Automated Time Series Forecasting, Backtesting, and Optimization

Marcello Tomasini
Sr. Data Scientist

VERITAS™
The truth in information.

Agenda

1	Forecasting 101
2	The problem of automation
3	Validation a.k.a. Backtesting
4	Tuning

VERITAS™

Definition

Forecasting is the process of making *predictions* of the future based on past and present data

Disclaimer…

Nobody/Nothing can predict the future.

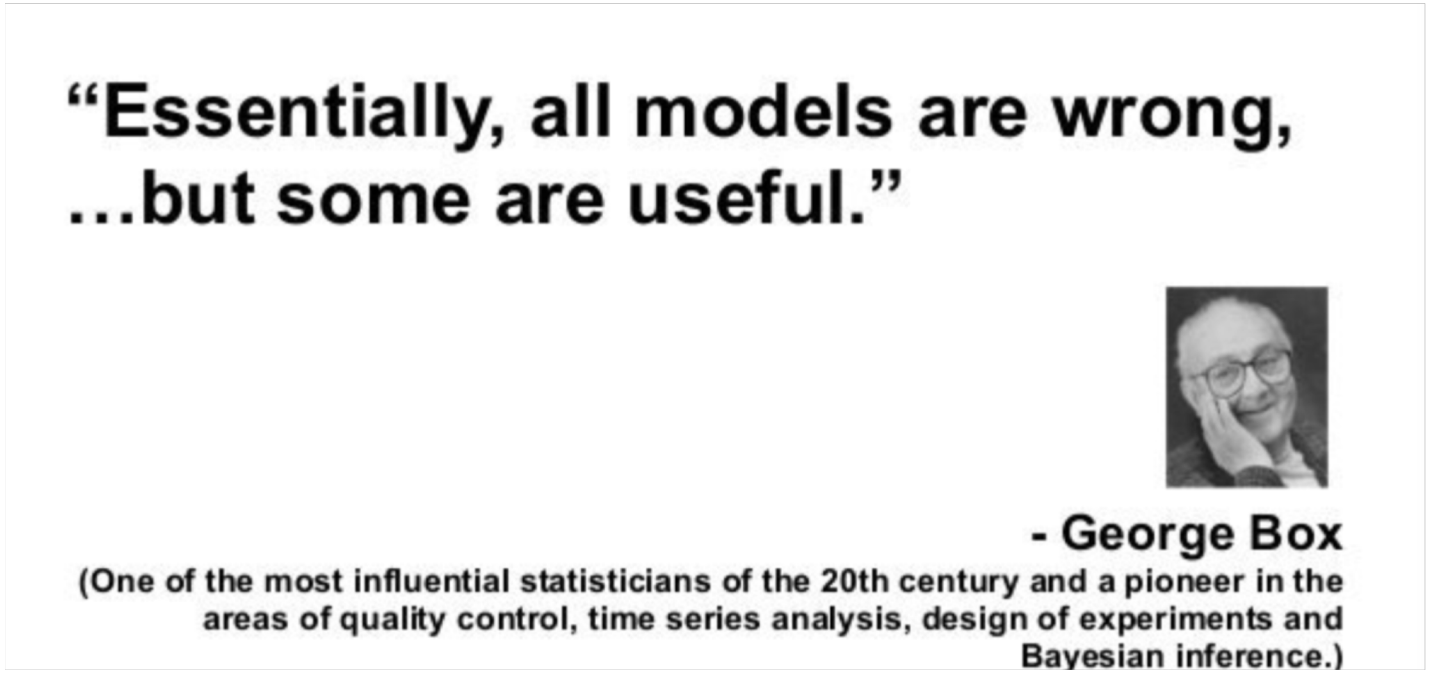
Disclaimer…

Nobody/Nothing can predict the future.

- *I think there is a world market for maybe five computers.* (Chairman of IBM, 1943)
- *Computers in the future may weigh no more than 1.5 tons.* (Popular Mechanics, 1949)
- *There is no reason anyone would want a computer in their home.* (President, DEC, 1977)
- *640K ought to be enough for anyone*. (President, Microsoft, 1986)

Disclaimer…

Nobody/Nothing can *exactly* predict the future.



What can be forecasted?

The predictability of an event or a quantity depends on several factors:

- how well we understand the factors that contribute to it;
- how much data are available;
- whether the forecasts can affect the thing we are trying to forecast.

What is normally assumed is that *the way in which the environment is changing will continue into the future*.

VERITAS

Forecasting Methods

Depends mostly on the data that are available:

- **qualitative forecasting** if no data. These methods are not purely guesswork—there are well-developed structured approaches to obtaining good forecasts without using historical data;
- **quantitative forecasting** can be applied when numerical information about the past is available.

Quantitative forecasting problems:

- **time series**: data collected at regular intervals over time;
- **cross-sectional**: data collected at a single point in time.

