Virgo detector characterization and data quality : tools

arXiv:2210.15634v2 [gr-qc] 25 Mar 2023

Virgo Detector Characterization and Data Quality: tools

F Acernese[1,2], M Agathos[3], A Ain[4], S Albanesi[5,6],
A Allocca[7,2], A Amato[8], T Andrade[9], N Andres[10],
M Andrés-Carcasona[11], T Andrić[12], S Ansoldi[13,14],
S Antier[15,16], T Apostolatos[17], E Z Appavuravther[18,19],
M Arène[20], N Arnaud[21,22], M Assiduo[23,24],
S Assis de Souza Melo[22], P Astone[25], F Aubin[24], S Babak[20],
F Badaracco[26], M K M Bader[27], S Bagnasco[6], J Baird[20],
T Baka[28], G Ballardin[22], G Baltus[29], B Banerjee[12],
C Barbieri[30,31,32], P Barneo[9], F Barone[33,2], M Barsuglia[20],
D Barta[34], A Basti[35,4], M Bawaj[18,36], M Bazzan[37,38],
F Beirnaert[39], M Bejger[40], I Belahcene[21], V Benedetto[41],
M Berbel[42], S Bernuzzi[3], D Bersanetti[43], A Bertolini[27],
U Bhardwaj[16,27], A Bianchi[27,44], S Bini[45,46], M Bischi[23,24],
M Bitossi[22,4], M-A Bizouard[15], F Bobba[47,48], M Boër[15],
G Bogaert[15], M Boldrini[49,25], L D Bonavena[37], F Bondu[50],
R Bonnand[10], B A Boom[27], V Boschi[4], V Boudart[29],
Y Bouffanais[37,38], A Bozzi[22], C Bradaschia[4], M Branchesi[12,51],
M Breschi[3], T Briant[52], A Brillet[15], J Brooks[22], G Bruno[26],
F Bucci[24], T Bulik[53], H J Bulten[27], D Buskulic[10], C Buy[54],
G S Cabourn Davies[55], G Cabras[13,14], R Cabrita[26],
G Cagnoli[8], E Calloni[7,2], M Canepa[56,43], S Canevarolo[28],
M Cannavacciuolo[47], E Capocasa[20], G Carapella[47,48],
F Carbognani[22], M Carpinelli[57,58,22], G Carullo[35,4],
J Casanueva Diaz[22], C Casentini[59,60], S Caudill[27,28],
F Cavalier[21], R Cavalieri[22], G Cella[4], P Cerdá-Durán[61],
E Cesarini[60], W Chaibi[15], P Chanial[22],
E Chassande-Mottin[20], S Chaty[20], F Chiadini[62,48],
G Chiarini[38], R Chierici[63], A Chincarini[43], M L Chiofalo[35,4],
A Chiummo[22], S Choudhary[64], N Christensen[15],
G Ciani[37,38], P Ciecielag[40], M Cieślar[40], M Cifaldi[59,60],
R Ciolfi[65,38], F Cipriano[15], S Clesse[66], F Cleva[15], E Coccia[12,51],
E Codazzo[12], P-F Cohadon[52], D E Cohen[21],
A Colombo[30,31], M Colpi[30,31], L Conti[38],
I Cordero-Carrión[67], S Corezzi[36,18], D Corre[21], S Cortese[22],
J-P Coulon[15], M Croquette[52], J R Cudell[29], E Cuoco[22,68,4],

M Curyło[53], P Dabadie[8], T Dal Canton[21], S Dall'Osso[12],
G Dálya[39], B D'Angelo[56,43], S Danilishin[69,27], S D'Antonio[60],
V Dattilo[22], M Davier[21], D Davis[70], J Degallaix[71],
M De Laurentis[7,2], S Deléglise[52], F De Lillo[26],
D Dell'Aquila[57], W Del Pozzo[35,4], F De Matteis[59,60],
A Depasse[26], R De Pietri[72,73], R De Rosa[7,2], C De Rossi[22],
R De Simone[62], L Di Fiore[2], C Di Giorgio[47,48],
F Di Giovanni[61], M Di Giovanni[12], T Di Girolamo[7,2],
A Di Lieto[35,4], A Di Michele[36], S Di Pace[49,25],
I Di Palma[49,25], F Di Renzo[35,4], L D'Onofrio[7,2],
M Drago[49,25], J-G Ducoin[21], U Dupletsa[12], O Durante[47,48],
D D'Urso[57,58], P-A Duverne[21], M Eisenmann[10], L Errico[7,2],
D Estevez[74], F Fabrizi[23,24], F Faedi[24], V Fafone[59,60,12],
S Farinon[43], G Favaro[37], M Fays[29], E Fenyvesi[34,75],
I Ferrante[35,4], F Fidecaro[35,4], P Figura[53], A Fiori[4,35],
I Fiori[22], R Fittipaldi[76,48], V Fiumara[77,48], R Flaminio[10,78],
J A Font[61,79], S Frasca[49,25], F Frasconi[4], A Freise[27,44],
O Freitas[80], G G Fronzé[6], B U Gadre[81,28], R Gamba[3],
B Garaventa[43,56], F Garufi[7,2], G Gemme[43], A Gennai[4],
Archisman Ghosh[39], B Giacomazzo[30,31,32], L Giacoppo[49,25],
P Giri[4,35], F Gissi[41], S Gkaitatzis[4,35], B Goncharov[12],
M Gosselin[22], R Gouaty[10], A Grado[82,2], M Granata[71],
V Granata[47], G Greco[18], G Grignani[36,18], A Grimaldi[45,46],
S J Grimm[12,51], P Gruning[21], D Guerra[61], G M Guidi[23,24],
G Guixé[9], Y Guo[27], P Gupta[27,28], L Haegel[20], O Halim[14],
O Hannuksela[28,27], T Harder[15], K Haris[27,28], J Harms[12,51],
B Haskell[40], A Heidmann[52], H Heitmann[15], P Hello[21],
G Hemming[22], E Hennes[27], S Hild[69,27], D Hofman[71],
V Hui[10], B Idzkowski[53], A Iess[59,60], P Iosif[83], T Jacqmin[52],
P-E Jacquet[52], S P Jadhav[64], J Janquart[28,27], K Janssens[84,15],
P Jaranowski[85], V Juste[74], C Kalaghatgi[28,27,86],
C Karathanasis[11], S Katsanevas[22]‡, F Kéfélian[15],
N Khetan[12,51], G Koekoek[27,69], S Koley[12], M Kolstein[11],
A Królak[87,88], P Kuijer[27], P Lagabbe[10], D Laghi[54],
M Lalleman[84], A Lamberts[15,89], I La Rosa[10],
A Lartaux-Vollard[21], C Lazzaro[37,38], P Leaci[49,25], A Lemaître[90],
M Lenti[24,91], E Leonova[16], N Leroy[21], N Letendre[10],
K Leyde[20], F Linde[86,27], L London[16], A Longo[92],
M Lopez Portilla[28], M Lorenzini[59,60], V Loriette[93],
G Losurdo[4], D Lumaca[59,60], A Macquet[15], C Magazzù[4],

‡ Deceased, November 2022.

M Magnozzi[43,56], E Majorana[49,25], I Maksimovic[93], N Man[15],
V Mangano[49,25], M Mantovani[22], M Mapelli[37,38],
F Marchesoni[19,18,94], D Marín Pina[9], F Marion[10],
A Marquina[67], S Marsat[20], F Martelli[23,24], M Martinez[11],
V Martinez[8], A Masserot[10], S Mastrogiovanni[20], Q Meijer[28],
A Menendez-Vazquez[11], L Mereni[71], M Merzougui[15],
A Miani[45,46], C Michel[71], L Milano[7§], A Miller[26],
B Miller[16,27], E Milotti[95,14], Y Minenkov[60], Ll M Mir[11],
M Miravet-Tenés[61], M Montani[23,24], F Morawski[40],
B Mours[74], C M Mow-Lowry[27,44], S Mozzon[55],
F Muciaccia[49,25], Suvodip Mukherjee[16], R Musenich[43,56],
A Nagar[6,96], V Napolano[22], I Nardecchia[59,60], H Narola[28],
L Naticchioni[25], J Neilson[41,48], C Nguyen[20], S Nissanke[16,27],
E Nitoglia[63], F Nocera[22], G Oganesyan[12,51], C Olivetto[22],
G Pagano[35,4], G Pagliaroli[12,51], C Palomba[25], P T H Pang[27,28],
F Pannarale[49,25], F Paoletti[4], A Paoli[22], A Paolone[25,97],
G Pappas[83], D Pascucci[27,39], A Pasqualetti[22],
R Passaquieti[35,4], D Passuello[4], B Patricelli[22,4],
R Pedurand[48], M Pegoraro[38], A Perego[45,46], A Pereira[8],
C Périgois[10], A Perreca[45,46], S Perriès[63], D Pesios[83],
K S Phukon[27,86], O J Piccinni[25], M Pichot[15],
M Piendibene[35,4], F Piergiovanni[23,24], L Pierini[49,25],
V Pierro[41,48], G Pillant[22], M Pillas[21], F Pilo[4], L Pinard[71],
I M Pinto[41,48,98], M Pinto[22], K Piotrzkowski[26], A Placidi[18,36],
E Placidi[49,25], W Plastino[99,92], R Poggiani[35,4], E Polini[10],
E K Porter[20], R Poulton[22], M Pracchia[10], T Pradier[74],
M Principe[41,98,48], G A Prodi[100,46], P Prosposito[59,60],
A Puecher[27,28], M Punturo[18], F Puosi[4,35], P Puppo[25],
G Raaijmakers[16,27], N Radulesco[15], P Rapagnani[49,25],
M Razzano[35,4], T Regimbau[10], L Rei[43], P Rettegno[5,6],
B Revenu[20], A Reza[27], F Ricci[49,25], G Riemenschneider[5,6],
S Rinaldi[35,4], F Robinet[21], A Rocchi[60], L Rolland[10],
M Romanelli[50], R Romano[1,2], A Romero[11], S Ronchini[12,51],
L Rosa[2,7], D Rosińska[53], S Roy[28], D Rozza[57,58], P Ruggi[22],
J Sadiq[101], O S Salafia[32,31,30], L Salconi[22], F Salemi[45,46],
A Samajdar[31], N Sanchis-Gual[102], A Sanuy[9], B Sassolas[71],
S Sayah[71], S Schmidt[28], M Seglar-Arroyo[10], D Sentenac[22],
V Sequino[7,2], Y Setyawati[28], A Sharma[12,51],
N S Shcheblanov[90], M Sieniawska[26], L Silenzi[18,19],
N Singh[53], A Singha[69,27], V Sipala[57,58], J Soldateschi[91,103,24],

K Soni[64], V Sordini[63], F Sorrentino[43], N Sorrentino[35,4],
R Soulard[15], V Spagnuolo[69,27], M Spera[37,38], P Spinicelli[22],
C Stachie[15], D A Steer[20], J Steinlechner[69,27],
S Steinlechner[69,27], N Stergioulas[83], G Stratta[104,25],
M Suchenek[40], A Sur[40], B L Swinkels[27], P Szewczyk[53],
M Tacca[27], A J Tanasijczuk[26], E N Tapia San Martín[27],
C Taranto[59], A E Tolley[55], M Tonelli[35,4], A Torres-Forné[61],
I Tosta e Melo[58], A Trapananti[19,18], F Travasso[18,19],
M Trevor[105], M C Tringali[22], L Troiano[106,48], A Trovato[20],
L Trozzo[2], K W Tsang[27,107,28], K Turbang[108,84], M Turconi[15],
A Utina[69,27], M Valentini[45,46], N van Bakel[27],
M van Beuzekom[27], M van Dael[27,109],
J F J van den Brand[69,44,27], C Van Den Broeck[28,27],
H van Haevermaet[84], J V van Heijningen[26],
N van Remortel[84], M Vardaro[86,27], M Vasúth[34],
G Vedovato[38], D Verkindt[10], P Verma[88], F Vetrano[23],
A Viceré[23,24], V Villa-Ortega[101], J-Y Vinet[15], A Virtuoso[95,14],
H Vocca[36,18], R C Walet[27], M Was[10], A R Williamson[55],
J L Willis[70], A Zadrożny[88], T Zelenova[22], and J-P Zendri[38]

[1] Dipartimento di Farmacia, Università di Salerno, I-84084 Fisciano, Salerno, Italy
[2] INFN, Sezione di Napoli, Complesso Universitario di Monte S. Angelo, I-80126 Napoli, Italy
[3] Theoretisch-Physikalisches Institut, Friedrich-Schiller-Universität Jena, D-07743 Jena, Germany
[4] INFN, Sezione di Pisa, I-56127 Pisa, Italy
[5] Dipartimento di Fisica, Università degli Studi di Torino, I-10125 Torino, Italy
[6] INFN Sezione di Torino, I-10125 Torino, Italy
[7] Università di Napoli "Federico II", Complesso Universitario di Monte S. Angelo, I-80126 Napoli, Italy
[8] Université de Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-69622 Villeurbanne, France
[9] Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona, C/ Martí i Franquès 1, Barcelona, 08028, Spain
[10] Univ. Savoie Mont Blanc, CNRS, Laboratoire d'Annecy de Physique des Particules - IN2P3, F-74000 Annecy, France
[11] Institut de Física d'Altes Energies (IFAE), Barcelona Institute of Science and Technology, and ICREA, E-08193 Barcelona, Spain
[12] Gran Sasso Science Institute (GSSI), I-67100 L'Aquila, Italy
[13] Dipartimento di Scienze Matematiche, Informatiche e Fisiche, Università di Udine, I-33100 Udine, Italy
[14] INFN, Sezione di Trieste, I-34127 Trieste, Italy
[15] Artemis, Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, F-06304 Nice, France
[16] GRAPPA, Anton Pannekoek Institute for Astronomy and Institute for High-Energy Physics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands

[17]Department of Physics, National and Kapodistrian University of Athens, School of Science Building, 2nd floor, Panepistimiopolis, 15771 Ilissia, Greece

[18]INFN, Sezione di Perugia, I-06123 Perugia, Italy

[19]Università di Camerino, I-62032 Camerino, Italy

[20]Université de Paris, CNRS, Astroparticule et Cosmologie, F-75006 Paris, France

[21]Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France

[22]European Gravitational Observatory (EGO), I-56021 Cascina, Pisa, Italy

[23]Università degli Studi di Urbino "Carlo Bo", I-61029 Urbino, Italy

[24]INFN, Sezione di Firenze, I-50019 Sesto Fiorentino, Firenze, Italy

[25]INFN, Sezione di Roma, I-00185 Roma, Italy

[26]Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

[27]Nikhef, Science Park 105, 1098 XG Amsterdam, Netherlands

[28]Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, Princetonplein 1, 3584 CC Utrecht, Netherlands

[29]Université de Liège, B-4000 Liège, Belgium

[30]Università degli Studi di Milano-Bicocca, I-20126 Milano, Italy

[31]INFN, Sezione di Milano-Bicocca, I-20126 Milano, Italy

[32]INAF, Osservatorio Astronomico di Brera sede di Merate, I-23807 Merate, Lecco, Italy

[33]Dipartimento di Medicina, Chirurgia e Odontoiatria "Scuola Medica Salernitana", Università di Salerno, I-84081 Baronissi, Salerno, Italy

[34]Wigner RCP, RMKI, H-1121 Budapest, Konkoly Thege Miklós út 29-33, Hungary

[35]Università di Pisa, I-56127 Pisa, Italy

[36]Università di Perugia, I-06123 Perugia, Italy

[37]Università di Padova, Dipartimento di Fisica e Astronomia, I-35131 Padova, Italy

