Statistics, Technology and Data Science for Economic and Social Development

Book of short papers of the ASA Rome Conference
18 to 20 September 2024

"Measuring and Interpreting World Changes with Statistics, Data Science and AI"

Book of Short Papers of the
Sapienza University of Rome Conference

Supplement to Volume 37/2

Guest Editors:
Leonardo S. Alaimo, Fabio Crescenzi,
Luigi Fabbris, Filomena Maggino, Maurizio Vichi

Statistica Applicata – ITALIAN JOURNAL OF APPLIED STATISTICS

A.S.A CONTACTS
Principal Contact
Francesco Palumbo (Editor in Chief)
editor@sa-ijas.org

Support Contact
Domenico Vistocco (Editorial Manager)
ijas@sa-ijas.org

JOURNAL WEBPAGE
https://www.sa-ijas.org/ojs/index.php/sa-ijas

Statistica Applicata – Italian Journal of Applied Statistics is associated to the following Italian and international journals:

QTQM – Quality Technology & Quantitative Management (http/web.it.nctu.ew/~qtqm/)

SINERGIE – Italian Journal of Management

Pubblished: April 2025

Published by Firenze University Press
Powered by Firenze University Press

Firenze University Press
Università degli Studi di Firenze
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

Statistica Applicata – Italian Journal of Applied Statistics is a four-monthly journal published by the Associazione per la Statistica Applicata (A.S.A.), Largo Gemelli 1 – 20123 Milano, Italy (phone + 39 02 72342904). Advertising:Cleup sc, via G. Belzoni, 118/3 – 35128 Padova, Italy (phone +39 049 8753496 – Fax +39 049 9865390), email: info@cleup.it.
Rules for manuscript submission: https://www.sa-ijas.org/ojs/index.php/sa-ijas/about/submissions
Subscription: yearly €103.30; single copy €40.00; A.S.A. associates €60.00; supporting institutions: €350.00. Advertisement lower than 70%. Postal subscription Group IV, Milan. Forum licence n. 782/89. CLEUP SC on behalf of ASA, 7 March 2023.

doi.org/10.26398/asaproc.0091

Volatility decomposition of traffic flow time series with complex seasonality using GARCH models

Maurizio Carpita[a], Rodolfo Metulini[b], Manlio Migliorati[a]
[a] Department of Economics and Management, University of Brescia, Brescia, Italy.
[b] Department of Economics, University of Bergamo, Bergamo, Italy.

1. Introduction

Accurately forecasting people's movements between specific areas is essential for urban planning, transportation optimization, and emergency management. Consequently, identifying the most effective statistical models is critical for policymakers to make well-informed decisions (Xie et al., 2020; WHO, 2022). In previous studies, we considered the problem of forecasting the traffic flow time series from September 2020 to August 2021 of "origin-destination" mobile phone hourly data, from an area with high flooding risk in the province of Brescia, Italy; in particular, for each of the 38 sub-areas of interest, the three (In, Out and Internal) flows were considered (Metulini and Carpita, 2023; Perazzini et al., 2023).

By applying a trivariate Vector AutoRegressive model with eXogenous regressors (VARX; Tsay, 2013) and Dynamic Harmonic Regression (DHR; Hyndman and Athanasopoulos, 2021) components to these time series, we found a discrete forecasting accuracy but also non-normal residuals, with heavy tails and time-varing heteroschedasticity (Carpita et al., 2024).

To explore this evidence for the 38×3 =114 residual's time series of our VARX+DHR model, the Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) model (Bollerslev, 1986) has been used in this short paper. More specifically, the *multiplicative component standard* (mcs) GARCH model has been applied to the 114 residuals time series, to decompose volatility in *Daily*, *Hourly* and *Intradaily* (Engle and Sokalska, 2012).

The structure of the paper is as it follows: Section 2 describes the mcsGARCH model, Section 3 explain its estimation strategy and Section 4 presents some results and future developments.

2. The multiplicative component standard (mcs) GARCH model

Since the publication of the seminal paper published in 1982 by the 2003 Nobel Laurate Robert Engle (Engle, 1982), today the broad family of the GARCH models is well known and used in the financial literature (Bollerslev, 1986; Palm, 1996; Engle, 2001), but has rare applications in the transport field (Zhang et al., 2015; Ali et al., 2023).

