646 0.02531646 0.01265823 0.        ]
 [0.         0.01712329 0.10958904 0.24657534 0.3630137  0.14726027
  0.06506849 0.03767123 0.00342466 0.01027397]
 [0.         0.         0.03305785 0.19008264 0.35123967 0.28099174
  0.07438017 0.06198347 0.00413223 0.00413223]
 [0.         0.         0.         0.01098901 0.12087912 0.36263736
  0.2967033  0.13186813 0.05494505 0.02197802]]
```

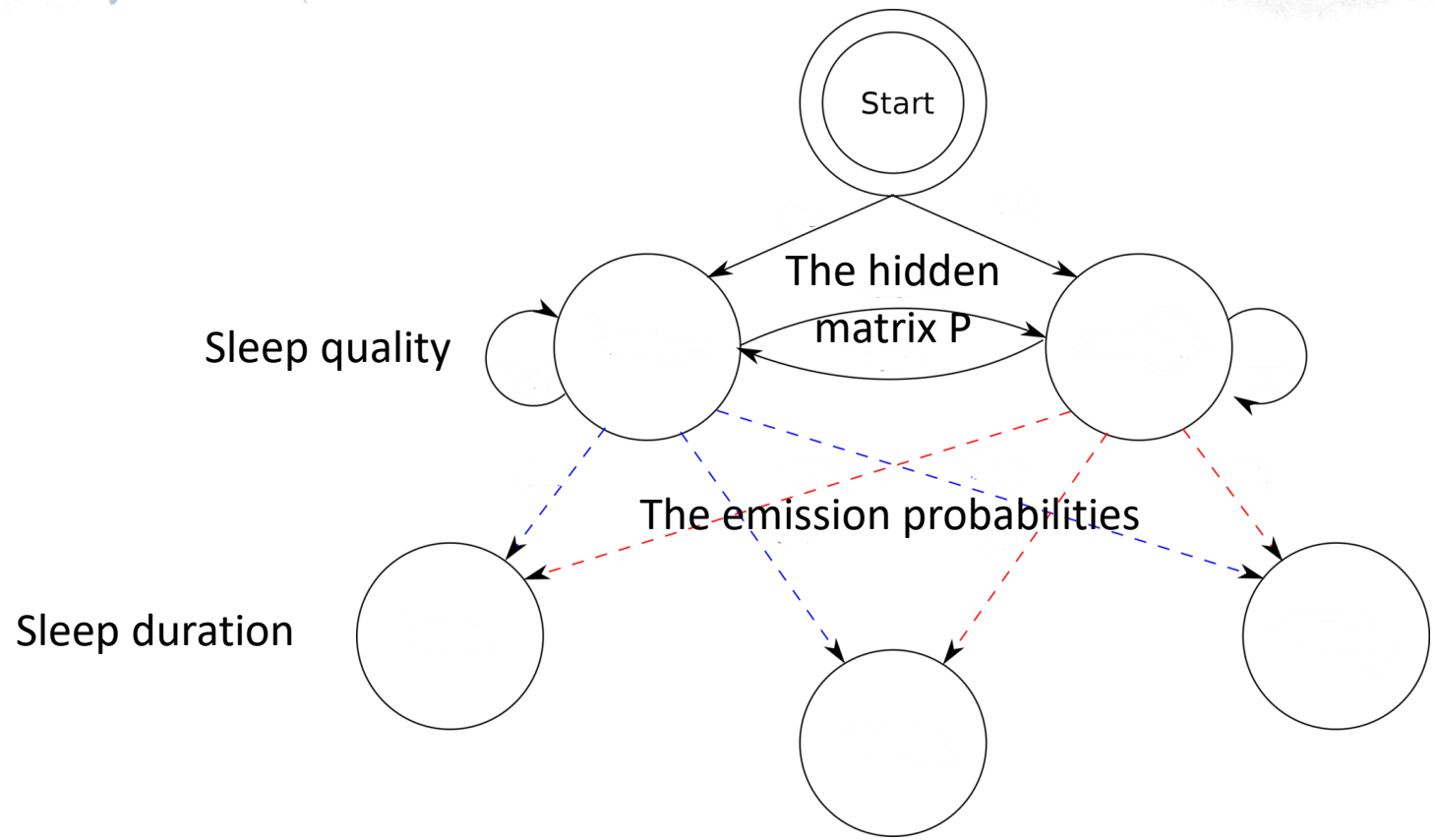
Past data (long term)

today

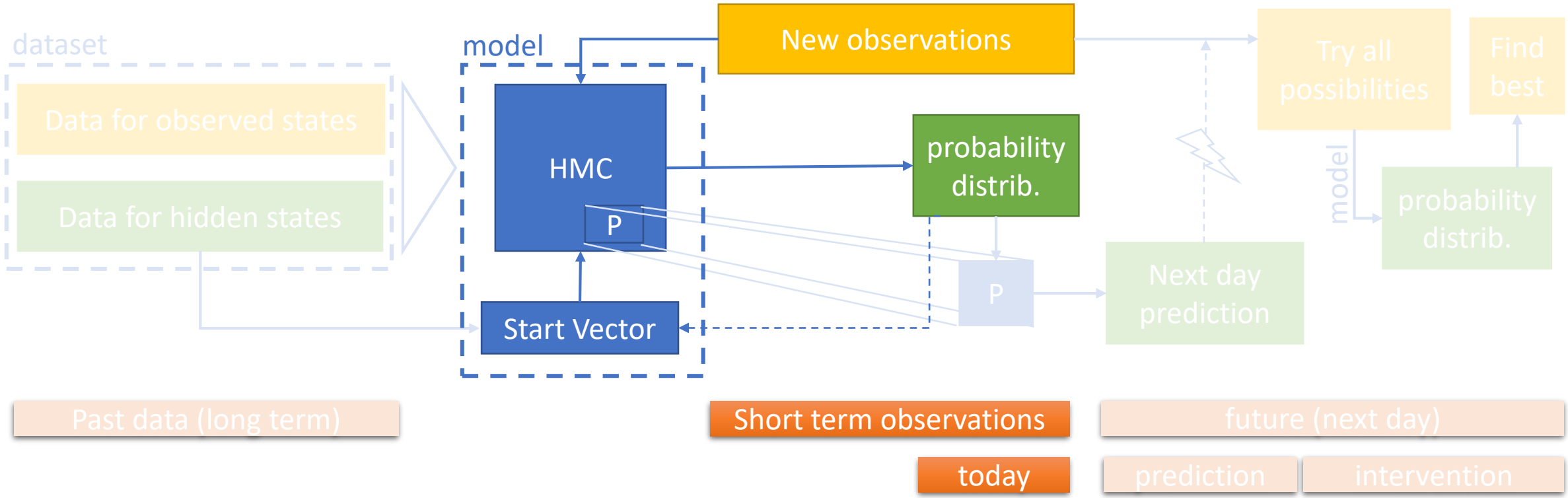
prediction

intervention

Our model


Start
The hidden
matrix P
Sleep quality
The emission probabilities
Sleep duration

Analyze data with the model



Analyze data with the model

Users enters their sleep duration for the last few days:

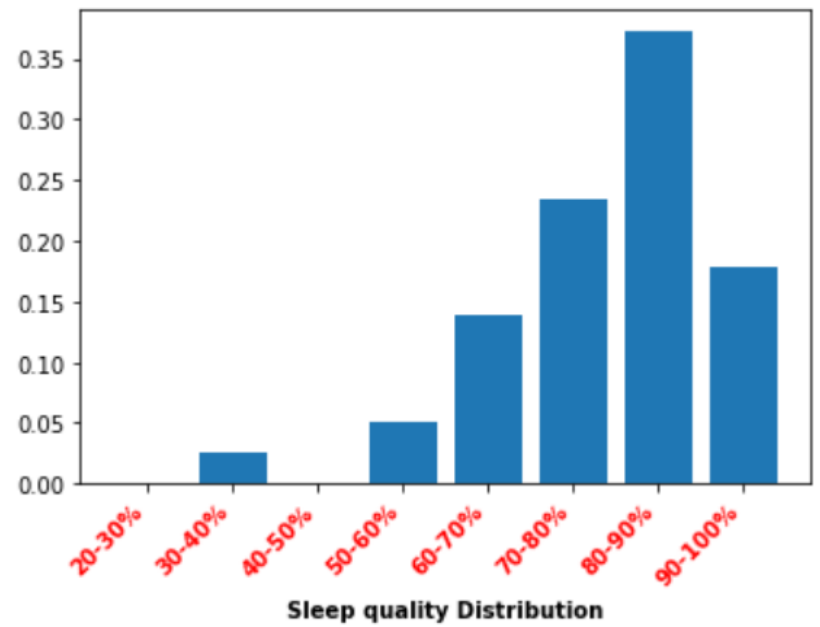
e.g. 6:45, 7:45, 6:40, 8:10, 8:10

```
sequence = ['390-420', '450-480', '390-420', '480-510', '480-510']
```

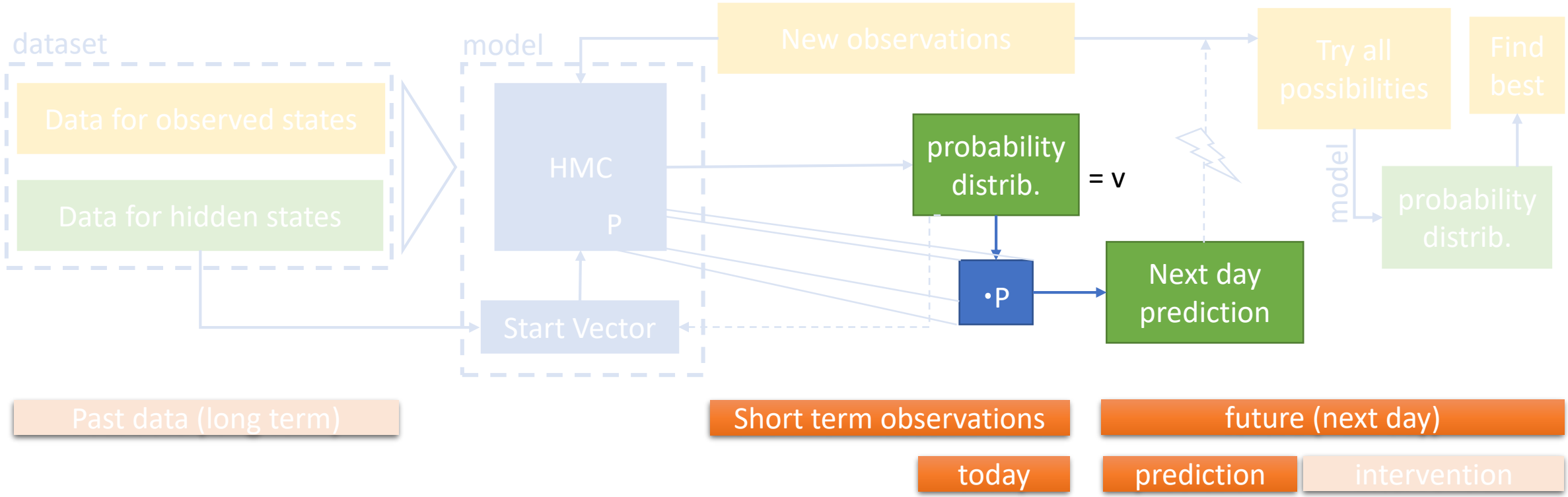
With the help of the forward algorithm we obtain the probability of each observation being aligned to each state by going forward through a sequence of sleep durations.