[38]INFN, Sezione di Padova, I-35131 Padova, Italy

[39]Universiteit Gent, B-9000 Gent, Belgium

[40]Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, 00-716, Warsaw, Poland

[41]Dipartimento di Ingegneria, Università del Sannio, I-82100 Benevento, Italy

[42]Departamento de Matemáticas, Universitat Autònoma de Barcelona, Edificio C Facultad de Ciencias 08193 Bellaterra (Barcelona), Spain

[43]INFN, Sezione di Genova, I-16146 Genova, Italy

[44]Vrije Universiteit Amsterdam, 1081 HV Amsterdam, Netherlands

[45]Università di Trento, Dipartimento di Fisica, I-38123 Povo, Trento, Italy

[46]INFN, Trento Institute for Fundamental Physics and Applications, I-38123 Povo, Trento, Italy

[47]Dipartimento di Fisica "E.R. Caianiello", Università di Salerno, I-84084 Fisciano, Salerno, Italy

[48]INFN, Sezione di Napoli, Gruppo Collegato di Salerno, Complesso Universitario di Monte S. Angelo, I-80126 Napoli, Italy

[49]Università di Roma "La Sapienza", I-00185 Roma, Italy

[50]Univ Rennes, CNRS, Institut FOTON - UMR6082, F-3500 Rennes, France

[51]INFN, Laboratori Nazionali del Gran Sasso, I-67100 Assergi, Italy

[52]Laboratoire Kastler Brossel, Sorbonne Université, CNRS, ENS-Université PSL, Collège de France, F-75005 Paris, France

[53]Astronomical Observatory Warsaw University, 00-478 Warsaw, Poland

[54]L2IT, Laboratoire des 2 Infinis - Toulouse, Université de Toulouse, CNRS/IN2P3, UPS, F-31062 Toulouse Cedex 9, France

[55]University of Portsmouth, Portsmouth, PO1 3FX, United Kingdom

[56]Dipartimento di Fisica, Università degli Studi di Genova, I-16146 Genova, Italy

[57]Università degli Studi di Sassari, I-07100 Sassari, Italy

[58]INFN, Laboratori Nazionali del Sud, I-95125 Catania, Italy

[59]Università di Roma Tor Vergata, I-00133 Roma, Italy

[60]INFN, Sezione di Roma Tor Vergata, I-00133 Roma, Italy

[61]Departamento de Astronomía y Astrofísica, Universitat de València, E-46100 Burjassot, València, Spain

[62]Dipartimento di Ingegneria Industriale (DIIN), Università di Salerno, I-84084 Fisciano, Salerno, Italy

[63]Université Lyon, Université Claude Bernard Lyon 1, CNRS, IP2I Lyon / IN2P3, UMR 5822, F-69622 Villeurbanne, France

[64]Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411 007, India

[65]INAF, Osservatorio Astronomico di Padova, I-35122 Padova, Italy

[66]Université libre de Bruxelles, Avenue Franklin Roosevelt 50 - 1050 Bruxelles, Belgium

[67]Departamento de Matemáticas, Universitat de València, E-46100 Burjassot, València, Spain

[68]Scuola Normale Superiore, Piazza dei Cavalieri, 7 - 56126 Pisa, Italy

[69]Maastricht University, P.O. Box 616, 6200 MD Maastricht, Netherlands

[70]LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

[71]Université Lyon, Université Claude Bernard Lyon 1, CNRS, Laboratoire des Matériaux Avancés (LMA), IP2I Lyon / IN2P3, UMR 5822, F-69622 Villeurbanne, France

[72]Dipartimento di Scienze Matematiche, Fisiche e Informatiche, Università di Parma, I-43124 Parma, Italy

[73]INFN, Sezione di Milano Bicocca, Gruppo Collegato di Parma, I-43124 Parma, Italy

[74]Université de Strasbourg, CNRS, IPHC UMR 7178, F-67000 Strasbourg, France

[75]Institute for Nuclear Research, Bem t'er 18/c, H-4026 Debrecen, Hungary

[76]CNR-SPIN, c/o Università di Salerno, I-84084 Fisciano, Salerno, Italy

[77]Scuola di Ingegneria, Università della Basilicata, I-85100 Potenza, Italy

[78]Gravitational Wave Science Project, National Astronomical Observatory of Japan (NAOJ), Mitaka City, Tokyo 181-8588, Japan

[79]Observatori Astronòmic, Universitat de València, E-46980 Paterna, València, Spain

[80]Centro de Física das Universidades do Minho e do Porto, Universidade do Minho, Campus de Gualtar, PT-4710 - 057 Braga, Portugal

[81]Max Planck Institute for Gravitational Physics (Albert Einstein Institute), D-14476 Potsdam, Germany

[82]INAF, Osservatorio Astronomico di Capodimonte, I-80131 Napoli, Italy

[83]Department of Physics, Aristotle University of Thessaloniki, University Campus, 54124 Thessaloniki, Greece

[84]Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen, Belgium

[85]University of Białystok, 15-424 Białystok, Poland

[86]Institute for High-Energy Physics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands

[87]Institute of Mathematics, Polish Academy of Sciences, 00656 Warsaw, Poland

[88]National Center for Nuclear Research, 05-400 Świerk-Otwock, Poland

[89]Laboratoire Lagrange, Université Côte d'Azur, Observatoire Côte d'Azur, CNRS, F-06304 Nice, France

[90]NAVIER, École des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France

[91]Università di Firenze, Sesto Fiorentino I-50019, Italy

[92]INFN, Sezione di Roma Tre, I-00146 Roma, Italy
[93]ESPCI, CNRS, F-75005 Paris, France
[94]School of Physics Science and Engineering, Tongji University, Shanghai 200092, China
[95]Dipartimento di Fisica, Università di Trieste, I-34127 Trieste, Italy
[96]Institut des Hautes Etudes Scientifiques, F-91440 Bures-sur-Yvette, France
[97]Consiglio Nazionale delle Ricerche - Istituto dei Sistemi Complessi, Piazzale Aldo Moro 5, I-00185 Roma, Italy
[98]Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi", I-00184 Roma, Italy
[99]Dipartimento di Matematica e Fisica, Università degli Studi Roma Tre, I-00146 Roma, Italy
[100]Università di Trento, Dipartimento di Matematica, I-38123 Povo, Trento, Italy
[101]Instituto Galego de Física de Altas Enerxías, Universidade de Santiago de Compostela, 15782, Santiago de Compostela, Spain
[102]Departamento de Matemática da Universidade de Aveiro and Centre for Research and Development in Mathematics and Applications, Campus de Santiago, 3810-183 Aveiro, Portugal
[103]INAF, Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy
[104]Istituto di Astrofisica e Planetologia Spaziali di Roma, Via del Fosso del Cavaliere, 100, 00133 Roma RM, Italy
[105]University of Maryland, College Park, MD 20742, USA
[106]Dipartimento di Scienze Aziendali - Management and Innovation Systems (DISA-MIS), Università di Salerno, I-84084 Fisciano, Salerno, Italy
[107]Van Swinderen Institute for Particle Physics and Gravity, University of Groningen, Nijenborgh 4, 9747 AG Groningen, Netherlands
[108]Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
[109]Eindhoven University of Technology, Postbus 513, 5600 MB Eindhoven, Netherlands

Abstract. Detector characterization and data quality studies — collectively referred to as *DetChar* activities in this article — are paramount to the scientific exploitation of the joint dataset collected by the LIGO-Virgo-KAGRA global network of ground-based gravitational-wave (GW) detectors. They take place during each phase of the operation of the instruments (upgrade, tuning and optimization, data taking), are required at all steps of the dataflow (from data acquisition to the final list of GW events) and operate at various latencies (from near real-time to vet the public alerts to offline analyses). This work requires a wide set of tools which have been developed over the years to fulfill the requirements of the various DetChar studies: data access and bookkeeping; global monitoring of the instruments and of the different steps of the data processing; studies of the global properties of the noise at the detector outputs; identification and follow-up of noise peculiar features (whether they be transient or continuously present in the data); quick processing of the public alerts. The present article reviews all the tools used by the Virgo DetChar group during the third LIGO-Virgo Observation Run (O3, from April 2019 to March 2020), mainly to analyse the Virgo data acquired at EGO. Concurrently, a companion article focuses on the results achieved by the DetChar group during the O3 run using these tools.

Contents

1. Introduction

1.1. Background: DetChar inputs to detect and study gravitational waves

GWTC-3 [1], the most recent edition of the gravitational wave (GW) Transient Catalog edited by the Laser Interferometer Gravitational-wave Observatory (LIGO) [2], Virgo [3], and now Kamioka Gravitational-Wave Detector (KAGRA) [4] scientific collaborations includes 90 GW signals recorded between 2015 and 2020 during three successive data-taking campaigns called Observation Runs (in short O_n with $n = 1, 2, 3$). These discoveries are the result of the joint work of hundreds of scientists worldwide, bringing together a wide range of expertise, ranging from the instrumental side to data analyses. Important components of this global effort are the DetChar activities [5, 6] that focus on studying the detector noises in all their variety: detector characterization on the one hand and data quality studies on the other.

Indeed, the Virgo GW strain channel $h(t)$, reconstructed from its raw data, is dominated by noises of various origins (fundamental, technical or environmental [7]), with different and time-varying characteristics (amplitude and frequency contents). Detector characterization targets the smooth and usually stationary noise floor, which makes the envelope of the sensitivity curve. Beyond that, two main categories of noise artifacts are studied in detail as they can impact the performances of the instrument in detecting genuine GW signals. The first one includes all noise transients, also called *glitches*, while the second one gathers all long-lasting noise excesses, the *spectral noises* (i.e. *lines* or *bumps* depending on their bandwidth type, narrow or wide, in the frequency domain).

To help understand the source of some of the noises that affect the Virgo sensitivity, hundreds of *auxiliary channels* monitor continuously the detector control systems as well as its local environment [8]. They are also useful to vet the GW candidates by assessing whether or not they seem to be of terrestrial origin.

1.2. The LIGO-Virgo O3 run

The third LIGO-Virgo Observing Run (O3) lasted about 11 months in total. It was divided into two parts: O3a, from April 1st, 2019 to October 1st, 2019; O3b, from November 1st, 2019 to March 27th, 2020, separated by a one-month commissioning break in October 2019. O3b should have lasted one more month but the worldwide Covid-19 pandemic forced the LIGO Scientific and Virgo Collaborations to end the data taking prematurely. A few days later, all three detectors were shutdown to cope with the various lock-down constraints.

The Virgo detector takes data in a configuration called *Science mode*. It corresponds to periods during which the instrument is controlled at its nominal working point, with that control stable and accurate enough to assume that the recorded data are of good quality

and suitable for physics analysis. This assumption is checked in real time against online data quality checks and the corresponding dataset is further refined by offline studies to make the final Virgo dataset.

1.3. Running the Virgo DetChar analysis tools

All DetChar analyses rely on dedicated software frameworks, generically called *tools* in the following. Some have been designed and set up within Virgo to meet the goals of the DetChar group, while others have been developed partly or totally by LIGO colleagues. In fact, any DetChar tool can potentially be used by the three groups (Virgo, LIGO, and KAGRA) thanks to the long-lasting collaborations among them.

More than 100 computing servers have been used in real-time during O3 to control and monitor the Virgo detector, run various data quality checks and perform specific DetChar tasks. Data are processed by the tools described in the following sections. Their outputs are included in the live data streams if they are available with a latency low-enough (about 15 s), or stored on disk otherwise. The end products of these analyses are converted into information for the control room and live summary plots that are updated with a latency of a few minutes at most, and regularly archived for reuse during offline analyses.

All this software framework is controlled using the Virgo Process Monitoring (VPM) software interface, that allows to configure, start/stop and monitor processes running on Virgo online servers. These include detector control, data transfer to and from Virgo, as well as the analysis of the reconstructed $h(t)$ stream by the online GW search algorithms running in the European Gravitational Observatory (EGO) computing center. All actions performed using the VPM interface are logged and recorded, in order to reconstruct as accurately as possible the running conditions at any given time, should this need arise.

The most important DetChar tools used by the Virgo group during the O3 run are classified in a few main categories depending on their usage or target: monitoring, generic data analysis, glitches and spectral noise investigations, or database management. Yet, they are not independent: they are often combined to characterize specific features of the detector, or to provide a complete overview of the quality of the Virgo data around a GPS time of interest. Such GPS ranges are called *segments*.

The flowchart in figure 1 presents an overview of the main analyses carried out by the DetChar group and shows the corresponding tools described in the following. The arrows follow the dataflow which starts from the detector raw data (top left corner) and goes all the way down to the final consumers of DetChar products: the on-duty crew in the Virgo control room, the broad community of DetChar users and the data analysts.
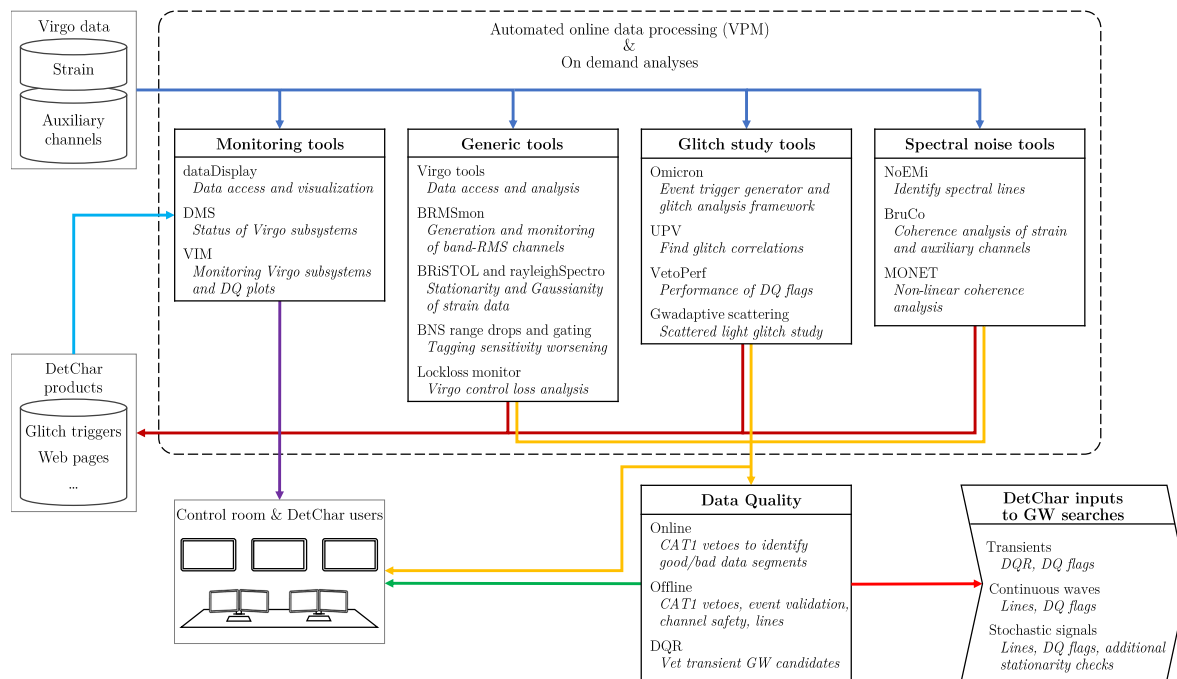
Figure 1. Flowchart of the various tools and monitors used for the Virgo DetChar studies during the O3 run — the results of these analyses are presented in the companion article [9]. The dataflow starts from the raw data acquired from the detector. These are then analysed by a wide set of DetChar tools, which in turn produce outputs of various types, used at different analysis levels. Either to monitor the detector and the data taking, or to construct data quality (in short "DQ") estimators, to be used by GW search pipelines.

