Preparing for InSight: Evaluation of the Blind Test for Martian Seismicity

Martin van Driel, Savas Ceylan, John Francis Clinton, Domenico Giardini, Hector Alemany, Amir Allam, David Ambrois, Julien Balestra, Bruce Banerdt, Dirk Becker, et al.

▶ To cite this version:

Martin van Driel, Savas Ceylan, John Francis Clinton, Domenico Giardini, Hector Alemany, et al.. Preparing for InSight: Evaluation of the Blind Test for Martian Seismicity. Seismological Research Letters, 2019, 90 (4), pp.1518-1534. 10.1785/0220180379 . hal-02365912

Preparing for InSight: Evaluation of the Blind Test for Martian Seismicity

Martin van Driel[1], Savas Ceylan[1], John Francis Clinton[2], Domenico Giardini[1], Hector Alemany[12], Amir Allam[19], David Ambrois[12], Julien Balestra[12], Bruce Banerdt[6], Dirk Becker[13], Maren Böse[1,2], Marc S. Boxberg[10], Nienke Brinkman[1], Titus Casademont[13], Jérôme Chèze[12], Ingrid Daubar[6], Anne Deschamps[12], Fabian Dethof[13], Manuel Ditz[10], Melanie Drilleau[5], David Essing[13], Fabian Euchner[2], Benjamin Fernando[18], Raphael Garcia[7], Thomas Garth[18], Harriet Godwin[18], Matthew P. Golombek[6], Katharina Grunert[13], Celine Hadziioannou[13], Claudia Haindl[18], Conny Hammer[2], Isabell Hochfeld[13], Kasra Hosseini[18], Hao Hu[14], Sharon Kedar[6], Balthasar Kenda[5], Amir Khan[1], Tabea Kilchling[13], Brigitte Knapmeyer-Endrun[16, 17], Andre Lamert[10], Jiaxuan Li[14], Philippe Lognonné[5], Sarah Mader[13, 20], Lorenz Marten[13], Franziska Mehrkens[13], Diego Mercerat[4], David Mimoun[7], Thomas Möller[10], Naomi Murdoch[7], Paul Neumann[13], Robert Neurath[13], Marcel Paffrath[10], Mark P. Panning[6], Fabrice Peix[12], Ludovic Perrin[8], Lucie Rolland[12], Martin Schimmel[15], Christoph Schröer[13], Aymeric Spiga[9], Simon Christian Stähler[1], René Steinmann[13], Eleonore Stutzmann[15], Alexandre Szenicer[18], Noah Trumpik[13], Maria Tsekhmistrenko[18], Cédric Twardzik[12], Renee Weber[3], Philipp Werdenbach-Jarklowski[13], Shane Zhang[11], and Yingcai Zheng[14]

[1]Institute of Geophysics, ETH Zürich Sonneggstrasse 5, 8092 Zürich, Switzerland

[2]Swiss Seismological Service, ETH Zürich Sonneggstrasse 5, 8092 Zürich, Switzerland

[3]NASA Marshall Space Flight Center, ST13/NSSTC 2047, Huntsville, Alabama 35805 U.S.A.

[4]CEREMA Méditerranée, project team MOUVGS, 500 route des Lucioles, 06903, Sophia Antipolis, France

[5]Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Université Paris Diderot, 75013 Paris, France

[6]Jet Propulsion Laboratory, California Institute of Technology Pasadena, California 91109 U.S.A.

[7]ISAE-SUPAERO, Université de Toulouse, DEOS/ SSPA, 10 av E. Belin, 31400 Toulouse, France

[8]Centre National d'Etudes Spatiales, 18 Avenue Edouard Belin, 31400 Toulouse, France

[9]Laboratoire de Météorologie Dynamique (LMD/IPSL), Sorbonne Université, Centre National de la Recherche Scientifique, École Polytechnique, École Normale Supérieure, Paris, France

[10]Ruhr University Bochum, Faculty of Geosciences, Institute of Geology, Mineralogy and Geophysics, 44780 Bochum, Germany

[11]Department of Physics, University of Colorado Boulder, Boulder, Colorado 80309, U.S.A.

