Forecasting tourism frequentation series using regional grouped time series

The case of the canton of Valais in Switzerland

Miriam Scaglione[(a)]

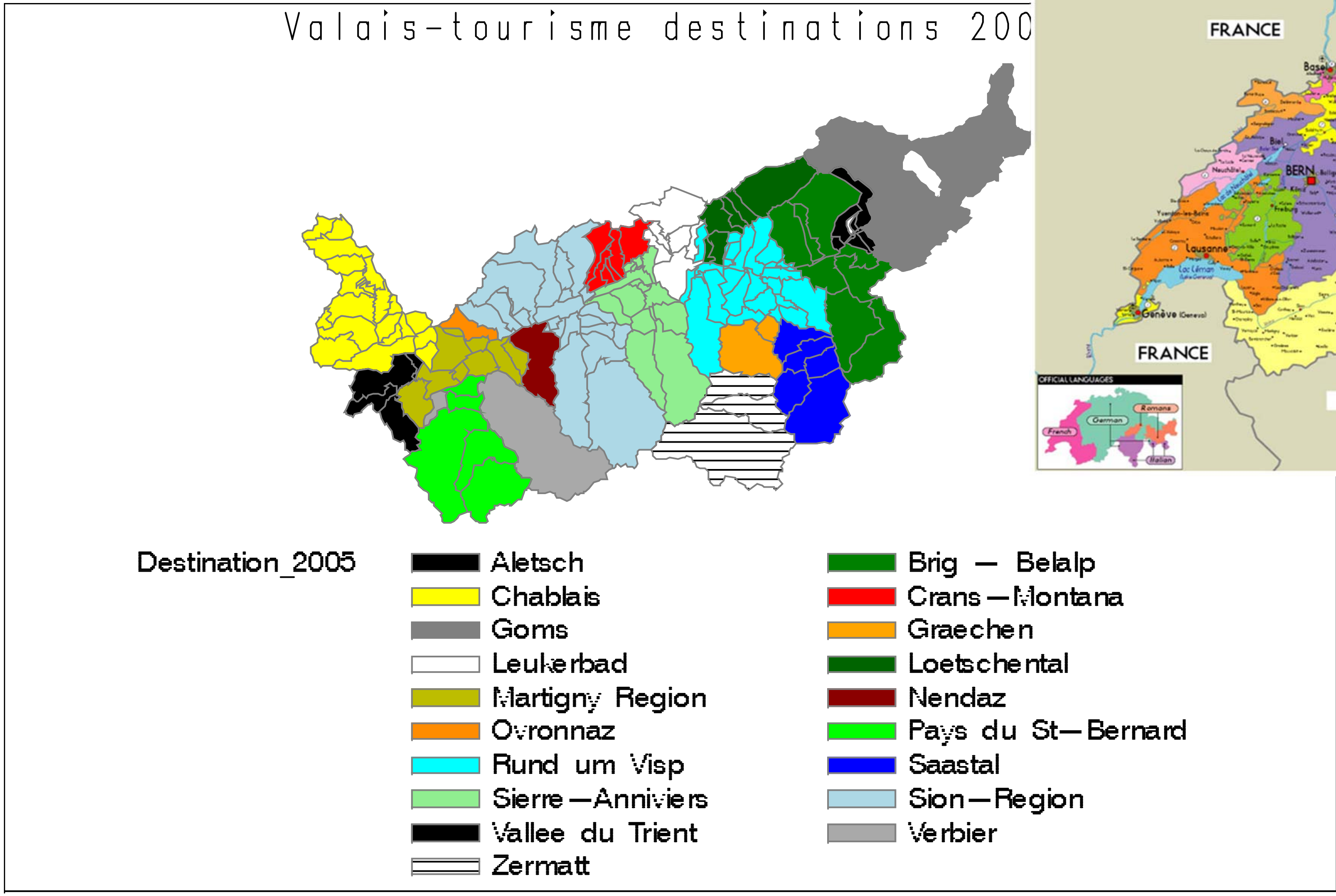
Michele Hibon[(b)]

Pascal Favre[(a)]

[(a)]University of Applied Sciences and Arts Western Switzerland Valais (HES-SO Valais-Wallis)-
Institute of Tourism Techno-Ark-Pôle sierre 3,
CH 3960 Sierre, Switzerland,
email: {miriam.scaglione, pascal.favre}@hevs.ch

[(b)]Michele Hibon
INSEAD, Boulevard de Constance, F77305 Fontainebleau, France.
email: michele.hibon@insead.edu

The 19 destinations in the Valais



Valais tourism organization

- Tourism sector in the canton of Valais is organized in 19 destinations management organization (DMO). Some of them are very well know, like the so called big five : Zermatt, Verbier, Leukerbad, Saastall et Crans-Montana.
- The canton is divided in three different regions: Bas Valais, Haut Valais and Valais Romand.
- The lowest level of aggregation for the overnights statistics is the town level.

