Data Science in Search for Best Predictions of **Ski Tour Difficulties**

KSFE e.V.

KSFE virtuell - Tipps & Tricks für SAS User

Dienstag, 5. Apr. 2022 von 10:30 bis 12:15

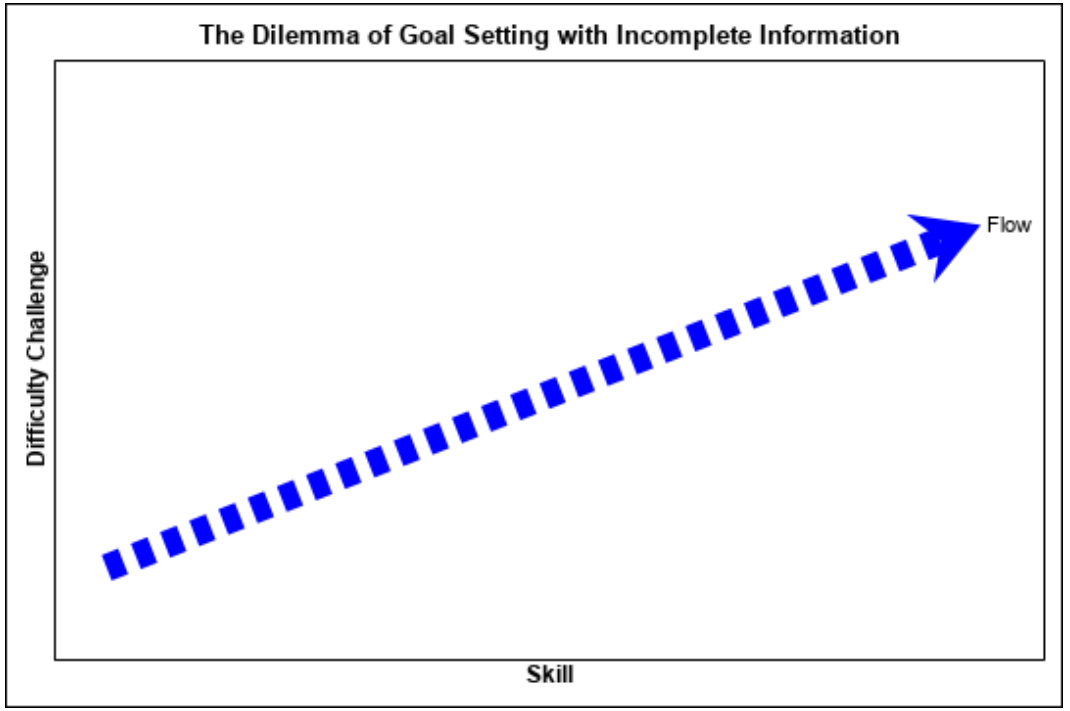
Günter Schmudlach, Skitourenguru.ch
Ulrich Reincke, SAS Institute

Difficulty of a ski tour

Why is it important?

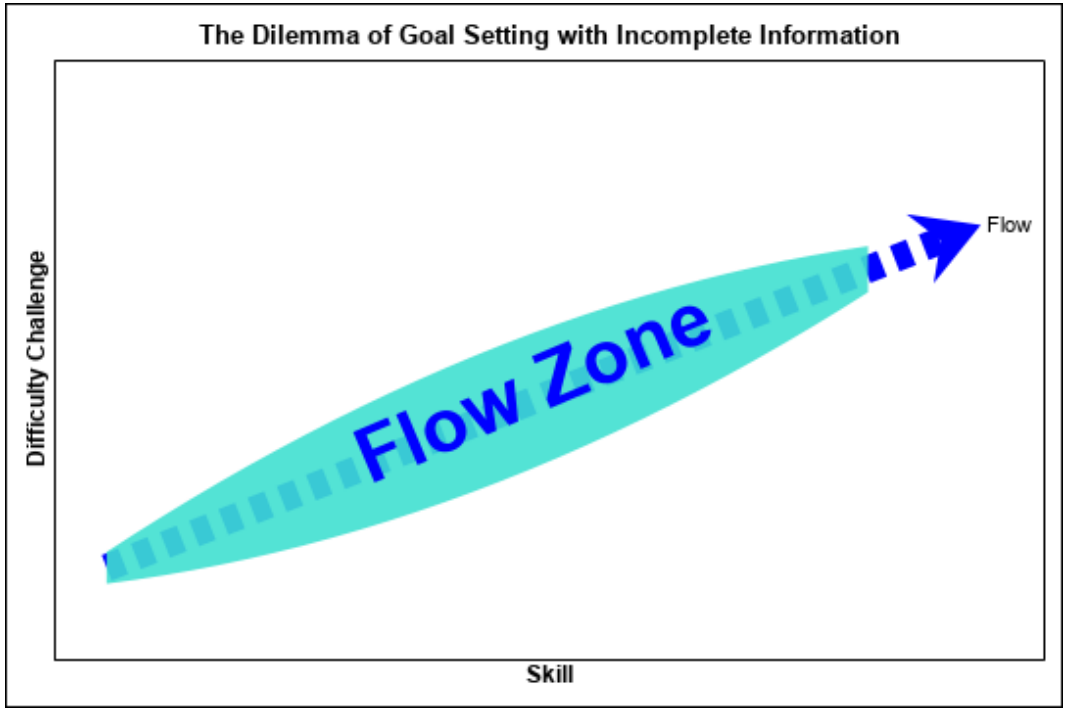
Difficulty of a ski tour

Why is it important?



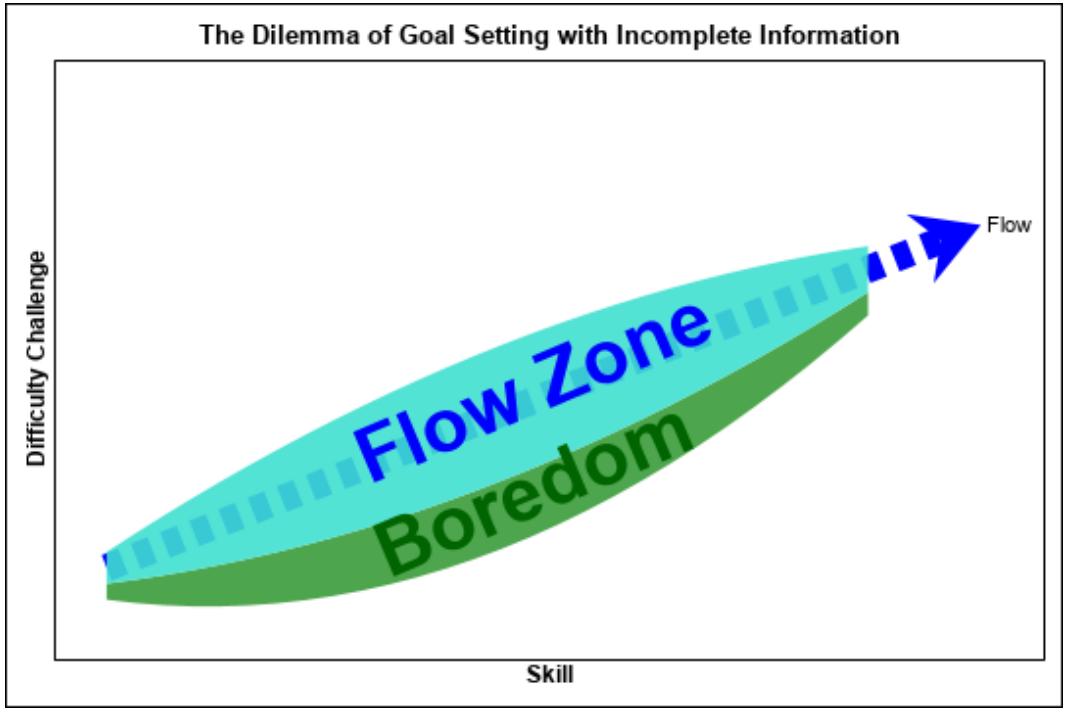
Difficulty of a ski tour

Why is it important?



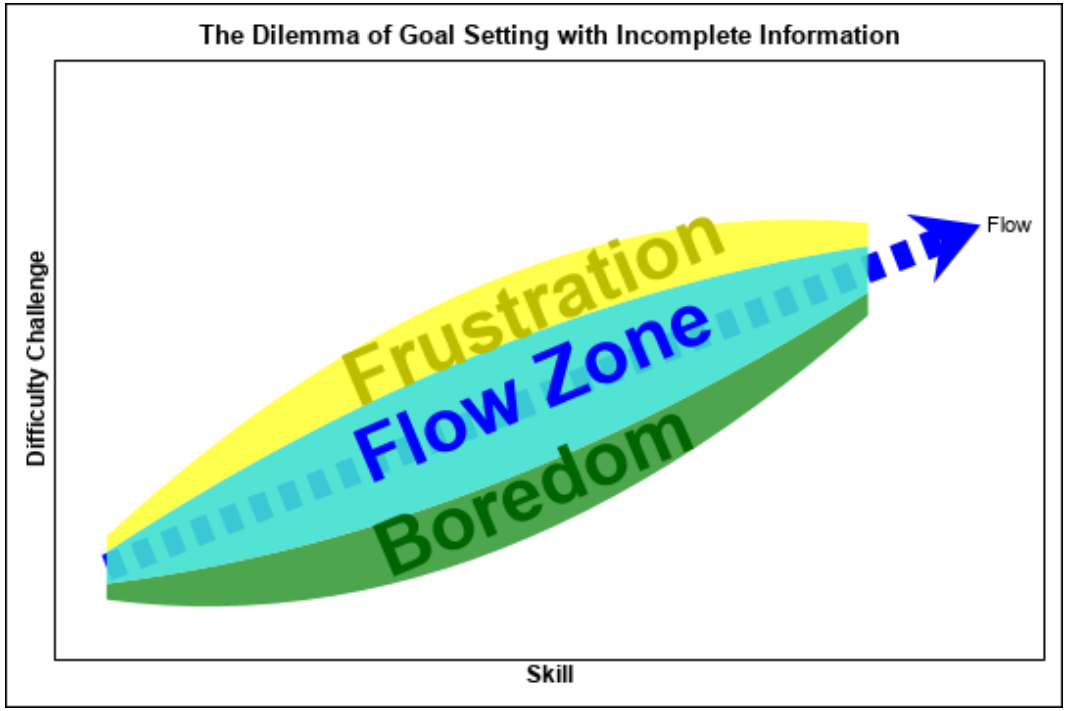
Difficulty of a ski tour

Why is it important?



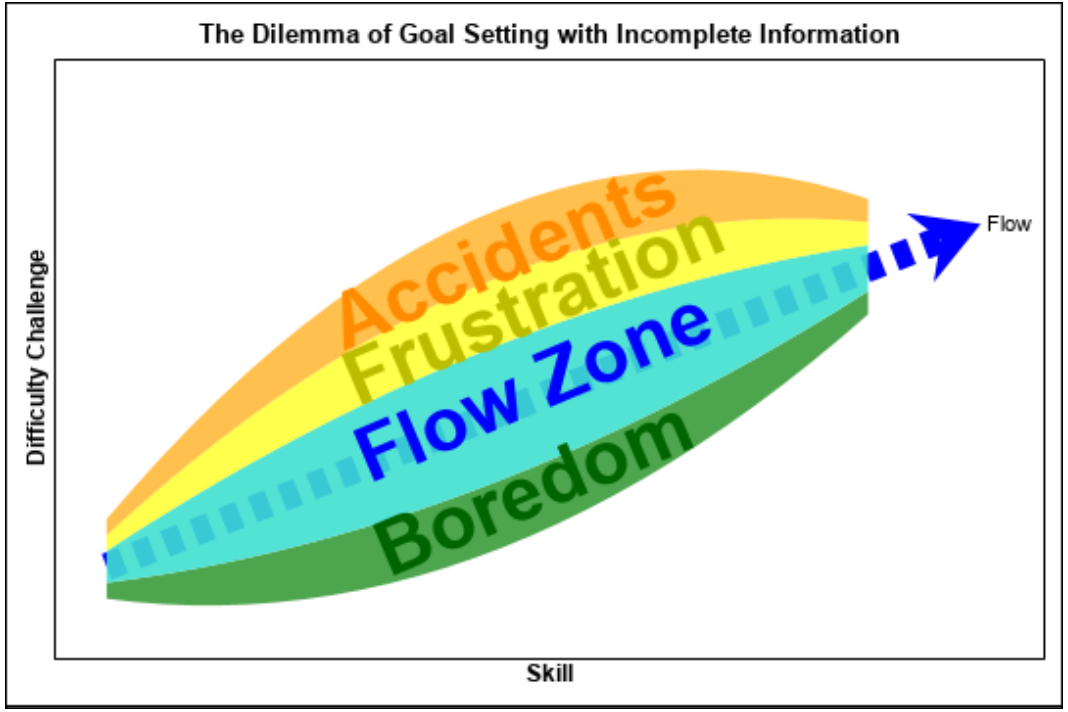
Difficulty of a ski tour

Why is it important?



Difficulty of a ski tour

Why is it important?