[12]Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, IRD, Géoazur, France

[13]Institute of Geophysics, University of Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany

[14]Department of Earth and Atmospheric Sciences, University of Houston, Houston, Texas, U.S.A.

[15]Institut de Physique du Globe de Paris, 1 rue Jussieu, 75252 Paris, Cedex 5, France

[16]Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany

[17]now at: Institute of Geology and Mineralogy, University of Cologne, Vinzenz-Pallotti-Str. 26, 51429 Bergisch Gladbach, Germany

[18]Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, UK

[19]Department of Geology & Geophysics, University of Utah, Salt Lake City, Utah, U.S.A.

[20]Karlsruhe Institute of Technology (KIT), Geophysical Institute, Hertzstr. 16, 76187 Karlsruhe, Germany

April 15, 2019

Abstract

In December 2018, the NASA InSight mission deployed a seismometer on the surface of Mars. In preparation for the data analysis, in July 2017 the Mars Quake Service initiated a blind test, in which participants were asked to detect and characterize seismicity embedded in a one Earth year long synthetic dataset of continuous waveforms. Synthetic data were computed for a single station, mimicking the streams that will be available from InSight as well as the expected tectonic and impact seismicity, and noise conditions on Mars (Clinton et al. 2017). In total, 84 teams from 20 countries registered for the blind test and 11 of them submitted their results in early 2018. The collection of documentations, methods, ideas and codes submitted by the participants exceeds 100 pages. The teams proposed well established as well as novel methods to tackle the challenging target of building a global seismicity catalogue using a single station. This paper summarizes the performance of the teams, and highlights the most successful contributions.

Introduction

The National Aeronautics and Space Administration (NASA) discovery-class mission InSight (Interior exploration using Seismic Investigations, Geodesy and Heat Transport, Banerdt et al. 2013, `http://insight.jpl.nasa.gov`) to Mars was launched on May 5th, 2018 and landed successfully on November 26th. It is dedicated to determining the constitution and interior structure of Mars. For this purpose, InSight deployed a single seismic station with both broadband and short-period seismometers on the surface of Mars, together with a number of other geophysical (Folkner et al. 2018; Spohn et al. 2018) and meteorological (Spiga et al. 2018) sensors. The seismic instrument package (SEIS) is specifically designed for martian conditions to record marsquakes as well as meteoroid impacts, and transmit data back to Earth for analysis (Lognonné et al. 2019, `www.seis-insight.eu`).

The Marsquake Service (MQS, Clinton et al. 2018) is tasked with the prompt review, detection and location of all martian seismicity recorded by InSight. It will also manage the seismicity catalogue, refining locations using the best available Mars models as they are developed during the project. To prepare the InSight science team and the wider seismological community for the data return, the MQS sent an open invitation to participate in a blind test to detect and locate seismic events hidden in a synthetic data set, which was published in SRL in July 2017 (Clinton et al. 2017). The data set was made available at `http://blindtest.mars.ethz.ch/` in August 2017 with mandatory registration. Following the submission deadline in February 2018, the true model and event catalogue together with the original waveform data are now openly available online.

Purpose of the Test

The blind test was initiated with the main purpose of improving and extending the set of methods for event location, discrimination and magnitude estimation as well as phase identification and source inversion to be applied in routine analysis of the InSight data set by collecting ideas from outside the InSight science team. It also helped to raise the profile of the InSight mission and to familiarize interested scientists with the data set to be expected from Mars.

Beyond this, the test also initiated a major effort to generate a single, consistent, temporal, synthetic data set that collected all best pre-landing estimates of seismicity, impacts, synthetic seismograms, atmospheric pressure variations and related noise, instrument self-noise and 1D structure models. The data set was made available in the same formats, and using similar web services as are now available for the real data from Mars. For this reason, the data set was also used for various operational readiness tests as well as scientific testing purposes in preparation for data return.