The research problem

Our task consists on the forecast of overnigths.

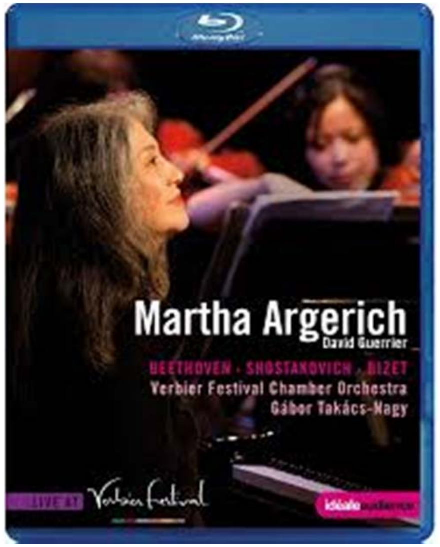
- The areas of forecast interests are aggregated levels such as destinations and cantonal level.
- The question is whether the use of the aggregated or hierarchical methods will yield better forecasting accuracy than the forecasting at the level of destinations and/or canton.

Rationale of the research question (I)

Better representation of the impact on overnights of local events at the aggregated level (destination or towns).

Examples:

- Omega European Masters (golf)
- Patrouille de Glacier
- Verbier festival
- etc.



Rationale of the research question (II)

Better representation behavior of the demand due to their taste or the schedule of their (winter) holidays.

Domestic market: two facts

The 26 cantons set school vacation independently but try to stagger them as much as possible.

1. There are moving-holidays as the date of beginning is linked to the end Carnival season.
2. It is not very clear which ski resort is preferred by the different Swiss cantons.

It is not possible or very complicated to treat this facts using intervention variables.

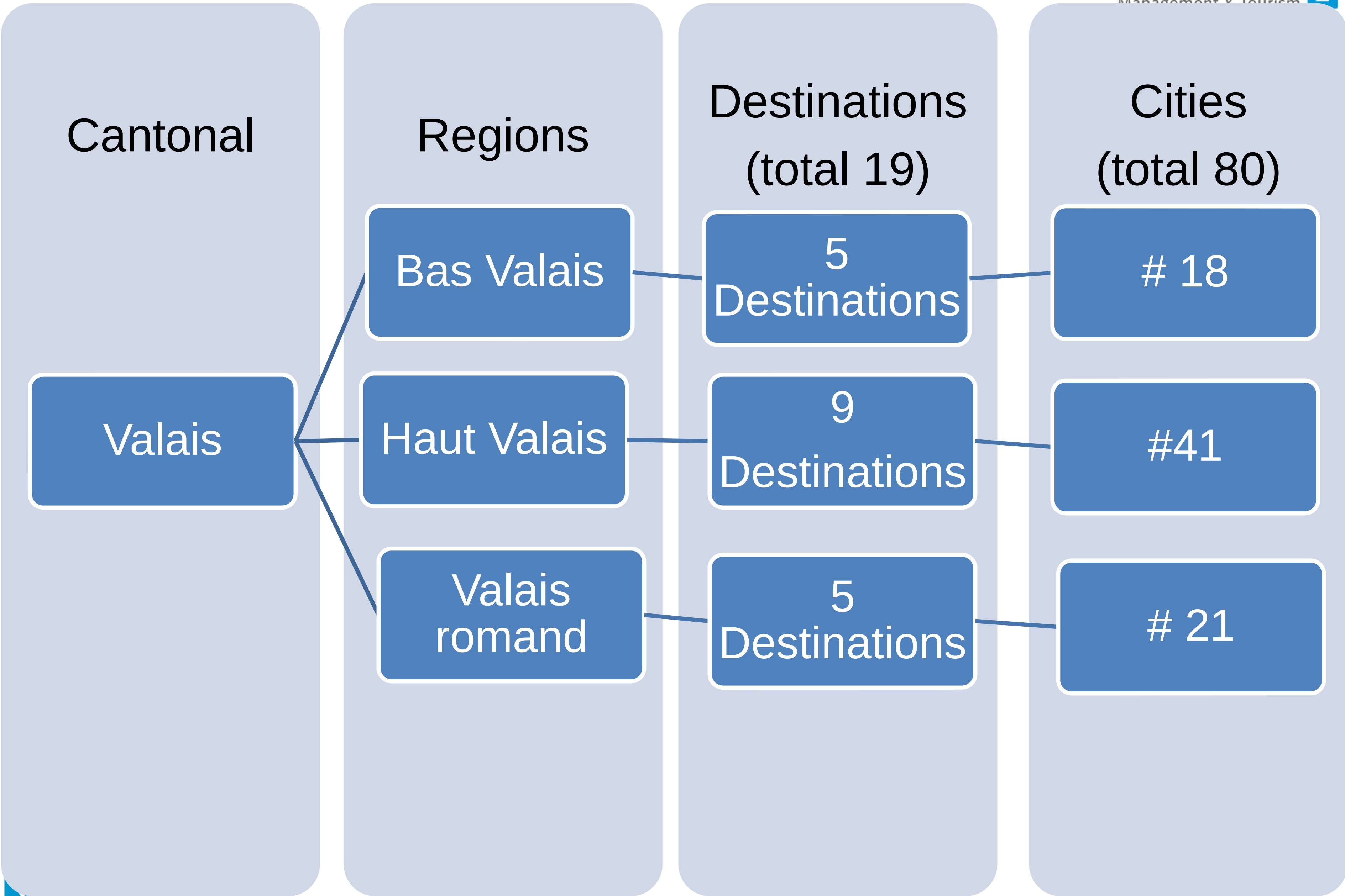
Up to some extent, the same considerations apply to the foreign demand.