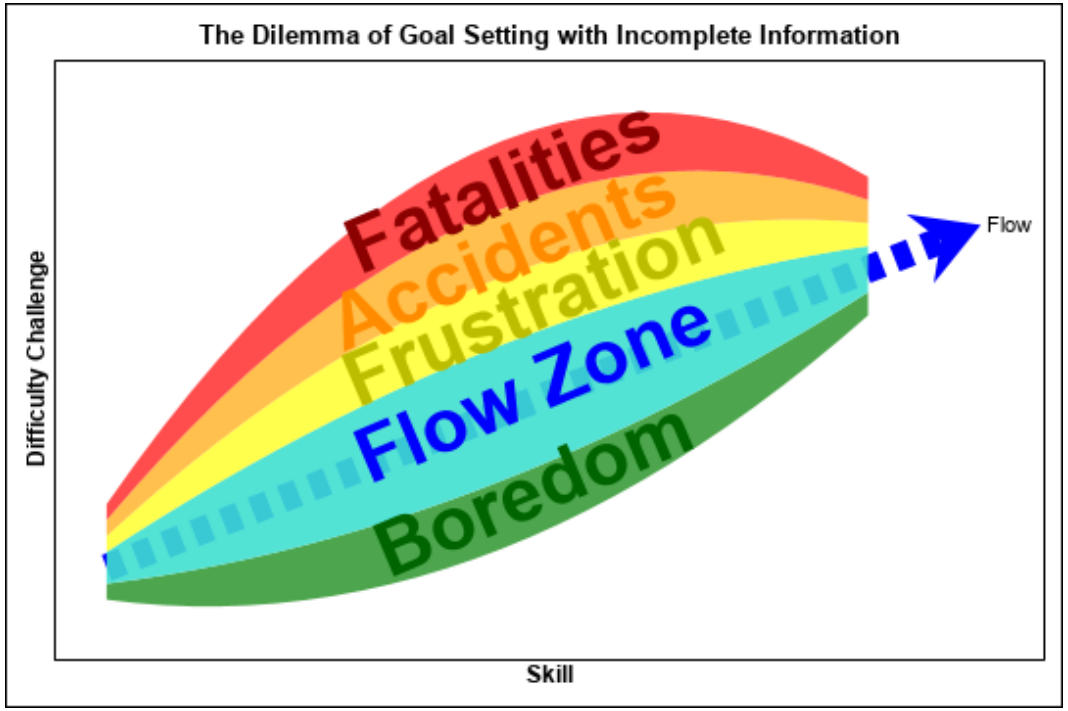
Difficulty of a ski tour

Why is it important?

Yearly official alpine accident statistics:

~70 fatalities of ~400 severe avalanche accidents

~25 fatalities of ~1000 severe ski tour accidents (non-avalanche)



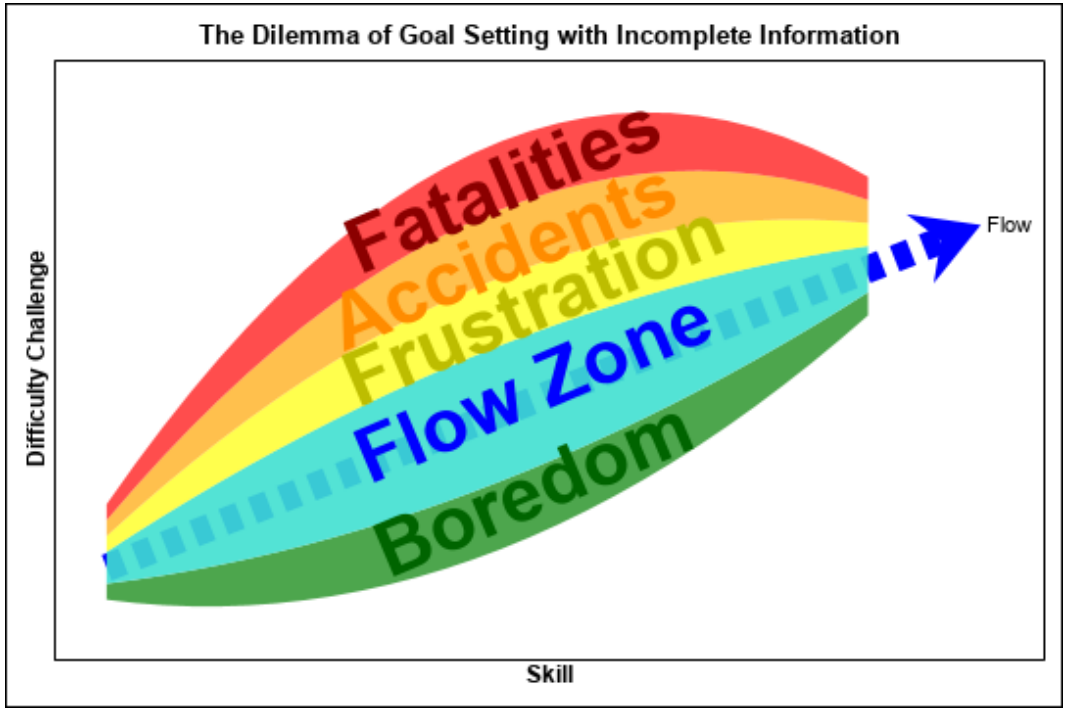
Knowledge of a tour‘s difficulties is important for

Better tour preparation, reduction of accidents and fatalities

Yearly official alpine accident statistics:

~70 fatalities of ~400 severe avalanche accidents

~25 fatalities of ~1000 severe ski tour accidents (non-avalanche)



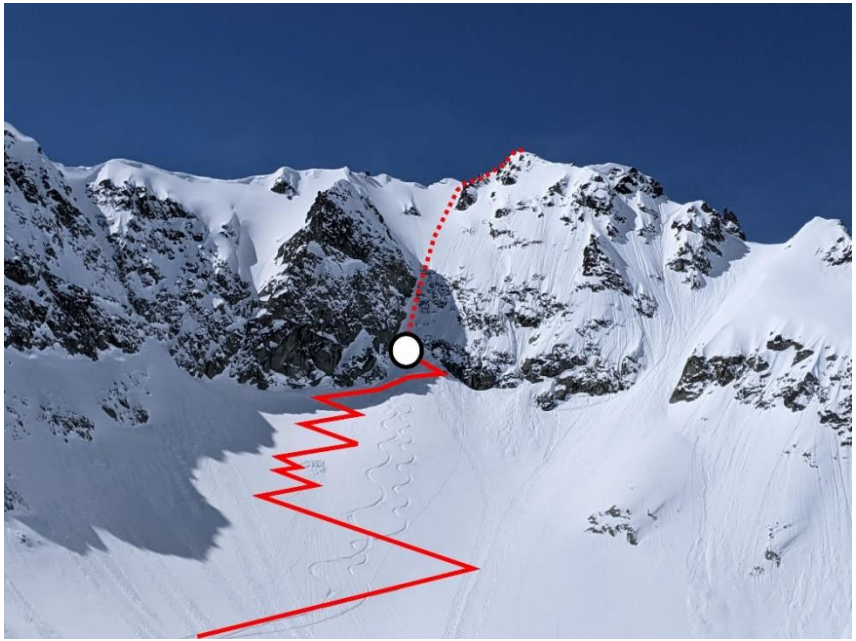
$$DIFFICULTY = f(SlopeAngle, SpeedMax, Curvature, Forestation, \ldots.)$$

Dependent Variable: Difficulty

N=1307 Swiss Ski Tours,

Published in Swiss ski touring literature:

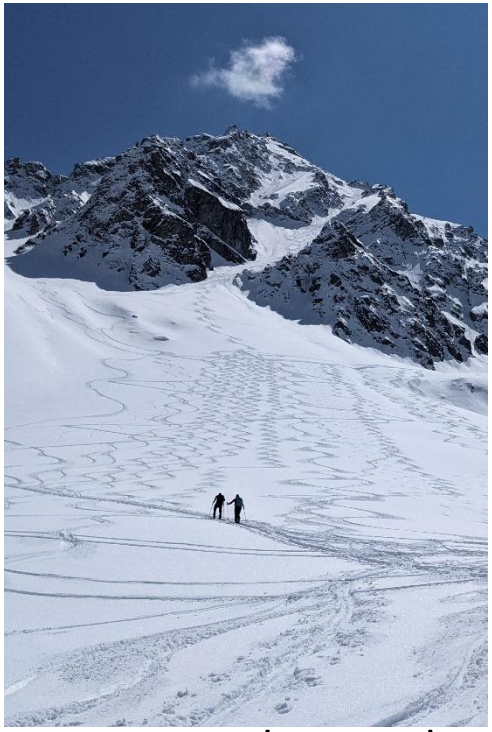
DIFFICULTY	DIFFICULTY LABEL
1	Easy
2	Easy (+)
3	Less Difficult(-)
4	Less Difficult
5	Less Difficult (+)
6	Quite Difficult (-)
7	Quite Difficult
8	Quite Difficult (+)
9	Difficult (-)
10	Difficult
11	Difficult (+)
12	Very Difficult (-)
13	Very Difficult
14	Very Difficult (+)
15	Extremely Difficult (-)
16	Extremely Difficult
17	Extremely Difficult (+)
18	Extremely Difficult



According to the SAC methodology, the difficulty level should only reflect the ski section of a tour up to the ski depot

Main criteria for the SAC difficulty scale

steepness, exposure to fall down, space conditions



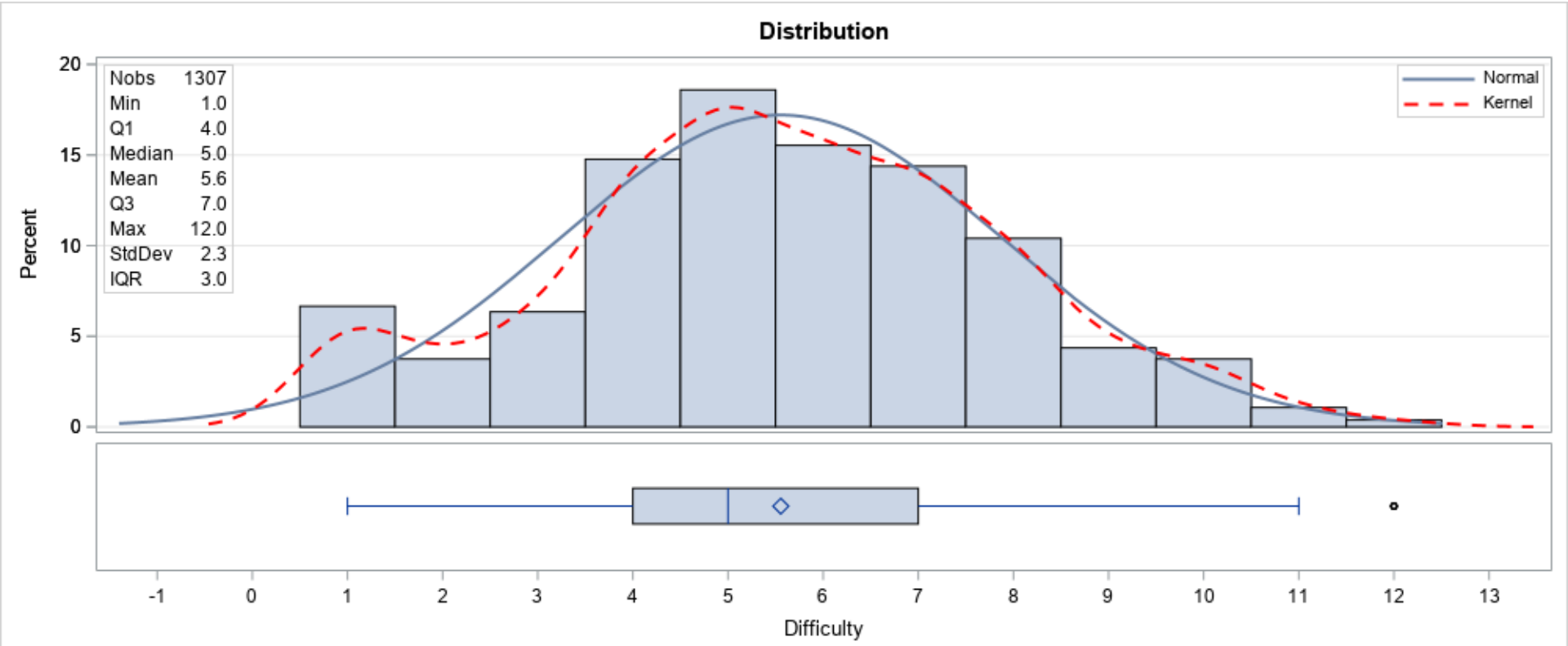
steepness: slope angle

exposure to fall: speed max

space conditions: corridor width

target variable / dependent variable

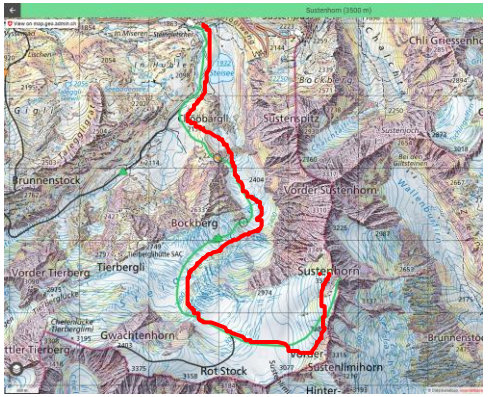
ski tour difficulty from SAC literature



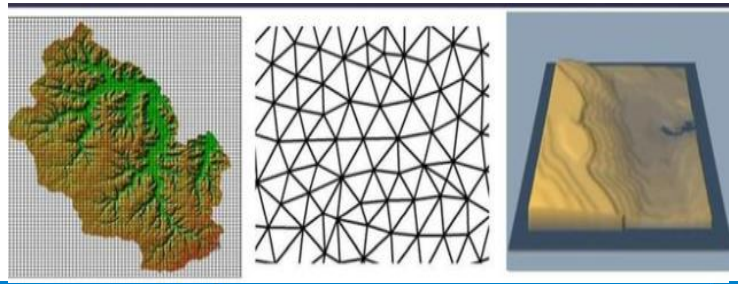
Data preparation: from properties to prediction features

N=1307 Swiss tours, ~9.3 mill. track meters

Local properties along each Track:



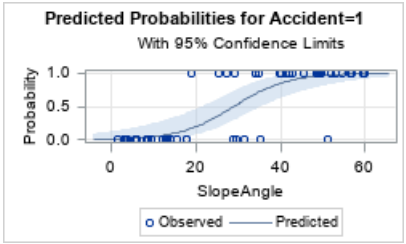
Digital Landscape Model 10m*10m



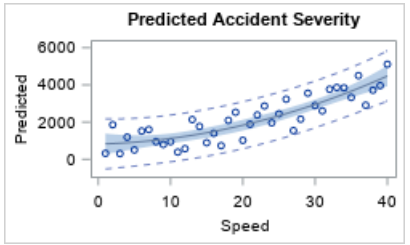
Properties:

-SlopeAngle (x,y) „steepness“
-SpeedMax (x,y) „exposure to fall“
-Width (x,y) „space conditions“
-Forestation (x,y)
-Curvature (x,y)
-Fold (x,y)