1.4. Article contents

This article categorizes the different DetChar tools used during the O3 run. It aims at becoming the main reference for those which had not been documented yet, while it puts all of them into perspective, by showing how they interact and complement each other.

First, section 2 presents the main monitoring tools commonly used to make quick and basic analysis of the Virgo data, or to get a digest of the detector status and of its performance at any given time. Then, Section 3 highlights a wide set of DetChar-generic tools: a software layer to ease access to the Virgo data for the users; band-limited root mean square (RMS)-based estimates as quantities to identify noise contributions and follow their evolution over time; different methods to test the noise stationarity and Gaussianity, two hypothesis which are often explicitly or implicitly assumed by many data analyses; finally, some tools detecting and studying peculiarities during data taking phases, that is sudden drops of the sensitivity and losses of the detector global working point.

Section 4 focuses on transient noise bursts, or glitches. It includes well-established tools to identify them in time-frequency representations and to search for their origins. In

addition, more recent scattered light monitors are described as well, as this kind of noise appears to be a real nuisance for all ground-based GW detectors. The following section 5 is dedicated to the tools used to identify spectral noise features and to search for their origin. For the latter part of the investigations, coherence in the frequency domain is the key quantity used to connect spectral noise in the GW strain $h(t)$ with auxiliary channels.

Then, the short section 6 introduces two databases jointly used by LIGO and Virgo, one to gather data quality inputs for the analyses, and the other to store information about all GW candidates found by those searches. Finally, section 7 presents the Data Quality Report (`DQR`), the new DetChar framework developed specifically for the O3 run in order to vet the GW candidates, in particular those found in low latency by the online data analyses.

Section 8 concludes this article by providing the outlines of the DetChar technical work to prepare the O4 run[1], using all the experience accumulated during O3, and later on to analyse its dataset.

The main results of the Virgo DetChar analyses performed on data from O3 are presented in the companion article [9].

The abbreviations and acronyms used in this article are defined in a dedicated section at the end.

2. Monitoring tools

2.1. dataDisplay

The `dataDisplay` software [10] allows the user to read (online or offline) Virgo data and to visualize various types of plots for all the channels available from the data acquisition system (DAQ). For instance, it helps to investigate quickly the time evolution of a noise artifact, the coherence between two control signals or the time-frequency characteristics of a transient noise, etc. It has been used extensively during the O2[2] and O3 runs and all over the Advanced Virgo (AdV) detector commissioning in between. Figure 2 shows an example of the `dataDisplay` interface and output.

2.2. DMS: the Detector Monitoring System

The Detector Monitoring System (`DMS`) [11, 12] provides a detailed live status of all the components that make the Virgo detector operate, from the hardware parts to the online

---

1 The fourth LIGO-Virgo-KAGRA Observing Run O4, is currently expected to start in Spring 2023.
2 O2 was the second LIGO-Virgo Observing Run. It started on November 30th, 2016 only with the two LIGO detectors. Virgo joined O2 on August 1st, 2017 and the three detectors took data jointly until August 25th.
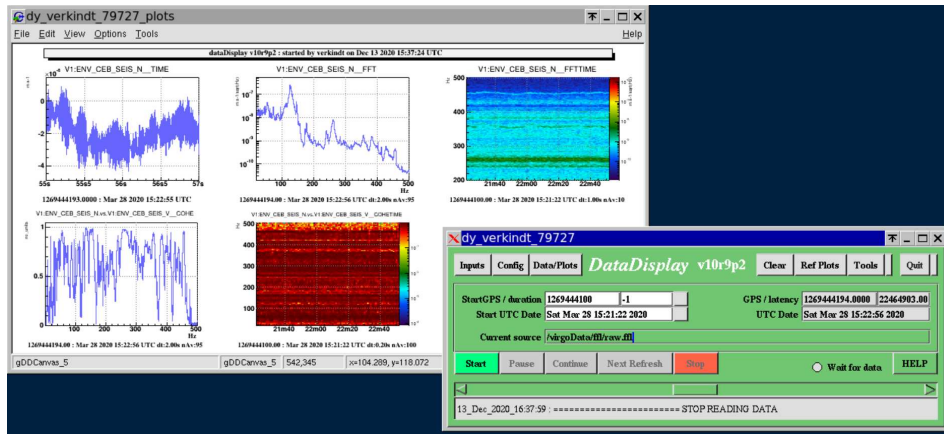
Figure 2. Example of the plots produced by the `dataDisplay` (left) and main panel of its graphical user interface (right).

software used to control the instrument, acquire the data and process it. It also includes the monitoring of environmental sensors from around the experimental areas. Each of the many `DMS` monitors uses a set of DAQ channels, combines them by performing mathematical and logical operations on their outputs and produces a flag whose value can take four severity levels, each associated with a color for visual display. A web interface is used to display and browse the `DMS` monitor flags with a few second-latency, both in the Virgo control room and remotely.

For instance, figure 3 shows the Virgo detector status about 4 s after the detection of the GW event GW190412 [13]. The `DMS` web interface looks like a checkerboard. Each row, labeled in the most-left column, corresponds to a different part of the instrument (mirror suspensions, vacuum system, etc.). Moving to the right of the screenshot, that part is broken down in smaller sets that are each associated with a cell on the web interface. Each cell can contain many `DMS` flags and its color reflects the highest severity among all these flags (green ↔ no alarm; yellow ↔ warning; red ↔ alarm (not present on that particular snapshot); grey ↔ some information is missing). Clicking on a cell gives access to the flag individual information: their values and associated severities.

In addition to the live global detector status, a new `DMS` archival system has been set up for the O3 run: complete `DMS` snapshots are taken every ∼10 s and archived. They can be retrieved later at any time, by running a playback application that uses the same interface as the live `DMS`. This functionality is particularly convenient to check the status of the detector a posteriori, when a GW candidate or a particular feature in the data have been identified.

2.3. VIM: the Virgo Interferometer Monitor

The Virgo Interferometer Monitor (`VIM`) [14, 15] manages a collection of automated scripts that update every few minutes a wide set of plots and tables; all these monitoring

DMS | ITF Mode: **Science** (0d 19h 27m 7s) | ITF State: **LOW_NOISE_3_SQZ** (0d 19h 50m 44s) | UTC: 2019-04-12 05:30:48

Injection	SIB1_IP	SIB1_BENCH	SIB1_BR	SIB1_Vert	SIB1_TE	SIB1_Guard	SIB1_Electr
	MC_IP	MC_PAY	MC_BR	MC_Vert	MC_TE	MC_Guard	MC_Electr
	Laser	LaserAmpli	LaserChiller	SL_TempController	RFC	LNFS	PC
	MC_Power	PSTAB	IMC_AA	IMC_AA_GALVO	MC_F0_z	BPC	BPC_Electr
Detection	PD	QPD_B1p	QPD_B2	QPD_B5	OMC	PicoDisable	Shutter
	SDB1_IP	SDB1_LC	SDB1_BR	SDB1_Vert	SDB1_TE	SDB1_Guard	SDB1_Electr
ISC	B2_8MHz_DPHI	B4_56MHz_DPHI	DARM_UGF	UNLOCK	SSFS_UGF	FmodErr	GIPC
	B1p_DC	B4_112MHz_MAG	B7_DC	B8_DC	LSC_rms	ASC_rms	
Suspensions	BS_IP	BS_F7	BS_PAY	BS_BR	BS_Vert	BS_TE	BS_Guard	BS_Electr
	NI_IP	NI_F7	NI_PAY	NI_BR	NI_Vert	NI_TE	NI_Guard	NI_Electr
	NE_IP	NE_F7	NE_PAY	NE_BR	NE_Vert	NE_TE	NE_Guard	NE_Electr
	PR_IP	PR_F7	PR_PAY	PR_BR	PR_Vert	PR_TE	PR_Guard	PR_Electr
	SR_IP	SR_F7	SR_PAY	SR_BR	SR_Vert	SR_TE	SR_Guard	SR_Electr
	WI_IP	WI_F7	WI_PAY	WI_BR	WI_Vert	WI_TE	WI_Guard	WI_Electr
	WE_IP	WE_F7	WE_PAY	WE_BR	WE_Vert	WE_TE	WE_Guard	WE_Electr
Environment	CB_Hall	MC_Hall	TCS_zones	NE_Hall	WE_Hall	WindActivity	Seismon	BRMSMon
	INJ_Area	DET_Area	EE_Room	DAQ_Room	External	DeadChannel	Lights	SeaActivity	WAB
Infrastructures	ACS_CB_Hall	ACS_TB	ACS_DAQ_Room	ACS_EE_Room	ACS_MC	ACS_INJ	ACS_DET	ACS_NE	ACS_WAB
	UPS_TB	UPS_CB	UPS_MC	UPS_NE	UPS_WE	FlatChannel	ExistChannel	ACS_WE	ACS_CB_CR	ACS_COB
SBE	EIB_SBE	SDB2_SBE	SDB2_LC	SNEB_SBE	SNEB_LC	SWEB_SBE	SWEB_LC	SPRB_SBE	SPRB_LC
TCS	NE_RH	WE_RH	NI_CO2_Laser	WI_CO2_Laser	Chillers		
SQZ	PLL	Squeezer	SQZ_AA	SQZ_Shutter	Cohe_CTRL	SQZ_Inj	Rack_TE
Vacuum	LargeValves	Clean_Air	TubeStations	TubePumps	MiniTowers	TurboLinks	RemDryPMP	VAC_SERVOS
	Pressure	CompressedAir	TowerServers	TowerPumps	CryoTrap	O2_Sensors	Tank
VPM	DetectorSEnvironmentM	ControlRoom	Minitowers	ISC	Injection	TCS	Suspension	Vacuum	Metatron
	DetectorMonitoring	DataCollection	Storage	DataAccess	Automation	DetChar	
DAQ-Computing	Latency	Disk	Timing	Timing_rtpc	Timing_dsp	Fast_DAC	ADCs_TE	Daq_Boxes_TE
	DMS_machines	DetOp_machines	olservers	rtpcs	CoilSwitchBoxes	INF_devices	ENV_devices	VAC_devices
Calib_Hrec	CalNE	CalWE	CalINJ	CalBS	CalPR	PCalNE	PCalWE	HOFT	NCAL	NoiseInjection
ITFOnCall	SoftwareAI	TemperaturesAI	InjectionAI	UpsAI	GeneratorAI	TcsAI	
DetChar	Hrec_RANGE_BNS	flag_AlertGraceDB	STATE_VECTOR			

Figure 3. `DMS` snapshot closest in time to the GW190412 GW event, showing the detailed status of the Virgo detector about 4 s after the arrival of that signal.

products are permanently archived on a daily basis. A web interface allows users to browse that database, both for live monitoring of the experiment and for offline investigations. `VIM` is an essential tool that provides a direct access to a detailed status of the various Virgo detector components and of related frameworks, such as calibration and online data processing, data transfer or online data analyses. A snapshot of the `VIM` web interface is shown in figure 4.

3. Generic tools

3.1. The VirgoTools utilities

In-depth studies of a particular feature observed in the data or analyses scanning a significant fraction of the dataset require the use of dedicated software. Common and key building blocks of these codes have access to the DAQ channels and to the detector component configurations. Thus, dedicated packages have been developed over the years to provide simplified and generic interfaces to these data: they rely on low-level core packages like the `FrameLib` software library [16] but calls to these functions are hidden to the users. These packages interact with the software, hardware and data of the Virgo interferometer: they are widely used within the collaboration, from

Figure 4. Screenshot of a `VIM` web page displaying information about the Virgo detector on Tuesday February 18th, 2020. Top left plot, stripchart of the detector status: the weekly maintenance, preceded by a planned calibration and followed by a short commissioning period, interrupts the data taking that restarts in the evening. Top right plot: daily sensitivity compared with references. Bottom left plot: glitch monitoring provided by the `Omicron` analysis described in section 4.1. Bottom right plot: spectrogram of the GW strain $h(t)$ in the 20-210 Hz frequency range.

daily use in the control room to DetChar studies. The two main collections of such functions are `PythonVirgoTools` [17] and `MatlabVirgoTools`, targeting Python and Matlab developers, respectively.

3.2. Computing Band-limited RMS

Band-limited RMSss (BRMSs) of DAQ channels computed in specific frequency ranges after Fourier-transforming the time series are useful indicators for transient disturbances or new features in the data – see Eq. 1 in Sec. 3.3.1 below. For instance, low-frequency BRMS of seismometer data allow to separate different contributions to the seismic noise at EGO [7]. Going from low to high frequencies, one can isolate successively: distant and potentially strong earthquakes; sea activity on the Tuscany coastline; anthropogenic contributions with day/night and weekly periodicities; finally, on-site activities. In addition, BRMS is used to monitor the excitation of the violin modes, i.e. the resonances of the mirror suspensions.

In Virgo, various software frameworks can compute BRMS. One worth-mentioning is `BRMSMon`, a dedicated software that is widely used by the environmental monitoring team and in data quality studies. In addition to generating BRMS, `BRMSMon` can compare their values to thresholds (either fixed or adaptive) and logically combine the outputs of these

comparisons into binary channels called *flags*. For instance, assuming a collection of 9 sensors installed in different EGO buildings, one can create a flag that is active (value equal to 1) if at least 5 of these 9 sensors exceed their own threshold and inactive (value 0) otherwise. The `BRMSMon` output channels, sampled at 1 Hz, are included in the DAQ.

3.3. Testing stationarity and Gaussianity

Several tools have been implemented to perform statistical tests to verify the stationarity and Gaussianity of the data. These properties are indeed the typical assumptions at the base of most of the statistical analyses of GW search pipelines, such as `MBTA` [18], `PyCBC` [19] and `GstLAL` [20] that are based on the matched filter technique [21, 22]. Moreover, the onset of a non-stationary behavior of the detector can be the symptom of some hardware malfunction or some contamination from environmental noises. In any case, it requires prompt investigations of its causes and, possibly, the actuation of adequate mitigation strategies.

3.3.1. BRiSTOL The detector output records can be described as the realization of a stochastic process. This process is said to be (strict-sense) stationary if its statistical properties are left unchanged by shifts in time, which, in most practical situations, allows us to estimate them from a sufficiently long realization of the process (*ergodic hypothesis*), and this estimate does not depend on when, in time, it has been performed. For most of the analysis, it is sufficient to consider a *weaker* form of stationarity that involves only the first distribution moments, namely the mean and the covariance function. Moreover, if the process is stationary, we can define the power spectral density (PSD) of the process as the Fourier transform of its covariance [22].

The Band-limited RMS Stationarity Test Tool (`BRiSTOL`) statistical test [23] verifies the hypothesis of weak-sense stationarity by verifying that subsequent PSD estimates, in predefined frequency bands, are compatible with the same probability distribution. The corresponding test statistics are based on a set of BRMS time series, estimated on an equal number of bands:

$$BRMS_t(b) = \sqrt{\int_{f \in b} \hat{S}_t(f)\,df}, \qquad \text{for} \quad b \in \left\{ [f_1^{\max}, f_1^{\min}], \ldots, [f_K^{\max}, f_K^{\min}] \right\} \qquad (1)$$

for data x_{t_n} recorded at Nyquist rate f_S, $t_n = t + n/f_S$, where $\hat{S}_t(f)$ is a PSD estimate referred to time t, and obtained with the *periodogram* method [24]:

$$\hat{S}_t(f_k) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x_{t_n} e^{-2\pi i n f_k/f_S} \right|^2, \qquad f_k = \frac{k\,f_S}{N}, \quad \text{for} \quad k = 0, \ldots, N-1 \qquad (2)$$

Two modifications have been implemented to make equation (1) more suitable for the study of transient noise, in particular to identify "slow non-stationarities", namely changes in the statistical properties of the data over time scales longer than a second. First, frequencies corresponding to spectral lines (refer to section 5.2 for more details) have been removed from the integral in (1). These lines are narrow features in the PSD of the data, originating from resonances in various parts of the interferometer and their harmonics. Their intensities can be orders of magnitude larger than the neighboring noise floor. Hence, if a line is present in a band where we are about to compute the BRMS, it is likely to dominate the final estimate, and also the corresponding fluctuations, preventing us from probing the features of the underlying noise floor. To remove these lines, we identify them with an algorithm similar to the one developed for the Noise Frequency Event Miner (NoEMi) pipeline [25], and based on the *prominence* of their PSD [26].