Research questions

The accuracy of the forecast can be affected by:

- The depth of the tree or hierarchy
- The kind of forecasting method (ARIMA, random walk or ets)
- The method used for reconciling the base forecasts (Optimal combination forecasts, bottom-up, top-down, combination, Middle-out forecasts).

The data (I)



Cantonal
Regions
Destinations (total 19)
Cities (total 80)
Valais
Bas Valais
Haut Valais
Valais romand
5 Destinations
9 Destinations
5 Destinations
18
#41
21

The data (II)

- The source of the data is **Swiss Federal Statistical Office** in Neuchâtel
- The series are monthly and count Hotels' Overnights from January 2005 to March 2014, 111 observations.
- The series are download at the city level for the Valais. The numbers of series was greater than the final 80 used in this research. The authors had to aggregate some towns given their political merge during the period under study or for confidentiality issues.
- Finally the set is composed of 80 series of overnigths at the city level.

Methodology (I) : software

The software used is R's times series facilities, more specifically, the package hts (hierarchical time series).

Details

R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos and H.L. Shang (2011) Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, **55**(9), 2579–2589.http://robjhyndman.com/papers/hierarchical/

https://www.otexts.org/fpp/9/4

As a benchmark the series at the aggregate level of canton, regions and destinations have been forecasted using Stamp software, using 2 intervention variables, one for Easter and the other for Ascension and Whitsunday.

Methodology (II): hierarchies

Different trees or hierarchies having different depths (1, 2 and 3) were tested.

1. **Full tree**: depth of the tree is 3 (cf. slide “The data”); called model “Total”
2. **Depth 2** : Cantonal (root) ->destinations->cities; called model “destinations”; and

   Cantonal (root)->Region->cities ;

   called mode, “Regions”.
3. **Depth 1**: Cantonal (root)->cities; called model “city”; .

Methodology (III): forecasting model

The forecasting models used for the base forecasts are the ones allowed in R and the hts package.

1. Exponential smoothing state space model (ets)
2. ARIMA
3. Random walk (rw)

Methodology (III): forecasting model (cont')

The models were run before the middle of June 2014 et re-run after that date.

Rob Hyndman has made a number of changes on the forecasting algorithms.

Moreover, in order to take into account the Easter and Ascension and Pentecost, two intervention variables, used in the benchmark estimation (STS) were included in the latter re-running process for ARIMA.

From hereafter, the earlier process will be named "Without" and the later "With".

Methodology (IV): reconciling methods for the base forecasts

We obtained 2’280 forecasts for each of the following conciliation methods, in total 13’348, for the benchmark only 103. The available methods in the hts package are the following :

1. Optimal combination forecasts (comb)
2. Bottom-up forecasts (bu)
3. Bottom-up forecasts Middle-out forecasts where the level used is specified by the level argument (mo)
4. Top-down forecasts based on the average historical proportions (Gross-Sohl method A) (tdgsa)
5. Top-down forecasts based on the proportion of historical averages (Gross-Sohl method F) (tdgsf)
6. Top-down forecasts using forecast proportions (tdfp).