Explanatory Model vs Time Series Model

- Explanatory model

$$y_{t+h} = f(\mathrm{VAR}_1, \mathrm{VAR}_2, \ldots, \mathrm{VAR}_n, error)$$

- Time series model

$$y_{t+h} = f(y_t, y_{t-1}, \ldots, y_{t-n}, error)$$

VERITAS

Trend and Seasonality

Additive Decomposition

$$y_t = S_t + T_t + R_t$$

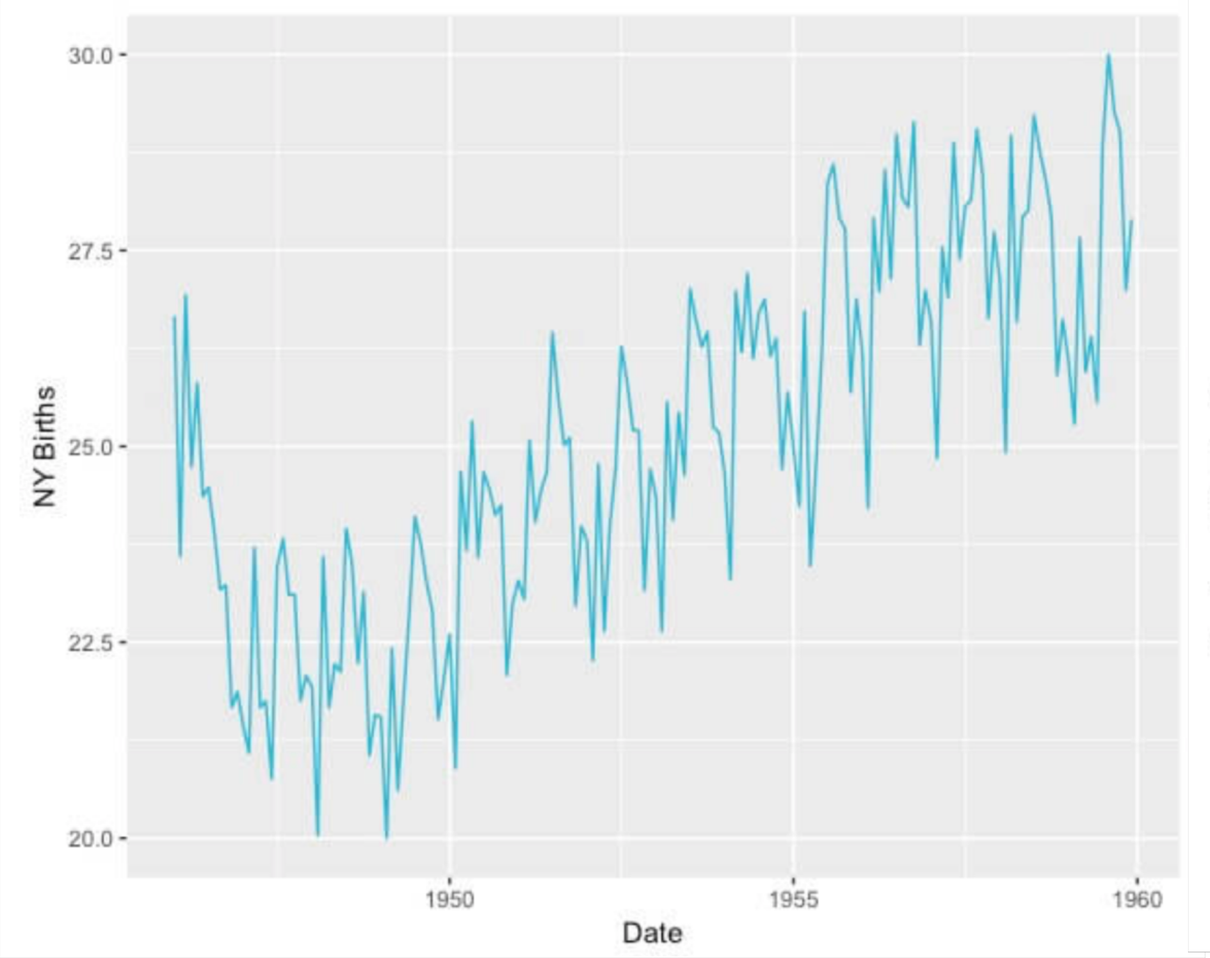
Multiplicative Decomposition

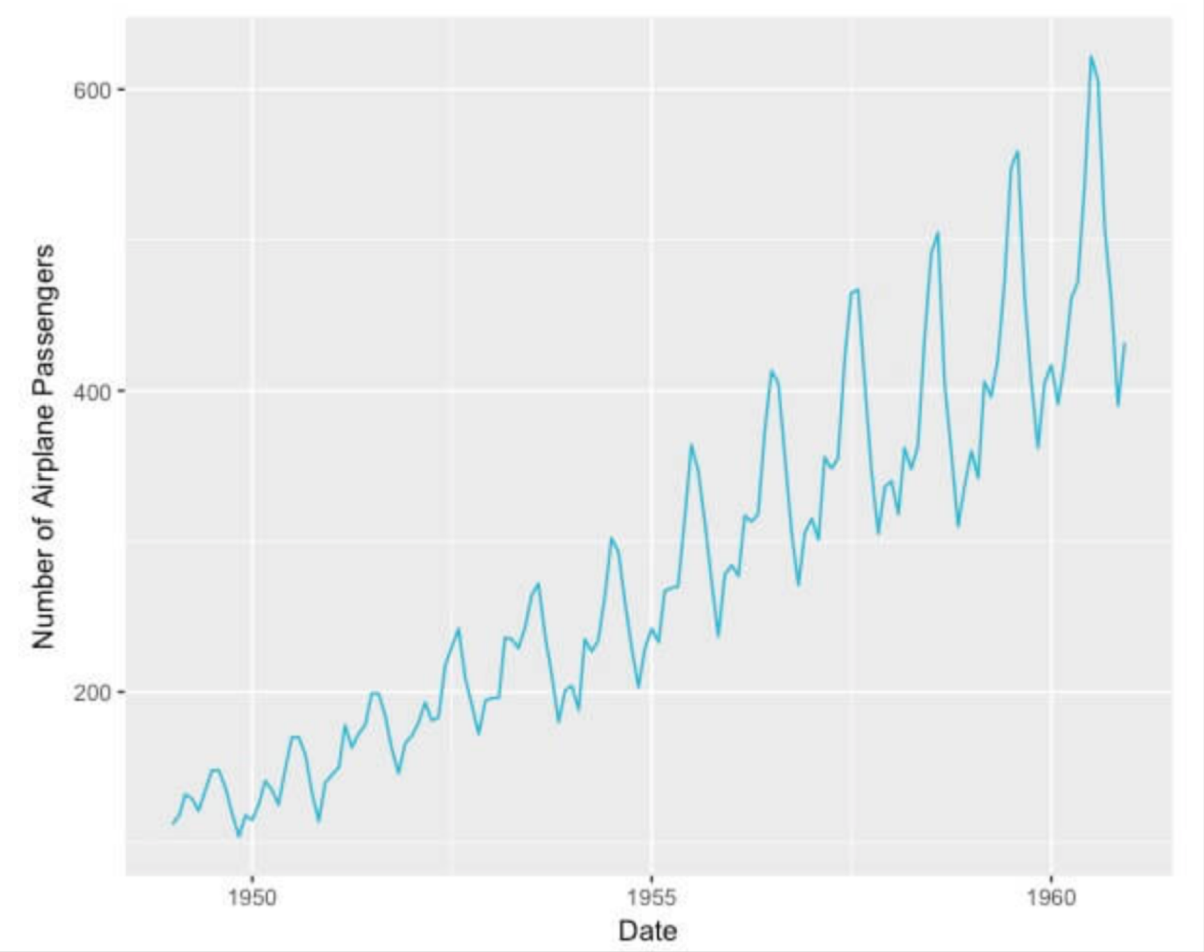
$$y_t = S_t \times T_t \times R_t$$

y_t is the data, S_t is the seasonal component, T_t is the trend-cycle component, and R_t is the remainder component, all at period t