*-Risk(x,y):=SlopeAngle (x,y)*SpeedMax (x,y)*



X

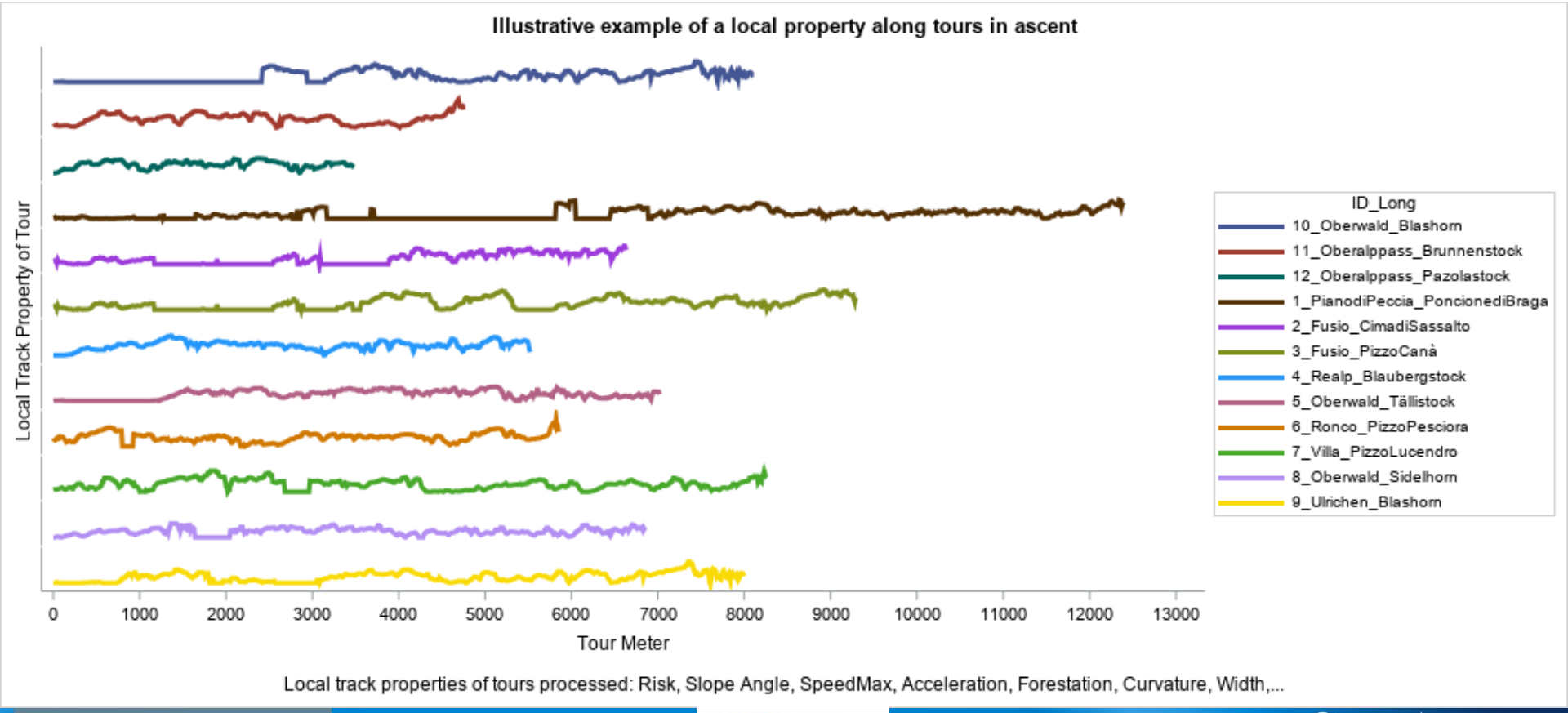


Data preparation

from local track properties to unique tour features

Data preparation

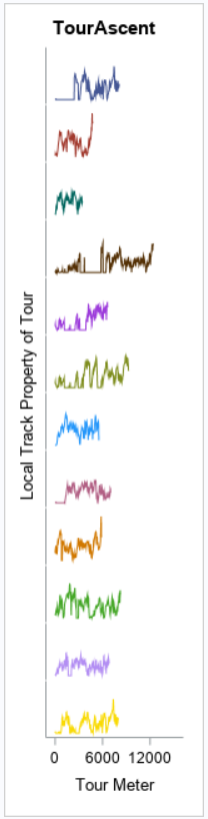
from local track properties to unique tour features



Local track properties of tours processed: Risk, Slope Angle, SpeedMax, Acceleration, Forestation, Curvature, Width,...