Furthermore, the submitted catalogues allow to derive detection and location thresholds as a function of magnitude and distance, that are not based on simple signal to noise ratio assumptions, but include the whole complexity of identifying and locating events in the time series. It is important to note though, that this data set included randomly distributed events over the sphere. Compared to the global fault distribution (Knapmeyer et al. 2006), this model may have too many events near the landing site, so the total number of detectable events in this dataset may be higher than predicted by recent seismicity models of similar total activity (Plesa et al. 2018). This needs to be accounted for if the detection threshold determined in this test is used for constraining seismic activity rates.

In the invitation, we envisioned a quantitative scoring in different categories (event detection and localization accuracy in different magnitude classes, impact discrimination and focal mechanism), but this turned not to be feasible given the heterogeneity of the submissions and relatively small number of detectable events in the data. Instead, we decided to focus on visual comparisons of the performances and compare them to the level 1 (L1) requirements of the mission, i.e. the required accuracy to achieve InSight's science objectives. The L1 requirements for quake location are 25% in distance and 20 degrees in azimuth (Banerdt et al. 2013).

Overview of the Test Data Set

The event catalogue included a total of 204 tectonic marsquakes as well as 36 impacts (Fig. 1), with only a fraction of them producing seismic signals above the noise level. The events were randomly distributed over the whole planet where the depth distribution of tectonic events followed a skewed Gaussian distribution with a maximum allowed depth of 80 km. The maximum event size was $M_w = 5$ and the magnitude-frequency distribution approximates a Gutenberg-Richter distribution with $a = 4.88, b = 1$; events with $M_w < 2.5$ were neglected (see Fig. 2 and Ceylan et al. 2017).

The impact catalogue is based on Teanby (2015) and the size distribution of observed newly dated craters (Daubar et al. 2018), again assuming a globally random distribution. To restrict amplitudes to levels similar to $M_w 2.5$ events, we only include impacts with impactor mass larger than 100 kg and assume an impact velocity of 10 km/s.

The seismic signals were computed using AxiSEM (Nissen-Meyer et al. 2014) and Instaseis (van Driel et al. 2015) as solutions to the elastic-wave equation in radially symmetric planet models. Continuous time series were then created by superimposing the event based data with seismic noise that reflects the pre-landing estimates for the surface installed instruments at the landing site (Murdoch et al. 2017a; Murdoch et al. 2017b; Mimoun et al. 2017; Kenda et al. 2017). It includes noise generated by the sensors and systems themselves, as well as through sources in martian environment (such as fluctuating pressure-induced ground deformation, the magnetic field, and temperature-related noise) and nearby lander (such as wind-induced solar panel vibrations).

Synthetic data were generated from one of the 14 candidate models (Zharkov and Gudkova 2005; Rivoldini et al. 2011; Khan et al. 2016) which were published as part of the data set, but the model choice was not revealed to participants. The model used for creation of waveform data set is shown in Figure 3 which explains two prominent features observed by most participating teams: 1) Clear S-wave arrivals were absent in most events due to the low velocity region in the upper mantle, which made distance estimations based only on relative P and S travel times very difficult, and 2) at the same time, the bedrock layer at the surface acted as a wave guide and caused a prominent P-coda arrival, that could be used for estimating locations in this 1D setting (see Fig. 4 for an overview of the most visible events). Such a phase is observed over long distances in specific settings on Earth, such as oceanic crust of constant thickness (e.g. Kennett and Furumura 2013), but in this blind test, it should be considered an artifact from the simple 1D model. It is not expected to be observed as a global phenomenon on Mars due to attenuation from 3D scattering.

An overview of responsibilities for the generation of the data set can be found in Table 1; further details can be found in Clinton et al. (2017). Based on the experience gained and performance of the MQS in particular within this test, the MQS is currently refining the location strategies and running an ORT (operational readiness test) with synthetic data computed in a 3D model.

In the following sections, we first summarize the methods used by each team. Then, we compare the success of each submission in terms of event detection, as well as estimated event distance, back-azimuth and origin time against the true event parameters.