Second, also glitches are typically removed from the BRMS time series. As these glitches can manifest at a rate of about 10 per minute (see O3 glitch rates in [9]), every data segment longer than a few seconds is likely to contain one of them, hence leading us to reject stationarity over such time scales. Dedicated algorithms, based on excess noise identification, are typically used to find them, and will be described in section 4. However, these algorithms are in general not sensitive to slower non-stationarities. To make BRiSTOL specifically targeted at the latter, we have then proceeded to identify glitches on the BRMS time series by means of an algorithm based on a rolling *median absolute deviation* (defined as the median absolute difference from the median), and to remove the corresponding data points from the analysis.

For each frequency band, the resulting modified BRMS time series are divided into "chunks" of equal duration, which are used to test stationarity. This hypothesis is tested by means of a two-sample Kolmogorov–Smirnov test [27] for each pair of consecutive chunks, and the corresponding p-values are compared to a test significance α (to be decided in advance). The (*null*) hypothesis of stationarity is rejected when a p-value is less than or equal to α, meaning that the estimates in the two chunks are drawn from different distributions and the underlying process has changed.

There are two advantages in using BRMS-based quantities. First, averaging over the frequencies of each band has a similar variance reduction effect than the means in Welch's PSD estimation method [28]. This in turn allows a finer time resolution while maintaining a moderate variance for the test statistics, that is, the empirical distribution of the BRMS. Second, the various non-stationarities typically manifest in specific frequency bands, closely related to the noise source that generated them. For example, the main harmonic of scattered light is usually visible below 30 Hz; non-linear and non-stationary couplings of the angular controls with the 150 Hz harmonic line are characteristic of a tight region around it, etc. So, without losing much of resolution, we can perform the noise characterization directly on these bands instead of on each frequency bin comprising the spectrum of the signal.
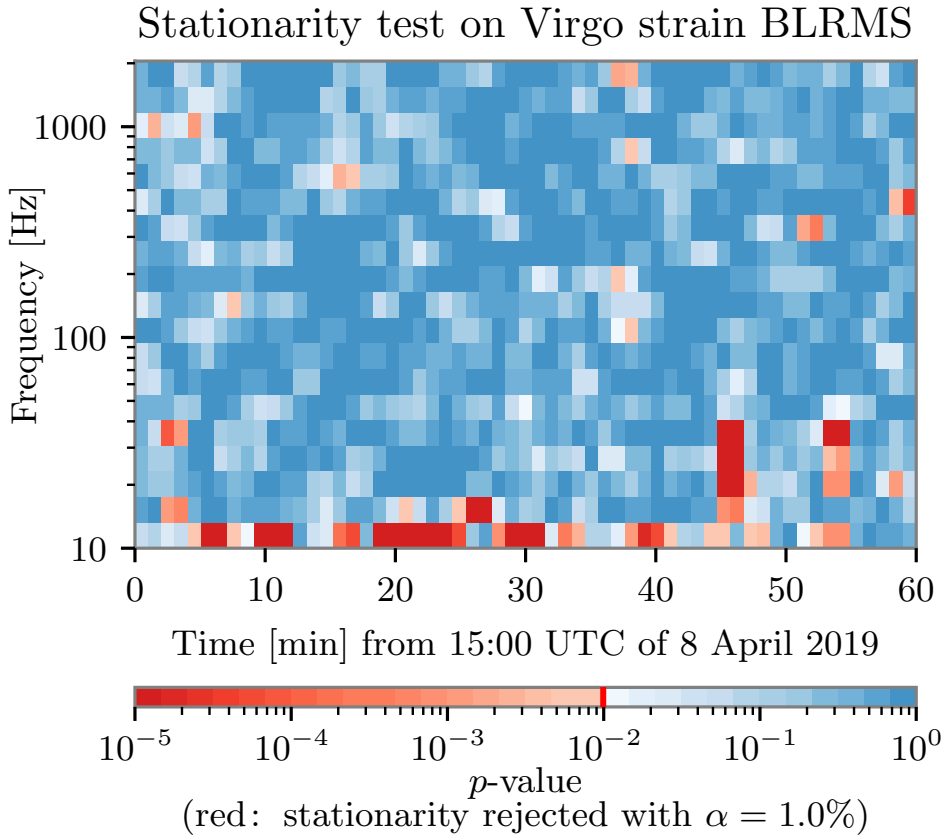
Figure 5. Example of stationarity time–frequency map obtained with `BRiSTOL`, where a significance $\alpha = 1\%$ has been chosen and regions rejecting the stationarity hypothesis are colored in shades of red.

The output of `BRiSTOL` can be visualized as a time–frequency map where, for each bin, the value of the test statistic p-value is reported. An example of such a map is shown in figure 5, where different color palettes are used to highlight those regions where stationarity should be rejected (red) and where not (blue). The time resolution of the test is given by the duration of each chunk and that of the PSD estimates, typically one minute and one second respectively; that in frequency is determined by the band division of the spectrum for computing the BRMS, which is conveniently done choosing exponentially spaced frequency intervals.

This tool has been developed in the commissioning phase preceding O3, and has been used during the run as well, to assess the quality of the data as part of the event validation procedure (refer to [9] for more details).

3.3.2. rayleighSpectro - Gaussianity test Similarly to what was discussed for the stationarity hypothesis, Gaussianity has to be tested separately in the different regions of the spectrum where noise sources can show up. `rayleighSpectro` [29], based on

the *Rayleigh test* [30], does this by means of a consistency check on the PSD estimated from the data with what is expected for stationary Gaussian noise. Indeed, if the data is compatible with the hypothesis of Gaussianity, the periodogram estimator in equation (2) is asymptotically (with N large) described by an *exponential distribution* of parameter $S(f_k)^{-1}$ [31], where $S(f_k)$ is the processed PSD. The corresponding amplitude spectral density (ASD) estimator, obtained as the square root of equation (2), is described by a Rayleigh distribution with parameter $\sqrt{S(f_k)/2}$. The scaling property of this distribution can be used to construct consistency tests. For example, the standard deviation of the ASD estimates obtained on non-overlapping segments provides an estimator of the standard deviation of this variable, which equals to $\sqrt{(4-\pi)\,S(f_k)/2}$. Similarly, the mean of these estimates provides an estimator of the mean: $\sqrt{\pi S(f_k)/2}$. The ratio of these two quantities gives a test statistic that, at each frequency f_k, is *asymptotically* equal to

$$\frac{\sqrt{4-\pi}}{\sqrt{\pi}} \simeq 0.52. \tag{3}$$

The actual value of the previous quantity for a finite number of averages and the corresponding critical values for performing statistical tests have been computed in [29, 32].

If the noise is not Gaussian, or its properties has changed while estimating the standard deviation and mean of its ASD, the resulting test statistic will take values different than what reported in (3). This means that the Rayleigh test is sensitive to both non-Gaussianities and non-stationarities in the data. Smaller values of the statistic are associated with data having smaller fluctuations than those expected for a Gaussian process; spectral lines usually behave in this way. Larger values are instead typical of non-stationary noises, such as glitches, that produce a larger variance of the ASD estimates.

By dividing the data into chunks of duration Δt, one can obtain a time–frequency map, similar to a spectrogram, showing with time resolution Δt the frequencies and times where the data significantly depart from the expected value of equation (3).

The output of `rayleighSpectro` are included in `VIM` and also used in the `DQR` for event validation (see [9] for more details).

The interplay between this tool and `BRiSTOL` for the assessment of the stationarity and Gaussianity of the data is the following. `BRiSTOL` assesses where the data is compatible with the hypothesis of wide sense stationarity, that is, the second order moments (i.e. the covariance or the RMS) are left unchanged by shifts in time. This corresponds also to a *strong sense stationarity* (the invariance of the probability distribution of the noise) if the data is also Gaussian, that is, completely characterized by its mean and covariance functions, as tested by `rayleighSpectro`. Conversely, deviations from these assumptions can be tested independently. This is useful for example with regions corresponding to spectral lines that are typically stationary in very good approximation but not Gaussian.
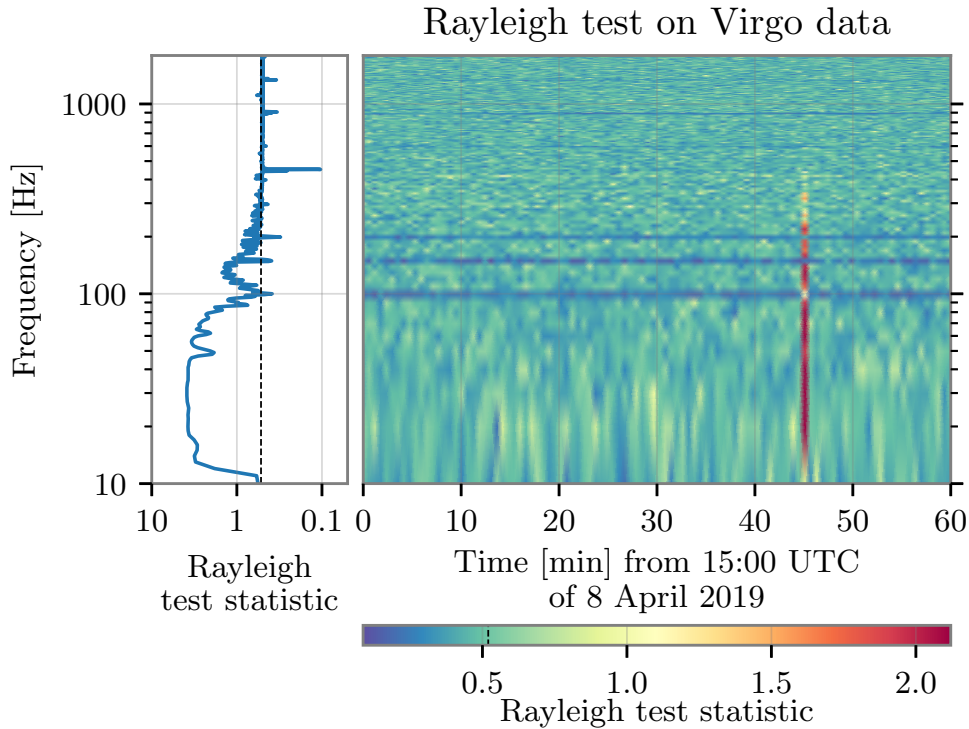
Figure 6. Example of application of the Rayleigh test, where the blue line in the left panel is the test statistic estimated over the entire hour of data, while the color map corresponds to ASD estimates over 10 s of data. The vertical dashed black line on the left plot indicates the 0.52 limit value. See text for details.

Figures 5 and 6 show examples of the application of these two tools to 1 h of data at the beginning of O3a. In the first plot, BRiSTOL highlights many slow non-stationarities at frequencies up to about 20 Hz, most likely due to high microseismic activity, as well as a loud glitch at about 15:45 UTC. This one is clearly identified by the Rayleigh test with values of the test statistic larger than what is expected for stationary and Gaussian noise. Moreover, in the color map of figure 6, spectral lines, in particular those associated with the 100, 150 and 200 Hz harmonics of the mains (the European power grid frequency is 50 Hz), are highlighted in blue, corresponding to values of the test statistic smaller than the asymptotical limit in the Gaussian case, given in equation (3). In the left-hand side panel of the same image, the 450 Hz frequency of the main test masses violin modes, and its first harmonic at about 900 Hz, are highlighted as well.

Figure 7 shows another example of a Rayleigh spectrum from O3, corresponding to a period during which a transient noise was present for several minutes between 10 Hz and 20 Hz.
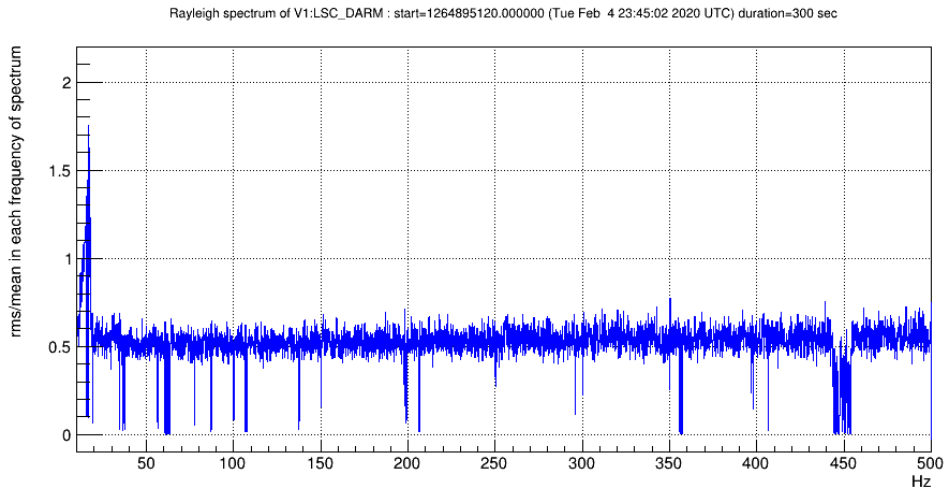
Figure 7. Example of Rayleigh spectrum averaged over 300 s, where any bin above 0.52 is a potential non-stationary or non-Gaussian noise present during those 300 s. Values well below 0.52 correspond to persistent frequency lines.

3.4. Monitoring BNS range drops and gating data

The binary neutron star (BNS) range is the average distance up to which the merger of a BNS system can be detected. The average is taken over the source location in the sky and the BNS system orientation, while a detection is defined by convention as a signal-to-noise ratio (SNR) of 8 or above. This quantity is related to the live detector sensitivity and can be used as a summary statistic to monitor the overall performance of the instrument in relation to transient searches.

Two useful high-level data quality monitors are based on the BNS range[3]ownwards excursions. One tags BNS range *drops*, that are significant and sudden decreases in this value, usually flagging data quality problems. The other automatically generates (logical) *gates* that are applied on the GW strain channel to smooth out to zero the data that are affected by a strong noise transient. BNS range drops and gates are related but not equivalent depending on the frequency contents of the noise burst.

3.4.1. BNS range drops A BNS range drop means that the live sensitivity of the detector is degrading significantly, at least in a given frequency band, possibly in the entire bandwidth of the instrument. Therefore, it is important to identify transient sensitivity worsenings and investigate their causes. BNS range drops are very diverse: the decrease goes from a few percents to almost the full range, while the drops can last from a few seconds to minutes.