Taking into account the strong hourly and daily seasonality observed in the traffic flow data, in this study we used a special version of the GARCH model, named *multiplicative component standard* (mcs), which allows to decompose the time series conditional volatility in three components (Engle and Sokalska, 2012). Considering, as in our case of study, a stationary and standardized (zero mean and unit variance) time series $\varepsilon_{t,i}$, where t denotes the day and i the hour, assuming the mcsGARCH model the *Total* conditional variance is the product of three components:

$$\varepsilon_{t,i} = \left(\sigma_t s_i h_{t,i}\right) z_{t,i} \qquad t = 1,2,..,T \quad \text{and} \quad i = 1,2,..,24 \tag{1}$$

where σ_t is a *Daily* (exogenously determined) volatility, s_i the *Hourly* volatility, $h_{t,i}$ the *Intradaily* volatility, and $z_{t,i}$ the i.i.d. standardized innovation, which conditional follows some appropriately chosen distribution. To estimate the mcsGARCH model with its conditional variance (volatility squared) components, the two-step procedure offered by the R package rugarch by Ghalanos (2024) can be used.

In the *first step*, the Daily volatility σ_t has to be preliminarily estimated: externally from a multifactor risk model as in Engle and Sokalska (2012), or predicted from a daily GARCH model as in Andersen and Bollerslev (1997). In our study, we adopted this second approach, using the very flexible *asymmetric power* (ap) GARCH(*q,p*) model, which allows for different positive and negative effects and includes six GARCH submodels (Ding et al., 1993):

$$\sigma_t^\delta = \omega + \sum_{l=1}^{q} \alpha_l (|\varepsilon_{t-l}| - \gamma_l \varepsilon_{t-l})^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta \tag{2}$$

with $\omega > 0, \alpha_l \geq 0$, and $\beta_j \geq 0$ with the stationarity constrain $\sum_{l=1}^{q} \alpha_l + \sum_{j=1}^{p} \beta_j < 1$; the parameter $\delta > 0$ allows for the Box-Cox transformation of σ_t, and γ_l is the parameter in the leverage term (Ghalanos, 2024, sec. 2.2.5).

In the *second step*, using the Daily volatility estimated in the first step $\hat{\sigma}_t$, the Hourly variance s_i^2 is estimated by simply applying the *T*-days average of the squared daily standardized residuals:

$$\hat{s}_i^2 = \frac{1}{T} \sum_{t=1}^{T} \frac{\varepsilon_{t,i}^2}{\hat{\sigma}_t^2} \qquad i = 1,2,..,24. \tag{3}$$

Then, scaling the residuals by the estimated Daily and Hourly volatility gives $\bar{\varepsilon}_{t,i} = \varepsilon_{t,i}/(\hat{\sigma}_t \hat{s}_i)$, which are used to estimate the Intradaily component of volatility $h_{t,i}$ assuming the hourly standard (s) GARCH(*q,p*) model (Bollerslev, 1986):

$$h_{t,i}^2 = \omega + \sum_{l=1}^{q} \alpha_l \bar{\varepsilon}_{t,i-l}^2 + \sum_{j=1}^{p} \beta_j h_{t,i-j}^2 \tag{4}$$

Obtained the constrained estimates $\hat{\omega} > 0, \hat{\alpha}_l \geq 0$ and $\hat{\beta}_j \geq 0$, the persistence parameter of the hourly sGARCH model (that is a measure of how quickly past shocks decay over time assuming value in (0,1): if close to 1 volatility shocks have a long-lasting impact on future volatility) is estimated by $\hat{P} = \sum_{l=1}^{q} \hat{\alpha}_l + \sum_{j=1}^{p} \hat{\beta}_j$, and the estimated *unconditional* volatility by this model is $\hat{\sigma} = \hat{\omega}/(1 - \hat{P})$.

To estimate any GARCH(*q,p*) model, a choice must be made about the order *q* and *p* of the two lagged components on the right-hand side of its equation. As stated in Engle and Sokalska (2012) for financial data, a GARCH(1,1) model proved to be an adequate and popular choice in most of the cases and, if any higher lags are included in the conditional variance specification, the models are not of a higher order than GARCH(1,2) or GARCH(2,1). Still, efficiency considerations favor models of a lower order.