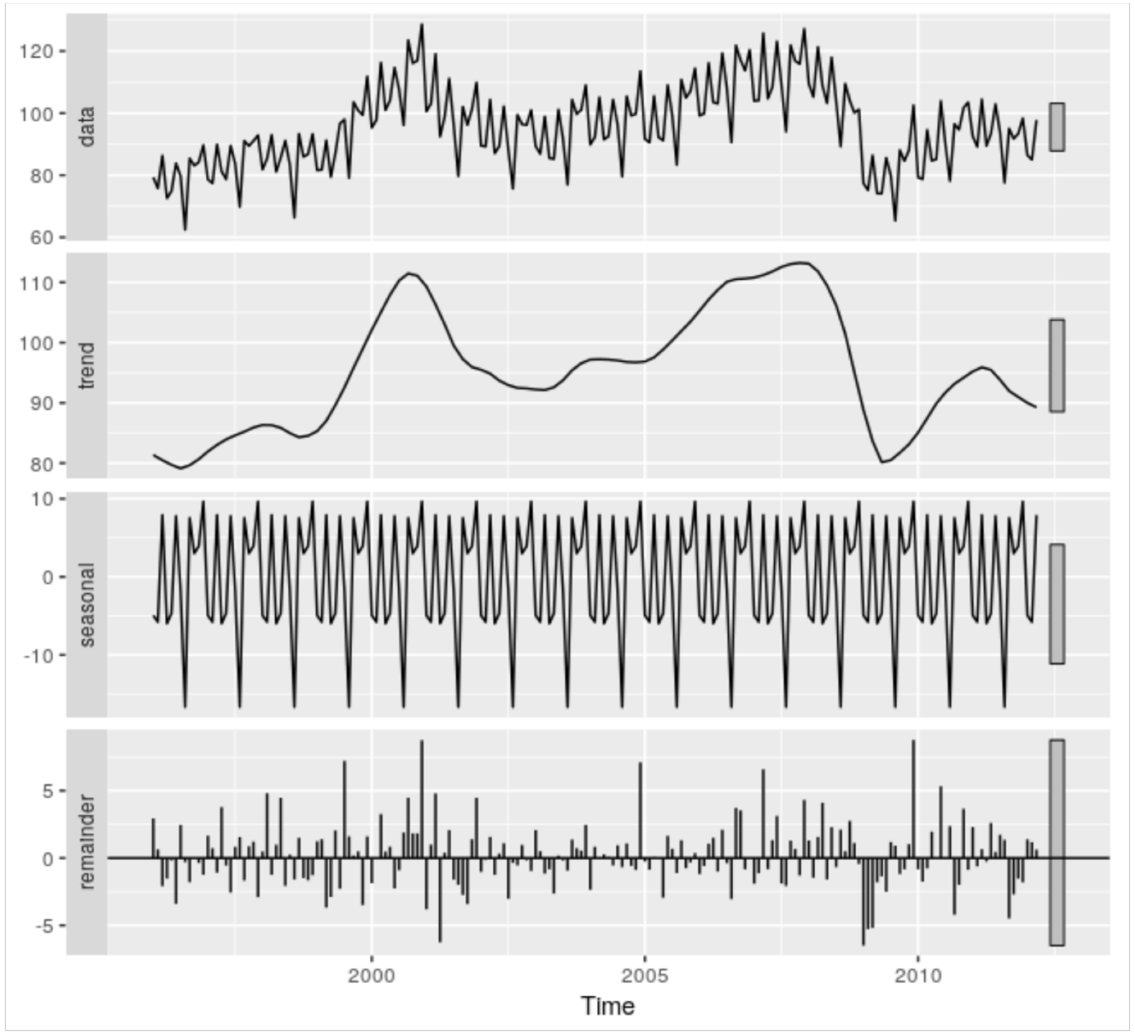
VERITAS

Additive vs Multiplicative Seasonality





Trend and Seasonality Decomposition



https://otexts.com/fpp2

Storage Usage Forecast @ Veritas Predictive Insights

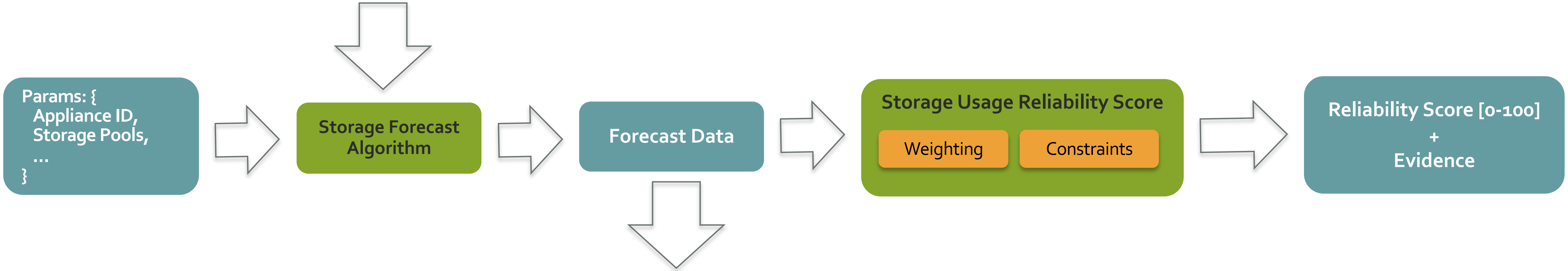
- 10,000+ NetBackup Appliances actively using Auto Support
- Telemetry data reported daily
  - 1 point every 15 minutes (96 points/day) per storage type (MSDP, Log, etc)
  - 2+ years of data (0.5M+ points/appliance)
- Storage forecast use cases
  - Resource planning
  - Workload anomalies
  - Possible data unavailability or SLA violations
  - Sales opportunity

https://www.veritas.com/product/backup-and-recovery/netbackup-appliances

VERITAS

Storage Usage Forecast @ Veritas Predictive Insights

Time	System	Log	Share	Advanced Disk	MSDP	MSDP Catalog	NBU Catalog	Config
2016-04-11T17:34:55.37602	14.0	15.00415	0.0	947.537800	6312.08	1.66180	0.0	21.13150
2016-04-11T17:49:51.58820	18.0	15.01245	0.0	947.542024	6312.14	1.66230	0.0	21.36500



Time	System	Log	Share	Advanced Disk	MSDP	MSDP Catalog	NBU Catalog	Config
2016-04-12	19.0	15.01345	0.0	947.5300	6312.16	1.66380	0.0	21.456
2016-04-13	20.5	15.01456	0.0	947.5424	6312.18	1.66430	0.0	21.635

Challenges of Automation

- Model selection
  – Which model is the best?
- Model Evaluation
  – How accurate is the model that is currently running in production?
- Model Update
  – How to keep tuning/improving the model?

Challenges of Automation

- Model selection
  – Which model is the best?
- Model Evaluation
  – How accurate is the model that is currently running in production?
- Model Update
  – How to keep tuning/improving the model?

Model Selection

- How do we deal with missing values?
- How do we deal with outliers?
- How do we deal with trend changepoints?
- How do we deal with trend & seasonality?
- How do we choose algorithm parameters?

Model Selection

- How do we deal with missing values?