Participation and Methods

In order to ensure effective communication with participants or anyone who wanted to experiment, registration for the test was mandatory for accessing the dataset. On the other hand, participation was completely voluntary; but we strongly encouraged all registrants to submit their results, particularly with event catalogues. In total, 84 teams registered and 11 of them submitted their analysis. Due to the lack of feedback, we do not have a further overview on how test data was used by other teams that downloaded the data but chose not to participate.

The participating teams were composed of researchers both from inside (IPGP, MQS, Max Planck) and outside (Colorado, Geoazur, Houston, Utah) the InSight science team. Participant profiles were rather diverse including senior researchers as well as PhD (Bochum, Oxford), masters (Hamburg) and even high school students (SEISonMars@school).

See Table 2 for a list of the teams and their members. In Table 3, we summarize the wealth of methods used by the participants with references to previous publications as much as possible, but a significant fraction of the methods applied by participants appears to have been developed specifically for this test.

Most teams inspected the waveforms visually or used spectrograms for event detection, while four teams (Bochum, Geoazur, Hamburg, Utah) also utilized STA/LTA algorithms with manual review for this purpose. In the case of a single station, event distance can be estimated using relative travel times between different body- and surface waves, and multi-orbit surface waves for the larger events. While the latter is independent of the model (Panning et al. 2017), body and minor arc surface wave travel times need a reference model for distance estimation. Hence, most teams tried to first determine the model from the 14 candidate models and then computed locations for that model. Three teams (Bochum, Colorado, MQS), however, used probabilistic methods to account for the inherent trade off between model and distance. Combining the distance estimate with the back-azimuths of the event and the known station location, an absolute location can be derived. The participants used a large variety of both P and Rayleigh polarization analysis methods for this purpose. Only two teams (Houston and MQS) attempted to determine depth, which was difficult as most events did not show clear depth-phases.

Only one team (Colorado) attempted to decorrelate the atmospheric pressure signals to reduce the noise; and one other team (Hamburg) classified pressure events automatically, while others relied on a visual check to exclude those from the catalogue. The Houston team was the only group to derive surface wave phase velocities. Two teams did not submit a catalogue but applied methods that facilitate event detection and phase recognition: IPGP focused on crustal structure and polarization analysis rather than event locations and Max Planck implemented an HMM (Hidden Markov Models) approach to detect events, which allowed them to provide only event detection times and no origin times.

None of the teams submitted information on the focal mechanisms within this test, but the method of Stähler and Sigloch (2014) has been applied successfully after the submission deadline by the MQS team for the largest 3 events (Clinton et al. 2018).

Performance

In the blind test announcement (Clinton et al. 2017), it was stated that it was mandatory to provide a location and origin time. A number of teams were only able to provide approximate detection times without locations and others only provided locations for parts of their catalogue. We decided to also show these results, though we understand that other teams that closely followed this rule may have left out detected events that they were not able to locate and hence the detection statistics needs to be interpreted with care.

Figure 5 gives an overview of the performance by different teams in detecting and locating events:

- The blue bars represent the total number of events in each catalogue, that besides true and false detections, may also include multiple detections for a single event. This was in particular the case for the fully automatic Hidden Markov Model (HMM) approach from the Max Planck team, since HMM is fundamentally a pattern matching approach operating on certain statistics that heavily relies on proper classification and representation of training events. In this application, only a single training event was used.

- The orange bars represent the number of events that could be associated with an event in the true catalogue solely based on the origin time and with duplicate detections removed. As we prevented event waveforms from overlapping in the seismicity catalogue, the association is straightforward. We assume any event time submitted that occurs within a window from 750 seconds before and 1500 seconds after the true origin time as correct. The three teams that performed best in detection (MQS, Hamburg, Bochum) all relied on a high degree of visual data inspection, while two of them (Hamburg, Bochum) assisted by STA/LTA triggering. Comparing seismic and pressure data visually allowed these teams to exclude most non-seismic events. MQS produced daily spectrograms that were visually scanned by different members of the team, which proved a very effective way to maximize event detection.
- The green bars represent the number of events for which full location information was provided (origin time, distance and azimuth).
- Finally, the red bars represents events that were located within the InSight mission L1 requirements for location accuracy.