The last choice is about the conditional distribution of the standardized innovation $z_{t,i}$: for the sake of flexibility, the Normal Inverse Gaussian (NIG) is used, which allows controlling for location, scale, skewness, and kurtosis (tail heaviness) of *z*'s, and includes as a special case the Normal distribution. To reduce the computational burden, the four parameters of the NIG can be estimated from a transformation of two parameters, γ_3 and γ_4, that estimate the skewness and the shape (scale and tail) of the NIG distribution respectively (Ghalanos, 2024, sec. 2.3.5).

3. The estimation strategy of the mcsGARCH model

To estimate the three volatility components of the mcsGARCH model described in the previous section, daily and hourly time series of the 114 VARX+DHR model residuals have been used. Each time series has been split into two parts: the first 4-month period (from September to December 2020) of daily data has been used to identify the best order (*q,p*) of the ap- GARCH model in the first step of the estimation procedure, and the last 8-month period (from January to August 2021) of hourly data has been used in the second step of the procedure to estimate the daily volatility and identify the best order (q,p) of the GARCH model for the Intradaily volatility. The hybrid solver estimation algorithm with the standardized time series was used (Ghalanos, 2024, sec. 3).

The estimation strategy adopted to apply the mcsGARCH model was as follows:

- Preliminary analysis was carried out on the 114 daily time series for the first 4-month period to assess the presence of ARCH effects ($q > 0$), using the classical Engle's Lagrange Multiplier test (Engle, 1982): the null hypothesis of variance stability (no ARCH effects) was rejected for all but 21 daily time series, therefore considered stationary in variance.
- For each of the 93 daily time series with significant ARCH effects ($q > 0$ for the Engle test), the first 4-month period was used to find the best order (q,p) of the apGARCH model, minimizing the AIC (Akaike Information Criterion) for all permutations of q in [1,2] and p in [0,1,2].
- The best apGARCH(q,p) model identified in the first 4-month period was used to forecast the daily autoregressive conditional volatility for the last 8-month period. For each of the 93 daily time series, a rolling forecast was used with 90-day length window.
- For each of the 114 hourly time series of the last 8-month period, scaled by forecast of Daily and estimate of Hourly volatility, the best order (q,p) of the sGARCH model for the Intradaily volatility was selected by the AIC for all permutations of q and p in [0,1,2].

In the end, for each of 114 time series of the VARX+DHR residuals, the total hourly volatility has been decomposed into its three components: Hourly, Daily, and Intradaily.

4. Some results and future developments

For the 93 daily time series with significant ARCH effects, the best orders (q,p) of the apARCH model (Section 3, second point) were the following: 57 (1,0), 21 (1,1), 5 (1,2), 9 (2,0) and 1 (2,2). For the 114 hourly time series, the best orders of the hourly sGARCH(q,p) model (Section 3, fourth point) were the following: 53 (1,1), 60 (1,2) and 1 (2,2).

Table 1 shows mean and standard deviation (std. dev.) of the parameter estimates for the two parts of the mcsGARCH model: results are on the left for the 93 daily apGARCH models estimated for the first 4-month period, and on the right for the 114 hourly sGARCH models estimated for the last 8-month period. Differences emerge in parameter estimates between Out and In flows versus Internal flows and between the daily and hourly GARCH models.

Table 1: Means (standard deviations) of the mcsGARCH model parameters by flow types.

	93 daily apGARCH Models (4 months)			**114 hourly sGARCH Models (8 months)**		
Parameters	*Out Flow*	*In Flow*	*Internal Flow*	*Out Flow*	*In Flow*	*Internal Flow*
Delta	1.899 (0.846)	1.986 (0.924)	2.999 (0.404)			
Gamma	-0.099 (0.341)	-0.110 (0.366)	0.484 (0.344)			
Omega	0.554 (0.334)	0.510 (0.354)	0.346 (0.229)	0.069 (0.089)	0.070 (0.080)	0.212 (0.019)
Persistence	0.564 (0.380)	0.610 (0.415)	0.841 (0.323)	0.937 (0.092)	0.937 (0.083)	0.798 (0.023)
Skew	-0.467 (0.480)	-0.480 (0.491)	-0.184 (0.190)	0.160 (0.116)	0.163 (0.128)	-0.123 (0.012)
Shape	3.165 (4.797)	3.089 (3.931)	1.707 (2.829)	4.105 (1.783)	5.124 (3.011)	1.189 (0.124)