During O3, BNS range drops were detected using an absolute threshold on the live value of that quantity. After the end of the run, adaptive methods able to follow the natural

3 d
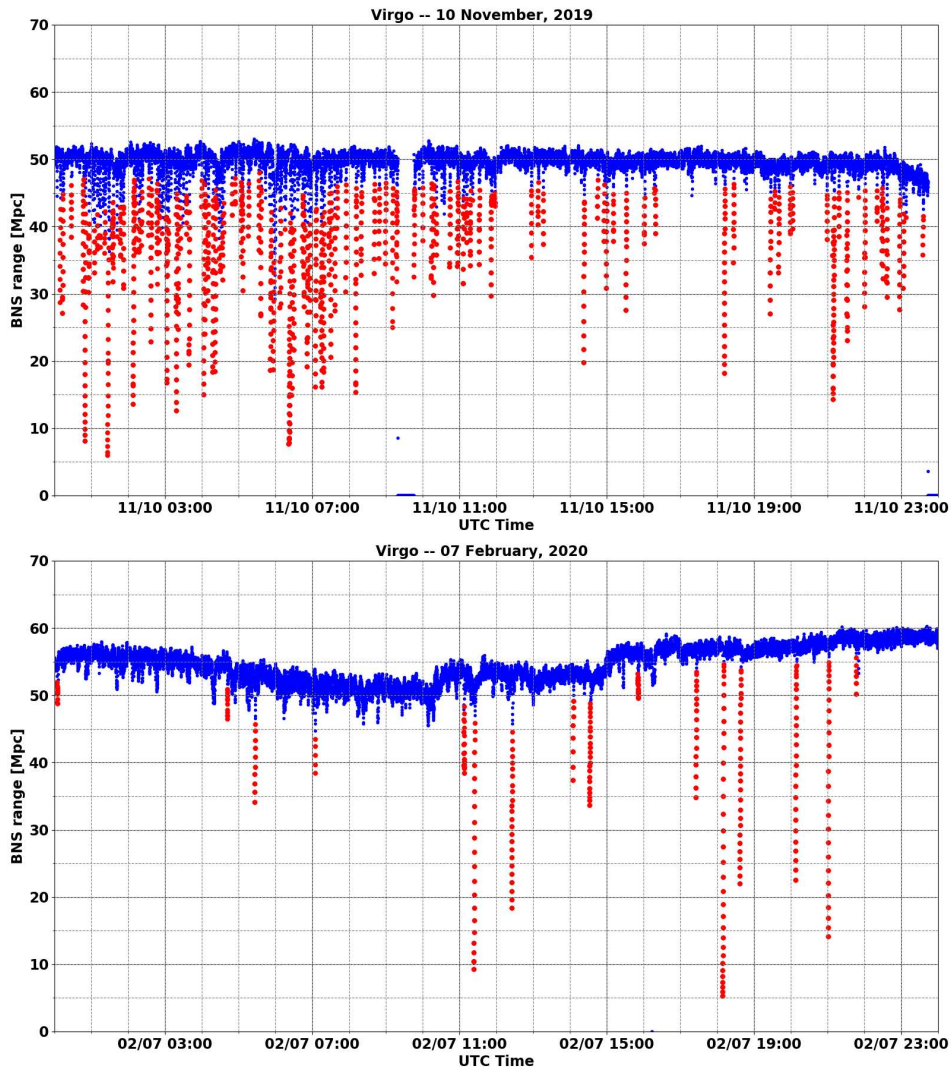
Figure 8. Performance of the BNS range drop locator during two days of O3. Top plot: November 10th, 2019, a day during which the duty cycle was quite high but the data taking conditions were not stable; many glitches and consequently BNS range drops were observed, mostly due to the laser power stabilization system in the morning and to a worsening of the weather conditions starting from the afternoon. Bottom plot: February 7th, 2020, a day with no global control loss but a BNS range baseline varying over time; actions took place during the afternoon to improve the Virgo performance, leading to visible improvements of the BNS range in steps. The blue traces show the range vs. time, while the red dots show the drops that have been identified. In both cases the BNS range drop locator is able to identify most, if not all, significant drops.

evolution of the BNS range and to locate all significant drops have been developed. Figure 8 shows examples of the output of the adaptive BNS range drop locator running on O3 data.

3.4.2. Gates If not removed from data, noise bursts can pollute the estimation of the noise spectrum for several seconds, hence limiting the sensitivity of the GW search

algorithms during that period. In Virgo, this problem is mitigated online by gating out (i.e. smoothing out to zero) glitchy chunks of data. The gating algorithm triggers on significant BNS range drops: at least 40% below its median value, computed over the last 10 s. On both sides of the gate, a weight is applied on the $h(t)$ strain channel during $10/32^{th}$ of a second, varying smoothly from 1 to 0 (0 to 1) before (after) the gate. The online gated $h(t)$ strain channel is included in the DAQ alongside the ungated one, and GW searches are free to use one or the other stream as input.

As gating is based on $h(t)$ variations, gated data cannot simply be removed from the physics-analysis dataset as this procedure could flag real GW signals, for instance loud high-mass binary black hole mergers. On the other hand, gating information can be used in a statistical way to help identify potential periods of bad data quality characterized by frequent gating usage. This can be measured using both the density of gates (number of gates per unit of time) and the fraction of the wall-clock time that is gated out.

During O3, the gating algorithm has produced more than 13,000 gates (corresponding to a few tens per day in average), adding up to about 4 h of gated data in total. The gate mean duration was around 1.1 s while the median was around 0.8 s, meaning that most gated glitches were very short as $20/32^{th}$ of a second were always added to the measured glitch duration to transition from non-gated data to the gate itself and back. The longest gate was about 10 s.

The segments that have been vetoed for offline data analyses [9] excluded from this online Science dataset led to a removal of about 20% of the gates and of about 30% of their total duration. However, this procedure only removed about 0.2% of the Virgo O3 dataset. As expected, more gates were generated when the quality of the data was degraded. Going one step further by requesting in addition that the baseline BNS range be greater than 35 Mpc. This excludes more than 50% of the remaining gates and more than 60% of the gated times, while that cut would remove about 1% of the data from the final dataset. Gates are generated more often when the data taking conditions are sub-optimal.

Finally, one can associate all gates with a glitch detected by `Omicron` (see section 4.1) whereas the opposite is not true: there are many glitches that have no impact on the BNS range. These glitches have a frequency range that is outside of the Virgo bandwidth for BNS GW waveforms: either because there is no significant signal contribution expected in that frequency range, or because the noise level is high enough to make that range contribute little if anything at all to the BNS range.

3.5. Monitoring global Control losses

Losses of the global working point of the Virgo interferometer (the mandatory configuration sensitive to passing GWs) do not just interrupt the data taking: they decrease the overall duty cycle as few tens of minutes are needed after such events to

restore the conditions for taking good-quality data sensitive to the passing of GWs. Therefore, categorizing control losses is important to understand their main causes and to get alerted when a new class of control losses appears, or when an already known category becomes more frequent.

An extensive offline study of the global control losses in Science data-taking mode during the O3 run has lead to the identification of the root cause of the control losses in most cases [7]. The experience gained with this work will be useful for the pre-O4 commissioning phase (noise hunting) and the subsequent data taking periods in two ways. First, the categories identified during O3 will be reused as a starting point to investigate new control losses. Then, an online monitor will analyse these global control losses within minutes of their occurrence; it will provide a set of automated plots for further human diagnosis and possibly point out their probable cause. This framework is currently under development and will reuse the approach (if not the proper software infrastructure) of the `DQR` (see section 7).

4. Glitch identification and characterization tools

4.1. Omicron

The `Omicron` [33] search algorithm is used to detect and characterize transient noises. The data is processed using a Q transform [34] which consists in decomposing a time series $x(t)$ onto a generic basis of complex-valued sinusoidal Gaussian functions centered on time τ and frequency f:

$$X(\tau, f, \sigma_t) = \int_{-\infty}^{+\infty} x(t) \frac{W}{\sigma_t \sqrt{2\pi}} \exp\left[-\frac{(t-\tau)^2}{2\sigma_t^2}\right] e^{-2i\pi ft} dt. \tag{4}$$

This transformation is a modification of the standard short Fourier transform in which the analysis window size σ_t varies inversely with the frequency and is characterized by a quality factor Q: $\sigma_t = Q/(\sqrt{8}\pi f)$. The parameter space (τ, f, Q) is tiled to guarantee both a high detection efficiency and an optimized processing speed. The noise of the input signal $x(t)$ is whitened prior to the Q transform such that all noise frequencies have the same weight. This is done through the normalization factor W which includes an estimate of the local stationary noise, such that the Q transform coefficient X directly measures the SNR associated to each individual tile (τ, f, Q). A glitch in the data is detected by `Omicron` as a collection of tiles with high-SNR values. An `Omicron` glitch is characterized by a set of parameters (τ, f, Q) given by the tile with the highest SNR value. `Omicron` offers a two-dimensional representation of glitches where the SNR distribution of tiles is plotted in one or several Q planes. GW events can also be visualized with `Omicron`: see figure 9.
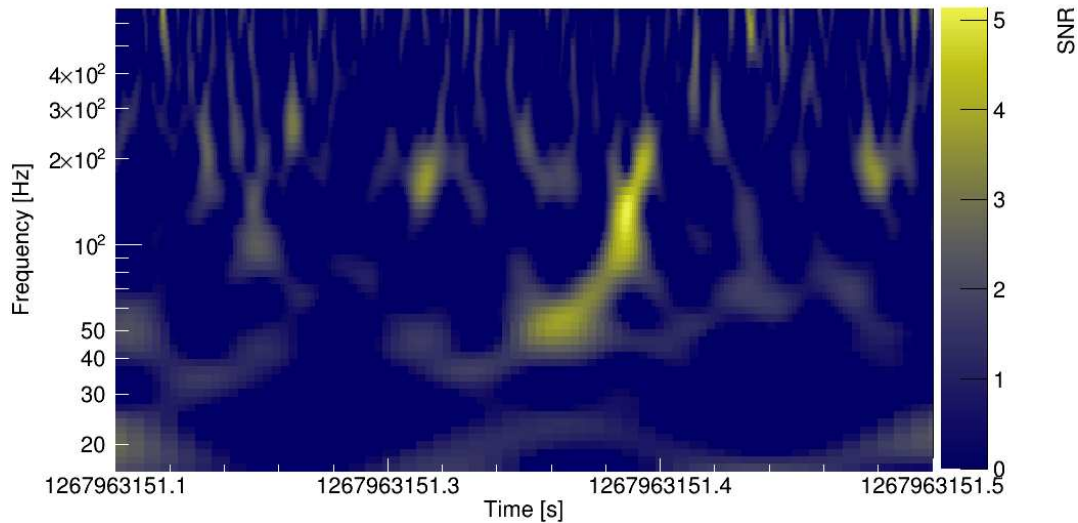
Figure 9. Spectrogram of GW200311_115853 [1] in the Virgo detector, as measured by `Omicron`. The time-frequency plane is tiled fixing $Q = 6.4$.

4.2. Use-Percentage Veto

The Use-Percentage Veto (`UPV`) algorithm [35] has been developed to detect and characterize noise correlations between two glitch data samples; one derived from the GW strain channel $h(t)$ and the other derived from an auxiliary channel. The algorithm tunes, considering `Omicron` triggers of a given auxiliary channel, a signal-to-noise ratio threshold such that, when a trigger is above threshold, there is a high probability to find a coincident glitch in the $h(t)$ data. In O3, the Virgo data were processed with the `UPV` algorithm on a daily basis to support the noise characterization effort; some auxiliary channels were identified by `UPV` as exhibiting glitches correlated with $h(t)$ glitches, providing hints about the noise coupling in the detector.

4.3. VetoPerf

The `VetoPerf` analysis tool measures the performance of a data quality flag. A data quality flag is defined as a list of time segments targeting transient noise events. `VetoPerf` counts the number of the $h(t)$ triggers detected by `Omicron` which are coincident with the data quality flag time segments. From this, it derives performance numbers and produces diagnostic plots characterizing that data quality flag.

4.4. Scattered light monitor

Scattered light is a non-linear, non-stationary noise affecting the sensitivity of the interferometer in the GW detection frequency band. As adaptive algorithms such as Empirical Mode Decomposition (EMD) [36, 37, 38] are suitable for the analysis of non-linear, non-stationary data, they can be used to quickly identify optical components

which are sources, i.e. culprits, of scattered light [39]. As part of the detector characterization effort, a tool was developed and applied to Virgo O3 data with the aim of identifying culprits of scattered light in the difference of the two arm cavity lengths (DARM) degree of freedom (DOF) of the detector [40]. The tool employs the recently developed time varying filter EMD algorithm (tvf-EMD) [41] which gives more accurate results compared to EMD [40]. When scattered light is affecting the detector, arches show up in DARM spectrograms. The arches frequency and their time of occurrence is given by the so called predictor (measured in Hz):

$$f_{arch}(t) = 2\frac{|v(t)|}{\lambda}, \tag{5}$$

where $v(t)$ is the velocity at which the optical component is moving and λ is the laser wavelength. Equation (5) is computed using the position data of several optics of the detector, such as for example the suspended west-end bench (SWEB). Having obtained predictors for several optical components, the tool computes the instantaneous amplitudes $IA(t)$, i.e. the envelope of DARM oscillatory modes which are extracted by tvf-EMD. $IA(t)$ can then be correlated with the list of predictors. The optical component with the highest correlation among its predictor and the $IA(t)$ of DARM is considered to be the culprit of the scattered light noise witnessed in DARM. Visual counterproof can be performed (see figure 6 from [40]) overlapping the culprit predictor on the DARM spectrogram [42]. The methodology of [39, 40] was extended and integrated in the *gwadaptive-scattering* pipeline, an automated Python code which allowed to characterize the origin of scattered light glitches in LIGO during the O3 run [43]. Furthermore, adaptive analysis can be used to monitor daily the onset and time evolution of scattered light noise in connection with microseismic noise variability [44]. These daily analyses have been integrated in the *gwadaptive-scattering* pipeline as well.

5. Spectral noise identification and characterization tools

5.1. Spectrograms and injected lines identification

Within the `VIM` (see section 2.3), spectrograms spanning periods from one day to a week are regularly updated using the custom `Spectro` software [29]. This framework is based on a set of ROOT [45, 46] scripts that provide various indicators (BRMS, Rayleigh spectra, etc.), useful to help investigating non-stationary spectral lines or intermittent noises. The `Spectro` tool has also been used during O3 to probe the time-frequency pattern of the glitches associated with BNS range drops.

5.2. NoEMi and the (known) lines database

The `NoEMi` tool [25, 47] tracks on a daily basis spectral lines, both stationary and wandering ones, and searches for coincidences between the lines found in a main channel — typically the GW channel $h(t)$ — and in a list of auxiliary channels. The `NoEMi` configuration defines several parameters and thresholds, like for instance: the threshold on the critical ratio[4] for peak selection in the spectra, the frequency resolution (linked to the time length of the data segments over which the fast Fourier transform (FFT) is computed), the name of the main channel, the list of auxiliary channels to search for coincidences. During O2, the `NoEMi` software produced daily results and looked for peaks in the spectra using a frequency resolution of 1 mHz. With this configuration, `NoEMi` looked for coincident spectral peaks between the DARM channel and approximately 40 auxiliary channels.

During the break between the O2 and O3 runs, the `NoEMi` software has been intensively modified, resolving the main issues identified in the old version. The original code was not well-structured (and hence difficult to modify) and also not fully-efficient CPU-wise. Furthermore, the original version produced several static files which were unessential for the final output. As a further improvement, the MySQL database which stores all parameters of each spectral line found during the run has been normalized, meaning that useless or redundant data have been removed and that the data storage is now more coherent. The database scalability has been improved as well, in order to allow storing more data and handling a higher load of requests. Additionally, a more dynamic interaction with the web interface used to browse the results has been introduced. The new version of the code has been used for the first time in O3.

During O3, `NoEMi` used the same set of ~40 auxiliary channels as in O2, plus an additional set of ~140 environmental channels, e.g. seismic, magnetic, and acoustic probes. The coincidence between a line in the GW strain channel $h(t)$ and the signal of one of the environmental monitor, suggests that the noise line originates from a physical source such as a vacuum pump, a cooling fan, an electronic device, etc. This information helps to identify the instrumental origin of detected lines in $h(t)$, and it has been included in the official Virgo-O3 line list publicly released by the Gravitational Wave Open Science Center (GWOSC) [48]. Figure 10 illustrates the lines identified in the Virgo GW strain channel during O3.