We calculate from this result the probabilities of sleep quality for the last night!

```
Last night
The probability of having a quality of sleep between 80% and 100% is:
55 %
You can expect to have a sleeping quality of 78 %
```



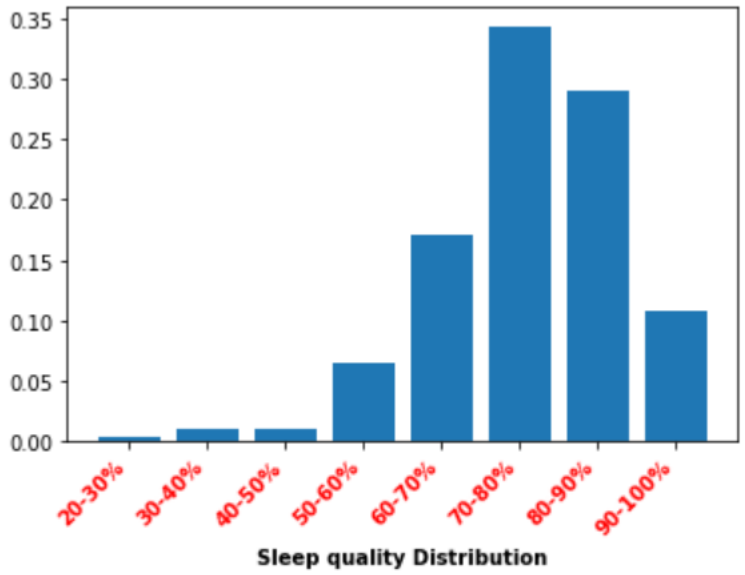
Make a prediction with the model



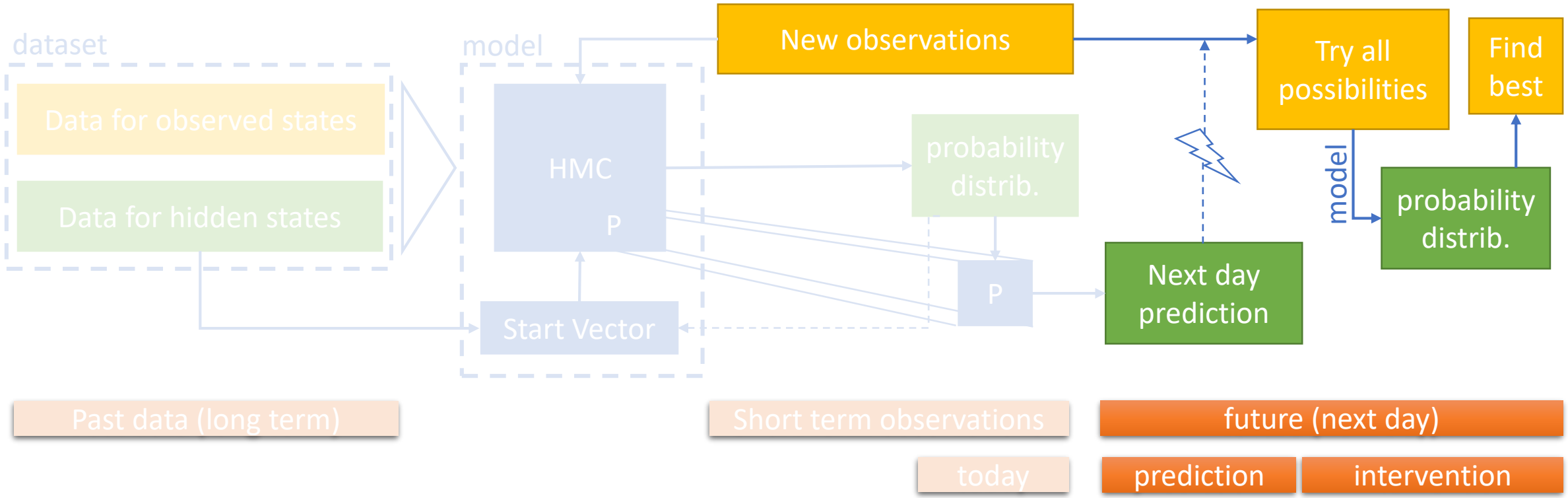
Make a prediction with the model

- In the previous state we have found a probability distribution for the sleep quality of last night, let's denote as vector v
- $prediction = v \cdot P$

```
For this night
The probability of having a quality of sleep between 80% and 100% is:
39 %
You can expect to have a sleeping quality of 76 %
```



Intervention based on the model



Intervention based on the model

- We basically use the forward algorithm again with all possible observable states

```
sequence = ['390-420', '450-480', '390-420', '480-510', '480-510']
```



```
['390-420', '450-480', '390-420', '480-510', '480-510','1-360', ']
```


HMC – forward algorithm

Prediction 1

Intervention based on the model

- We basically use the forward algorithm again with all possible observable states

```
sequence = ['390-420', '450-480', '390-420', '480-510', '480-510']
