

**Agarwal, Piyush; Tangirala, Arun K.**

**Reconstruction of missing data in multivariate processes with applications to causality analysis.** (English) [Zbl 06882421](#)

*Int. J. Adv. Eng. Sci. Appl. Math.* 9, No. 4, 196-213 (2017).

Summary: Recovery of missing observations in time-series has been a century-long subject of study, giving rise to two broad classes of methods, namely, one that reconstructs data and the other that directly estimate the statistical properties of the data, largely for univariate processes. In this work, we present a data reconstruction technique for multivariate processes. The proposed method is developed in the framework of sparse optimization while adopting a parametric approach using vector auto-regressive (VAR) models, where both the temporal and spatial correlations can be exploited for efficient data recovery. The primary purpose of recovering the missing data in this work is to develop a directed graphical or a network representation of the multivariate process under study. Existing methods for data-driven network reconstruction are built on the assumption of data being available at regular intervals. In this respect, the proposed method offers an effective methodology for reconstructing weighted causal networks from missing data. The scope of this work is restricted to linear, jointly stationary multivariate processes that can be suitably represented by VAR models of finite order and missing data of the random type. Simulation studies on different data generating processes with varying proportions of missing observations illustrate the efficacy of the proposed method in recovering the multivariate signals and thereby reconstructing weighted causal networks.

**MSC:**

**62M10** Time series, auto-correlation, regression, etc. in statistics (GARCH)  
**62-07** Data analysis (statistics) (MSC2010)

**Keywords:**

[missing data](#); [causal graph](#); [random sampling](#); [sparse optimization](#)

**Software:**

[iVAR](#)

**Full Text:** [DOI](#)

**References:**

- [1] Intiaz, S; Shah, S, Treatment of missing values in process data analysis, *Can. J. Chem. Eng.*, 86, 838-858, (2008)
- [2] Lakshminarayan, K; Harp, SA; Samad, T, Imputation of missing data in industrial databases, *Appl. Intell.*, 11, 259-275, (1999)
- [3] Lomb, NR, Least-squares frequency analysis of unequally spaced data, *Astrophys. Space Sci.*, 39, 447-462, (1976)
- [4] Kasam, A.A., Lee, B.D., Paredis, C.J.: Statistical methods for interpolating missing meteorological data for use in building simulation. In: *Building Simulation*, vol. 7, pp. 455-465. Tsinghua University Press, Springer (2014). <https://doi.org/10.1007/s12273-014-0174-7>
- [5] Ferrari, GT; Ozaki, V, Missing data imputation of climate datasets: implications to modeling extreme drought events, *Rev. Bras. Meteorol.*, 29, 21-28, (2014)
- [6] Kourti, T; MacGregor, JF, Process analysis, monitoring and diagnosis, using multivariate projection methods, *Chemom. Intell. Lab. Syst.*, 28, 3-21, (1995)
- [7] Scargle, JD, Studies in astronomical time-series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data, *Astrophys. J.*, 263, 835-853, (1982)
- [8] Warga, A, Bond returns, liquidity, and missing data, *J. Financial Quant. Anal.*, 27, 605-617, (1992)
- [9] Babu, P; Stoica, P, Spectral analysis of nonuniformly sampled data-a review, *Digit. Signal Process.*, 20, 359-378, (2010)
- [10] Scargle, JD, Studies in astronomical time-series analysis. iii-Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data, *Astrophys. J.*, 343, 874-887, (1989)
- [11] Hocke, K; Kämpfer, N, Gap filling and noise reduction of unevenly sampled data by means of the lomb-scargle periodogram,

- Atmos. Chem. Phys., 9, 4197-4206, (2009)
- [12] Hocke, K.: Phase estimation with the lomb-scargle periodogram method. In: *Annales Geophysicae*, vol. 16, pp. 356-358. Copernicus (1998)
  - [13] Schafer, JL; Olsen, MK, Multiple imputation for multivariate missing-data problems: a data analyst's perspective, *Multivar. Behav. Res.*, 33, 545-571, (1998)
  - [14] Isaksson, AJ, Identification of arx-models subject to missing data, *IEEE Trans. Autom. Control*, 38, 813-819, (1993) · [Zbl 0785.93028](#)
  - [15] Waele, S; Broersen, PMT, Error measures for resampled irregular data, *IEEE Trans. Instrum. Meas.*, 49, 216-222, (2000)
  - [16] Liu, S; Molenaar, PC, Ivar: a program for imputing missing data in multivariate time-series using vector autoregressive models, *Behav. Res. Methods*, 46, 1138-1148, (2014)
  - [17] Junger, W; Leon, AP, Imputation of missing data in time-series for air pollutants, *Atmos. Environ.*, 102, 96-104, (2015)
  - [18] Donoho, DL, Compressed sensing, *IEEE Trans. Inf. Theory*, 52, 1289-1306, (2006) · [Zbl 1288.94016](#)
  - [19] Baccalá, LA; Sameshima, K, Partial directed coherence: a new concept in neural structure determination, *Biol. Cybern.*, 84, 463-474, (2001) · [Zbl 1160.92306](#)
  - [20] Gigi, S; Tangirala, A, Reconstructing plant connectivity using directed spectral decomposition, *IFAC Proc. Vol.*, 45, 481-486, (2012)
  - [21] Granger, CW, Investi gating causal relations by econometric models and cross-spectral methods, *Econometrica*, 37, 424-438, (1969) · [Zbl 1366.91115](#)
  - [22] Gigi, S; Tangirala, AK, Quantitative analysis of directional strengths in jointly stationary linear multivariate processes, *Biol. Cybern.*, 103, 119-133, (2010) · [Zbl 1266.92002](#)
  - [23] Eichler, M, A graphical approach for evaluating effective connectivity in neural systems, *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 360, 953-967, (2005)
  - [24] Eichler, M, On the evaluation of information flow in multivariate systems by the directed transfer function, *Biol. Cybern.*, 94, 469-482, (2006) · [Zbl 1138.62048](#)
  - [25] Bahadori, M.T., Liu, Y.: Granger causality analysis in irregular time-series. In: *SDM*, pp. 660-671. SIAM (2012)
  - [26] Elad, M.: *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st edn. Springer, Berlin (2010) · [Zbl 1211.94001](#)
  - [27] Candes, EJ, The restricted isometry property and its implications for compressed sensing, *Comptes Rendus Math.*, 346, 589-592, (2008) · [Zbl 1153.94002](#)
  - [28] Perepu, SK; Tangirala, AK, Reconstruction of missing data using compressed sensing techniques with adaptive dictionary, *J. Process Control*, 47, 175-190, (2016)
  - [29] Wiener, N, The theory of prediction, *Mod. Math. Eng.*, 1, 125-139, (1956)
  - [30] Granger, CW, Some recent development in a concept of causality, *J. Econom.*, 39, 199-211, (1988)
  - [31] Lütkepohl, H.: *New Introduction to Multiple Time-Series Analysis*. Springer, Berlin (2005) · [Zbl 1072.62075](#)
  - [32] Garg, A; Tangirala, AK, Interaction assessment in multivariable control systems through causality analysis, *IFAC Proc. Vol.*, 47, 585-592, (2014)
  - [33] Ambat, S.K., Hari, K., et al.: Fusion of sparse reconstruction algorithms for multiple measurement vectors. arXiv preprint arXiv:1504.01705 (2015) · [Zbl 1365.94068](#)
  - [34] Wooten, R.: Statistical analysis of the relationship between wind speed, pressure and temperature. In: *Sixth International Conference on Dynamic Systems and Applications* (2011)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.