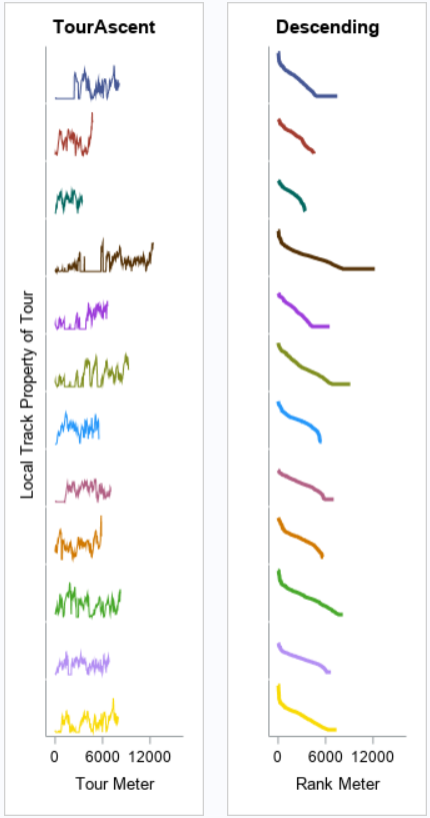
Data preparation

from local track properties to unique tour features



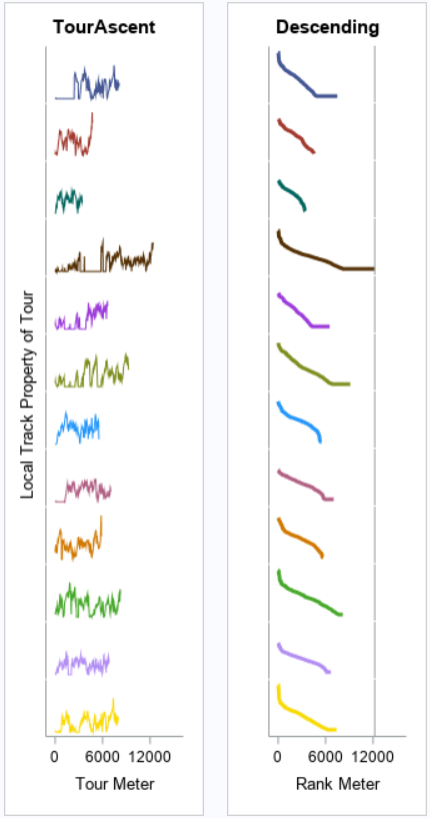
Data preparation

from local track properties to unique tour features



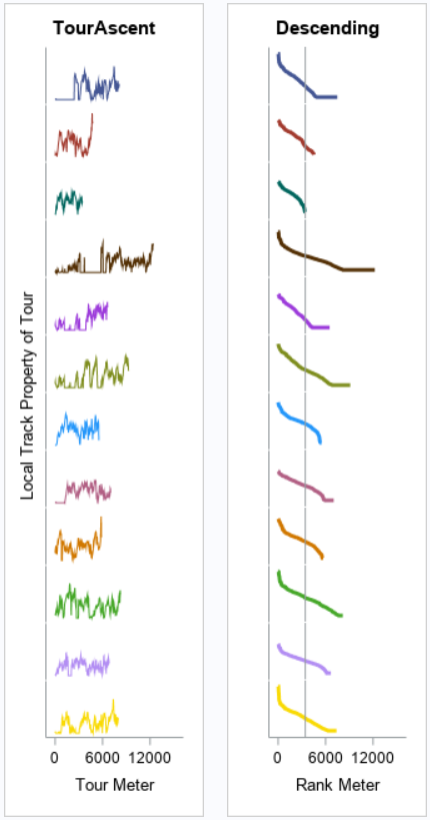
Data preparation

from local track properties to unique tour features



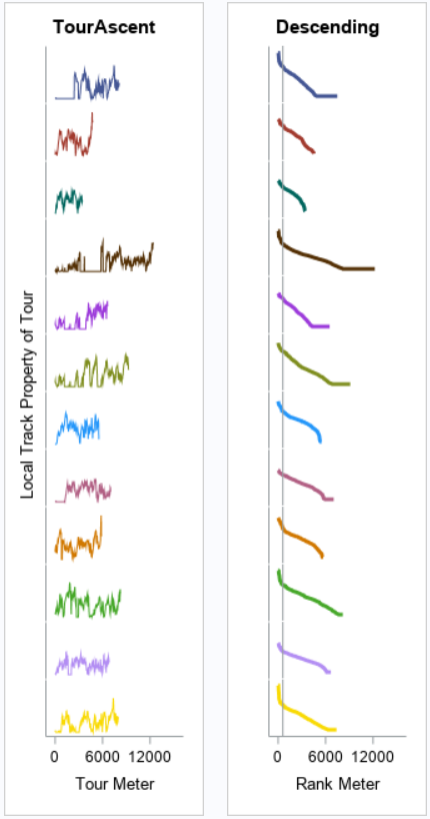
Data preparation

from local track properties to unique tour features



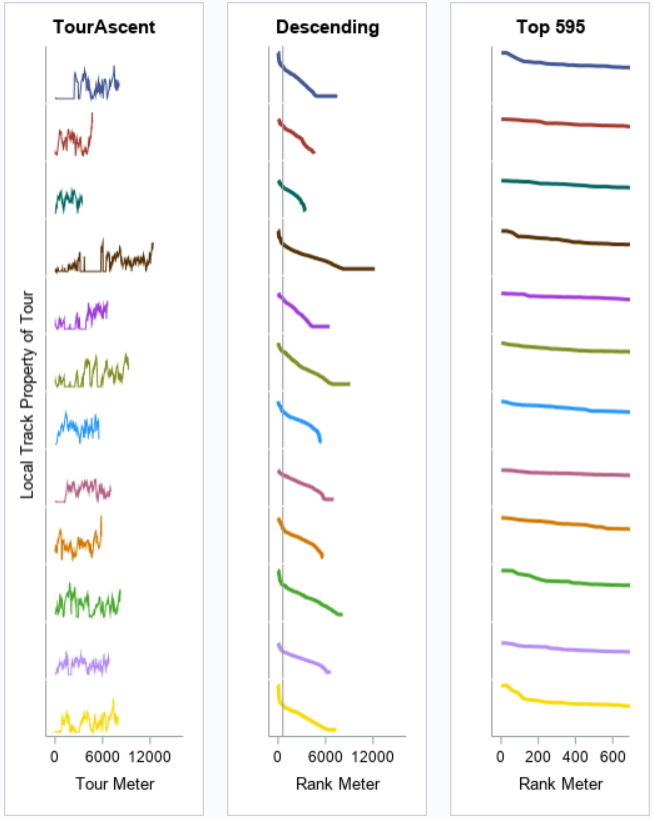
Data preparation

from local track properties to unique tour features



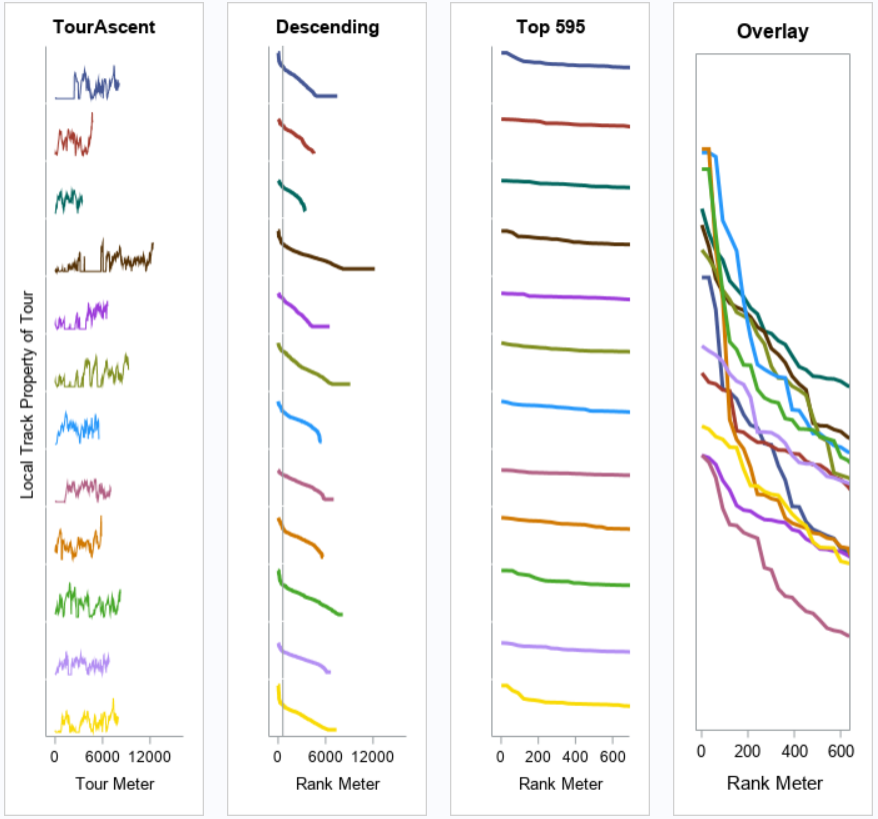
Data preparation

from local track properties to unique tour features



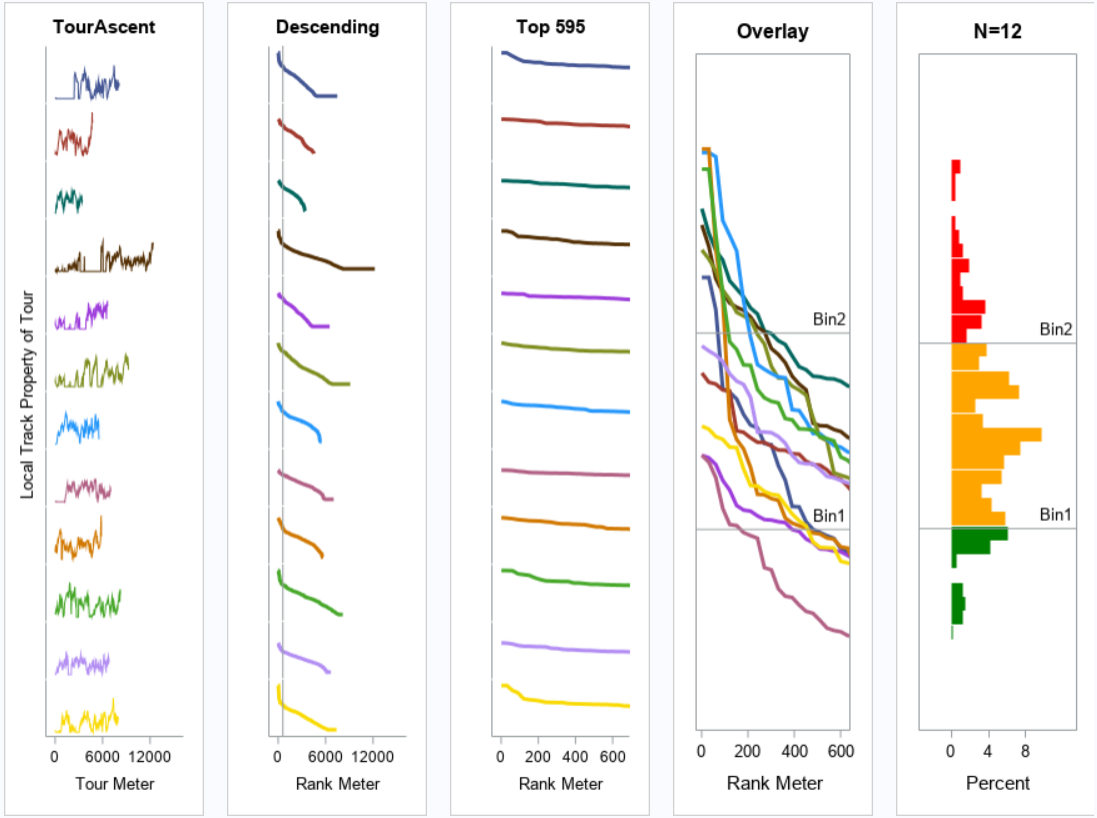
Data preparation

from local track properties to unique tour features



Data preparation

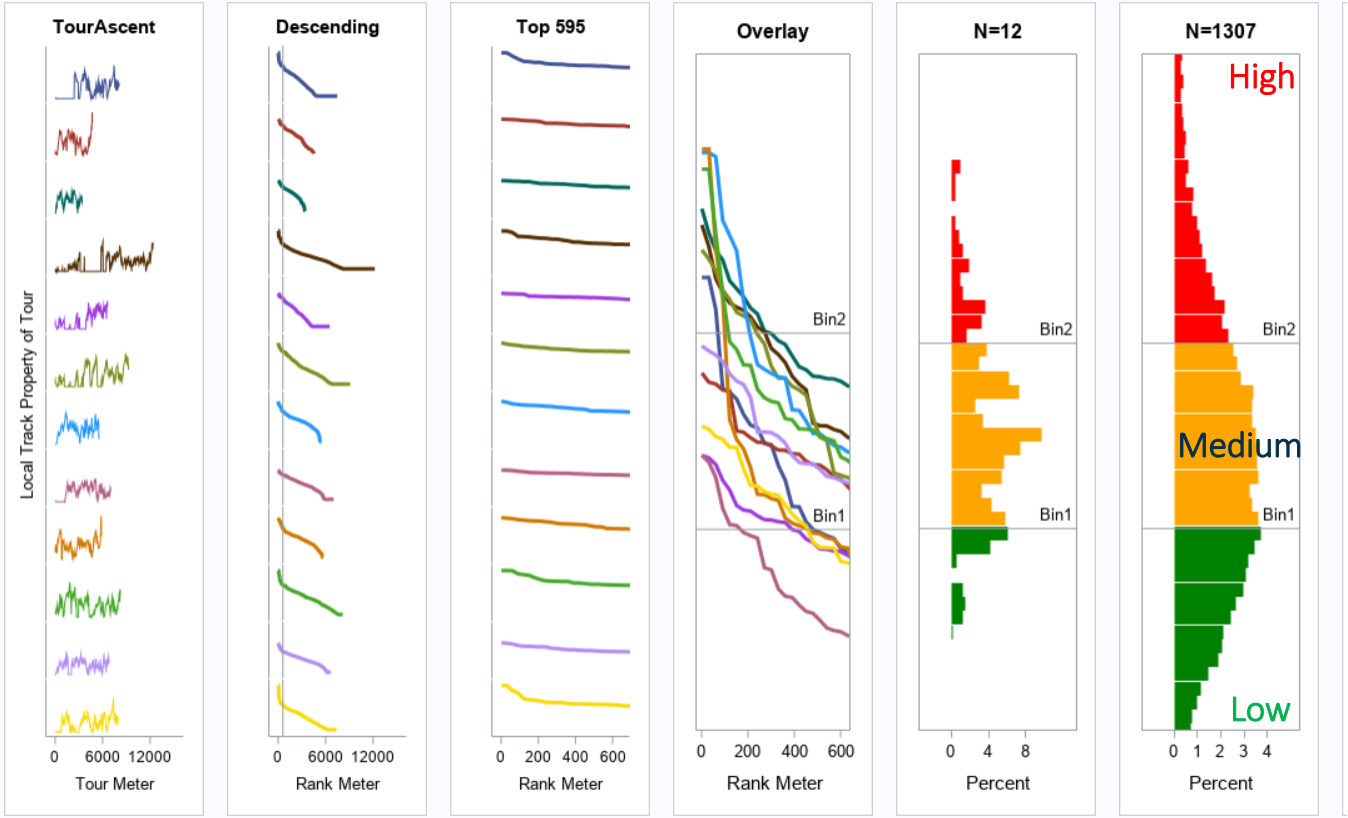
from local track properties to unique tour features



Local track properties of tours processed: Risk, Slope Angle, SpeedMax, Acceleration, Forestation, Curvature, Width,...

Data preparation

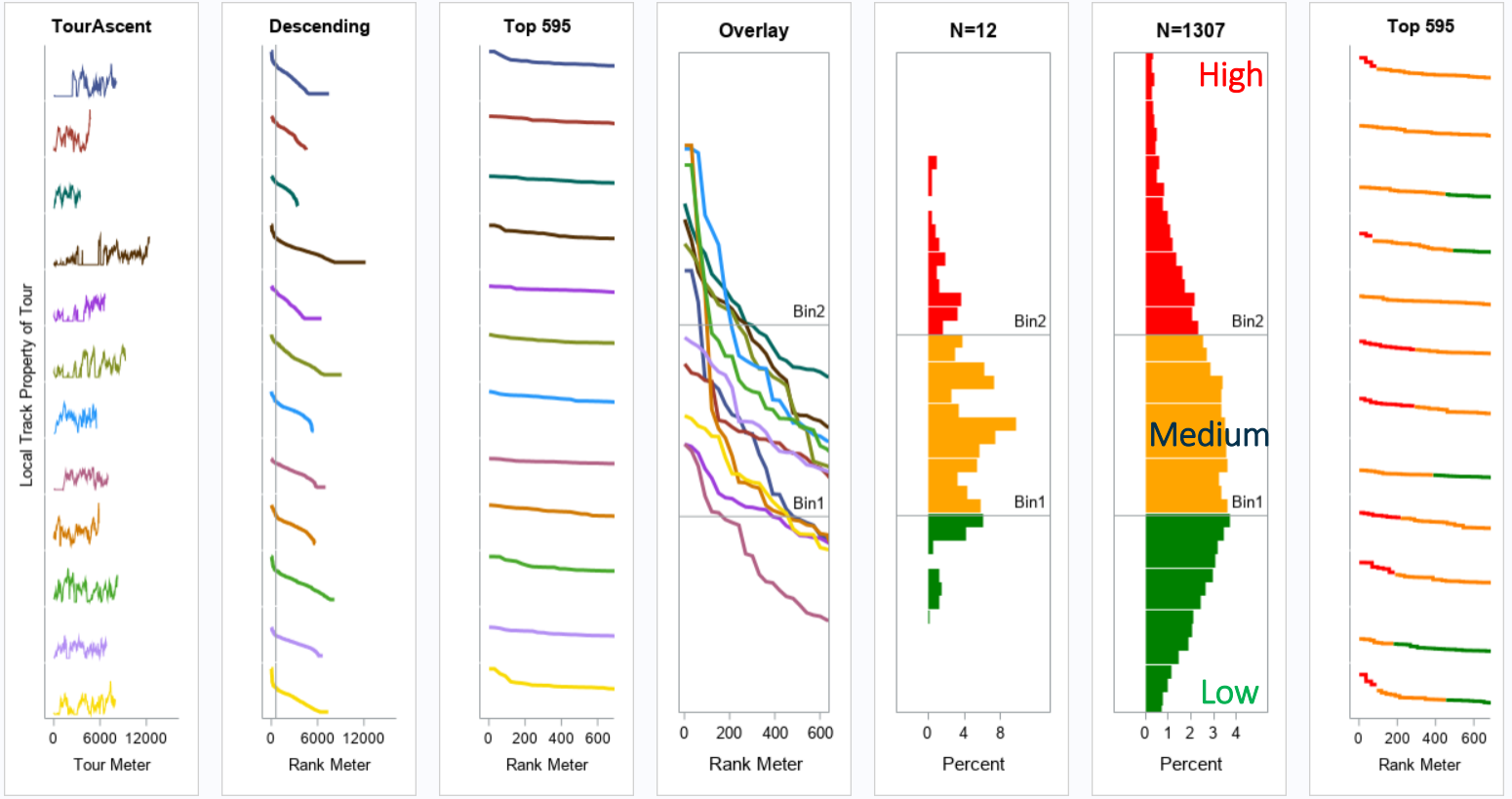
from local track properties to unique tour features



Local track properties of tours processed: Risk, Slope Angle, SpeedMax, Acceleration, Forestation, Curvature, Width,...

Data preparation

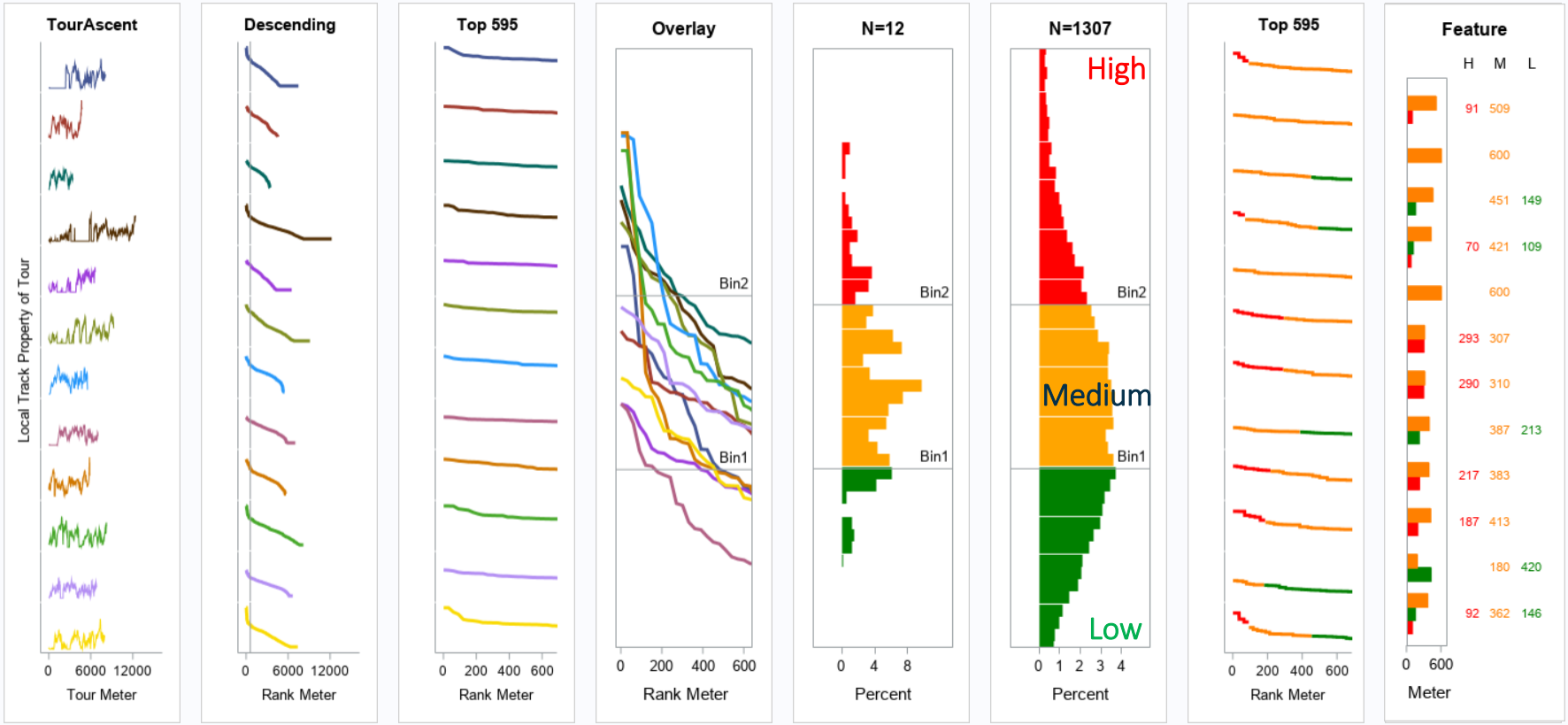
from local track properties to unique tour features



Local track properties of tours processed: Risk, Slope Angle, SpeedMax, Acceleration, Forestation, Curvature, Width,...