Data imputation

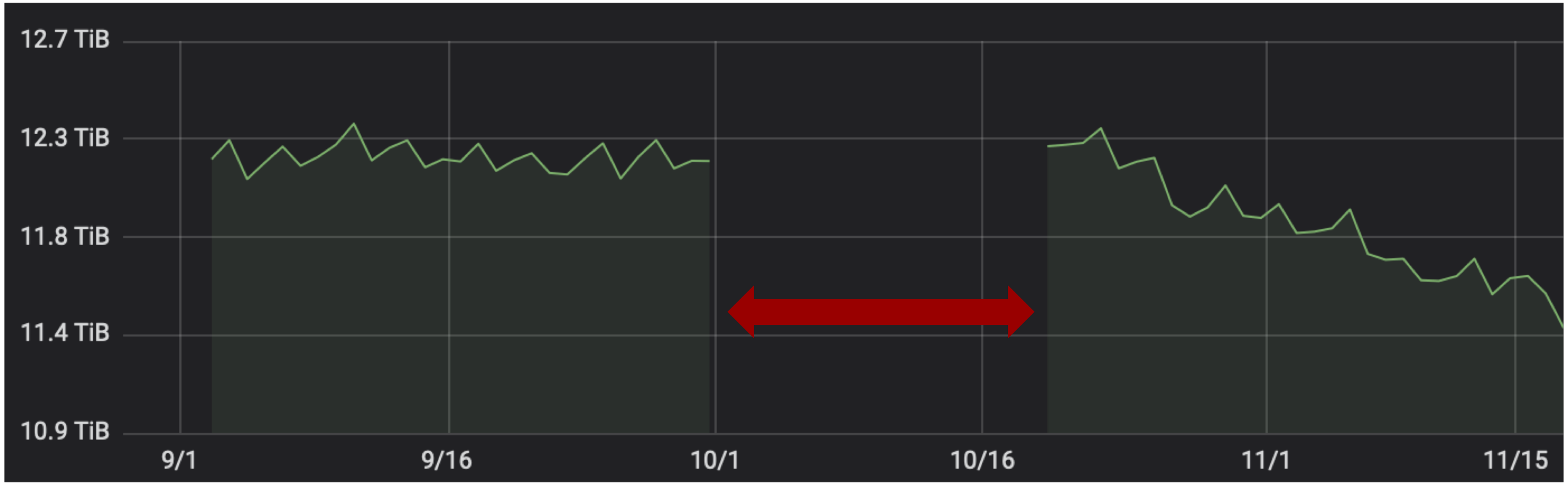
- How do we deal with outliers?
- How do we deal with trend changepoints?
- How do we deal with trend & seasonality?
- How do we choose algorithm parameters?

VERITAS

Model Selection

- How do we deal with missing values?

Data imputation → Problem: "holes" in the data



VERITAS

Model Selection

- How do we deal with missing values?
- How do we deal with outliers?

Just remove them

- How do we deal with trend changepoints?
- How do we deal with trend & seasonality?
- How do we choose algorithm parameters?

Model Selection

- How do we deal with missing values?
- How do we deal with outliers?

Just remove them → Problem: automated outliers detection?

- How do we deal with trend changepoints?
- How do we deal with trend & seasonality?
- How do we choose algorithm parameters?

Model Selection

- How do we deal with missing values?
- How do we deal with outliers?
- How do we deal with trend & seasonality?

Trend/Seasonality decomposition or seasonal model

- How do we deal with trend change points?
- How do we choose algorithm parameters?

Model Selection

- How do we deal with missing values?
- How do we deal with outliers?
- How do we deal with trend & seasonality?

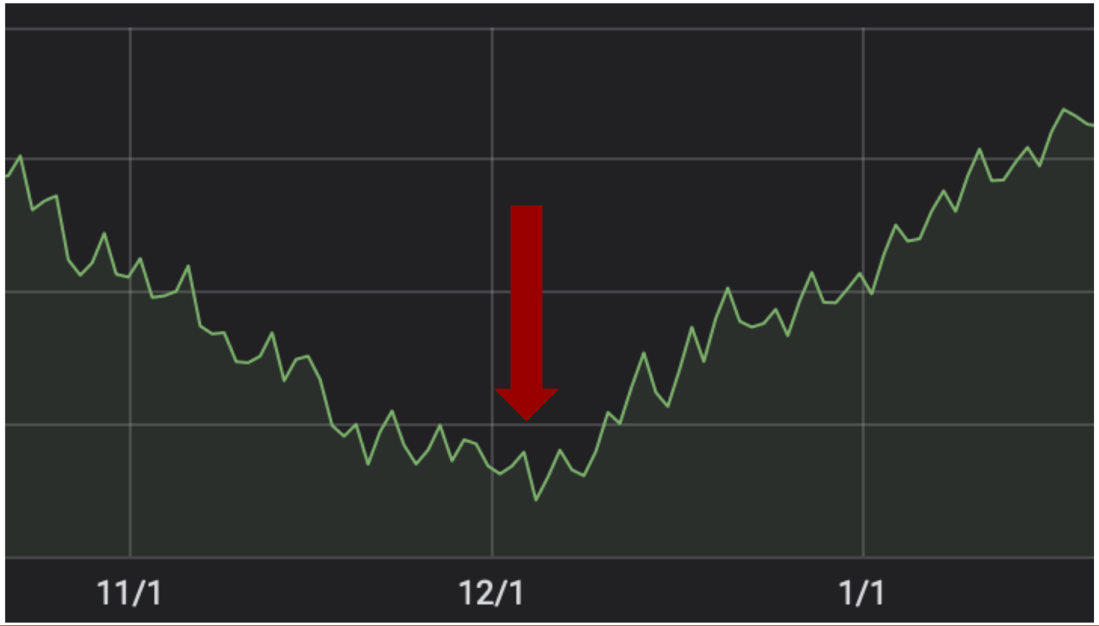
Trend/Seasonality decomposition or seasonal model

→ Problem: seasonality period & multiple seasonalities

- How do we deal with trend change points?
- How do we choose algorithm parameters?

Model Selection

- How do we deal with missing values?
- How do we deal with outliers?
- How do we deal with trend & seasonality?
- How do we deal with trend change points?



Model Selection

- How do we deal with missing values?
- How do we deal with outliers?
- How do we deal with trend & seasonality?
- How do we deal with trend change points?
- How do we choose algorithm parameters?

Cross Validation (Backtesting)

Model Selection

- How do we deal with missing values?
- How do we deal with outliers?
- How do we deal with trend & seasonality?
- How do we deal with trend change points?
- How do we choose algorithm parameters?

Cross Validation (Backtesting) → Problem: thousands of models…

Model Selection

- Bad News:
  - Real world data is messy
  - We must tackle each case, no way around it
  - Impossible deadlines

DS/ML/AI in production is hard!

