Forecasting Traffic Flows with Complex Seasonality using Mobile Phone Data

Rodolfo Metulini[1], Maurizio Carpita[2]

1. Department of Economics and Statistics - University of Salerno
2. Data Methods and Systems Statistical Laboratory - Department of Economics and Management, University of Brescia

IES 2022 - Innovation ans Society 5.0: Statistical and Economic Methodologies for quality assessment

January 27th, 2022

The project

MoSoRe@UniBS:

Project of Lombardy Region, Italy (Infrastrutture e servizi per la Mobilità Sostenibile e Resiliente) - CallHub ID 1180965, bit.ly/2Xh2Nfr

Scientific collaboration:

The *Department of Civil, Environmental, Architectural Engineering and Mathematics, UNIBS* (Prof. Roberto Ranzi)

Agenda:

We will present a work at European Geosciences Union (EGU) General Assembly 2022 in Wien (with B. Razdar and R. Ranzi)

Scientific output:

- Carpita, M., Metulini, R. (2021). Modelling the spatio-temporal dynamic of traffic flows with gravity models and mobile phone data, **ASA 2021 Statistics and Information Systems for Policy Evaluation**, Edited by: Bertaccini, B.; Fabbris, L.; Petrucci, A.
- Metulini, R., Carpita, M. (2021). A Spatio-Temporal Indicator for City Users based on Mobile Phone Signals and Administrative Data - **Social Indicator Research**, 156, 761–781.
- Balistrocchi, M., Metulini, R., Carpita, M., and Ranzi, R. (2020). Dynamic maps of people exposure to floods based on mobile phone data. **Natural Hazards and Earth System Sciences**, 20, 3485–3500.

Introduction

- Monitoring and forecasting people mobility is a relevant aspect for metropolitan areas.
- In smart cities, Information & Communication technologies (ICT) with big data are massively used to support the optimization of traffic flows and the study of urban systems (**Albino et al., 2015; Benevolo et al., 2016; Bibri & Krogstie, 2017**).
- Mobile phone network data allow to obtain dynamic information on people's presences (**Metulini & Carpita, 2021**) and movements (**Tettamanti & Varga, 2014; Carpita & Metulini, 2021**)
  - used, e.g., to develop dynamic exposure to flood risk maps for areas with hydrogeological criticity (**Balistrocchi et al., 2020**).
- Using data on origin-destination (O-D) flows of mobile phone signals, we implement a framework to model such a flows for the *Mandolossa* area (a critical zone with flood episodes in the north-west of Brescia)
  - to predict the "real time" amount of traffic flows,
  - for smart cities emergency management plans.

Data & area's delimitation

- Numer of humans' SIM provided by Olivetti and Fasternet (both italians and foreigners, 85% of the total. 15% are machine SIM) from/to 235 *aree di censimento* (ACE of ISTAT), recorded hourly basis for 365 days (September 2020 to August 2021)
  - **Limitation 1**: locations are retrieved each 5 minutes, but time to travel from A to B may be less than 5 minutes (after 5 minutes the SIM is already at C). This underestimation leads to problems in making inference to the population.
  - **Limitation 2**: flows from two areas whose travel time is higher than 5 minutes should be 0 (unless machine's measurement errors).
- Four ACEs which intersect with the flooding-risk area (*Brescia Mandolossa*, *Gussago*, *Cellatica*, *Rodengo Saiano*).
- Other 38 neighboring ACE (aggregated in 4 macro areas) having a minimum outflow from / inflow to the four ACE of the Mandolossa.
  - The total flows counted between the 4 Mandolossa's ACE and the 38 selected neighboring ACE counts for about the 84% of the total outflows from or to the Mandolossa's ACE.

Empirical evidences

- We have studied the time series of the flows from the ACE of Brescia Mandolossa to the ACE of Gussago
- Also by means of the autocorrelation function (ACF) and the partial ACF (PACF), which we display of 1 week and of 1 month length, to show, respectively, daily and weekly patterns.
- ... and by means of an additive decomposition of the time series in trend and daily and weekly seasonality, obtained using *Seasonal-Trend decomposition with LOESS* (STL) (**Hyndman & Athanasopoulos, 2021**).
  - According to the height of the grey bars, the daily pattern (*season*_24) is the most important component, whereas the importance of the trend is smaller than all the other components.