Data preparation

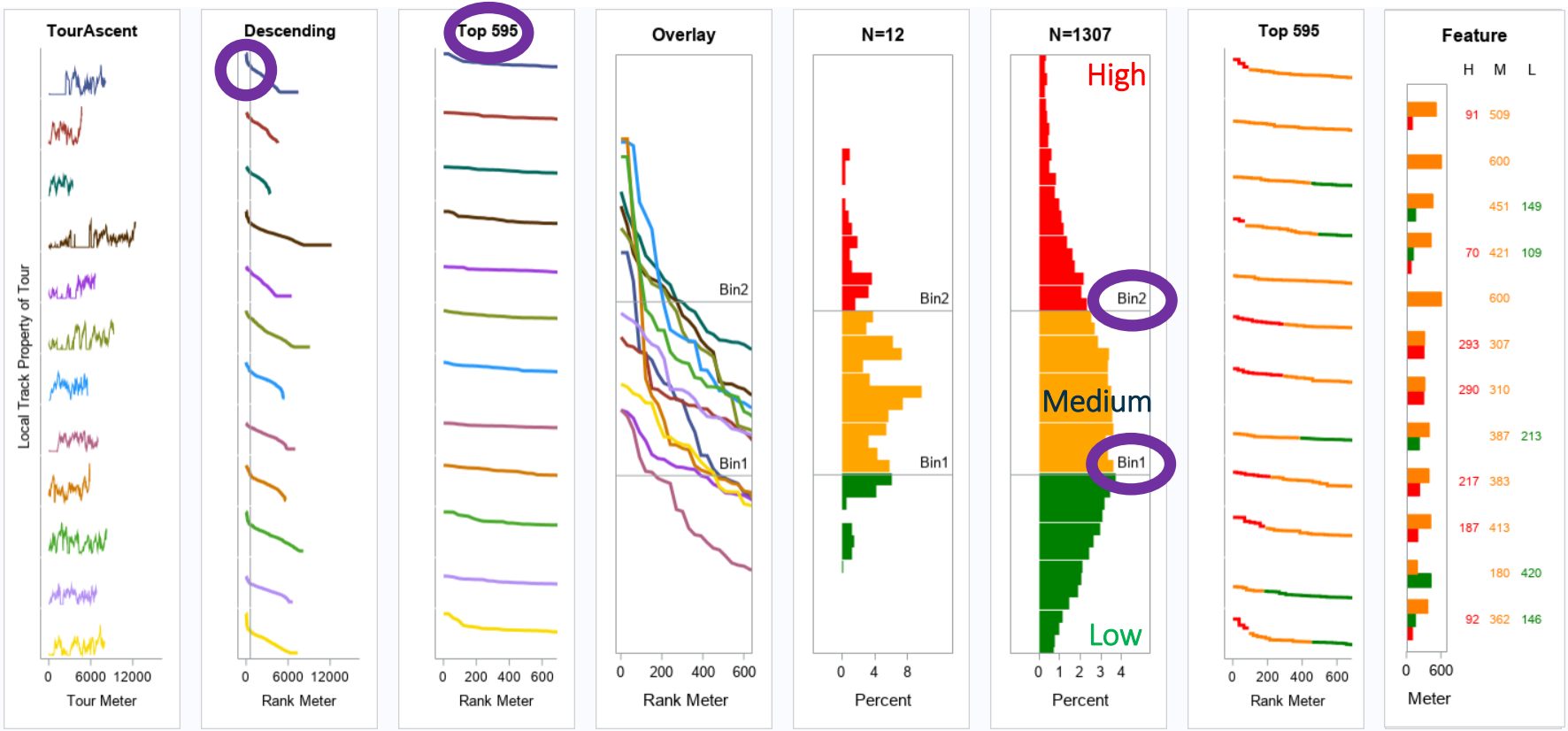
from local track properties to unique tour features



Local track properties of tours processed: Risk, Slope Angle, SpeedMax, Acceleration, Forestation, Curvature, Width,...

Data preparation

from local track properties to unique tour features



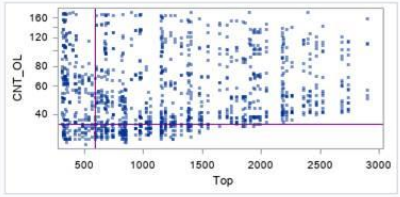
Local track properties of tours processed: Risk, Slope Angle, SpeedMax, Acceleration, Forestation, Curvature, Width,...

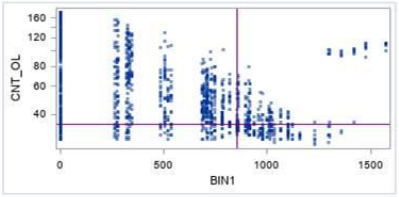
How to find good segmentation parameters: Top, Bin1, Bin2

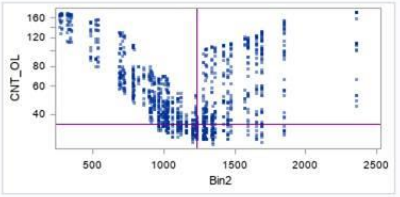
„Trial and Error“ minimizing Mean Absolute Prediction Error MAE

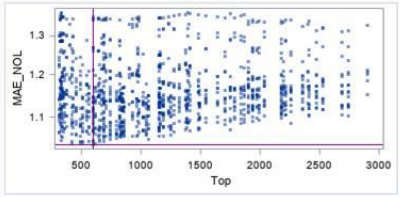
Optimal quantile regression model with best segmentation parameter TOP, BIN1, BIN2 (out of 5000 trials)

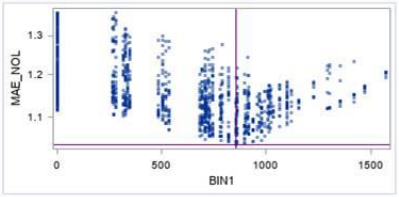
Top	BIN1	Bin2	Opt	MAE_all	MAE_NOL	CNT_OL	Intercept	RiskCnt_3	RiskCnt_3f	RiskCnt_2	RiskCnt_2f	SAC3_BEE_BEW_BVS_FRV	SAC3_TI	SAC3_ZS_GRN_GRS_GL_V
595	855	1229	*	1.12596	1.03841	35	1.83060	0.00866	0.00686	0.00464	.	-0.83060	0.67489	0

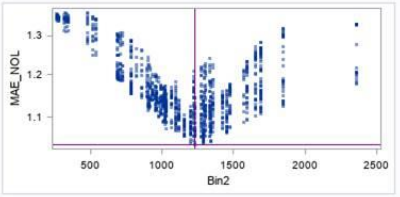


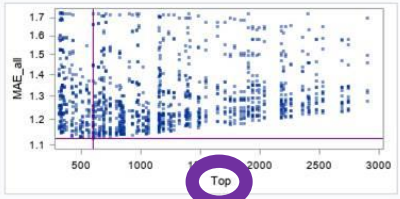


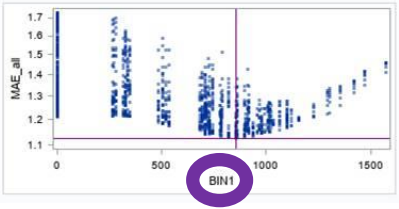


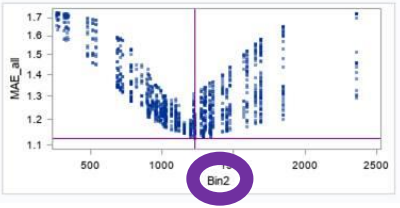






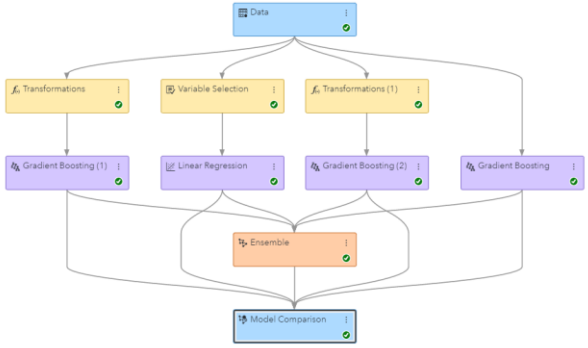


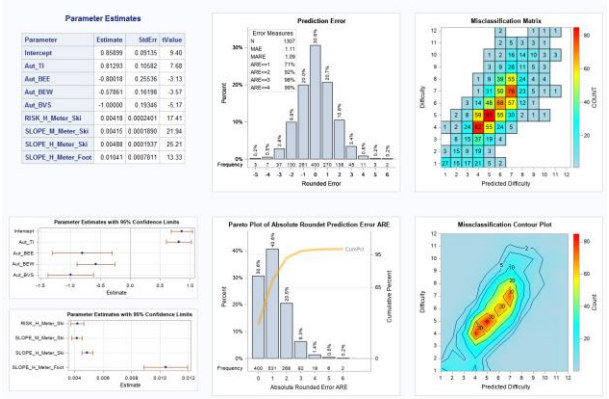




What predictive modeling approach did we take?

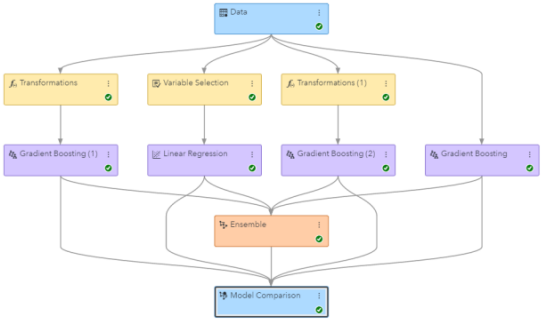
Machine Learning vs. Statistical Model

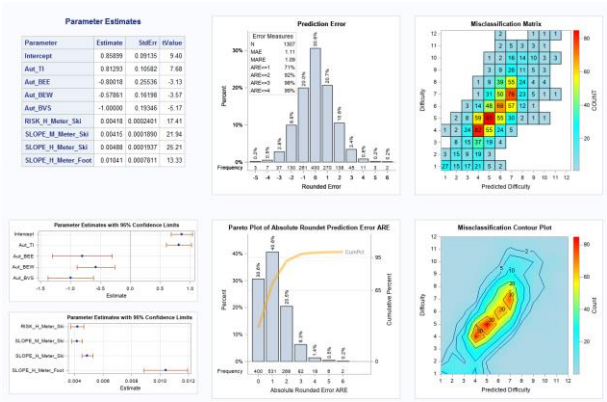


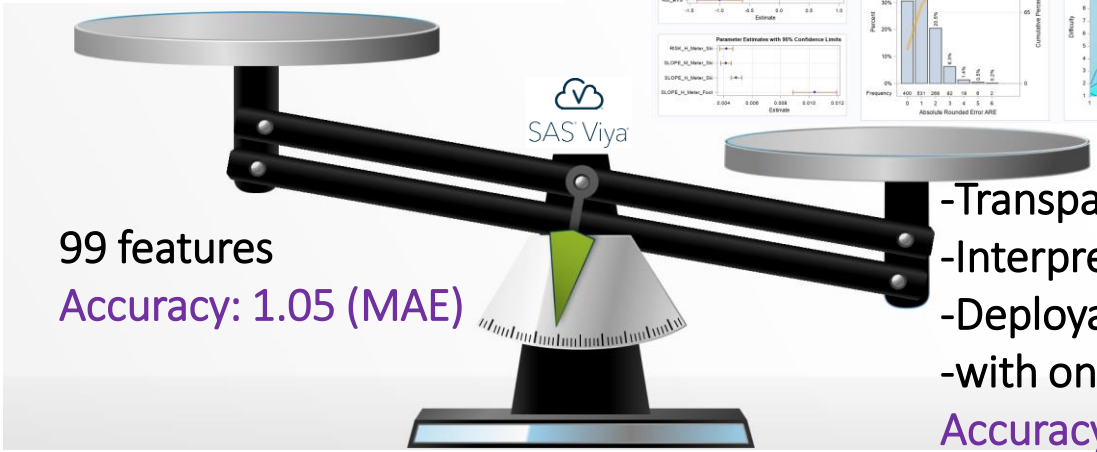


What Results did we get?

Transparency, Interpretability, Deployability outweighted Accuracy







99 features

Accuracy: 1.05 (MAE)

-Transparency

-Interpretability,

-Deployability

-with only 4 features

Accuracy: 1.11 (MAE)

Variable selection with quantile regression

For median of difficulty

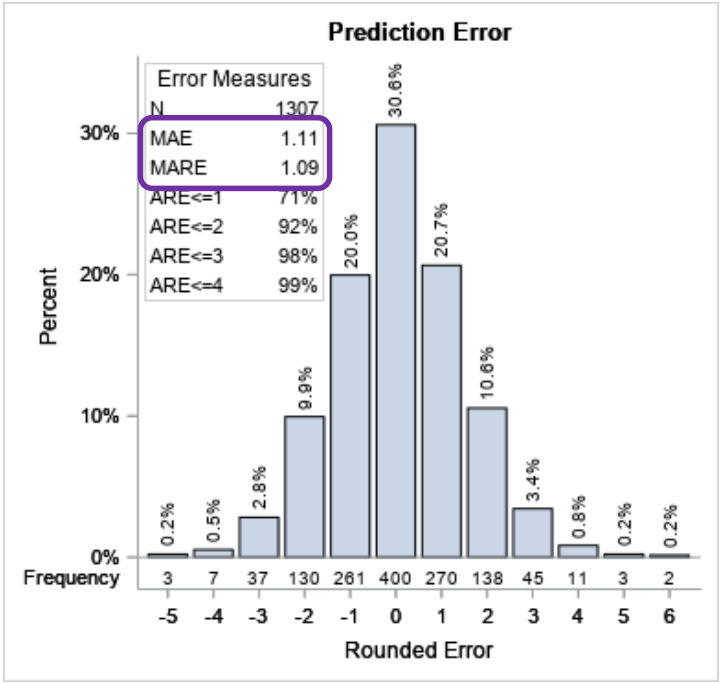
The HPQUANTSELECT Procedure
Quantile Level = 0.5

Selection Summary

Step	Effect Entered	Number Effects In	AIC	SBC
0	Intercept	1	-185.5716	-180.3961
1	RISK_H_Meter_Ski	2	-984.6951	-974.3442
2	SLOPE_H_Meter_Foot	3	-1216.9940	-1201.4675
3	Aut_BVS	4	-1246.9930	-1226.2911
4	Aut_BEW	5	-1277.7180	-1251.8405
5	SLOPE_H_Meter_Ski	6	-1307.2939	-1276.2410
6	SLOPE_M_Meter_Ski	7	-1484.6233	-1448.3949
7	Aut_TI	8	-1510.2989	-1468.8950
8	Aut_BEE	9	-1522.1704*	-1475.5910*