Methodology (V): Accuracy assessment

Hes·so // VALAIS WALLIS
School of Management & Tourism

RMSE Root Mean Squared Error $$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(F_t - A_t)^2}$$	**MAE Mean Absolute Error** $$MAE = \frac{1}{n}\sum_{t=1}^{n}\lvert F_t - A_t \rvert$$	**MAPE Mean Absolute Percentage Error** $$\text{MAPE} = \frac{\sum_{t=1}^{n}\left\lvert \frac{A_t - F_t}{A_t} \right\rvert}{n} \times 100$$
MPE Mean Percentage Error $$MPE = \frac{\sum_{t=1}^{n}\frac{A_t - F_t}{A_t}}{n} \times 100$$	MASE Mean Absolute Scaled Error $$\text{MASE} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\lvert A_t - F_t \rvert}{\frac{1}{T-m}\sum_{i=m+1}^{T}\lvert A_i - A_{i-m} \rvert}\right)$$	

A scaled error is less than one if it arises from a better forecast than the average naïve forecast computed on the training data. Conversely, it is greater than one if the forecast is worse than the average naïve forecast computed on the training data.

Σ π ≈ &

Valais excellence

Methodology (VI):Error calculi

Except for the benchmark, all the errors were calculated using hts package.

For the benchmark, a customerized routine was programmed in SAS Institute V9.4.

The graphs and tables of this presentation and the significant tests were calculated using Sphinx V5.

Warning: The ANOVA test has only an informatif value because the sample composed by hts methods and the benchmark is unbalanced (hts #13'348 vs STS #103)

Accuracy leaves = towns

Errors Towns

Error measure

	ME		RMSE		MAE		MAPE		MPE		MASE	
	Mean	Frequency	Mean	Frequency	Mean	Frequency	Mean	Frequency	Mean	Frequency	Mean	Frequency
arima/With	-262.18	1920	1021.99	1920	827.42	1920	88.68	1752	-66.31	1752	4.85	1872
arima/Without	-202.18	1920	966.19	1920	764.37	1920	81.61	1752	-53.28	1752	4.60	1872
ets/With	-368.15	1920	1037.33	1920	833.13	1920	84.08	1752	-60.14	1752	5.10	1872
ets/Without	-377.15	1921	1056.11	1921	848.64	1921	86.24	1753	-59.56	1753	5.16	1873
rw/With	261.41	1920	2189.97	1920	1879.77	1920	219.45	1778	-174.76	1778	7.04	1872
rw/Without	328.53	1960	2521.67	1960	2175.23	1960	223.98	1820	-177.86	1820	7.20	1912
sts/bench	1014.78	80	3667.45	80	3170.77	80	3371.82	80	-208.36	80	4.05	80
Total	-94.14	11641	1484.27	11641	1238.07	11641	155.74	10687	-100.16	10687	5.65	11353

	arima /With	arima/W ithout	ets/With	ets/Wi thout	rw/With	rw/Wi thout	sts/b ench	**Total**
MASE>1	1450	1425	1481	1480	1747	1800	69	9452
MASE<1	470	495	439	441	173	160	11	2189
Total	1920	1920	1920	1921	1920	1960	80	11641

p = **0.0%** ; chi2 = **411.45** ; dof = **6** (**VS**)

Arima outperforms all the others. But in general the performance is not as good as we would have liked to.

Arima and ets tend to overestimate (ME<0) whereas the contrary happens with RW and STS (ME>0).

Accuracy Valais

rror measure

	ME		RMSE		MAE		MAPE		MPE		MASE	
	Mean	Frequency	Mean	Frequency	Mean	Frequency	Mean	Frequency	Mean	Frequency	Mean	Frequency
arima/With	-20974.14	24	34499.63	24	29217.86	2[illegible]	9.34	24	-7.41	24	1.16	24
arima/Without	-16174.21	24	27587.62	24	21482.97	2[illegible]	7.05	24	-4.72	24	0.85	24
ets/With	-29451.73	24	38394.41	24	31609.15	2[illegible]	10.18	24	-9.21	24	1.25	24
ets/Without	-28881.31	23	37967.60	23	31182.87	2[illegible]	10.05	23	-9.01	23	1.23	23
rw/With	20912.89	24	138827.52	24	126143.56	2[illegible]	52.33	24	-21.33	24	4.99	24
rw/Without	22997.00	22	143390.18	22	130401.80	2[illegible]	54.25	22	-21.91	22	5.16	22
sts/bench	-26043.88	1	33676.85	1	28549.83	[illegible]	8.24	1	-6.40	1	1.38	1
Total	-9020.23	142	69048.84	142	60686.48	14[illegible]	23.42	142	-12.11	142	2.40	142