Figure 6 shows a more detailed view of the 10 submitted catalogues, highlighting false detections (blue vertical lines) as well as detection and location of quakes (circles) impacts (star symbols). The rate of correct detection and location as well as false detections varies significantly over the time span of the dataset. This may be related to sharing of the workload between multiple operators; for example MQS split the initial detection on a monthly bases between team members.

In the following, we focus on the six teams that provided the most complete results in terms of the number of events correctly located within L1 requirements: Bochum, Geoazur, Hamburg, Houston, MQS and Oxford. MQS submitted two catalogues (focusing on absolute and relative distances, respectively), but as they are of very similar quality and were built iteratively using information from both approaches, we treat them as one for the purpose of this paper.

Distance Magnitude Trade-off

Figure 7 provides an overview of the six most complete catalogues with respect to distance and magnitude. It also reveals that although MQS had the highest number of correct detections, a handful of events were missed that other teams were able to detect, and some detected events were located more precisely by other teams. MQS carefully analyzed each of these events again to identify the root cause of these mislocations and unidentified events. Besides mislabeled seismic phases, several issues in the MQS workflow were recognized and resolved, with the most important improvement being the increase of the overlap in the daily plots used for visual screening.

Most of the six teams detected all events above magnitude 4, globally. Between magnitude 3 and 4, several teams detected all events until approximately 40 degree distance, even though they could not locate them within the L1 requirements. MQS detected all events above magnitude 3.5 and all events above magnitude 2.5 within 30 degree distance, which suggests that the detection threshold may be even lower than 2.5 for regional events. The detection curve for MQS is only distance/magnitude dependent, without an indication of an effect of different focal mechanisms.

Distance Estimation

Distance estimation (Fig. 8) was complicated by the low velocity layers in the upper mantle, which made S-waves very hard to identify in the data with the given noise. An easy estimate based only on the traveltime difference between P and S phase could hence not be applied to most events. On the other hand, Rayleigh wave group arrival times could be used with unrealistically high accuracy in this 1D model, which is one reason for running the current ORT with 3D synthetics. This new test suggests that including estimates of crustal thickness variations from gravity (Wieczorek and Zuber 2004), topography from MOLA (Mars Orbiting Laser Altimeter), and ellipticity lateral variations of surface wave arrival times of up to a few hundred seconds should be expected.

An additional simplification was employed by most teams by determining the correct model from the 14 candidate models based on the biggest event in the dataset (see table 3) and then using that model to locate the smaller events. In practice, a number of small events are expected to be seen in the data before any event that is big enough to constrain the model. To add this complexity to the problem, the data in the new 3D test was released in weekly chunks.

The MQS catalogue included a data quality classification, where reliable locations where classified as quality A, unreliable locations as quality B, and very unreliable/unconstrained locations as quality C. This figure indicates that only class C and a few class B events could not be located correctly (Clinton et al. 2018).

Back-Azimuth Estimation

The back-azimuth estimation in Figure 9 reveals that some methods suffer from a 180° ambiguity, which can however be resolved by either assuming retrograde Rayleigh motion or including the incidence angle in P-wave azimuth estimates (Panning et al. 2015; Böse et al. 2016). Like for the distance estimate, all MQS quality A and the majority of quality B location estimates meet the L1 requirement.

Origin Time Estimation

The error in origin time estimation is closely related to distance estimation by the fixed model set that was provided for this test, and this can also be observed in the strong correlation in performance for distance and origin time (Fig. 10). Similar arguments as in the distance estimation apply for the model complexities and 3D effects.

Impact Discrimination

Only one team (MQS) classified the event type as quake/impact in their catalogue. Only a single event was identified as an impact, which was correct, and no other event was mis-labeled as impact. MQS did miss the biggest impact event of the dataset in the detection stage. Hence we cannot evaluate the distinction capability in this test and just document the three strongest impact events together with three quakes for reference in Figure 11: If the signal is above the noise, the waveforms appear very distinct from quakes due to trapped energy in the high Q shallow layers of the 1D model as well as very short period surface waves excited by the surface source. In contrast, quakes at depth neither excite trapped waves in the shallow layers in this 1D model due to Snel's law nor the very short period surface waves due to their limited penetration depth.