Fit Statistics

Objective Function	725.08515
R1	0.40396
Adj R1	0.40028
AIC	-1522.17042
AICC	-1522.03163
SBC	-1475.59101
ACL	0.55477

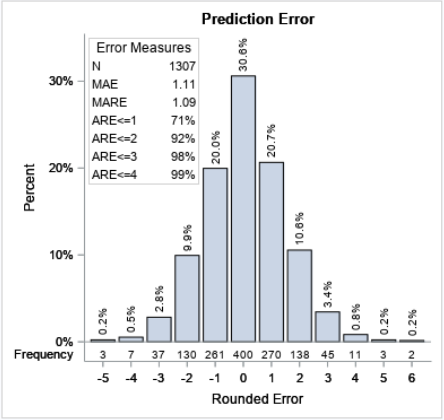


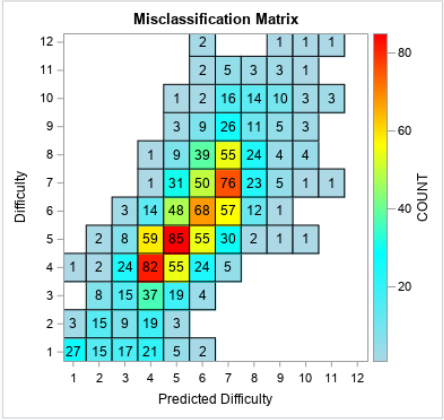
Risk:=SlopeAngle*SpeedMax

Method: Quantile Regression - PROC HPQUANTSELECT

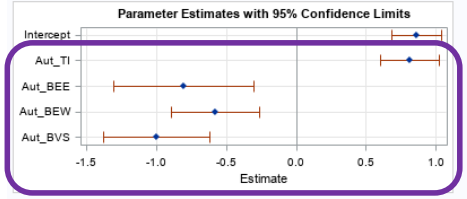
Parameter Estimates

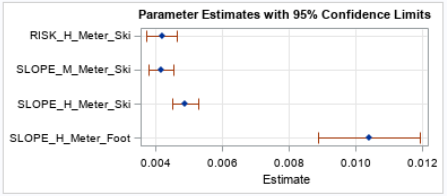
Parameter	Estimate	StdErr	tValue
Intercept	0.85899	0.09135	9.40
Aut_TI	0.81293	0.10582	7.68
Aut_BEE	-0.80018	0.25536	-3.13
Aut_BEW	-0.57861	0.16198	-3.57
Aut_BVS	-1.00000	0.19346	-5.17
RISK_H_Meter_Ski	0.00418	0.0002401	17.41
SLOPE_M_Meter_Ski	0.00415	0.0001890	21.94
SLOPE_H_Meter_Ski	0.00488	0.0001937	25.21
SLOPE_H_Meter_Foot	0.01041	0.0007811	13.33

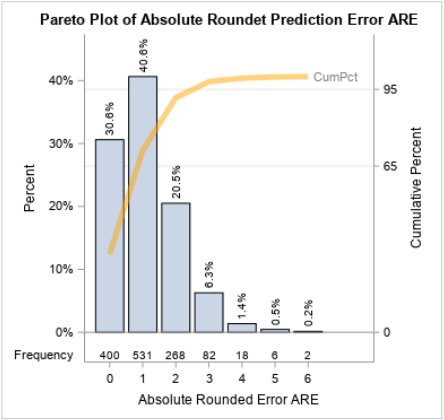


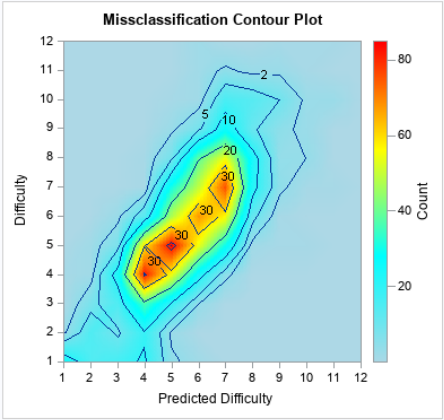


Four out of 12 selected author dummy variables







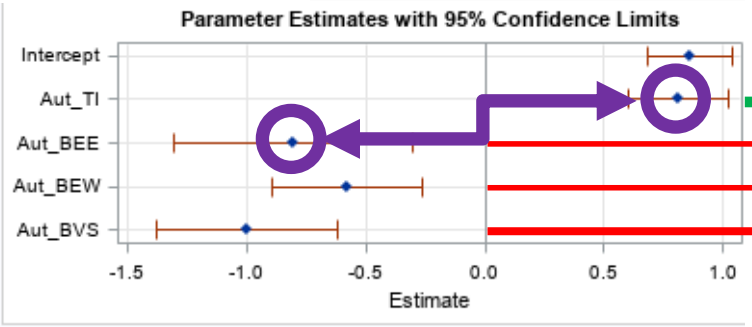


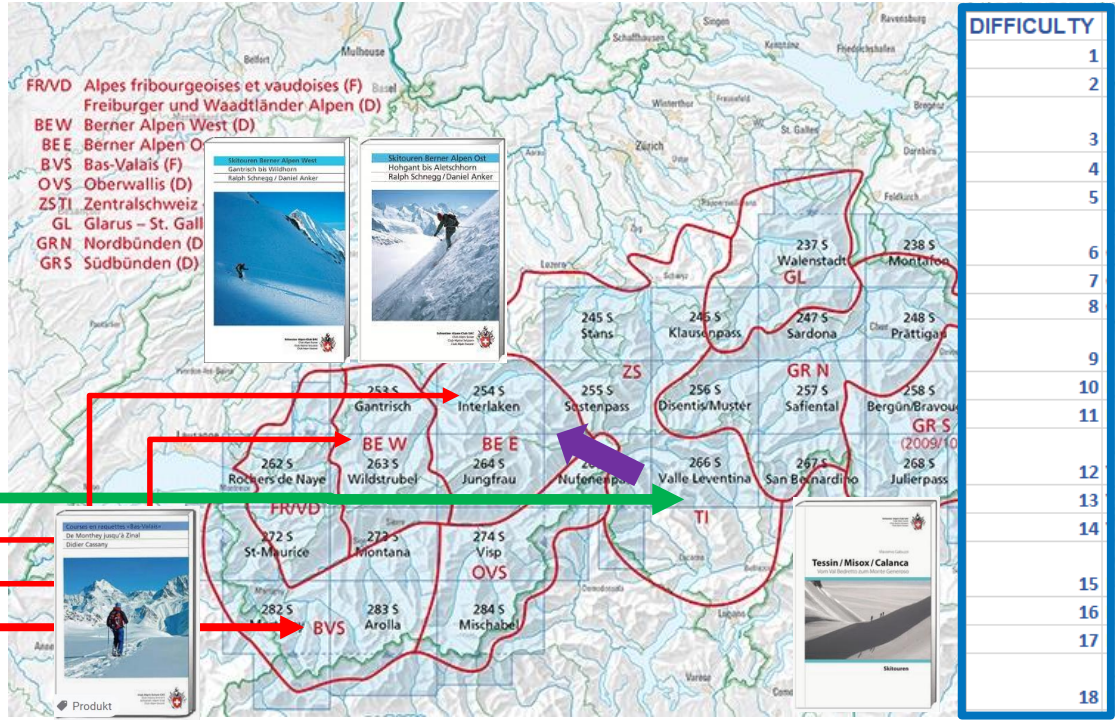
Significant author dummy variables

Systematic Overrating vs Underrating bias detected for difficulty

Error Measures	
N	1307
MAE	1.11
MARE	1.09
ARE<=1	71%
ARE<=2	92%
ARE<=3	98%
ARE<=4	99%

Parameter	Estimate	StdErr	tValue
Intercept	0.85899	0.09135	9.40
Aut_TI	0.81293	0.10582	7.68
Aut_BEE	-0.80018	0.25536	-3.13
Aut_BEW	-0.57861	0.16198	-3.57
Aut_BVS	-1.00000	0.19346	-5.17

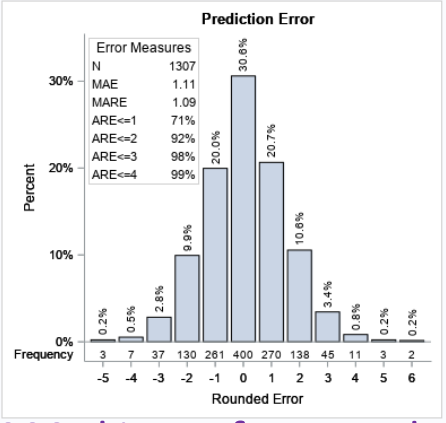


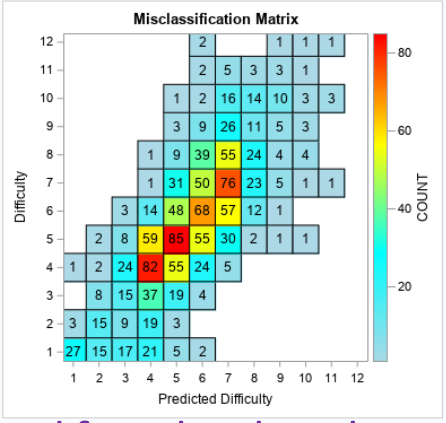


Method: Quantile Regression - PROC HPQUANTSELECT

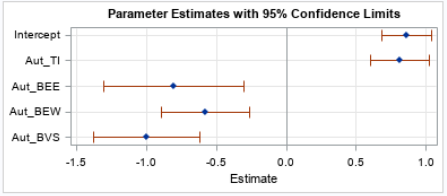
Parameter Estimates

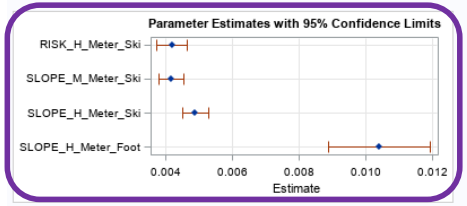
Parameter	Estimate	StdErr	tValue
Intercept	0.85899	0.09135	9.40
Aut_TI	0.81293	0.10582	7.68
Aut_BEE	-0.80018	0.25536	-3.13
Aut_BEW	-0.57861	0.16198	-3.57
Aut_BVS	-1.00000	0.19346	-5.17
RISK_H_Meter_Ski	0.00418	0.0002401	17.41
SLOPE_M_Meter_Ski	0.00415	0.0001890	21.94
SLOPE_H_Meter_Ski	0.00488	0.0001937	25.21
SLOPE_H_Meter_Foot	0.01041	0.0007811	13.33

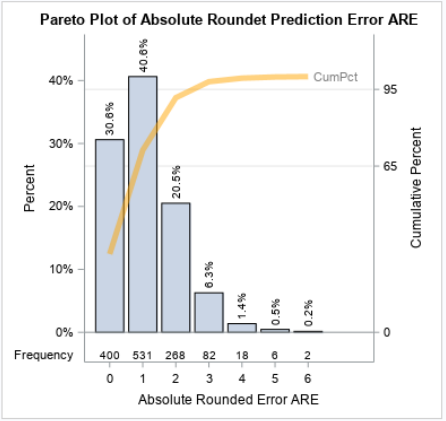


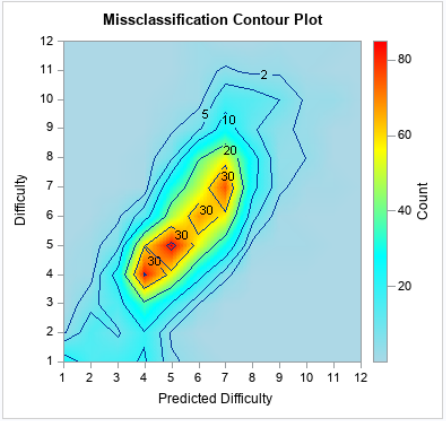


Selected four out of ~20 000 ski tour features derived from local track properties









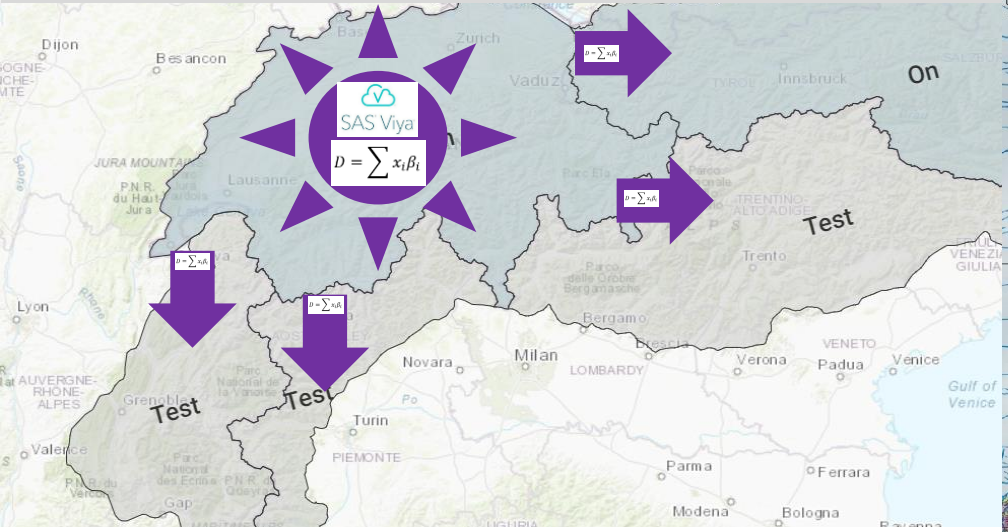
Model deployment to expand services of skitourenguru.ch