Model Selection

- Bad News:
  - Real world data is messy
  - We must tackle each case, no way around it
  - Impossible deadlines

DS/ML/AI in production is hard!

- Good News:
  - Somebody took care of most of the issues already…

VERITAS

Facebook Prophet

- Decomposable time series model

$$y_t = g(t) + s(t) + h(t) + \epsilon_t$$

- Incorporate trend changes in the growth model by explicitly defining changepoints
- Fourier series to provide a flexible model of periodic effects (seasonality)
- Holidays as an additional regressor
- Observations do not need to be regularly spaced
- Handle missing values
- Robust to outliers

https://research.fb.com/prophet-forecasting-at-scale/

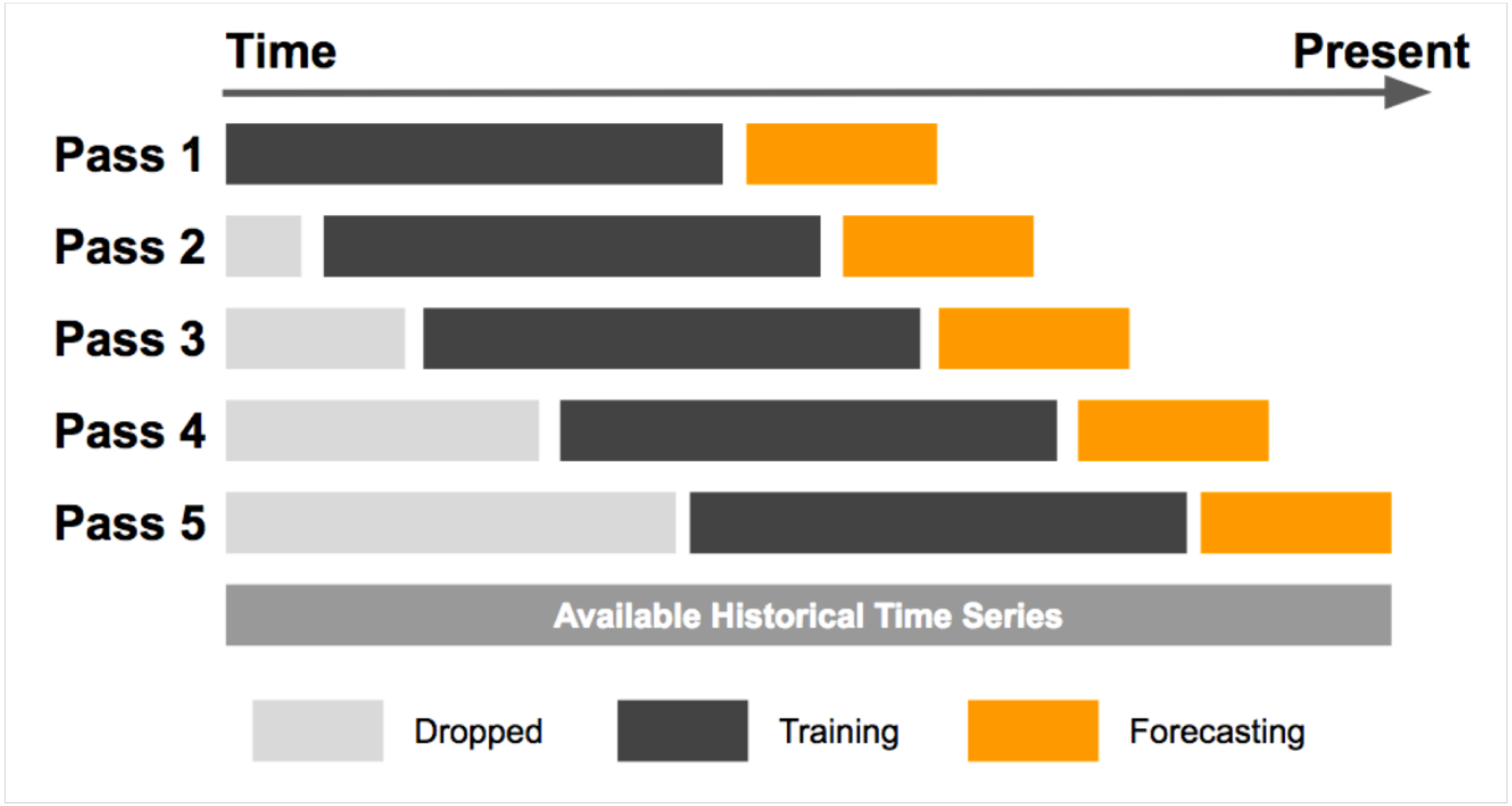
VERITAS

Challenges of Automation

- Model selection
  - Which model is the best?
- Model Evaluation
  - How accurate is the model that is currently running in production?
- Model Update
  - How to keep tuning/improving the model?