First of all, let's consider the special parameters of the daily apGARCH model: the mean of the estimated delta is almost 2 (std. dev. almost 1) for the Out and In flows, and about 3 (std. dev. 0.4) for the Internal flows; the mean of the estimated gamma (only the first one is considered, as only 10 time series have q = 2) is about – 0.1 (std. dev. about 0.3) for the Out and In flows, and about 0.5 (std. dev. about 0.3) for the Internal flows. The means for the omega (intercept) parameter estimates are much higher for the 93 daily apGARCH models (0.5 for Out and In flows and 0.3 for Internal flow, with std. dev. 0.3 and 0.2 respectively), with respect to the 114 hourly sGARCH models (0.07 for Out and In flows and 0.2 for Internal flow, with std. dev. about 0.09 and 0.02 respectively). For both Out and In flows, the mean estimated daily persistence (sum of the alpha and beta parameters) is lower (about 0.6, with std. dev. 0.4) than the hourly persistence (about 0.9, with std. dev. 0.1). In contrast, for Internal flows, persistence estimates are similar across both the two periodicities: around 0.8, but with a daily std. dev. of 0.3 and an hourly std. dev. of only 0.02. The means for the estimated skew of the NIG distribution for the GARCH residuals are negative for all daily flows (about 0.5, with analogous standard deviation) and for hourly Internal flows (lower than 0.2, and smaller standard deviation), while they are positive but very close to zero for the other two hourly flow types. The means of the estimated shape parameter of the NIG distribution are positive, ranging from 3 to 5 for Out and In flows, and from 1.2 to 1.7 for both daily and hourly Internal flows: this means that the autoregressive part of the mcsGARCH model doesn't fully capture the heavy tails observed for the VARX+DHR residual (leptokurtic) distributions.

Figure 1 is a visualized example of mcsGARCH decomposition of the estimated volatility of one of the 114 analyzed time series: under the graph of the VARX+DHR model's residual for the last 8-month period Out flow from the area of Paderno-Franciacorta, the estimated Total volatility and it's three multiplicative components (Daily, Hourly and Intradaily) are represented. In the considered period, the range of the estimated Total volatility is between 0.5 and 2, but some clusters with peaks between 4 and 6 are observed. The base level of the estimated Daily volatility is around 1, and some clusters with higher variability (up to 3) at the beginning of January, April, June, and at the end of July are observed. The estimated Hourly volatility profile, which repeats each 24 hours, shows low (below 1) variability during the night, moderate variability (about 1.5) in the diurnal part of the day, and high (peaks above 2) at 8 and 18 o'clock. The last graph for the Intradaily volatility, estimated with the hourly apGARCH model, shows smaller clusters and therefore, for this time series, can be considered as residual.

To explore relations between the traffic flows variability in the various areas under consideration, the linear correlation coefficients between the time series of the Total volatility estimates and each of its three components (93 Daily, 114 Hourly and 114 Intradaily) obtained from the mcsGARCH model were computed, and their boxplots by flow types (Out, In and Internal) are in Figure 2. Distributions for the Internal flow correlations of the Total and each estimated volatility component are very different with respect to Out and In flows ones. Considering the Total volatility estimates, the boxplots of the correlations for Out and In flows are (excluding outliers) roughly symmetric with median near 0.6 and 0.7 respectively and interquartile range about 0.1, whereas for the Internal flow is near to 1, with a very small interquartile range. Analogous correlation distributions are observed for the three estimated volatility components of the Internal flow, whereas for the Out and In flows boxplots these are lower (median between 0.3 and 0.6) for the Daily and Intradaily volatility estimates, and higher (median about 0.8 and 0.9 respectively) with negative asymmetry for the Hourly one.

This preliminary study, carried out to explore the volatility structure of the VARX+DHR model's residuals from a previous analysis, could be extended in some directions. The ap-GARCH model with asymmetric effects could be used to modeling the hourly volatility too, with the aim to capture with more flexibility the residual heteroskedasticity highlighted by the two parameter estimates of the NIG distribution. Moreover, volatility estimated components could be grouped using a cluster analysis approach, also to analyze the spatial distribution of traffic flows in the areas of interest. Finally, predicted autoregressive conditional volatility with a trivariate GARCH model (one equation for each flow type) for the VARX+DHR model's residuals could be used to check if the forecast of the traffic flows can be improved.