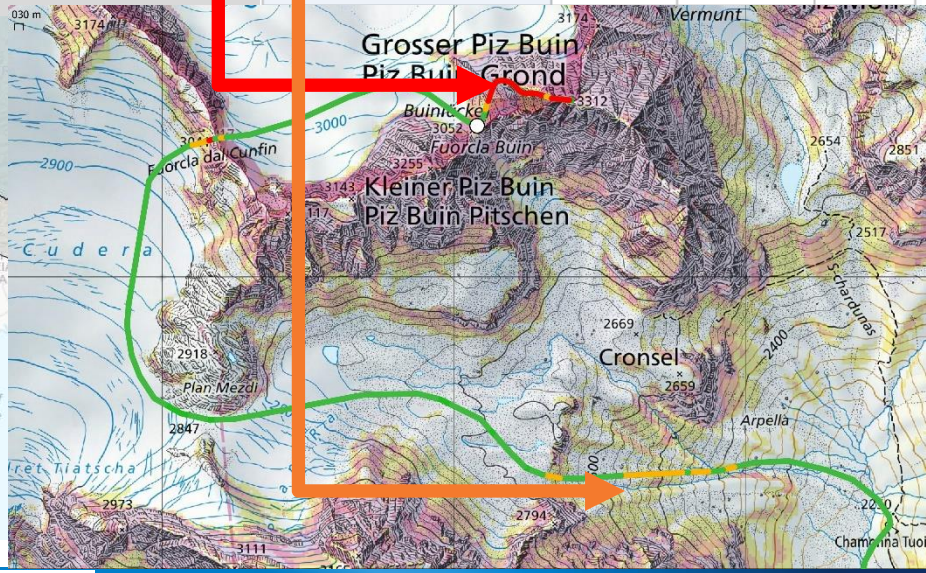
to 4 neighboring countries with ~10 000 additional unrated ski tours

SKITOURENGURU

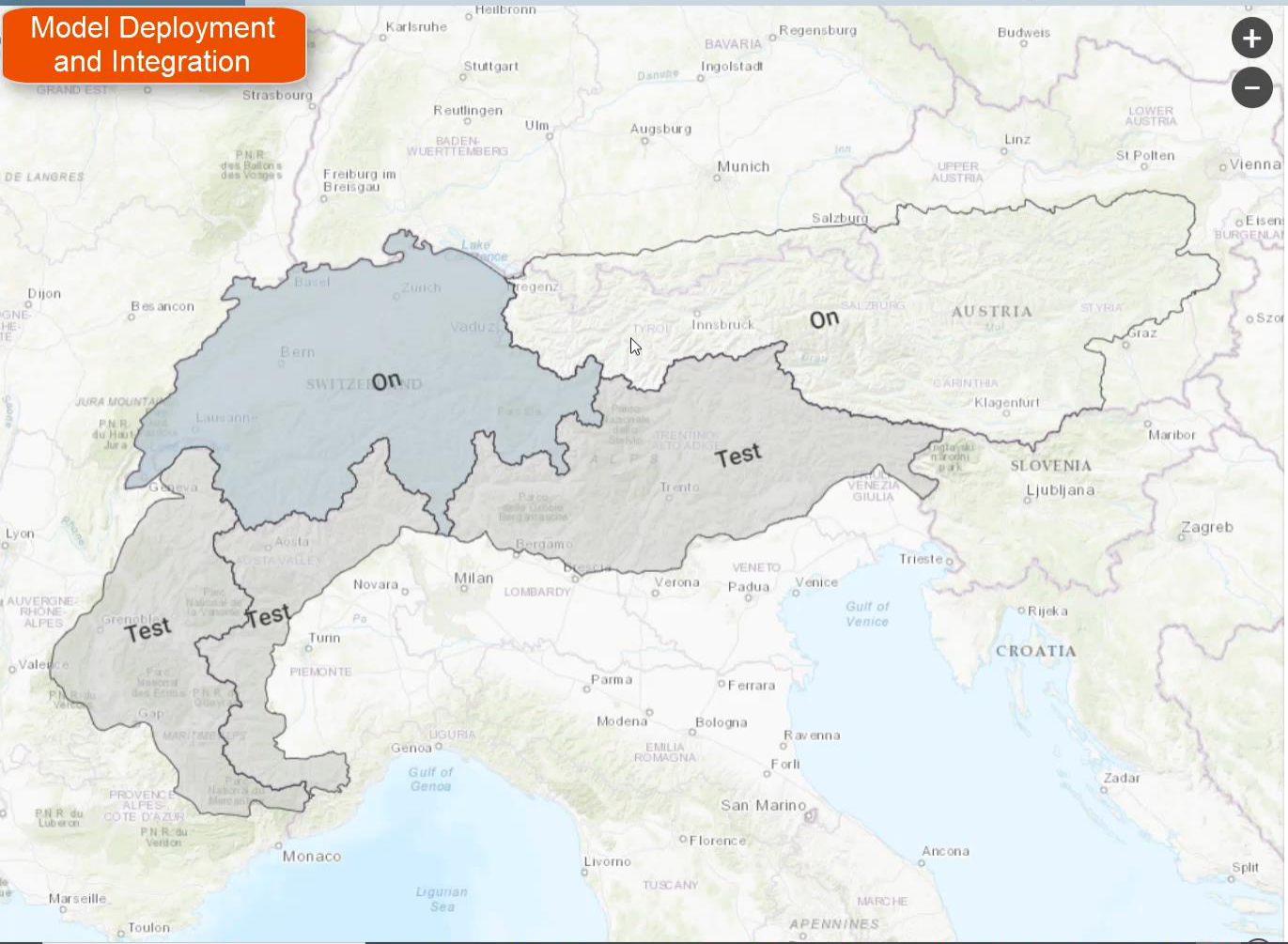
P_Difficulty=0.859+(418*RISK_H_Meter_Ski
+ 415*Slope_M_Meter_Ski
+ 488*Slope_H_Meter_Ski
+1041*Slope_H_Meter_Foot)/100000

Parameter	Estimate	StdErr	tValue
Intercept	0.85899	0.09135	9.40
RISK_H_Meter_Ski	0.00418	0.0002401	17.41
SLOPE_M_Meter_Ski	0.00415	0.0001890	21.94
SLOPE_H_Meter_Ski	0.00488	0.0001937	25.21





Model Deployment and Integration



What's Skitourenguru

Skitourenguru supports you in the selection and planning of a suitable ski tour with low avalanche risk. For this purpose, Skitourenguru assigns daily an avalanche risk to thousands of ski tours in the alpine region:

Symbol	Values	Avalanche risk
▲	0–1	Low avalanche risk
▼	1–2	Elevated avalanche risk
●	2–3	High avalanche risk

In addition Skitourenguru marks static route cruxes with grey rings:

Symbol	Class	Meaning
○	1	Avalanche terrain
◎	2	Typical avalanche terrain
◎	3	Very typical avalanche terrain

On site and in the individual slope usually information becomes accessible that is not available to Skitourenguru. The information presented on Skitourenguru is subject to uncertainties (see Handbook). Therefore Skitourenguru must not be the only criterion to access a slope.

Choose a region

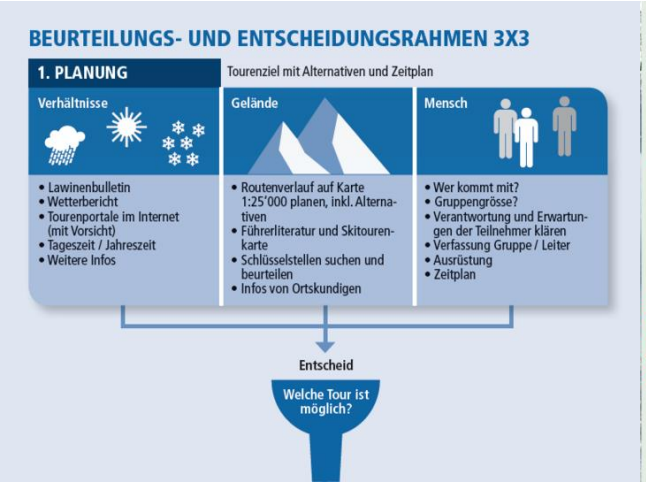
Region	State	Edition	Valid
Switzerland	On	17.30 h	16.4.2021–17.00 h
East Alps	On	18.30 h	16.4.2021–18.00 h
France	Test	16.30 h	16.4.2021–18.00 h
Nothwest-Italy	Test	16.30 h	16.4.2021–16.00 h
Notheast-Italy	Test	17.30 h	16.4.2021–16.00 h

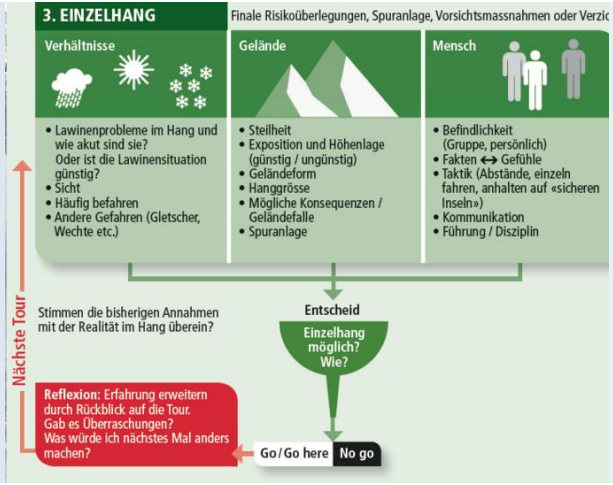
Partners

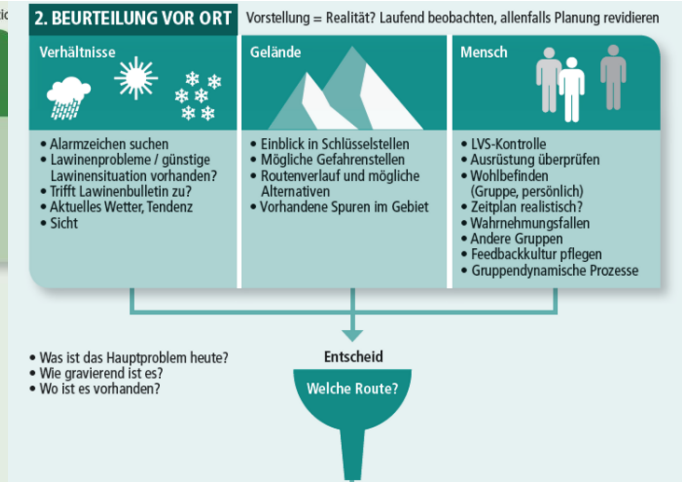
Skitourenguru is supported among others by the following partners

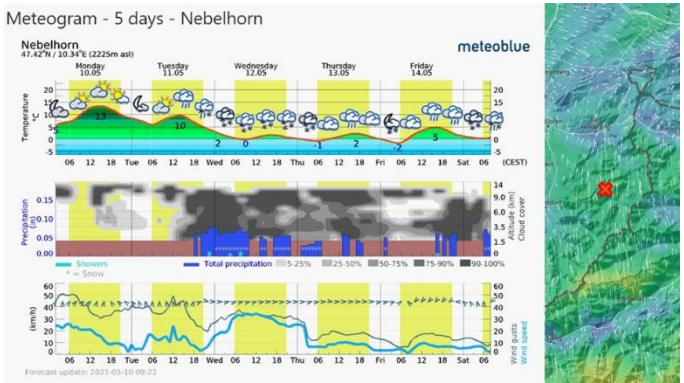
Off course, skitourenguru does not exempt you from

applying the recommended avalanche and risk assessment strategies



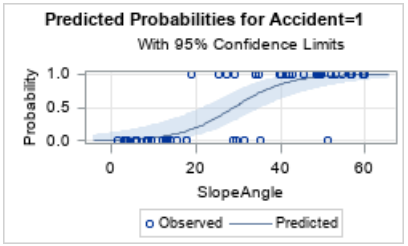




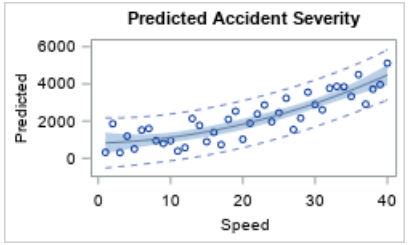


Takeaways:
What did we achieve?

Difficulty ~



X



DIFFICULTY	DIFFICULTY LABEL
1	Easy
2	Easy (+)
3	Less Difficult(-)
4	Less Difficult
5	Less Difficult (+)
6	Quite Difficult (-)
7	Quite Difficult
8	Quite Difficult (+)
9	Difficult (-)
10	Difficult
11	Difficult (+)
12	Very Difficult (-)
13	Very Difficult
14	Very Difficult (+)
15	Extremely Difficult (-)
16	Extremely Difficult
17	Extremely Difficult (+)
18	Extremely Difficult

- We are proposing a new definition of **difficulty** metric derived from interaction of two local track properties: **slope angel and speed of falling** acting as proxies for accident probability and severity
- Overall, this metric is consistent with the unique human and cultural expertise published in the extensive SAC ski touring literature from which our model was trained.
- The discretionary range of the SAC methodology and prediction error margin is in the range of 1.1 to 1.8 levels of the 18-step SAC difficulty scale (i.e. “+” or “-”)
- An additional advantage of this methodology is its independent from prevailing weather and snow conditions at the moment of manual rating.
- We still have ongoing discussions with incorporation of the foot section in this model.
- The model provides the basis for fast and automatic bulk scoring prediction for up to ~10000 tours throughout the alps in AT, DE, IT, FR. It will support the expansion of Skitourenguru’s services.