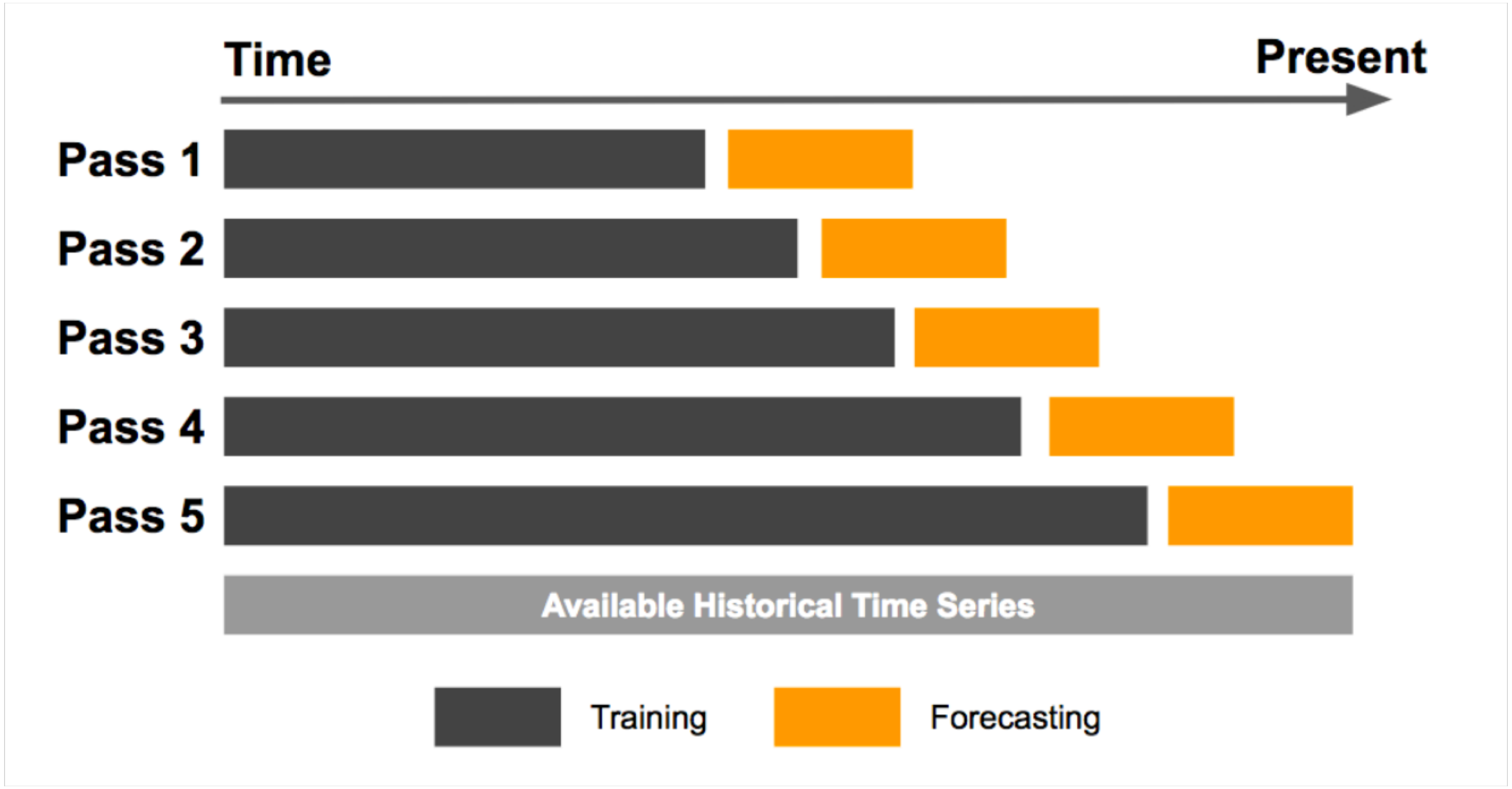
VERITAS

Cross Validation: Sliding Window Backtesting



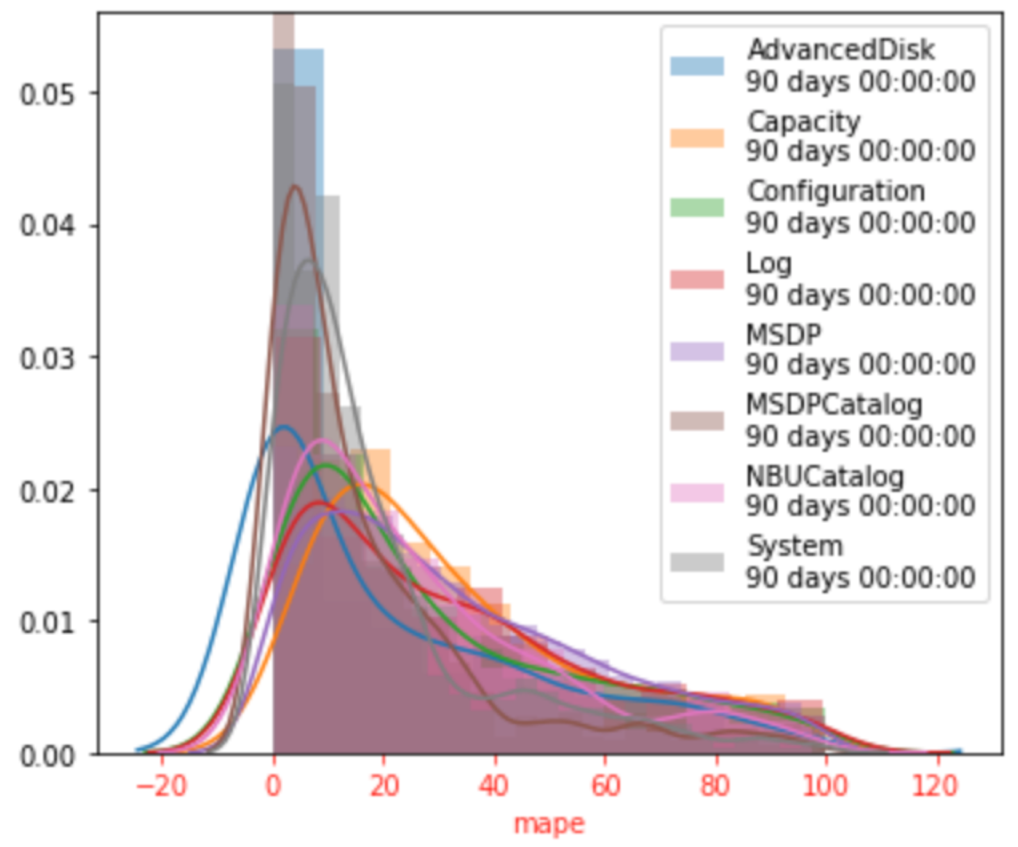
https://eng.uber.com/omphalos/

Cross Validation: Expanding Window Validation



https://eng.uber.com/omphalos/

Cross Validation: Experimental Results



	sample_periods	pool_type	horizon	coverage	mape	mdape	mpe	support
0	150	AdvancedDisk	1 days	0.874991	3.723209	2.104240	-0.224229	1511
1	150	AdvancedDisk	7 days	0.896520	4.729747	2.977387	-0.442343	1490
2	150	AdvancedDisk	15 days	0.923884	6.030256	3.805062	-0.219382	1467
3	150	AdvancedDisk	30 days	0.934661	7.500707	5.203543	-0.700334	1429
4	150	Log	1 days	0.623440	2.785619	1.372949	-0.128965	2082
5	150	Log	7 days	0.872525	4.780381	2.302635	-0.074960	2072
6	150	Log	15 days	0.927999	7.034532	3.682870	-0.695416	2046
7	150	Log	30 days	0.951800	9.242708	6.172322	-1.077535	1993
8	150	MSDP	1 days	0.516166	4.336129	2.572391	0.230806	1956
9	150	MSDP	7 days	0.816177	7.785473	5.101999	0.094574	1929
10	150	MSDP	15 days	0.920877	11.518986	8.098703	-0.353916	1887
11	150	MSDP	30 days	0.957575	16.805184	12.581191	-1.025039	1798
12	150	System	1 days	0.719698	1.456294	0.723867	0.013275	2087
13	150	System	7 days	0.885093	2.899766	1.545664	-0.070124	2085
14	150	System	15 days	0.936821	4.782500	2.554065	-0.281044	2078
15	150	System	30 days	0.939398	7.145672	4.700071	-0.535361	2059

Backtesting vs Online Testing

- Problems:
  - Thousands of series
  - Possibly more than one model per time series
  - One-off/batch run means results are always outdated