	arima /With	arima/Without	ets/With	ets/Without	rw/With	rw/Without	sts/bench	**Total**
MASE>1	13	1	24	22	24	22	1	107
MASE<1	11	23	0	1	0	0	0	35
Total	24	24	24	23	24	22	1	142

p = **0.0%** ; chi2 = **99.61** ; dof = **6** (**VS**)

		bench	bu	comb	mo	tdfp	tdgsa	tdgsf
ME	Mean	-26043.88	-6619.81	-9032.60	-7673.58	-10010.33	-10010.33	-10010.33
	Frequency	1	24	22	23	24	24	24
RMSE	Mean	33676.85	67818.22	68030.28	64725.17	71642.73	71642.73	71642.73
	Frequency	1	24	22	23	24	24	24
MAE	Mean	28549.83	59763.68	59817.79	56861.05	62927.87	62927.87	62927.87
	Frequency	1	24	22	23	24	24	24
MAPE	Mean	8.24	23.44	23.03	22.12	24.17	24.17	24.17
	Frequency	1	24	22	23	24	24	24
MPE	Mean	-6.40	-11.89	-11.89	-11.27	-12.60	-12.60	-12.60
	Frequency	1	24	22	23	24	24	24
MASE	Mean	1.38	2.36	2.37	2.25	2.49	2.49	2.49
	Frequency	1	24	22	23	24	24	24

ARIMA without interventions seems to outperform all the others, but it follows by the benchmark.

The conciliation methods do not yield a significant test, giving the impression that accuracy is related only to forecast method.

Accuracy Region (HV, VC; VR)

Error measure

	ME		RMSE		MAE		MAPE		MPE		MASE	
	Mean	Frequency	Mean	Frequency	Mean	Frequenc	Mean	Frequency	Mean	Frequency	Mean	Frequency
arima/With	-7628.53	36	11271.01	36	8953.21	3[illegible]	8.65	36	-6.21	36	0.88	36
arima/Without	-5227.02	36	10714.69	36	8280.65	3[illegible]	8.87	36	-3.56	36	0.84	36
ets/With	-9936.67	36	13844.10	36	11039.95	3[illegible]	12.24	36	-9.67	36	1.22	36
ets/Without	-9973.91	36	13878.66	36	11075.15	3[illegible]	12.36	36	-9.81	36	1.23	36
rw/With	7665.67	36	48179.62	36	43550.12	3[illegible]	54.09	36	-22.20	36	4.13	36
rw/Without	7665.67	36	48179.62	36	43550.12	3[illegible]	54.09	36	-22.20	36	4.13	36
sts/bench	-6934.85	3	11385.11	3	9395.49	[illegible]	7.68	3	-1.76	3	1.07	3
Total	**-2960.99**	**219**	**24167.09**	**219**	**20914.88**	**21[illegible]**	**24.81**	**219**	**-12.13**	**219**	**2.06**	**219**

	arima /With	arima/Without	ets/With	ets/Without	rw/With	rw/Without	sts/bench	**Total**
MASE>1	6	4	21	21	36	36	2	126
MASE<1	30	32	15	15	0	0	1	93
Total	36	36	36	36	36	36	3	219

p = **0.0%** ; chi2 = **109.63** ; dof = **6** (**VS**)

		bench	bu	comb	mo	tdfp	tdgsa	tdgsf
ME	Mean	-6934.85	-2206.60	-2892.87	-2325.01	-3336.78	-3336.78	-3336.78
	Frequency	3	36	36	36	36	36	36
RMSE	Mean	11385.11	23727.17	24257.76	23831.14	24757.38	24746.09	24748.16
	Frequency	3	36	36	36	36	36	36
MAE	Mean	9395.49	20682.43	21045.87	20660.88	21410.22	21340.77	21309.03
	Frequency	3	36	36	36	36	36	36
MAPE	Mean	7.68	23.98	24.82	24.56	25.34	25.71	25.89
	Frequency	3	36	36	36	36	36	36
MPE	Mean	-1.76	-10.66	-12.05	-11.66	-12.78	-13.07	-13.44
	Frequency	3	36	36	36	36	36	36
MASE	Mean	1.07	1.96	2.07	2.04	2.13	2.11	2.11
	Frequency	3	36	36	36	36	36	36

STS seems to outperform all the others, but it follows by the ARIMA with interventions.