The Harmonic Dynamic Regression (HDR) model

- Dynamic harmonic regression (**Hyndman & Athanasopoulos, 2021**) is a time series autoregressive model based on the principle that a combination of *sine* and *cosine* functions can approximate any periodic pattern.

$$y_t = \beta_0 + \sum_{k=1}^{K} [\alpha_k s_k(t) + \gamma_k c_k(t)] + \epsilon_t$$

- $s_k(t) = sin(\frac{2\pi kt}{m})$ and $c_k(t) = cos(\frac{2\pi kt}{m})$, where m is the seasonal period (e.g., 1 week: $m = 168$), α_k and γ_k are regression coefficients and ϵ_t is modeled as an $ARIMA_{p,d,q}$ process
  - For construction, K can not be greater than $m/2$.

- Actually, HDR may account for more than one periodic function. In our case we have two summation terms (one for the daily periodicity, another for the weekly periodicity)

Modelling the seasonal patterns

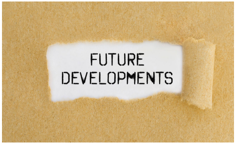
- After a proper data pre-processing, for the following flows:
  1. Gussago → Brescia Mandolossa
  2. Brescia Mandolossa → Gaussago
- We model the following HDR with multiple seasonal periods:

$$Flow_t = \beta_0 + \sum_1^{K_d} [\alpha_{k_d} s_{k_d}(t) + \gamma_{k_d} c_{k_d}(t)] + \sum_1^{K_w} [\alpha_{k_w} s_{k_w}(t) + \gamma_{k_w} c_{k_w}(t)] + month + \epsilon_t$$

  - We also include a set of dummies to control for the possible presence of changes in levels by month (e.g., higher traffic flows in march).
- The number of Fourier are selected by minimising the Akaike Information Criterion (AIC) following a two step approach:
  1. We find K_d by minimizing the AIC for the model with no ARIMA error and no weekly pattern;
  2. We find K_w by minimizing the AIC for the model with K_d bases and no ARIMA error.
- For both directions we obtain $K_d = 7$ and $K_w = 4$ (note that we have 2 parameters to be estimated for each Fourier basis, that of *sin* and that of *cos*)

- Results of the HDR with the eror process modeled with an $ARIMA_{5,0,0}$.
- Daily and weekly Fourier bases as well as most of the monthly dummies are statistical significant.
- One day forecast with confidence bands estimated with the HDR model and the boostrap are presented for both flows.

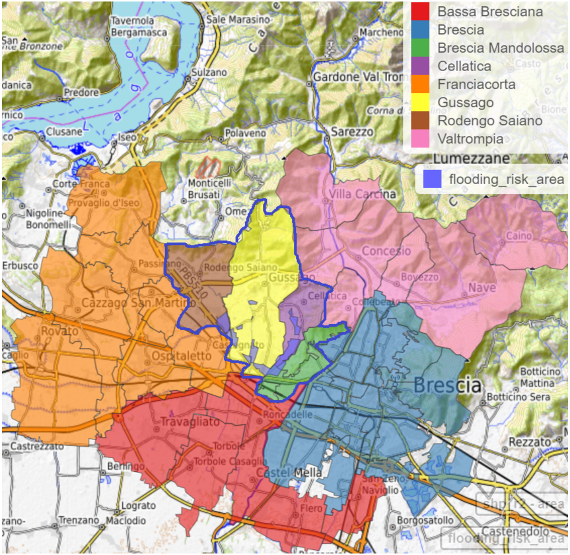
- In this short paper, after studying the seasonal structure of the traffic flows recorded between two ACE of the Mandolossa area in the north of Brescia using the hourly TIM mobile phone data from September 2020 to August 2021, the HDR model with ARIMA error is estimated.
- Preliminary results show the statistical significance of daily, weakly and monthly effects, and the opportunity offered by this model to forecast future traffic flows.



- Robustness check of the model using (k-fold) cross validation (month by month?) and *Root Mean Square Error* (RSME)
- To include additional explanatory features available "real-time" (e.g. precipitation data)
- Refine the quality of original data, by just considering the in- and out-flows transiting through the flooding risk areas.