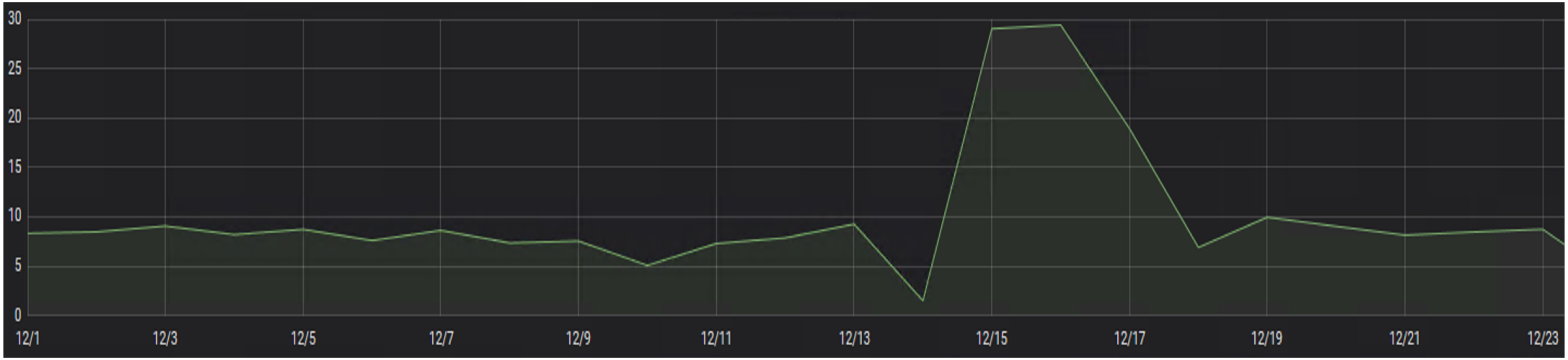
Backtesting vs Online Testing

- Problems:
  - Thousands of series
  - Possibly more than one model per time series
  - One-off/batch runs mean results are always outdated
- Solution:
  - Save previous forecasts
  - Do multi-horizon forecasts
  - Compute error as we generate new forecasts

VERITAS

Backtesting vs Online Testing

- Forecast error as a time series
  - Per appliance
  - Per horizon
  - Per storage type



Challenges of Automation

- Model selection
  - Which model is the best?
- Model Evaluation
  - How accurate is the model that is currently running in production?
- Model Update
  - How to keep tuning/improving the model?

Model Update Problems

- Thousands of models
- Each model needs to be tuned for a specific time series
- Need to adapt to changes in underlying process
- Backtesting is computationally too expensive but we have the online validation data!

Sequential Model-Based Optimization (SMBO)

- Hyperparameter optimization

$$x^* = \arg\min_{x \in \chi} f(x)$$

where *f(x)* represents an objective score to minimize— such as RMSE— evaluated on the validation set.

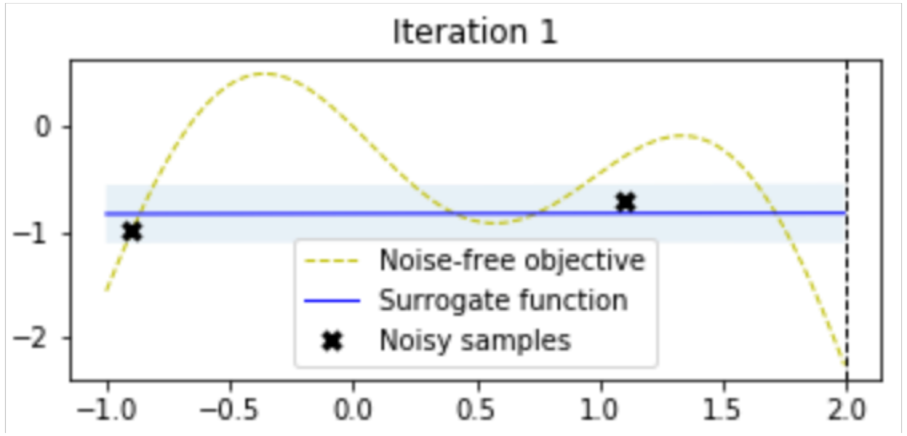
- Bayesian optimization uses a surrogate function:

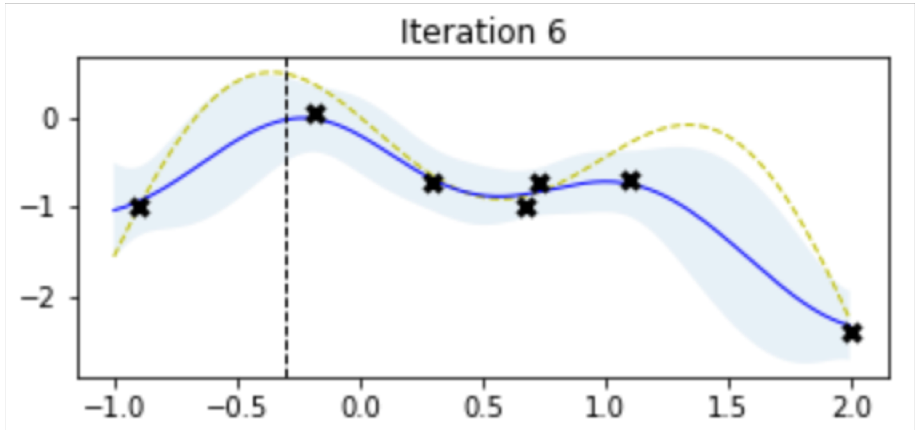
$$P(f(x)|x)$$

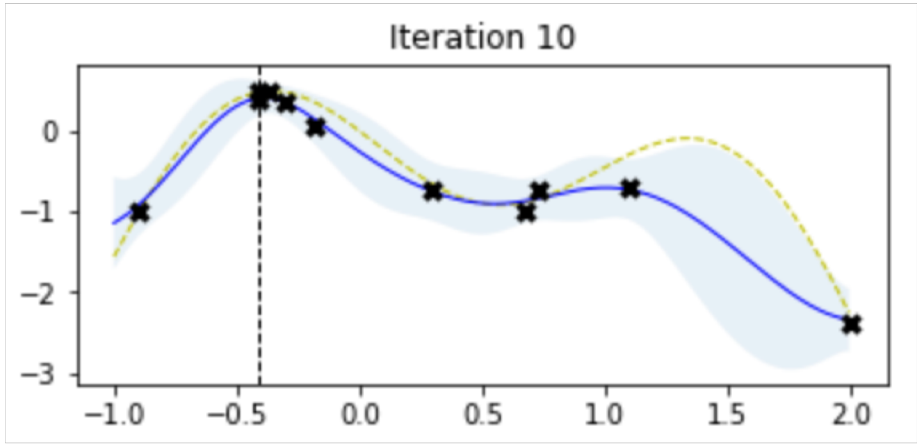
to build a probability model of the objective.