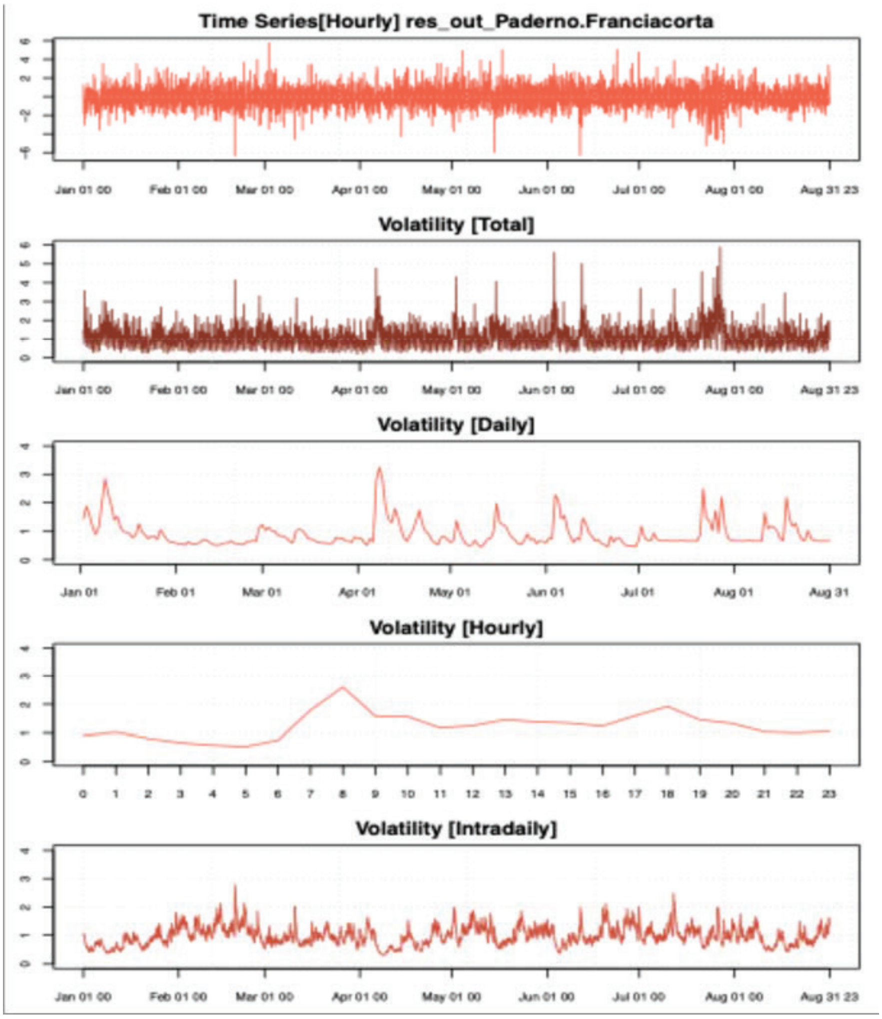

Figure 1: Example of mcsGARCH decomposition of the volatility of a time series (last 8-month period Out flow from the area of Paderno-Franciacorta). From top to bottom: The original residuals from the VARX+DHR model, the estimated Total, Daily, Hourly and Intradaily volatilities.