References

1. Albino, V., Berardi, U., D'angelico, R. M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. Journal of Urban Technology, 22(1): 3–21.
2. Balistrocchi, M., Metulini, R., Carpita, M., & Ranzi, R. (2020). Dynamic maps of human exposure to floods based on mobile phone data. Natural Hazards and Earth System Sciences, 20(12), 3485-3500.
3. Benevolo, C., Dameri, R.P., D'Auria, B. (2016). Smart mobility in smart city. In T. Torre et al. (Eds.), Empowering Organizations (13–28). Springer.
4. Bibri, S.E., Krogstie, J. (2017). Smart sustainable cities of the future: An extensive interdisciplinary literature review. Sustainable Cities & Society, 31: 183–212.
5. Carpita, M., & Metulini, R. (2021). Modelling the spatio-temporal dynamic of traffic flows with gravity models and mobile phone data. In ASA 2021 Statistics and Information Systems for Policy Evaluation: BOOK OF SHORT PAPERS of the on-site conference (Vol. 132). Firenze University Press.
6. Hyndman, R. J., Athanasopoulos, G. (2021). Forecasting: principles and practice. 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3.
7. Metulini, R., & Carpita, M. (2021). A spatio-temporal indicator for city users based on mobile phone signals and administrative data. Social Indicators Research, 156(2), 761-781.
8. Tettamanti, T., Varga, I. (2014). Mobile phone location area based traffic flow estimation in urban road traffic. Columbia International Publishing, Advances in Civil and Environmental Engineering, 1(1): 1-15.



Back to slide

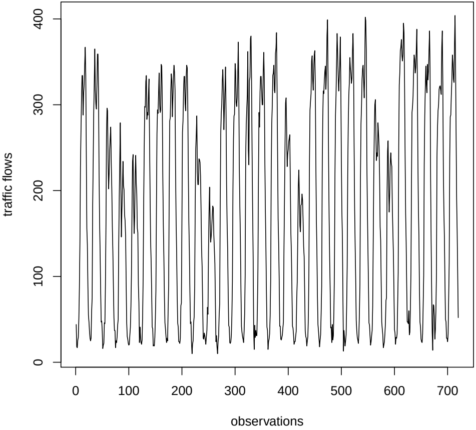


Figure: Traffic flows from Brescia Mandolossa to Gussago. Time series of the original data, full month of April 2021.

Back to slide

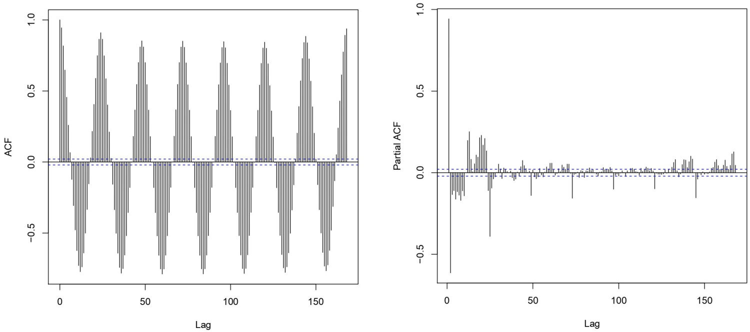


Figure: Traffic flows from Brescia Mandolossa to Gussago. ACF (**left**) and PACF (**right**), t = 1, ..., 168 (1 week).

Back to slide

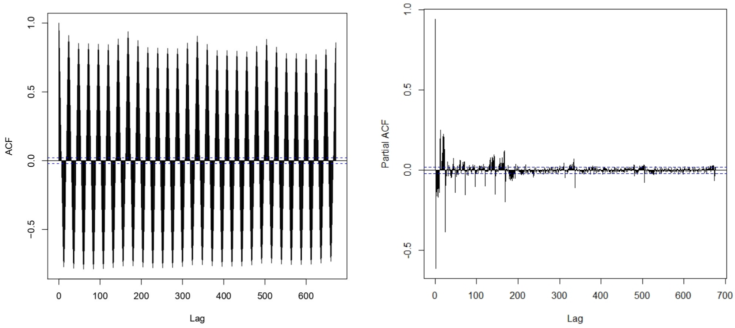


Figure: Traffic flows from Brescia Mandolossa to Gussago. ACF (**left**) and PACF (**right**), t = 1, ... , 720 (1 month).