```



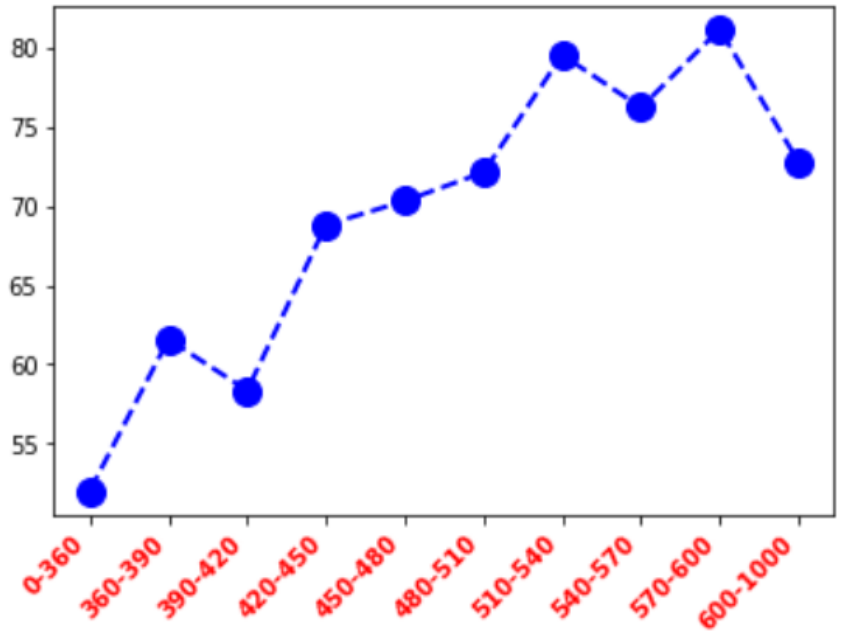
```
['390-420', '450-480', '390-420', '480-510', '480-510', '601-1000']
```


HMC – forward algorithm

Prediction 10
Prediction 9
Prediction 8
Prediction 7
Prediction 6

Intervention based on the model

We have computed all possibles sleeping duration for the next night
and found the ones that will increase your sleeping quality the most



Sleep quality (%) as a function of the duration of your next night (min)

Conclusion

- For most cases we get really similar graphs for our intervention suggestions

  it turns out that this person should aim always for a similar sleep duration, doesn't matter much how long he slept the days before. For other users this could be different.

- It's questionable how precise our dataset is and what sleep quality even mean

  but the point was anyway to generate a proof of concept

Generalization

$Data = \{P_1, P_2, \dots, P_n\}$

Goal: optimize some set S $\subseteq Data$

$$HS = P_i \times P_j \times \dots$$
$$OS = P_k \times P_l \times \dots$$

Trying out different combinations of properties in our dataset in both, the hidden and the observable, states

All properties of S should be in HS or OS