MQS' classification of the impact was purely based on the waveform's appearance, which they recognised as very different from all other events. With very few impact events ever seismically recorded and the distinct impact behaviour due to the atmosphere on Earth compared to the Moon, there is no well established discrimination technique. Gudkova et al. (2011) suggest a different spectral content of impacts compared to quakes for the Moon. Other criteria include the depth of the event, although the absence of depth phases is difficult to demonstrate. Additionally, newly detected craters on satellite images from Mars might help to discriminate impact events if they can be correlated in time and location.

Conclusions

The submissions to this blind-test have provided the InSight science team with a range of new ideas and brought the specific challenges of single station seismology on Mars to a broader range of seismologists from the general community. In practice, the main benefits of the test to the MQS was that it provided the opportunity to thoroughly test software and routines as well as benchmark the event detection and location capabilities on a previously unavailable quality data set; and to evaluate whether there are new or existing methodologies that were overlooked and could significantly improve MQS' performance.

Finally, various teams contributed to this 1D test with a number of useful and different ideas; however, the algorithms established in MQS produced comparable or better performance. Further evaluation in the light of the 3D effects from synthetics as well as the actual seismicity observed by the InSight seismometers will be necessary to decide if MQS will adopt any of the suggested methods from other teams. From the test it is also obvious that the best performances were produced by the teams that had the time to dedicate to the test – an important lesson for MQS for organizing routine operations: one team member is always on duty to analyze all new data for possible seismic events with another person as backup. Any suspected event is then analyzed carefully by the review team before communicating to the whole science team (see Clinton et al. 2018, for details on the operations).

The blind test experience has helped forming the basis for the currently running operational readiness tests with 3D synthetic data for both the MQS and MSS (Mars Structure Service Panning et al. 2017), which give an opportunity to the operational teams to train daily data review.

Data and Resources

The test data set is described in more detail by Clinton et al. (2017) and available online at `http://blindtest.mars.ethz.ch/` (last accessed December 2018). Figures are created using ObsPy (Krischer et al. 2015). Submissions (catalogues and documentation) by individual teams are not publicly available.

Acknowledgements

The co-author list of this paper includes contributors to the evaluation (up to and including D. Giardini), contributors to the data set and invitation paper (Table 1) as well as the participants of the blind test (Table 2).

This work was jointly funded by (1) Swiss National Science Foundation and French Agence Nationale de la Recherche (SNF-ANR project 157133 “Seismology on Mars”), (2) Swiss State Secretariat for Education, Research and Innovation (project “MarsQuake Service—Preparatory Phase”) and (3) ETH Zurich (project “Preparatory phase for Mars InSight Ground Segment Support”). Additional support came from the Swiss National Supercomputing Centre (CSCS) under project ID s682. Some of the research described in this article was supported by the InSight (Interior exploration using Seismic Investigations, Geodesy and Heat Transport) project, Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (NASA). The Houston team was partially funded by EAR-1621878. A. Spiga and L. Rolland acknowledge funding by CNES (Centre National d’Études Spatiales). This paper constitutes InSight Contribution Number 93.

References

Allam, A. A., Y. Ben-Zion, and Z. Peng (2014). Seismic Imaging of a Bimaterial Interface Along the Hayward Fault, CA, with Fault Zone Head Waves and Direct P Arrivals, *Pure Appl. Geophys.* **171.11** 2993–3011.

Allam, A. A., V. Schulte-Pelkum, Y. Ben-Zion, C. Tape, N. Ruppert, and Z. E. Ross (2017). Ten kilometer vertical Moho offset and shallow velocity contrast along the Denali fault zone from double-difference tomography, receiver functions, and fault zone head waves, *Tectonophysics* **721** 56–69.