Internally, lines that have been identified are stored in a dedicated database that includes detailed information about them: most notably their times of appearance, and links pointing to the associated documentation (logbook entries, studies, mitigation actions, etc.). The contents of the database can be compared with a new `NoEMi` processing, to find out quickly which lines identified by `NoEMi` are already known and which ones are not.

4 Defined as the number of standard deviations a given peak amplitude is different from the mean of the peaks amplitude distribution.
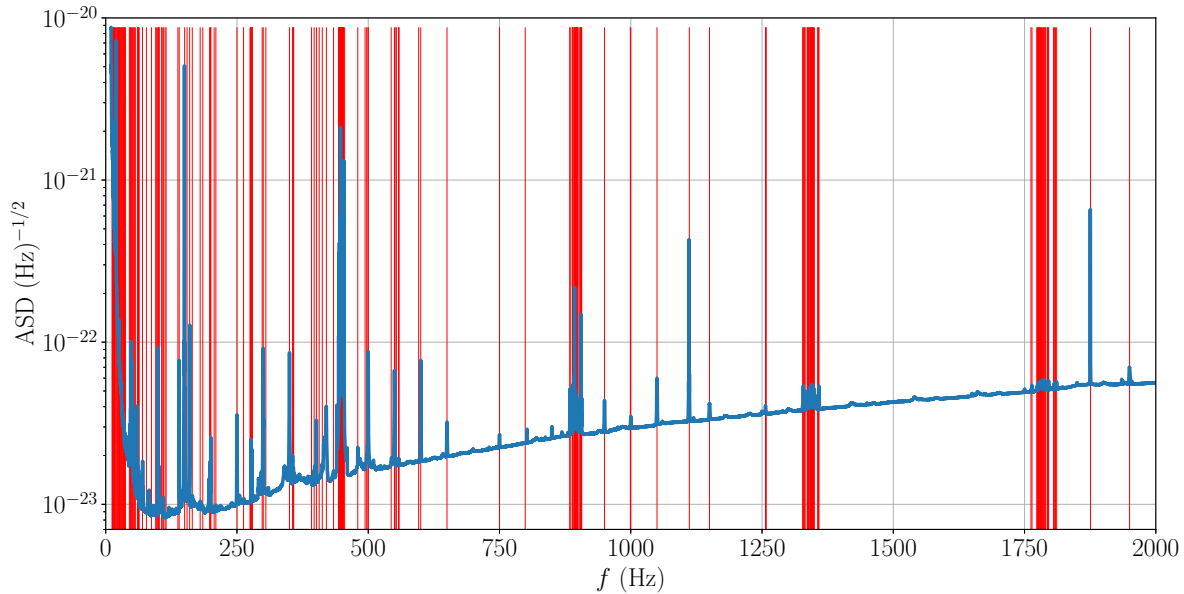
Figure 10. Virgo spectral lines identified during O3. The blue curve is the estimated ASD of the Virgo strain channel during O3. Red vertical bars mark the frequency of the identified spectral lines. Lines parameters are listed in Reference [48]. Most of the lines have been found by `NoEMi`.

5.3. Bruco

The brute-force coherence tool (`BruCo`) [49] is a python-based tool designed to search for correlated noise among channels. The code (version 2017-01-23) is publicly available from the git repository [50], which also provides a description of the argument list. An instance of the repository is kept with Virgo-specific data access features.

`BruCo` computes the magnitude-squared coherence between a main channel (typically, but not necessarily, the detector strain channel $h(t)$) and all auxiliary channels that, at the time of interest, are recorded by the DAQ system. Optionally, a set of redundant channels which are known a priori to be correlated with the main one, can be excluded. In Virgo, during the O3 run, there were approximately 3,000 non-redundant channels with a sampling frequency $\geq$1 kHz. To deal with the high computational load required by this analysis, `BruCo` implements the option of multi-core parallel processing in up to 10 threads.

In the `BruCo` implementation adopted for Virgo during O3, a continuous Science data segment of length $T = 800$ s is selected for the main channel $h(t)$ and, in turn, for each auxiliary channel $n(t)$. Each data segment is resampled to a targeted output frequency of 2 kHz with the Fourier resampler *scipy.signal.resample*, divided into $N_{ave} = 100$ sub-segments (8 s long) and the averaged magnitude-squared coherence is computed, as:

$$C_{h,n}(f_i) = \frac{|< FFT_n(f_i)^* FFT_h(f_i) >|^2}{<| FFT_h(f_i) |^2><| FFT_n(f_i) |^2>} \tag{6}$$

where FFT denotes the windowed fast Fourier transform, "$<>$" denotes the averaging operation, f_i is the ith frequency bin, and "$*$" the complex conjugate operation. With these parameters, the frequency resolution is $df = N_{ave}/T = 0.125$ Hz. Coherence is examined up to 1 kHz, and its value is deemed significant if it exceeds a threshold set to 0.03, a value corresponding to the 95% confidence level of the distribution of averaged coherence between random data [51], given the selected parameters.

BruCo jobs were run regularly and automatically during the whole O3 run. Daily results are HTML-formatted and made accessible in a dedicated VIM web page (see section 2.3). The BruCo VIM summary page allows to quickly spot noise paths contributing to the GW strain channel $h(t)$ in specific frequency bands. In addition, BruCo has often been used as an on-demand analysis tool to examine specific time periods.

BruCo main output is a table that contains, for each frequency bin, the ordered list of the auxiliary channels that are most coherent with the main channel. The cell background is color-coded in shades of red from full red (maximum coherence: 1) to white (no coherence) as shown in figure 11. For each auxiliary channel, a plot (see figure 13) of the *projected coherence* quantity, $h_n(f) =< FFT_h(f) > \sqrt{C_{h,n}(f)}$, is produced and linked to the table. In the hypothesis of linear coupling, this quantity estimates the contribution to the strain channel of the noise witnessed by the nth auxiliary channel [51]. Additionally, the VIM daily summary page contains the list of the top ranked channels in the frequency bins with coherence greater than 0.3 shown in table 1, and a plot of the combined projected coherence greater than 0.5 presented in figure 12.

Figure 13 shows BruCo daily plots illustrating one example of noise contamination spotted during O3, which triggered a more in-depth investigation [8, 53].

5.4. MONET

The interferometer noise spectrum can sometimes present peculiar structures as a consequence of non-linear couplings between different noise processes. Some of these structures consist of two pairs of sidebands around known spectral lines, which are not explained by means of the previously described linear coherence methods. One example of this kind of noise is the bilinear noise, generated by the coupling of two noise sources that jointly affect a third signal. In GW detectors, the main cause of this bilinear noise is the up-conversion of the low frequency seismic noise, that can affect the mirrors angular controls, which couples with some narrow-band noise processes, like power mains lines (50 Hz fundamental frequency or its harmonics) or calibration lines (see for example [23, 5]).

The Modulated NoisE Tool (MONET) [54] is a python-coded framework designed to investigate these sidebands. The main hypothesis at the basis of this tool is that the

166.88	SDB1_LC_TZ _fb (0.24)	SDB1_LC_TZ _corr (0.24)	SDB1_LC_TZ _err (0.24)	SDB1_LC_TZ (0.24)	SDB1_LC_COIL _FL_V (0.24)
167.00	SDB1_LC_TZ _fb (0.32)	SDB1_LC_TZ _corr (0.32)	SDB1_LC_COIL _FL_V (0.32)	SDB1_LC_COIL _BL_V (0.32)	SDB1_LC_COIL _BR_V (0.32)
167.12	SDB1_LC_COIL _FR_V (0.45)	SDB1_LC_COIL _BL_V (0.45)	SDB1_LC_COIL _FL_V (0.45)	SDB1_LC_COIL _BR_V (0.45)	SDB1_LC_TZ _err (0.45)
167.25	SDB1_LC_COIL _FR_V (0.44)	SDB1_LC_TZ _corr (0.44)	SDB1_LC_TZ _fb (0.44)	SDB1_LC_COIL _BR_V (0.44)	SDB1_LC_COIL _FL_V (0.44)
167.38	SDB1_LC_COIL _BL_V (0.40)	SDB1_LC_TZ (0.40)	SDB1_LC_COIL _FL_V (0.40)	SDB1_LC_TZ _err (0.40)	SDB1_LC_COIL _FR_V (0.40)
167.50	SDB1_LC_COIL _BL_V (0.41)	SDB1_LC_COIL _FL_V (0.41)	SDB1_LC_TZ _err (0.41)	SDB1_LC_TZ _fb (0.41)	SDB1_LC_TZ _corr (0.41)
167.62	SDB1_LC_TZ _err (0.42)	SDB1_LC_COIL _FL_V (0.42)	SDB1_LC_TZ (0.42)	SDB1_LC_COIL _BL_V (0.42)	SDB1_LC_COIL _BR_V (0.42)
167.75	SDB1_LC_TZ _err (0.34)	SDB1_LC_COIL _BL_V (0.34)	SDB1_LC_TZ _fb (0.34)	SDB1_LC_TZ _corr (0.34)	SDB1_LC_TZ (0.34)

Figure 11. An example of `BruCo` result web page displaying, for each frequency bin (leftmost column), the most coherent channels sorted by decreasing coherence value, also represented by the red shade intensity. These data are from November 11th, 2019. The large coherence detected at frequencies 155—170 Hz triggered some further investigations of the noise [52].
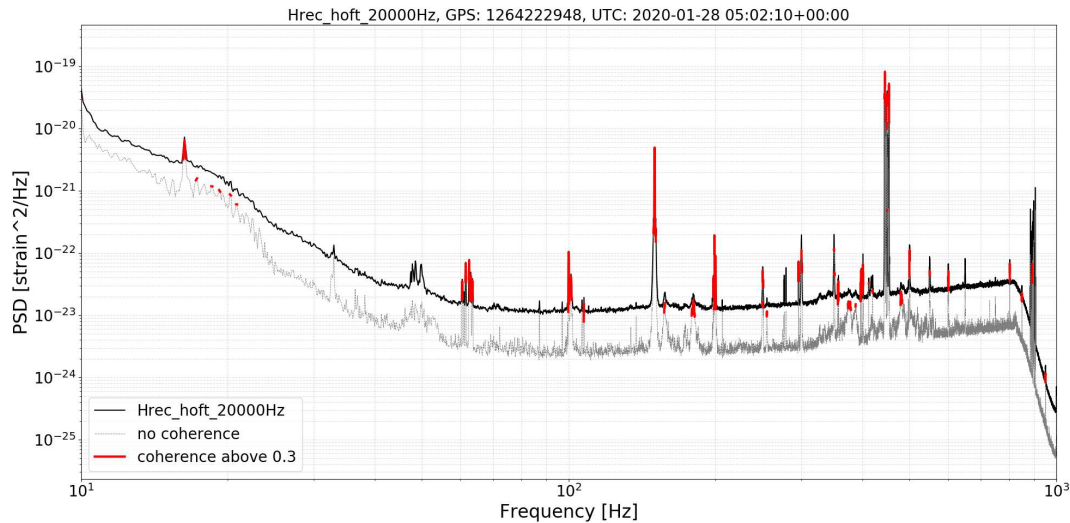
Figure 12. A `BruCo` `VIM` daily combined projection of the GW strain channel $h(t)$ at GPS time 1264222948 (2020/01/28 at 05:02:10 UTC). An anti-aliasing filter with cut-off frequency of 800 Hz is applied to the data. Frequencies where the coherence exceeds 0.3 are highlighted in red, while the dashed grey line shows the noise projection in those regions of the spectrum where the coherence value is below the threshold.

sidebands are due to the coupling of a carrier signal with the low-frequency (with a typical cutoff frequency f_c of a few Hz) part of an auxiliary channel, called the modulator

Table 1. An excerpt (15 top lines) of a `BruCo` summary table showing the top-ranked channels with coherence greater than 0.3 for GPS time 1264222948 (2020/01/28 at 05:02:10 UTC).

Frequency [Hz]	Coherence	Type of channel
10.9	0.329	Error signal of the second stage of laser frequency stabilization control loop
16.2	0.527	Longitudinal correction applied to the beam splitter payload
17.2	0.376	West arm transmitted light power
18.5	0.357	Same signal
19.2	0.342	Same signal
20.2	0.391	Same signal
20.8	0.318	Same signal
21.8	0.327	Same signal
33.0	0.417	Error signal of the second stage of laser frequency stabilization control loop
60.4	0.579	Calibration signal applied to the west end test mass
61.0	0.444	Signal of the phase camera on the external detection bench
61.4	0.817	Calibration signal applied to the beam splitter (BS) payload
61.5	0.944	Calibration signal applied to the north end test mass
62.500	0.929	Calibration signal applied to the west end test mass
99.800	0.356	Experimental power grid voltage monitor signal

signal. Under this hypothesis, `MONET` searches for coherence between a main channel (typically, but not necessarily, the detector strain channel or the DARM channel) and a new signal given by the product, in the time domain, of the chosen carrier signal and a modulator signal. That carrier signal can be a real channel or a simulated signal such as, for example, a sinusoidal signal.

Similarly to what is done with `BruCo`, continuous Science data segments of a specific time length T are selected for all the channels to be investigated with `MONET`. Then, each data segment is resampled to a targeted output frequency (f^{out}) and, finally, the magnitude-squared coherence is computed using equation (6). For the analysis of the O3 Virgo data, we typically used the following values: $f_c = 5$ Hz, $T = 1200$ s and $f^{\text{out}} = 1$ kHz.

`MONET` can be executed on demand, investigating dozens of auxiliary channels and spectral lines in every single processing. The outputs are made available through a hierarchical structure:

- A main directory, whose name is built from the main channel name, the initial GPS
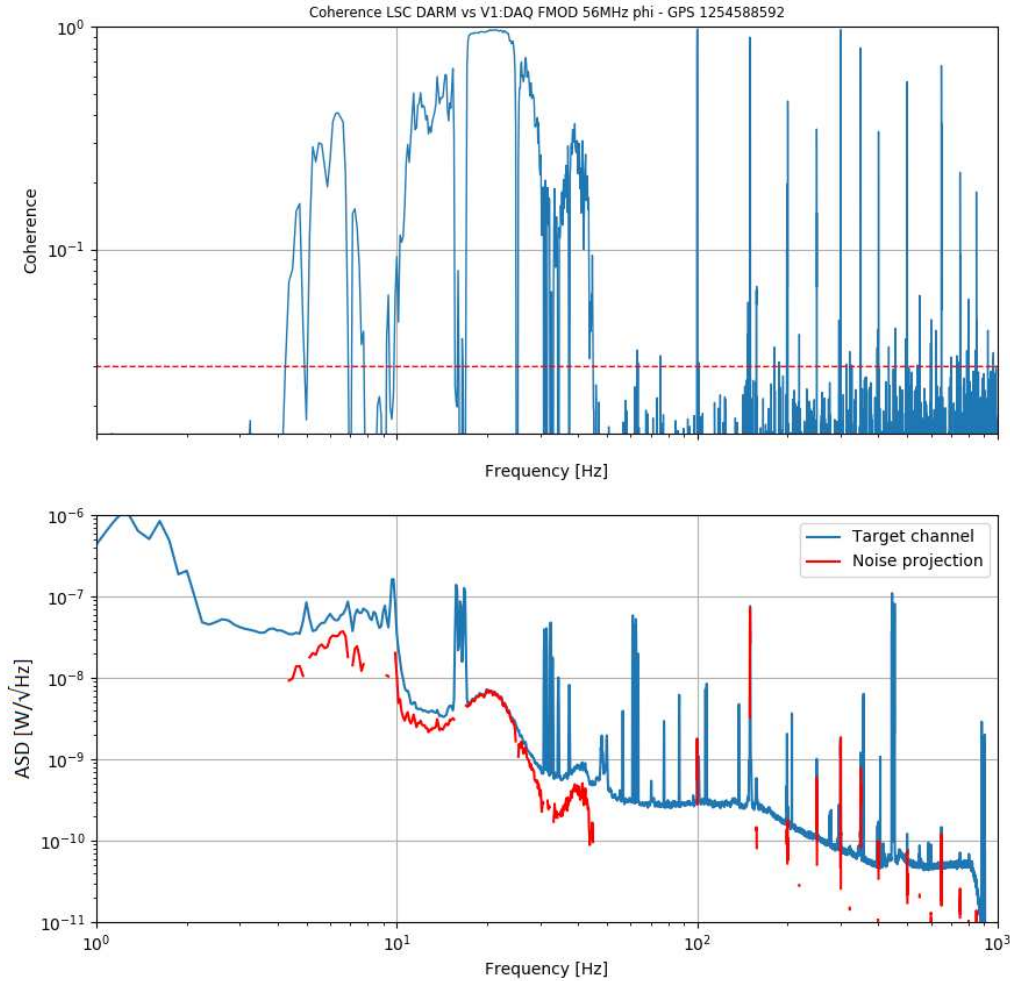
Figure 13. Selection of `BruCo VIM` daily plots evidencing noise contaminating the Virgo GW strain channel during O3. The top plot shows the coherence between the DARM channel and the laser electro-optical modulator which produces the 56 MHz signals used for the arms length control. The bottom plot shows the ASD of the DARM signal (blue line) and the corresponding projected coherence (red line) in the frequency ranges where it was found significant enough. The noise was then found to originate from back-reflected light onto the laser bench, most likely due to a damage on the electro-optical modulator for which the component has been removed after O3.

time and the time length of the segment of data to be analysed.