The conciliation methods do not yield a significant test, giving the impression that accuracy is related only to forecast method.

Accuracy Destinations

Error measure

	ME		RMSE		MAE		MAPE		MPE		MASE	
	Mean	Frequency	Mean	Frequency	Mean	Frequency	Mean	Frequency	Mean	Frequency	Mean	Frequency
arima/With	-1187.83	228	3185.79	228	2572.29	228	35.99	228	-26.27	228	3.60	228
arima/Without	-812.01	228	3206.97	228	2523.98	228	35.04	228	-22.12	228	3.47	228
ets/With	-1516.73	228	3449.46	228	2748.91	228	37.57	228	-28.31	228	3.80	228
ets/Without	-1463.87	228	3432.25	228	2732.23	228	37.72	228	-27.88	228	3.81	228
rw/With	1210.37	228	8694.86	228	7559.22	228	135.14	228	-97.74	228	6.95	228
rw/Without	1210.37	190	8698.17	190	7568.98	190	134.96	190	-98.27	190	6.96	190
sts/bench	-746.71	19	2215.35	19	1716.26	19	19.46	19	-3.94	19	1.14	19
Total	-477.24	1349	4969.42	1349	4155.57	1349	66.85	1349	-48.09	1349	4.65	1349

	arima /With	arima/Without	ets/With	ets/Without	rw/With	rw/Without	sts/bench	**Total**
MASE>1	133	130	151	149	228	190	12	993
MASE<1	95	98	77	79	0	0	7	356
Total	228	228	228	228	228	190	19	1349

p = **0.0%** ; chi2 = **225.03** ; dof = **6** (**VS**)

STS seems to outperform all the others, but it follows by both ARIMAs.

The conciliation methods yield a significant tests, benchmark shows significant better performances (RMSE, MSE,MAPE), follows by mo and comb.

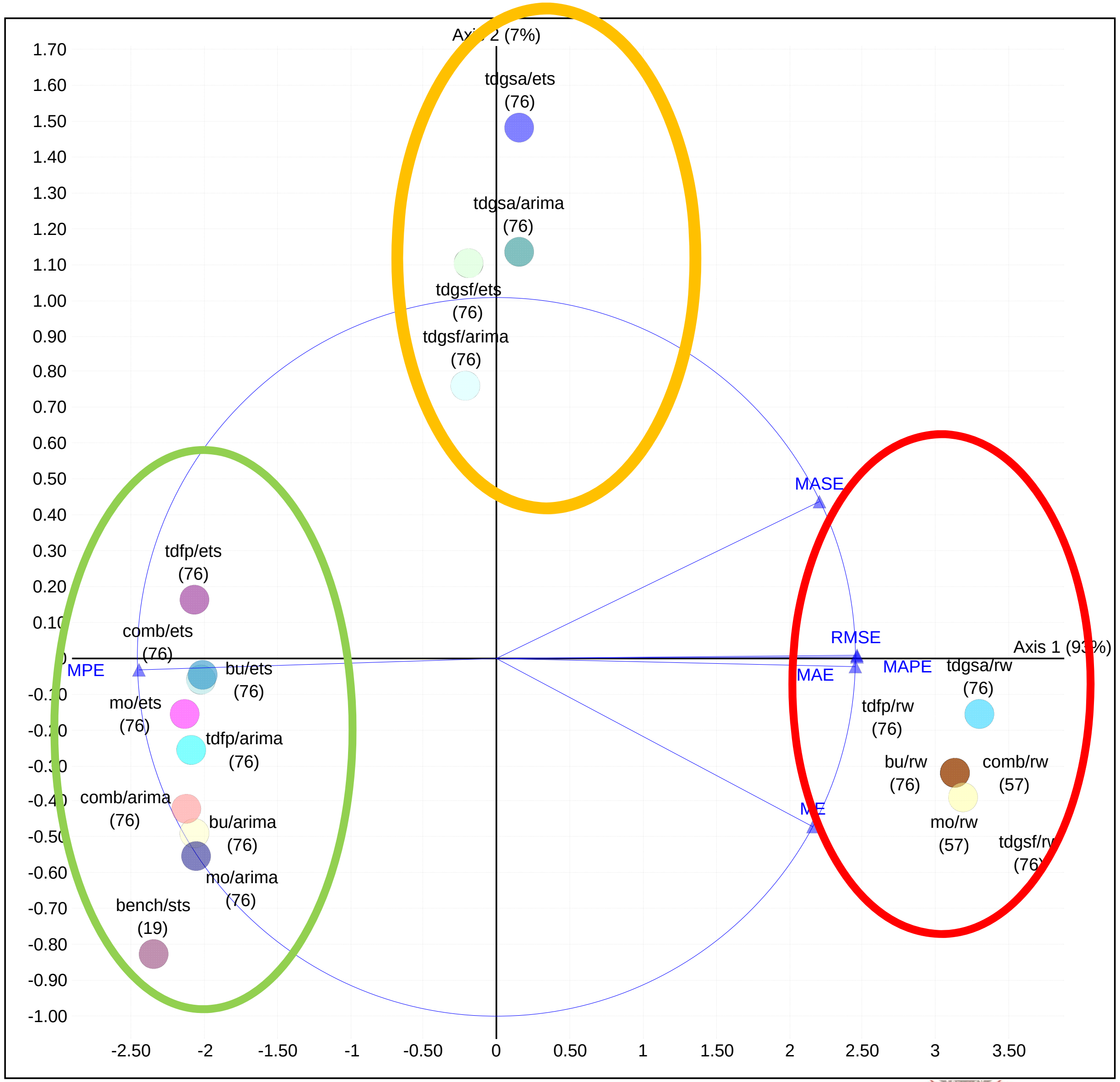
		bench	bu	comb	mo	tdfp	tdgsa	tdgsf
ME	Mean	-746.71	-348.41	-510.27	-397.84	-526.86	-526.86	-526.86
	Frequency	19	228	209	209	228	228	228
RMSE	Mean	2215.35	4561.07	4190.13	4132.07	4678.21	6463.92	5885.91
	Frequency	19	228	209	209	228	228	228
MAE	Mean	1716.26	3772.03	3427.21	3379.42	3864.89	5508.57	5059.22
	Frequency	19	228	209	209	228	228	228
MAPE	Mean	19.46	61.49	53.82	52.28	60.07	86.12	88.97
	Frequency	19	228	209	209	228	228	228
MPE	Mean	-3.94	-42.86	-36.63	-35.08	-41.69	-64.80	-69.11
	Frequency	19	228	209	209	228	228	228
MASE	Mean	1.14	3.77	3.55	3.49	3.92	7.22	6.06
	Frequency	19	228	209	209	228	228	228