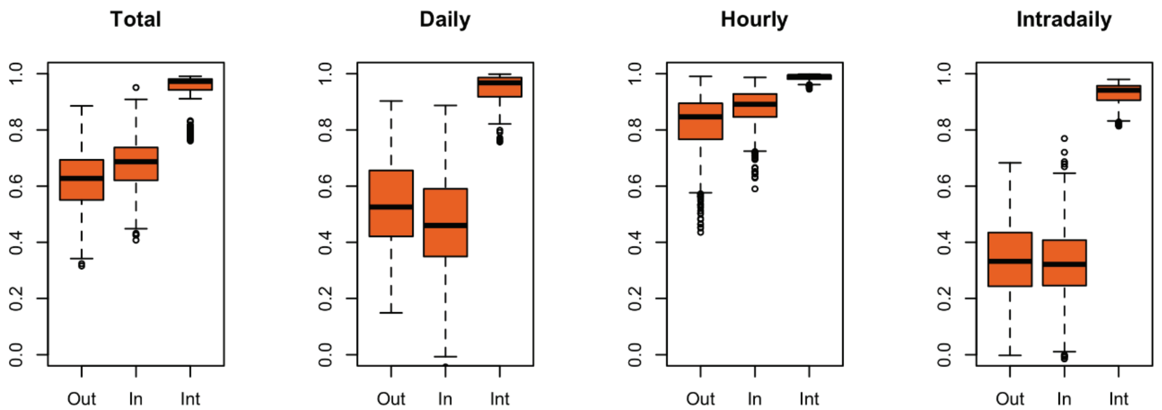

Figure 2: Boxplot of the correlations between the mscGARCH estimated Total volatility and its components (Daily, Hourly, Intradaily) by flow types (Out, In, Internal).

Acknowledgement This contribution has been developed for the project "Study of mobile phone siGNals for the evalUation of the interconnections between Mobility and the environment in Lombardia (SIGNUM)" CUP: F53D23010910001- PRIN 2022 PNRR M4C2 - financed by the European Union - Next Generation EU (DD MUR n. 1409 del 14/09/2022).

M. Carpita acknowledges the European Union (EU) and the Italian Ministry for Universities and Research (MUR), National Recovery and Resilience Plan (NRRP), within the project "Sustainable Mobility Center (MOST)", 2022-2026, CUP D83C22000690001, Spoke N. 7, "CCAM, Connected networks and Smart Infrastructures".

References

Ali M., Yusof K.M., Wilson B., Ziegelmueller C. (2023). Traffic speed prediction using GARCH-GRU hybrid model. *IET Intelligent Transport Systems*, 17(11), pp. 2300–2312.

Andersen, T.G., Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2), pp. 115–158.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, pp. 307–327.

Carpita, M., De Luca, G., Metulini, R., Zuccolotto, P. (2024). Traffic flows time series in a flood-prone area: modeling and clustering on extreme values with a spatial constraint. *Stochastic Environmental Research and Risk Assessment*, 1–17, online first.

Ding, Z., Granger, C. W., Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1), pp. 83–106.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50 (4): pp. 987–1007.

Engle, R. F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, 15(4), pp. 157–168.

Engle, R. F., Sokalska, M. E. (2012). Forecasting intraday volatility in the US equity market. Multiplicative component GARCH. *Journal of Financial Econometrics*, 10(1), pp. 54–83.

Ghalanos, A. (2024). rugarch: Introduction to the rugarch package. *R Package Version 1.4-3*.

Hyndman, R.J., and Athanasopoulos, G. (2021). *Forecasting: Principles and Practice.* 3rd edition, OTexts: Melbourne, Australia.

Metulini, R., Carpita, M. (2023). Modeling and forecasting traffic flows with mobile phone big data in flooding risk areas to support a data-driven decision making. *Annals of Operations Research*, 1–26, online first.

Palm, F. C. (1996). GARCH Models of Volatility. In: Maddala, G. and Rao, C., Eds., *Handbook of Statistics,* Elsevier Science, Amsterdam, pp. 209–240.

Perazzini, S., Metulini, R., Carpita, M. (2023). Integration of flows and signals data from mobile phone network for statistical analyses of traffic in a flooding risk area. *Socio-Economic Planning Sciences*, 90, 101747, 1–17.

Tsay, R. S. (2013). *Multivariate Time Series Analysis with R and Financial Applications*. John Wiley & Sons.

World Health Organization (2022). *Urban Planning, Design and Management Approaches to Building Resilience – An Evidence Review*. WHO Regional Office for Europe, Copenhagen.

Xie, P., Li, T., Liu J., Du, S., Yang, X., Zhang, J. (2020). Urban flow prediction from spatiotemporal data using machine learning: A survey. *Information Fusion*, 59, pp. 1–12.

Zhang, Y., Haghani, A., Zeng, X. (2015). Component GARCH models to account for sea- sonal patterns and uncertainties in travel-time prediction. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), pp. 719–729, April 2015.