- The main directory contains one sub-directory for each carrier signal used.
- Each sub-directory contains in turn several folders, one for each modulator channel.

The sub-directory contains a table in text format and a summary figure. The table (see table 2 for an excerpt example) is made of three columns and several rows. Each

Table 2. An excerpt from a `MONET` table with coherence values above the threshold, fixed in this case at 0.3. The initial GPS time of the analysed data is 1229237309 (2018/12/19 at 06:48:11 UTC). Main channel: the DARM one; secondary channel: a voltage monitor of the uninterruptible power supply in the central building.
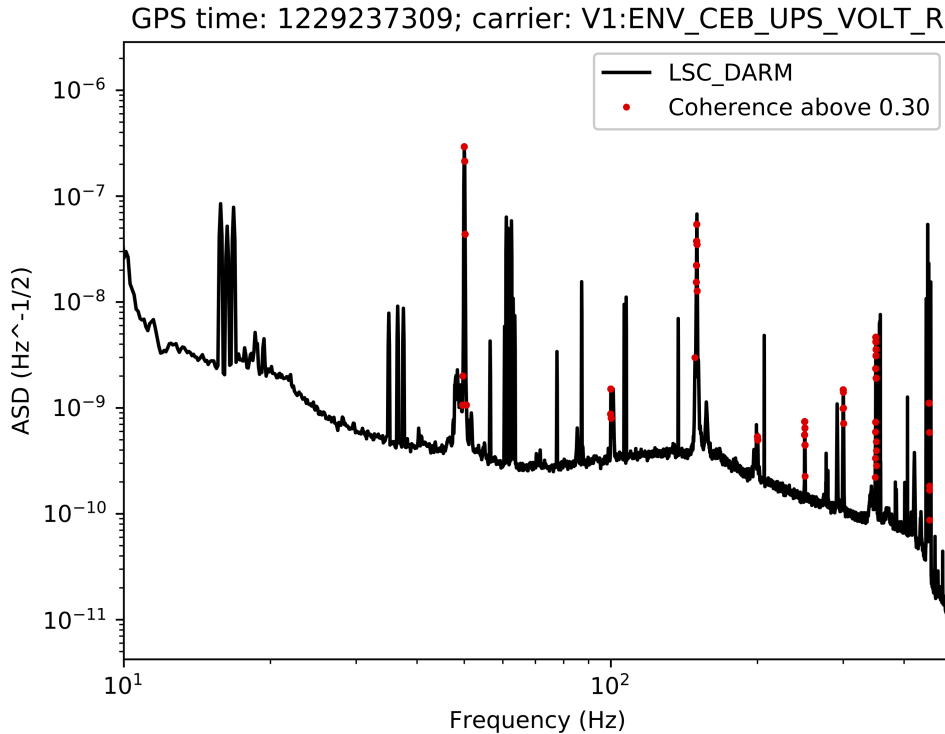
Frequency [Hz]	Coherence	Channels
200.0	0.313—0.347	Some control signals, mirror longitudinal corrections (applied by varying the current flowing into the coil-magnet actuators for the corresponding test masses), and DC laser powers measured in various locations of the interferometer.
200.2	0.379—0.437	Same channels as above.
200.3	0.456—0.496	Same channels as above.

row corresponds to a different frequency bin, indicated in the first column. The second and third columns display, for each modulator channel, the computed above-threshold coherence value and the name of that channel respectively. Figure 14 shows a `MONET` summary plot for the DARM channel during the pre-O3 commissioning phase. Red points marking the frequencies at which coherence above threshold is found with at least one auxiliary channel are superimposed to the main channel ASD. For instance, the red points at ∼200 Hz shown in figure 14 correspond to the frequency bins reported in table 2; at such frequencies, a coherence value above the chosen threshold was found for several modulator channels.

In each innermost folder, a table and a plot are generated, in which the coherence values associated with the specific modulator channel for each frequency bin are reported. Several other plots are also produced, in which the ASD of the main channel is reported, together with the noise projection based on the coherence values, around the specific spectral lines to be investigated as seen in figure 15.

The results of `MONET` can help identify the origin of the observed sidebands in the main channel. The modulating channels that, once multiplied by the carrier signal, exhibit the largest coherence with the main channel can witness those parts of the detector that are more sensitive to low frequency noise, usually due to microseismic activity, and its up-conversion through bilinear couplings.

During O3 and the preceding commissioning phase [55, 56, 57], `MONET` has been effectively used to characterize the spectral noise in the strain channel and guide the commissioning activity on how to improve it. This has been the case for the observed sidebands of the 150 Hz mains frequency harmonic line [58]. The increase in the noise around this frequency during intense microseismic activity has been found coherent with the angular controls of the beam splitter (BS) as modulating channels. Moreover, under similar noise conditions, the value of the coherence, hence the strength of the bilinear coupling and the observed overall noise level at 150 Hz, was proven to depend

Figure 14. ASD of the DARM channel (in black), together with red points that mark the frequencies at which the coherence found by `MONET` is above the threshold, fixed in this case at 0.3. The initial GPS time of the analysed data and the chosen carrier signal (a voltage monitor of the uninterruptible power supply in the central building) are indicated on the top of the figure.

on the way the BS is controlled ("full bandwidth" vs. "drift control"), convincing the team working on the interferometer alignment to prefer to switch the operational mode of the BS during high microseismic activity with an improvement of up to 2 Mpc of BNS range [23].

In addition to that, `MONET` has helped characterize the effect of the beam alignment on the power recycling (PR) mirror on the linear noise subtraction of the frequency noise coupling due to the arm asymmetry used to produce the reconstructed GW strain signal $h(t)$. Large values of coherence are found with the channels controlling the PR alignment at 1111 Hz, the frequency of the injected line for the laser frequency stabilization control loop, when the efficiency of the noise subtraction was smaller [59]. This has been interpreted as a consequence of the PR mirror transverse position modulating the frequency noise coupling, hence producing non-linear effects, not adequately taken into account by means of the linear subtraction method [60]. This effect has been quantified in a reduction of 2 Mpc of BNS range when the beam is transversely misaligned on the PR.
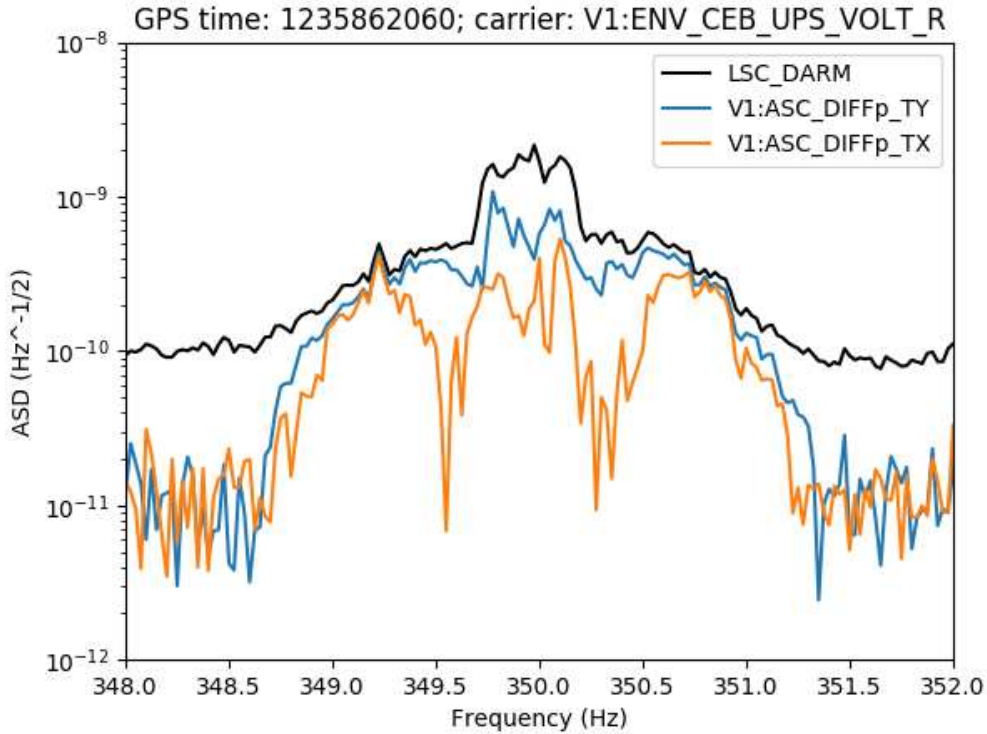
Figure 15. ASD of the DARM channel around the 350 Hz spectral line (in black), together with the noise projection, based on the `MONET` coherence values obtained with the modulator channels `ASC_Diffp_TY` and `ASC_Diffp_TX`, in blue and orange, respectively. These two channels are used to control the angular motion of the mirrors in the detector arms, in order to guarantee the proper recombination of the laser beams at the beam splitter. The initial GPS time of the analysed data and the chosen carrier signal are indicated on the top of the figure.

6. Common LIGO-Virgo tools

6.1. DQSEGDB

For each data quality flag, the Data Quality SEGment Database (`DQSEGDB`) [61] stores the segments (integer GPS ranges) during which that particular flag is active, meaning that the set of conditions it is based on is fulfilled. For instance, one such flag tags the GPS segments during which the Virgo detector is taking data in Science mode. There are two ways to fill this database with Virgo flags:

- Online, during the data taking, through the `SegOnline` server [9] that is compiling information provided by various data streams;
- Offline, by completing or fixing existing segment sets, or adding new data quality flags to monitor additional conditions.

A versioning system is used to keep track of changes in segment lists that can modify a particular flag, i.e. that impact offline analyses, by changing the contents of the dataset they are processing. By convention, the highest version number corresponds to the best (most recent) segment list and is the one queried by default.

6.2. GraceDB

During O3, the GRAvitational-wave Candidate Event DataBase (`GraceDB`) [62] has been the central place where information about transient GW candidates was uploaded and stored: online search triggers, source localization estimates in the sky, data quality information, other metadata, etc. In particular, `GraceDB` triggered automatically frameworks like the `DQR` through the LIGO-Virgo Alert System (`LVAlert`) [63] when candidate events of interest were identified; and, consequently, `DQR` results (see section 7) got uploaded back to `GraceDB` as soon as they became available. `GraceDB` has a public-faced portal that provides information about the public alerts shared with the astronomer community, while most of its data are private and reserved to the LIGO, Virgo and KAGRA collaborations.

7. Data Quality Reports

7.1. Introduction

The `DQR` is a framework developed by LIGO and Virgo for the O3 run, in order to quickly gather enough information to vet the significant triggers found by the online transient GW searches. The goal is to either confirm the associated public alert, or have it retracted at once. All 80 public alerts delivered during O3 [64] (of which 24 have been retracted) have used this input.

A `DQR` runs on a computing cluster where the $h(t)$ strain channel and the associated raw data auxiliary channels are available in low latency. Therefore, each collaboration (Virgo at EGO and LIGO with its two detectors) was responsible for the implementation, the operation, the monitoring and the upgrade of its own `DQR` framework. There was however an agreement on a common format for the check outputs, originally developed by LIGO [65].

The `DQR` framework is triggered by `GraceDB` through the `LVAlert` protocol. A `JSON` payload received from `GraceDB` allows for the generation and the configuration of a new `DQR`. The checks are then processed and their results are uploaded back to `GraceDB`, alongside all the records associated with that particular GW candidate. In order for the `DQR` to be triggered, a candidate event is required to have a false-alarm rate below 1/day. This is a conservative threshold, much higher than that required to release the candidate as a public alert, but still low enough to keep the computational cost of generating the `DQR`s under control. Therefore, in average, only a handful of `DQR`s were automatically
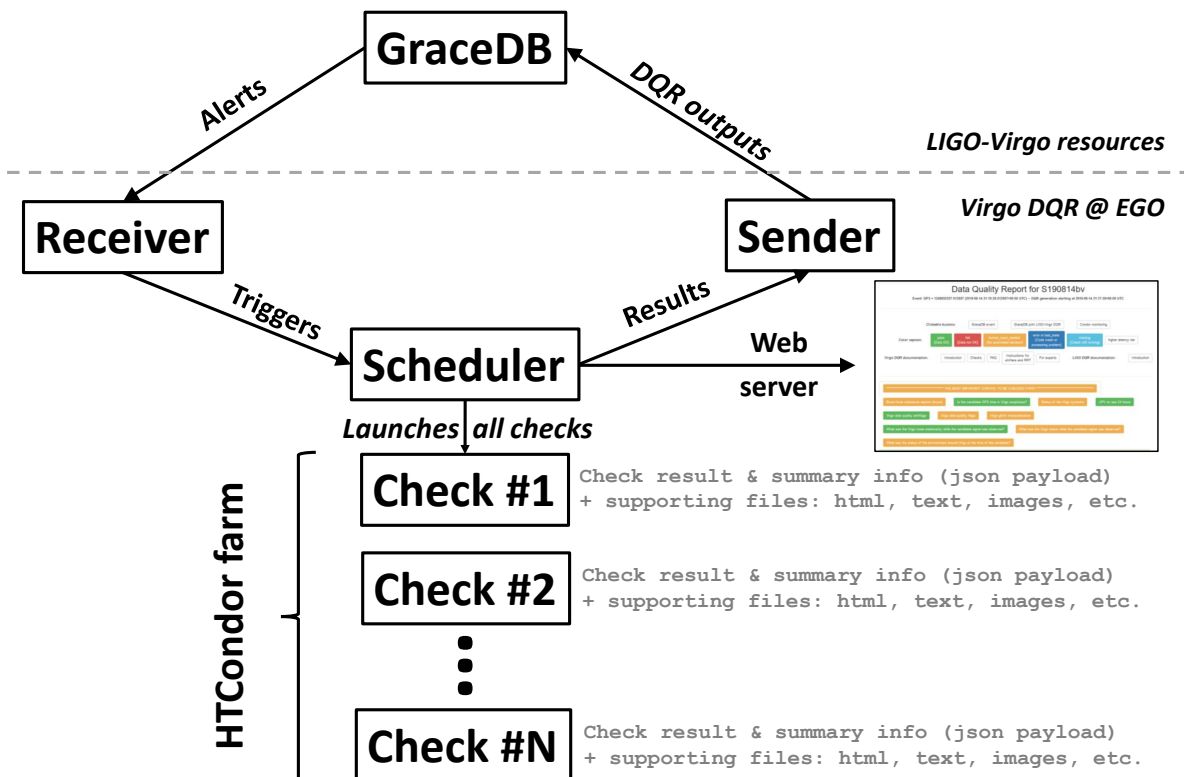
Figure 16. Schematics of the Virgo O3 `DQR` architecture. See text for a description of this workflow.

processed on a daily basis during the ∼330 days of the O3 run: not a high CPU load overall, but still about 20 times more `DQR`s than the number of public alerts that had to be vetted.