Methods / ME p = **97.8%** ; F = **0.19** (**NS**)
Methods / RMSE p = **<0.1%** ; F = **4.04** (**VS**)
Methods / MAE p = **<0.1%** ; F = **4.30** (**VS**)
Methods / MAPE p = **<0.1%** ; F = **7.03** (**VS**)
Methods / MPE p = **<0.1%** ; F = **6.64** (**VS**)
Methods / MASE p = **<0.1%** ; F = **13.38** (**VS**)

Conciliation and forecast method vs mean errors (with & without), including leaves

	ME	RMSE	MAE	MAPE	MPE	MASE	Total
bench/sts	-746.71	2215.35	1716.26	19.46	-3.94	1.14	533.59
bu/arima	-830.63	2288.90	1759.37	22.40	-14.12	2.10	538.00
bu/ets	-1424.98	2715.98	2046.31	26.00	-19.37	2.29	557.71
bu/rw	1210.37	8678.33	7510.41	136.08	-95.10	6.91	2907.83
comb/arima	-950.45	2358.56	1781.97	21.39	-12.03	2.10	533.59
comb/ets	-1360.57	2655.54	2010.06	24.56	-17.39	2.49	552.45
comb/rw	1210.37	8678.33	7510.41	136.08	-95.10	6.91	2907.83
mo/arima	-714.64	2258.01	1720.94	21.59	-12.36	2.14	545.95
mo/ets	-1287.21	2596.43	1939.64	20.13	-12.77	2.28	543.08
mo/rw	1210.37	8678.33	7510.41	136.08	-95.10	6.91	2907.83
tdfp/arima	-1167.93	2561.23	1930.65	22.31	-14.19	2.19	555.71
tdfp/ets	-1623.02	2795.07	2153.61	21.84	-15.80	2.66	555.73
tdfp/rw	1210.37	8678.33	7510.41	136.08	-95.10	6.91	2907.83
tdgsa/arima	-1167.93	5197.60	4295.96	62.37	-45.13	7.01	1391.65
tdgsa/ets	-1623.02	5265.40	4379.54	66.15	-50.42	7.22	1340.81
tdgsa/rw	1210.37	8928.76	7850.21	129.82	-98.87	7.45	3004.62
tdgsf/arima	-1167.93	4513.97	3799.94	63.01	-47.34	5.68	1194.56
tdgsf/ets	-1623.02	4616.68	3914.27	67.18	-52.80	5.88	1154.70
tdgsf/rw	1210.37	8527.09	7463.47	136.73	-107.18	6.62	2872.85
Total	-477.24	4969.42	4155.57	66.85	-48.09	4.65	1445.20