Günter Schmudlach, Skitourenguru GmbH, Zürich CH

Ulrich Reincke, Principal Data Scientist, SAS, Heidelberg, DE

Thank you for your attention. And don’t forget: Always put safety first

Outlyer list with absolute prediction error gt 3.5

Type=Overrating

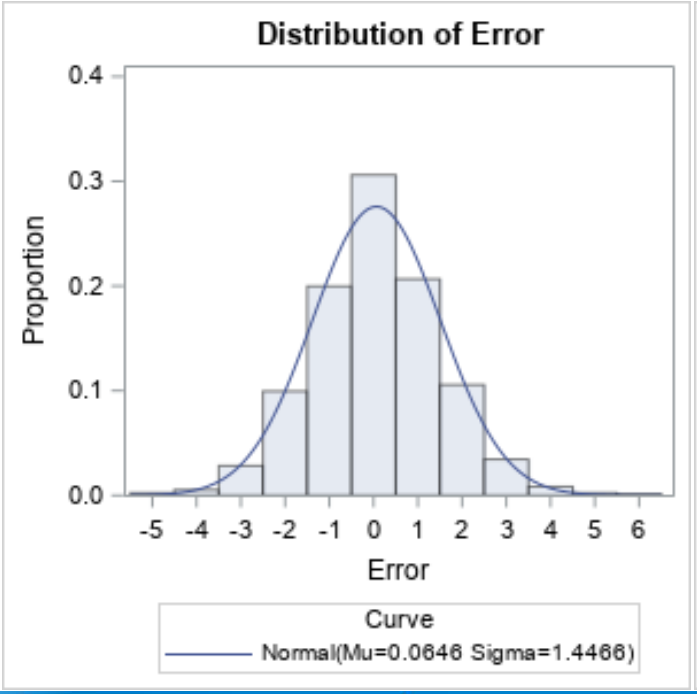
id_long	Difficulty	P	E	SAC0	Outlyer_Comment	StartEle	StopEle	Ele	RISK_H_Meter_Ski	SLOPE_H_Meter_Ski	SLOPE_M_Meter_Ski	SLOPE_H_Meter_Foot
1258_Hasen_Gotterli	1	5	-4	ZS		449	1394	945	0	748	0	0
171_Cons_PizTerri	7	11	-4	GRN		1468	3146	1789	595	748	0	360
564_Küblis_Chrüz	1	5	-4	GRN	Different Route	809	2190	1384	157	372	376	0
255_Furggels_Stelli	1	5	-4	GRN	Manual Underrat	1198	2047	976	147	643	105	0
912_Mühlebach_Ärnergale	1	5	-4	VSE	Different Route	1232	2621	1391	169	629	119	0
387_MittlerRossfal_Hochalp	1	5	-4	GL	Compromise	899	1527	650	234	264	484	0
535_Ladstafel_Mittaghorn	5	9	-4	VSE		1924	3004	1080	595	748	0	220
358_Latsch_CuolmdaLatsch	1	6	-5	GRS	Road above 1600	1609	2294	686	244	748	0	0
1035_HospizSimplonp_MonteLeone	5	10	-5	VSE		1998	3548	1657	508	698	50	290
1466_Sufers_VizanPintg	1	6	-5	GRN	Road above 1600	1413	2513	1120	423	748	0	0

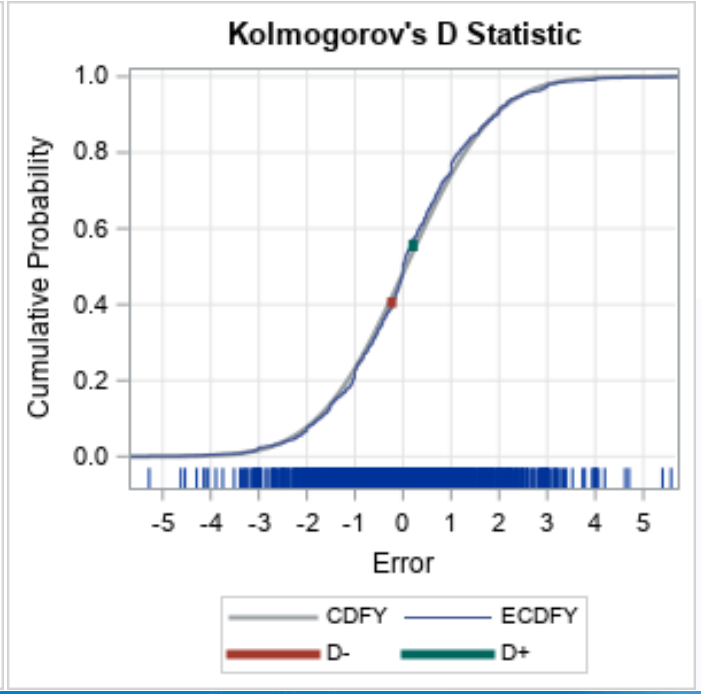
Type=Underrating

id_long	Difficulty	P	E	SAC0	Outlyer_Comment	StartEle	StopEle	Ele	RISK_H_Meter_Ski	SLOPE_H_Meter_Ski	SLOPE_M_Meter_Ski	SLOPE_H_Meter_Foot
903_MayensdeMérib_PointedeVouasso	12	6	6	BVS		1728	3481	1755	595	748	0	0
367_ZurEich_GrosBrun	12	6	6	BEW	Compromise	951	2098	1147	595	748	0	0
1231_Engi_Gufelstock	11	6	5	GL	Compromise	812	2434	1622	260	748	0	0
706_ChantSura_PizRadönt	10	5	5	GRS	Other Ski Depot	2330	3056	751	28	120	147	300
407_Urnerboden_Läckistock	11	6	5	ZS	Compromise	1376	2483	1107	455	697	51	0
725_Dürrboden_Leidhorn	9	5	4	GRS	Compromise	2006	2930	925	150	292	456	0
507_H.d'Allières_VanildesArtses	11	7	4	FRV	Other Ski Depot	1006	1986	980	127	0	707	250
613_Diavolezza_PizCambrena	11	7	4	GRS		2978	3595	855	595	748	0	0
736_Brigels_Bifertenstock	11	7	4	GRN		1285	3416	2173	595	748	0	0
818_Jochstock_ReissendNollen	11	7	4	ZS		2508	3002	493	595	748	0	0
886_LeFlon_Chambairy	10	6	4	BVS		1046	2198	1151	595	748	0	0
916_BourgSt.Berna_MontVélan	10	6	4	BVS		1916	3721	1805	595	748	0	0
1448_Münster_HejiZwächte	9	5	4	BEE	Compromise	1387	3083	1696	323	748	0	0
1236_Elm_Grüenenspitz	9	5	4	GL	Other Ski Depot	960	2354	1394	94	316	432	60
1227_Horb_Frümsel	11	7	4	GL	Other Ski Depot	887	2261	1374	22	113	635	300
591_Tschlin_Muttler	8	4	4	GRS	Other Ski Depot	1533	3290	1758	44	22	726	30

Prediction Residuals / Error

Test for normality (N=1307)





Kolmogorov's D Statistic

The UNIVARIATE Procedure
Fitted Normal Distribution for Error

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.03076998	Pr > D	0.168
Cramer-von Mises	W-Sq	0.33496573	Pr > W-Sq	0.110
Anderson-Darling	A-Sq	1.76279181	Pr > A-Sq	0.126

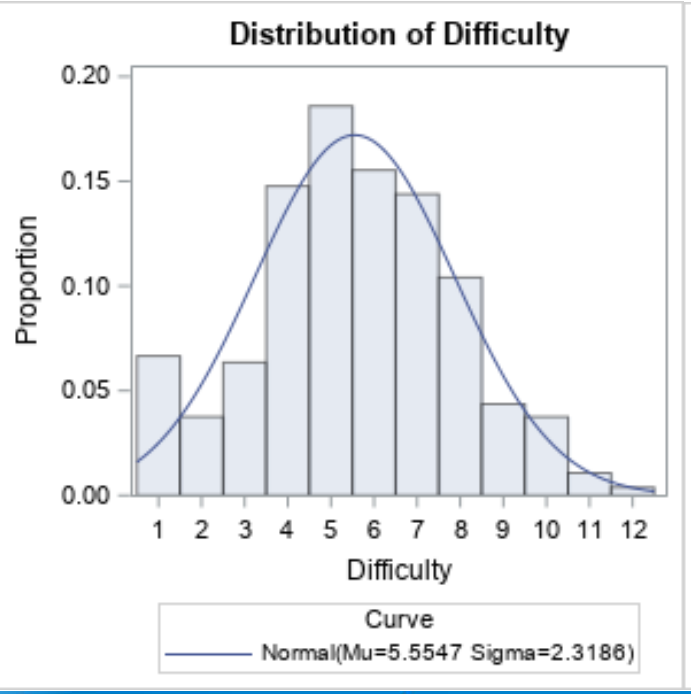
Kolmogorov's D Statistic

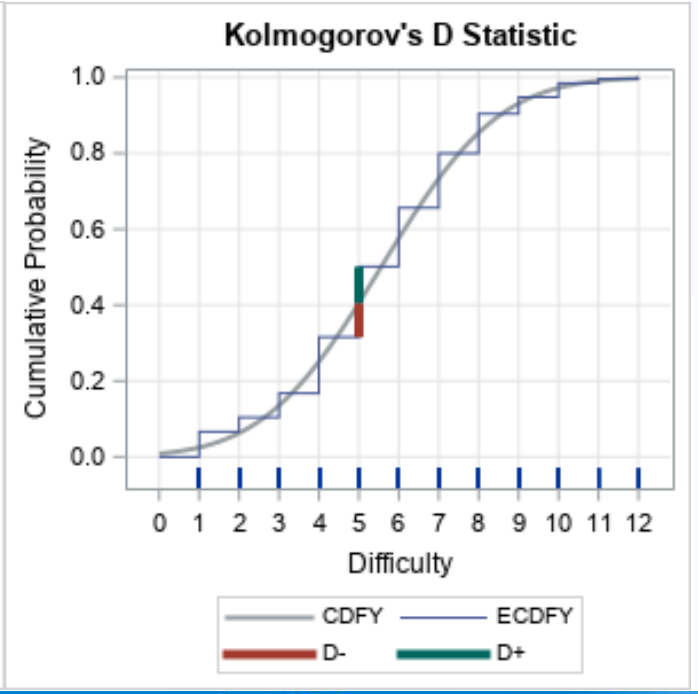
D
0.03077

Kolmogorov D				
	Error	Value	Low	High
D-	-0.232248	0.0269607	0.3917368	0.4186975
D+	0.2099216	0.03077	0.5400028	0.5707728

Target Variable Difficulty

Test for normality (N=1307)





Kolmogorov's D Statistic

The UNIVARIATE Procedure
Fitted Normal Distribution for Difficulty (diff)

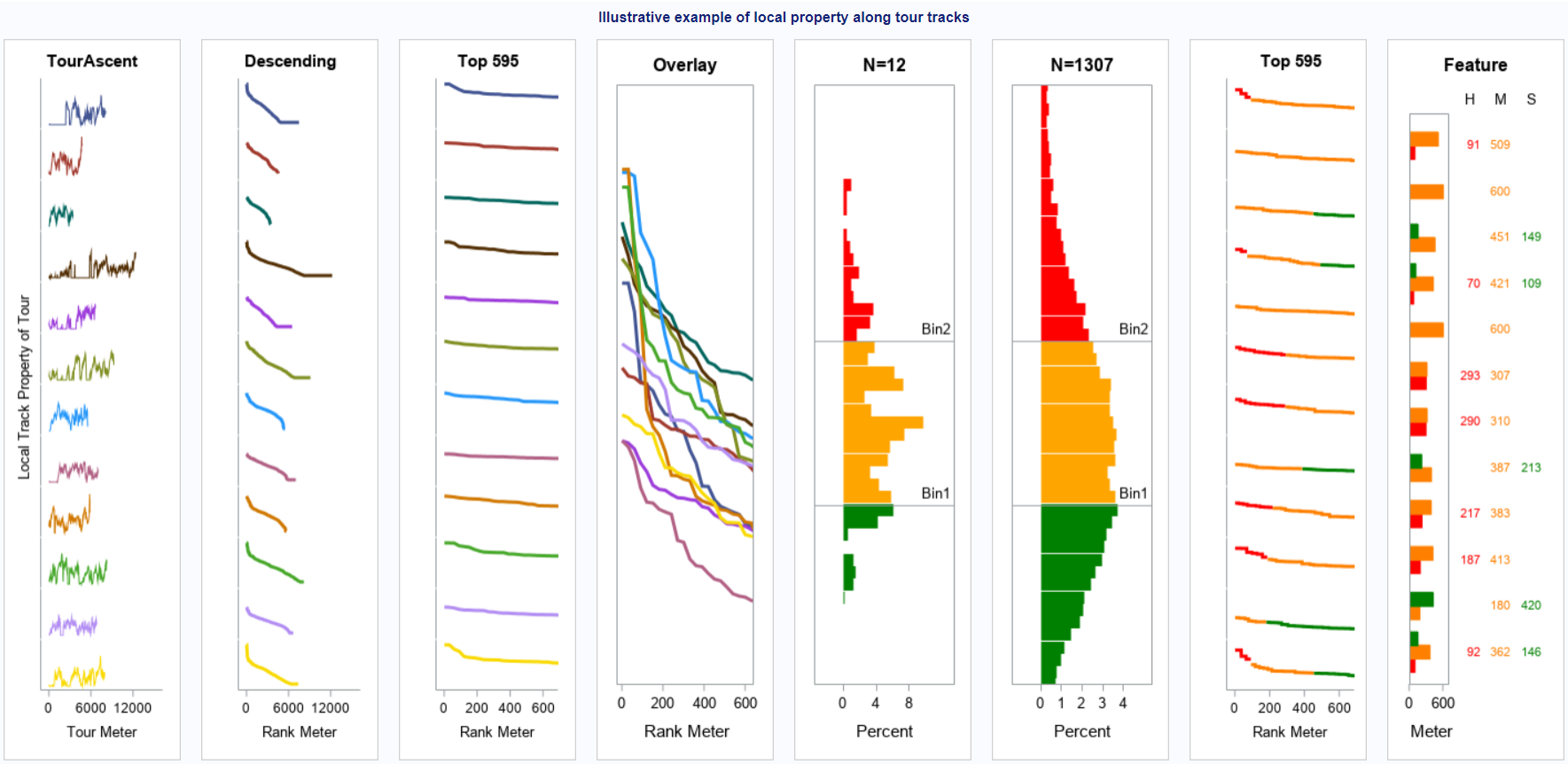
Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.0956892	Pr > D	<0.001
Cramer-von Mises	W-Sq	2.0483810	Pr > W-Sq	<0.001
Anderson-Darling	A-Sq	11.8013910	Pr > A-Sq	<0.001

Kolmogorov's D Statistic

D
0.0956892

Kolmogorov D				
	x	Value	Low	High
D-	5	0.0902327	0.3152257	0.4054584
D+	5	0.0956892	0.4054584	0.5011477

Data preparation: from properties to features



Local track properties of tours processed: Risk, Slope Angle, SpeedMax, Acceleration, Forestation, Curvature, Width,...