7.2. *Virgo implementation and contents*

Figure 16 summarizes the Virgo `DQR` architecture used during the O3 run. When a trigger with a low-enough false alarm rate is received, a new `DQR` is created and configured, using information from `GraceDB`. Then, the data quality checks are run in parallel on the EGO HTCondor [66] farm. As soon as a given check is complete, its results are uploaded back to `GraceDB`. In parallel, the `DQR` progress and results are immediately available for Virgo DetChar experts and on-duty people, through an EGO-internal web server. The `DQR` format [65], originally developed by LIGO, is lighter and more versatile than the `GraceDB` user interface: it ensures a direct access to the Virgo `DQR` outputs. The `DQR` web page URL is automatically sent to the relevant internal mailing list as soon as the newly-created `DQR` processing starts.

The Virgo `DQR` framework has evolved quite significantly over the course of O3. Partly to tune and improve the workflow based on the experience accumulated when stressing

the system during the actual run, but mainly to extend the scope of the `DQR` by adding additional data quality checks. These new checks were either tests that had been foreseen but could not have been implemented by the start of O3, or new procedures that brought additional information that was found missing or useful when vetting real O3 triggers.

Therefore, at the end of O3, the Virgo `DQR` included 34 checks, for a total of 99 jobs. There are roughly three jobs per check: the first, to run the code and process the data; the second, to post-process the check results and convert them to the LIGO-Virgo common `DQR` format; finally, the third to upload the results back to `GraceDB`. Some checks included a fourth job as an initial configuration phase while others, developed specifically for the `DQR`, produced directly check outputs in the required `DQR` output format, meaning that those checks required one job less.

The Virgo `DQR` checks have been categorized in the following way:

- Key checks
  They bring information mandatory to properly vet a candidate event. This includes: the top-level status of the detector at the time of the trigger; some time-frequency spectrograms of the GW strain data at different timescales around that time; finally, the scan of the main data quality flags available online, in order to look for any obvious problem in the data.

- Characterization of the Virgo detector noise around the time of the trigger
  The noise transients (glitches) are inventoried and their potential overlap with the time-frequency extent expected for the candidate is probed, if applicable. In addition, searches for noise correlations in the time domain and noise coherences in the frequency domain are run, such as tests of noise Gaussianity and stationarity.

- Detailed Virgo status
  Several different analysis contribute to this global picture of the instrument. All data quality flags available are checked. In addition, the `DMS` database is scanned to extract the snapshots closest in time to the trigger, to see what warnings or alarms were on, if any. Also, the logfiles of all the online servers running in the DAQ are scanned to spot errors that could be coincident with the trigger or impact it. Finally, various live data/reference comparison plots are generated to check the time series and distributions of a subset of the DAQ channels.

- Digest of the environment status
  This includes checking the seismic noise at EGO in various frequency bands corresponding to different sources (microseism related to sea activity on the Tuscany shoreline or local anthropogenic activities: see [7] for details), the sea activity and the weather.

8. Conclusion

This article has presented the diverse tools used by the Virgo DetChar group to fulfill its tasks during the O3 run. These tools will again form the basis of the DetChar framework for the next LIGO-Virgo-KAGRA observing run O4, currently scheduled to start in Spring 2023. As the sensitivity of the detectors is expected to be improved, a larger rate of detections should follow. To prepare this new challenge, work has been ongoing for the past two years and will continue in the coming months. Improvements are manifold: first on the core software itself, by improving the performance of the existing packages and extending their functionalities. A priority is to automate analyses further and to have them run more often, so that results are readily available and simpler to use. Another avenue being explored is the use on Virgo data of tools which have been developed by our LIGO colleagues: `iDQ` [67] to help identifying glitches in the GW strain channel $h(t)$ or `STAMP-PEM` [68], a framework similar to `BruCo` but designed to run on much longer stretches of data and useful for the analyses searching for continuous GW signals. Finally, new tools are being developed as well, one example being the hunt for scattering light-induced noise, one of the main limitations of the current ground-based GW detectors. These additional analyses will require dedicated computing resources which will be provided by the Virgo Tier1 computing centers, in particular the CC-IN2P3 in France.

List of Abbreviations

`BRiSTOL` Band-limited RMS Stationarity Test Tool

`BruCo` brute-force coherence tool

`DMS` Detector Monitoring System

`DQR` Data Quality Report

`DQSEGDB` Data Quality SEGment Database

`GraceDB` GRAvitational-wave Candidate Event DataBase

`LVAlert` LIGO-Virgo Alert System

`MONET` Modulated NoisE Tool

`NoEMi` Noise Frequency Event Miner

`UPV` Use-Percentage Veto

`VIM` Virgo Interferometer Monitor

AdV Advanced Virgo

ASD amplitude spectral density

BNS binary neutron star

BRMS Band-limited RMS

BS beam splitter

DAQ data acquisition system

DARM difference of the two arm cavity lengths

DOF degree of freedom

EGO European Gravitational Observatory

EMD Empirical Mode Decomposition

FFT fast Fourier transform

GW gravitational wave

GWOSC Gravitational Wave Open Science Center

KAGRA Kamioka Gravitational-Wave Detector

LIGO Laser Interferometer Gravitational-wave Observatory

PR power recycling

PSD power spectral density

RMS root mean square

SNR signal-to-noise ratio

SWEB suspended west-end bench

VPM Virgo Process Monitoring

References

[1] Abbott R *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, KAGRA Collaboration) 2021 GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run URL https://arxiv.org/abs/2111.03606

[2] Aasi J *et al.* (LIGO Scientific Collaboration) 2015 *Class. Quant. Grav.* **32** 074001

[3] Acernese F *et al.* (Virgo Collaboration) 2015 *Class. Quant. Grav.* **32** 024001

[4] Akutsu T *et al.* (KAGRA Collaboration) 2021 *Progress of Theoretical and Experimental Physics* **2021** ISSN 2050-3911 05A102

[5] Aasi J *et al.* (Virgo Collaboration) 2012 *Class. Quant. Grav.* **29** 155002

[6] Aasi J *et al.* (LIGO Scientific Collaboration, Virgo Collaboration) 2015 *Class. Quant. Grav.* **32** 115012

[7] Acernese F *et al.* 2022 *Class. Quant. Grav.* **39** 235009

[8] Fiori I *et al.* 2020 *Galaxies* **8** ISSN 2075-4434

[9] Acernese F *et al.* 2022 Virgo Detector Characterization and Data Quality: results from the O3 run URL https://arxiv.org/abs/2210.15633

[10] D Verkindt 2021 Advanced Virgo+ `dataDisplay` https://tds.virgo-gw.eu/ql/?c=16295

[11] Berni F *et al.* 2012 The Detector Monitoring System https://tds.virgo-gw.eu/ql/?c=9005

[12] F Berni 2020 `DMS` help manual https://tds.virgo-gw.eu/ql/?c=15469

[13] Abbott R *et al.* (LIGO Scientific Collaboration, Virgo Collaboration) 2020 *Phys. Rev. D* **102** 043015

[14] Hemming G and Verkindt D 2016 Virgo Interferometer Monitor (`VIM`) Web User Interface (WUI) User Guide Tech. Rep. VIR-0546A-16 URL https://tds.virgo-gw.eu/ql/?c=11869.

[15] Verkindt D 2019 *Proceedings of 54th Rencontres de Moriond, La Thuile* URL http://moriond.in2p3.fr/download/proceedings-moriond-gravitation-2019.pdf

[16] Frame Library https://git.ligo.org/virgo/virgoapp/Fr/-/tree/master/doc

[17] Virgo Collaboration 2021 `PythonVirgoTools` https://anaconda.org/conda-forge/pythonvirgotools

[18] Aubin F *et al.* 2021 *Class. Quant. Grav.* **38** 095004

[19] Dal Canton T *et al.* 2021 *Astrophys. J.* **923** 254

[20] Messick C *et al.* 2017 *Phys. Rev. D* **95** 042001

[21] Davis M H A 1989 *A Review of the Statistical Theory of Signal Detection* (Dordrecht: Springer Netherlands) pp 73–94 ISBN 978-94-009-1185-7

[22] Abbott B P *et al.* 2020 *Class. Quant. Grav.* **37** 055002

[23] Di Renzo F 2020 *Characterisation and Mitigation of Non-Stationary Noise in Advance Gravitational Wave Detectors* Ph.D. thesis Università di Pisa (Pisa, Italy) URL https://etd.adm.unipi.it/theses/available/etd-06192020-081853/

[24] Schuster A 1898 *Terrestrial Magnetism (Journal of Geophysical Research)* **3** 13

[25] Accadia T *et al.* 2012 *Journal of Physics: Conference Series* **363** 012037

[26] Acernese F *et al.* 2005 *Class. Quant. Grav.* **22** S1189

[27] Kolmogoroff A 1941 *The annals of mathematical statistics* **12** 461–463

[28] Welch P 1967 *IEEE Transactions on Audio and Electroacoustics* **15** 70–73

[29] Verkindt D 2019 Virgo spectrogram tools Tech. Rep. VIR-012A-19 URL https://tds.virgo-gw.eu/ql/?c=13835

[30] Finn L S Gonzalez G and Sutton P J Rayleigh Monitor (Gaussianity test for LIGO data) URL http://gallatin.physics.lsa.umich.edu/~keithr/lscdc/sutton_may01.html

[31] Kokoszka P and Mikosch T 2000 *Stochastic Processes and their Applications* **86** 49–79 ISSN 0304-4149

[32] Di Renzo F and Sorrentino N 2020 BRiSTOL and RAGoUT: two tools for investigating noise stationarity and Gaussianity, for Detchar studies and Data Analysis Tech. Rep. VIR-0089A-20 URL https://tds.virgo-gw.eu/ql/?c=15203

[33] Robinet F *et al.* 2020 *SoftwareX* 100620 ISSN 2352-7110

[34] Brown J C 1991 *The Journal of the Acoustical Society of America* **89** 425–434

[35] Isogai T (Virgo Collaboration, LIGO Scientific Collaboration) 2010 *J. Phys. Conf. Ser.* **243** 012005

[36] Huang N E *et al.* 1998 *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* **454** 903–995

[37] Huang N E, Shen Z and Long S R 1999 *Annual review of fluid mechanics* **31** 417–457

[38] Yang Y, Deng J and Wu C 2009 Analysis of mode mixing phenomenon in the empirical mode decomposition method *2009 Second International Symposium on Information Science and Engineering* (IEEE) pp 553–556

[39] Valdes G, O'Reilly B and Diaz M 2017 *Class. Quant. Grav.* **34** 235009

[40] Longo A *et al.* 2020 *Class. Quant. Grav.* **37** 145011

[41] Li H, Li Z and Mo W 2017 *Signal Processing* **138** 146–158

[42] Chatterji S K 2005 *The search for gravitational wave bursts in data from the second LIGO science run* Ph.D. thesis Massachusetts Institute of Technology URL http://hdl.handle.net/1721.1/34388

[43] Bianchi S *et al.* 2022 *Class. Quant. Grav.* **39** 195005

[44] Longo A *et al.* 2022 *Class. Quant. Grav.* **39** 035001

[45] Brun R and Rademakers F 1997 *Nucl. Instrum. Methods Phys. Res. A* **389** 81–86 ISSN 0168-9002 new Computing Techniques in Physics Research V

[46] Brun R *et al.* root-project/root: v6.18/02 URL https://doi.org/10.5281/zenodo.848818

[47] Hemming, G and Piccinni, O J NoEMi https://apps-online.virgo-gw.eu/noemi/docs

[48] LIGO Scientific Collaboration, Virgo Collaboration 2021 O3a Instrumental Lines https://www.gw-openscience.org/O3/o3aspeclines

[49] Vajente G 2008 *Analysis of sensitivity and noise sources for the Virgo gravitational wave interferometer* Ph.D. thesis Scuola Normale Superiore (Pisa, Italy) chapter 8 URL https://hdl.handle.net/11384/85869

[50] Vajente, G BRUte force COherence package https://github.com/gw-pem/bruco

[51] Bendat JS and Piersol AG 2010 *Random Data: Analysis and Measurement Procedures* Wiley Series in Probability and Statistics (Hoboken, New Jersey, USA: Wiley) ISBN 978-0-470-24877-5

[52] Di Renzo, F and Cieślar M 2019 Noise bumps between 155 and 170 Hz https://logbook.virgo-gw.eu/virgo/?r=47602

[53] Wąs, M 2019 PRCL bumps at 20 Hz and 40 Hz https://logbook.virgo-gw.eu/virgo/?r=44994

[54] Patricelli B and Cella G 2020 Tools for modulated noise study Tech. Rep. VIR-0575B-19 URL https://tds.virgo-gw.eu/?content=3&r=17154

[55] Fiori, I, Paoletti, F, Singh, N and Patricelli, B 2018 Coherence analysis of 50Hz harmonics and calibration lines https://logbook.virgo-gw.eu/virgo/?r=42880

[56] Fiori, I, Patricelli, B and Paoletti, F 2018 Comment to 1.5Hz sidebands https://logbook.virgo-gw.eu/virgo/?r=42630

[57] Fiori, I, Was, M, Patricelli, B and Paoletti, F 2018 Demodulation noise at mains lines https://logbook.virgo-gw.eu/virgo/?r=42158

[58] Di Renzo, F and Patricelli, B 2019 150 Hz side-bands https://logbook.virgo-gw.eu/virgo/?r=45974

[59] Was, M 2019 Comment to Range variations and subtraction efficiency https://logbook.virgo-gw.eu/virgo/?r=47675

[60] Was, M and Patricelli, B 2019 Comment to Range variations and subtraction efficiency https://logbook.virgo-gw.eu/virgo/?r=47852

[61] Fisher R P *et al.* 2021 *SoftwareX* **14** 100677

[62] The Gravitational-Wave Candidate Event Database (`GraceDB`) https://gracedb.ligo.org

[63] The LIGO-Virgo Alert system (`LVAlert`) https://gracedb.ligo.org/documentation/lvalert.html

[64] The LIGO-Virgo O3 public alerts https://gracedb.ligo.org/superevents/public/O3

[65] LIGO Scientific Collaboration V 2018 Data Quality Report User Documentation https://docs.ligo.org/detchar/data-quality-report

[66] Litzkow M, Livny M and Mutka M 1988 Condor - a hunter of idle workstations *Proceedings of the 8th International Conference of Distributed Computing Systems* pp 104–111

[67] Essick R, Godwin P, Hanna C, Blackburn L and Katsavounidis E 2020 *Machine Learning: Science and Technology* **2** 015004

[68] Meyers, Patrick Cross-correlation searches for persistent gravitational waves with Advanced LIGO and noise studies for current and future ground-based gravitational-wave detectors https://hdl.handle.net/11299/199038 STAMP-PEM is the ancronym of the Stochastic Transient Analysis Multi-detector Pipeline for Physical Environmental Monitors, described in pages 87-94. Dissertation thesis retrieved from the University of Minnesota Digital Conservancy.