- Selection function used to choose next set of hyperparameters.

Sequential Model-Based Optimization (SMBO)







http://krasserm.github.io/2018/03/21/bayesian-optimization

VERITAS

Am I Missing Something?

Am I Missing Something?

Where is Deep Learning!

Deep Learning

- Too many parameters for the data & cross-training leads to “average” models
- Needs careful data preparation (normalization, trend, seasonality, etc)
- Computational cost of training

R Algorithms	Median wMAPE	Overall sMAPE	Overall wMAPE	Backtesting Time per City (seconds)
Tbats	0.071	0.090	0.079	269.77
Auto ARIMA	0.078	0.098	0.089	84.35
Prophet	0.075	**0.083**	0.084	78.97
Nnetar	0.098	0.128	0.111	73.68
Holt-winters	0.070	0.093	0.080	56.88
Ets	0.069	0.088	0.078	26.34
Stlm	0.069	0.090	0.078	1.94
Thetam	**0.066**	0.086	**0.077**	0.71
Golang Algorithms	**Median wMAPE**	**Overall sMAPE**	**Overall wMAPE**	**Backtesting Time per City (seconds)**
AutoForecastor	0.069	0.095	0.081	0.46

https://eng.uber.com/omphalos/

M4 Forecasting Competition

Table 1
The performances of the 17 methods submitted to M4 that had OWAs at least as good as that of the Comb.

Type	Author(s)	Affiliation	sMAPE[c] Yearly (23k)	sMAPE[c] Quarterly (24k)	sMAPE[c] Monthly (48k)	sMAPE[c] Others (5k)	sMAPE[c] Average (100k)	OWA[d] Yearly (23k)	OWA[d] Quarterly (24k)	OWA[d] Monthly (48k)	OWA[d] Others (5k)	OWA[d] Average (100k)	Rank[a]	% improvement of method over the benchmark: sMAPE	% improvement of method over the benchmark: OWA
Benchmark for methods[b]			14.848	10.175	13.434	4.987	12.555	0.867	0.890	0.920	1.039	0.898	19	sMAPE	OWA
Hybrid	Smyl, S.	Uber Technologies	13.176	9.679	12.126	4.014	11.374	0.778	0.847	0.836	0.920	0.821	1	9.4%	8.6%
Combination	Montero-Manso, P., Talagala, T., Hyndman, R. J. & Athanasopoulos, G.	University of A Coruña & Monash University	13.528	9.733	12.639	4.118	11.720	0.799	0.847	0.858	0.914	0.838	2	6.6%	6.7%
Combination	Pawlikowski, M., Chorowska, A. & Yanchuk, O.	ProLogistica Soft	13.943	9.796	12.747	3.365	11.845	0.820	0.855	0.867	0.742	0.841	3	5.7%	6.3%
Combination	Jaganathan, S. & Prakash, P.	Individual	13.712	9.809	12.487	3.879	11.695	0.813	0.859	0.854	0.882	0.842	4	6.8%	6.2%
Combination	Fiorucci, J. A. & Louzada, F.	University of Brasilia & University of São Paulo	13.673	9.816	12.737	4.432	11.836	0.802	0.855	0.868	0.935	0.843	5	5.7%	6.1%
Combination	Petropoulos, F. & Svetunkov, I.	University of Bath & Lancaster University	13.669	9.800	12.888	4.105	11.887	0.806	0.853	0.876	0.906	0.848	6	5.3%	5.6%
Combination	Shaub, D.	Harvard Extension School	13.679	10.378	12.839	4.400	12.020	0.801	0.908	0.882	0.978	0.860	7	4.3%	4.2%
Statistical	Legaki, N. Z. & Koutsouri, K.	National Technical University of Athens	13.366	10.155	13.002	4.682	11.986	0.788	0.898	0.905	0.989	0.861	8	4.5%	4.1%
Combination	Doornik, J., Castle, J. & Hendry, D.	University of Oxford	13.910	10.000	12.780	3.802	11.924	0.836	0.878	0.881	0.874	0.865	9	5.0%	3.7%
Combination	Pedregal, D.J., Trapero, J. R., Villegas, M. A. & Madrigal, J. J.	University of Castilla-La Mancha	13.821	10.093	13.151	4.012	12.114	0.824	0.883	0.899	0.895	0.869	10	3.5%	3.2%
Statistical	Spiliotis, E. & Assimakopoulos, V.	National Technical University of Athens	13.804	10.128	13.142	4.675	12.148	0.823	0.889	0.907	0.975	0.874	11	3.2%	2.7%
Combination	Roubinchtein, A.	Washington State Employment Security Department	14.445	10.172	12.911	4.436	12.183	0.850	0.885	0.881	0.992	0.876	12	3.0%	2.4%
Other	Ibrahim, M.	Georgia Institute of Technology	13.677	10.089	13.321	4.747	12.198	0.805	0.890	0.921	1.079	0.880	13	2.8%	2.0%
Combination	Kull, M., et al.	University of Tartu	14.096	11.109	13.290	4.169	12.496	0.820	0.960	0.932	0.883	0.888	14	0.5%	1.1%
Combination	Waheeb, W.	Universiti Tun Hussein Onn Malaysia	14.783	10.059	12.770	4.039	12.146	0.880	0.880	0.927	0.904	0.894	15	3.3%	0.4%
Statistical	Darin, S. & Stellwagen, E.	Business Forecast Systems (Forecast Pro)	14.663	10.155	13.058	4.041	12.279	0.877	0.887	0.887	1.011	0.895	16	2.2%	0.3%
Combination	Dantas, T. & Oliveira, F.	Pontifical Catholic University of Rio de Janeiro	14.746	10.254	13.462	4.783	12.553	0.866	0.892	0.914	1.011	0.896	17	0.0%	0.2%
Best ML	Trotta, B.	Individual	14.397	11.031	13.973	4.566	12.894	0.859	0.939	0.941	0.991	0.915	23	−2.7%	−1.9%
2nd Best ML	Bontempi, G.	Université Libre de Bruxelles	16.613	11.786	14.800	4.734	13.990	1.050	1.072	1.007	1.051	1.045	37	−11.4%	−16.4%

[a] Rank: rank calculated considering both the submitted methods (50) and the set of benchmarks (10).
[b] Com: the benchmark is the average (combination) of Simple, Holt and Damped exponential smoothing.
[c] sMAPE: symmetric mean absolute percentage error.
[d] OWA: overall weighted average of the relative sMAPE and the relative MASE.

https://eng.uber.com/m4-forecasting-competition
https://doi.org/10.1016/j.ijforecast.2018.06.001

Thank you!

Marcello Tomasini

Marcello.Tomasini@veritas.com
https://www.linkedin.com/in/marcellotomasini

Q & A