Back to slide

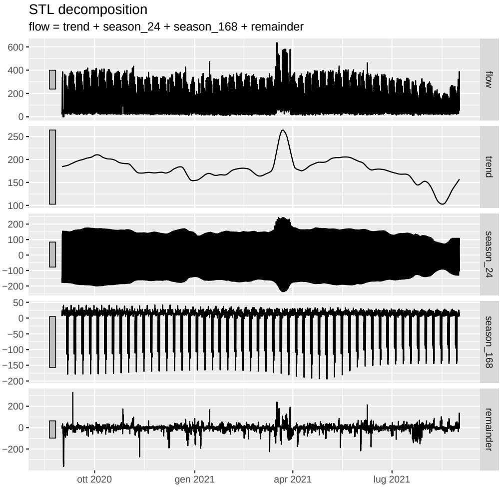


Figure: Traffic flows from Brescia Mandolossa to Gussago. Sesonal-Trend decomposition with LOESS (STL) with trend, daily and weekly patterns and error components. 1 year of data.

Regressors	Brescia Man. → Gussago	Gussago → Brescia Man.
Intercept	160***	160***
$fourier\_day(1)_{cos;sin}$	-132.0*** ; -62.8***	-135.0*** ; -54.7***
$fourier\_day(2)_{cos;sin}$	-9.6*** ; -12.1***	-21.2*** ; -23.5***
$fourier\_day(3)_{cos;sin}$	13.5*** ; 24.0***	19.8*** ; 20.3***
$fourier\_day(4)_{cos;sin}$	0.1 ; -4.1***	-2.1*** ; -0.2
$fourier\_day(5)_{cos;sin}$	-8.1*** ; 0.3	-8.9*** ; 0.7
$fourier\_day(6)_{cos;sin}$	3.0*** ; 1.7***	3.6*** ; 0.2
$fourier\_day(7)_{cos;sin}$	1.7*** ; 5.4***	3.8*** ; 5.3***
$fourier\_week(1)_{cos;sin}$	28.5*** ; -6.9***	29.5*** ; -7.7***
$fourier\_week(2)_{cos;sin}$	-15.9*** ; -11.0***	-15.8*** ; 11.8***
$fourier\_week(3)_{cos;sin}$	7.7*** ; -4.6***	8.0*** ; -5.1***
$fourier\_week(4)_{cos;sin}$	-6.7*** ; -4.7***	-7.2*** ; -4.4***
Month (ref: *January*)		
February	12.3	13.7
March	43.7***	44.2***
April	19.1***	20.9***
May	32.4***	35.9***
June	12.4	15.2
July	-11.4	-9.6
August	-30.3***	-30.0***
September	30.2***	32.7***
October	34.1***	37.5***
November	11.0	12.0
December	0.2	1.5
ARIMA(p,d,q)	5,0,0	5,0,0
AIC	78,885	80,093

Notes: Significance levels for t test: . $p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Back to slide

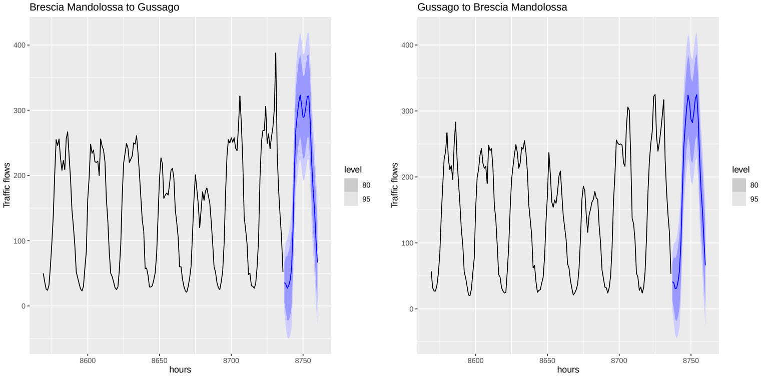
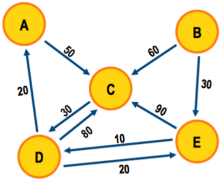


Figure: One day forecast of traffic flows from Brescia Mandolossa to Gussago (**left**) and from Gussago to Brescia Mandolossa (**right**) with the HDR model. Bootstrap confidence bands are reported for the confidence levels of 80% and 95%.

Back to slide

Example of a car, whose driver's SIM is retrieved by the antenna during his trip (**top**); a representation using an oriented graph (**botttom left**) and an origin-destination (O-D) matrix (**bottom right**)



	A	B	C	D	E	Ti
A	0	0	50	0	0	50
B	0	0	60	0	30	90
C	0	0	0	30	0	30
D	20	0	80	0	20	120
E	0	0	90	10	0	100
Tj	20	0	280	40	50	390

Back to slide