method_For_T / ME p = **<0.1%** ; F = **16.71** (**VS**)
method_For_T / RMSE p = **<0.1%** ; F = **12.59** (**VS**)
method_For_T / MAE p = **<0.1%** ; F = **13.13** (**VS**)
method_For_T / MAPE p = **<0.1%** ; F = **30.09** (**VS**)
method_For_T / MPE p = **<0.1%** ; F = **16.45** (**VS**)
method_For_T / MASE p = **<0.1%** ; F = **11.13** (**VS**)



Conciliation and forecast method vs mean errors (with intervention), including leaves

	ME	RMSE	MAE	MAPE	MPE	MASE	Total
bench/sts	195.60	3915.73	3330.17	2622.78	-162.67	3.40	1650.83
bu/arima	-437.54	1140.32	905.27	40.50	-20.19	2.54	279.65
bu/ets	-886.62	1582.98	1240.02	44.16	-21.61	3.15	336.47
bu/rw	749.90	5185.07	4565.31	201.33	-148.42	6.98	1798.62
comb/arima	-451.11	1949.48	1617.14	79.68	-55.91	3.61	539.13
comb/ets	-966.02	1625.93	1281.80	44.30	-22.57	3.26	337.27
comb/rw	478.06	4558.21	3993.89	167.96	-124.92	6.10	1554.94
mo/arima	-436.38	1149.43	904.93	39.85	-19.59	2.58	281.36
mo/ets	-883.96	1586.84	1236.15	42.94	-19.42	3.23	337.09
mo/rw	749.90	5185.07	4565.31	201.33	-148.42	6.98	1798.62
tdfp/arima	-964.31	1556.53	1228.47	42.55	-24.33	2.86	315.88
tdfp/ets	-1005.57	1662.90	1313.54	43.95	-21.13	3.36	342.47
tdfp/rw	749.90	5185.07	4565.31	201.33	-148.42	6.98	1798.62
tdgsa/arima	-964.31	2555.11	2108.19	143.10	-123.85	8.82	639.44
tdgsa/ets	-1005.57	2598.18	2148.46	145.52	-126.45	8.89	646.64
tdgsa/rw	749.90	5309.86	4712.47	226.47	-197.90	7.75	1854.80
tdgsf/arima	-964.31	2307.62	1928.06	137.02	-116.63	7.23	565.95
tdgsf/ets	-1005.57	2350.81	1971.38	139.59	-119.39	7.25	573.69
tdgsf/rw	749.90	5156.88	4575.76	234.01	-204.00	6.96	1805.17
Total	-311.21	2939.97	2505.13	162.43	-93.78	5.44	891.46

method_For / ME p = **<0.1%** ; F = **19.03** (**VS**)
method_For / RMSE p = **<0.1%** ; F = **8.08** (**VS**)
method_For / MAE p = **<0.1%** ; F = **8.46** (**VS**)
method_For / MAPE p = **<0.1%** ; F = **9.37** (**VS**)
method_For / MPE p = **<0.1%** ; F = **14.32** (**VS**)
method_For / MASE p = **<0.1%** ; F = **19.46** (**VS**)

Conclusions

- At the lowest levels (towns), hts perform better than the benchmark.
- At the aggregate levels (regions, destinations and canton) the performances of hts and benchmark are comparable.
- ARIMA outperforms all the other forecast methods of the hts package and the worst seems to be RW.
- The conciliation methods seems to have some effect at the destination level, mo and comb are superior.

Wishing list (I)

- The hts forecasting method could be quite interesting for tourism if the aim is to forecast at all the levels, from towns to the canton.

- If this is not the case, the benchmark is comparable with the results of hts with the economy of saving the data of the lowest levels.

Wishing list (II)

- A real asset would be to able to make assumption on the forecast of the trend at the lowest level in order to simulate/estimate the effect at the highest level.

For instance, a group planning a festival/event in a town could show the effect on overnight at the highest level (destination, region or canton). Also establish different scenarii using different methods of conciliation.

Hes·so VALAIS WALLIS
School of Management & Tourism

Thank very much!
Questions?

miriam.scaglione@hevs.ch