Banerdt, W. B., S. Smrekar, P. Lognonné, T. Spohn, S. Asmar, D. Banfield, L. Boschi, U. Christensen, V. Dehant, W. M. Folkner, D. Giardini, W. Goetze, M. P. Golombek, M. Grott, T. Hudson, C. Johnson, G. Kargl, N. Kobayashi, J. Maki, D. Mimoun, A. Mocquet, P. Morgan, M. P. Panning, W. Pike, J. Tromp, T. van Zoest, R. Weber, M. A. Wieczorek,

R. F. Garcia, and K. Hurst (2013). InSight: A Discovery Mission to Explore the Interior of Mars, *44th Lunar Planet. Sci. Conf.* 1915.

Bayer, B., R. Kind, M. Hoffmann, X. Yuan, and T. Meier (2012). Tracking unilateral earthquake rupture by P-wave polarization analysis, *Geophys. J. Int.* **188.3** 1141–1153.

Böse, M., J. F. Clinton, S. Ceylan, F. Euchner, M. van Driel, A. Khan, D. Giardini, P. Lognonné, and W. B. Banerdt (2016). A Probabilistic Framework for Single-Station Location of Seismicity on Earth and Mars, *Phys. Earth Planet. Inter.* **262** 48–65.

Böse, M., D. Giardini, S. Stähler, S. Ceylan, J. F. Clinton, M. van Driel, A. Khan, F. Euchner, P. Lognonné, and W. B. Banerdt (2018). Magnitude Scales for Marsquakes, *Bull. Seismol. Soc. Am.* **108.5A** 2764–2777.

Ceylan, S., M. van Driel, F. Euchner, A. Khan, J. F. Clinton, L. Krischer, M. Böse, S. C. Stähler, and D. Giardini (2017). From Initial Models of Seismicity, Structure and Noise to Synthetic Seismograms for Mars, *Space Sci. Rev.* **211.1-4** 595–610.

Clinton, J., D. Giardini, M. Böse, S. Ceylan, M. van Driel, F. Euchner, R. F. Garcia, S. Kedar, A. Khan, S. C. Stähler, B. Banerdt, P. Lognonne, E. Beucler, I. Daubar, M. Drilleau, M. Golombek, T. Kawamura, M. Knapmeyer, B. Knapmeyer-Endrun, D. Mimoun, A. Mocquet, M. Panning, C. Perrin, and N. A. Teanby (2018). The Marsquake Service: Securing Daily Analysis of SEIS Data and Building the Martian Seismicity Catalogue for InSight, *Space Sci. Rev.* **214.8**.

Clinton, J., D. Giardini, P. Lognonné, B. W. Banerdt, M. van Driel, M. Drilleau, N. Murdoch, M. P. Panning, R. Garcia, D. Mimoun, M. Golombek, J. Tromp, R. Weber, M. Böse, S. Ceylan, I. Daubar, B. Kenda, A. Khan, L. Perrin, and A. Spiga (2017). Preparing for InSight: An Invitation to Participate in a Blind Test for Martian Seismicity, *Seismol. Res. Lett.* **88.5** 1290–1302.

Daubar, I., P. Lognonné, N. A. Teanby, K. Miljkovic, J. Stevanović, J. Vaubaillon, B. Kenda, T. Kawamura, J. Clinton, A. Lucas, M. Drilleau, C. Yana, G. S. Collins, D. Banfield, M. Golombek, S. Kedar, N. Schmerr, R. Garcia, S. Rodriguez, T. Gudkova, S. May, M. Banks, J. Maki, E. Sansom, F. Karakostas, M. Panning, N. Fuji, J. Wookey, M. van Driel, M. Lemmon, V. Ansan, M. Böse, S. Stähler, H. Kanamori, J. Richardson, S. Smrekar, and W. B. Banerdt (2018). Impact-Seismic Investigations of the InSight Mission, *Space Sci. Rev.* **214.8**.

Eisermann, A. S., A. Ziv, and G. H. Wust-Bloch (2015). Real-Time Back Azimuth for Earthquake Early Warning, *Bull. Seismol. Soc. Am.* **105.4** 2274–2285.

Fernando, B., M. Tsekhmistrenko, and K. Hosseini (2018). Training martian seismologists for InSight, *Astron. Geophys.* **59.5** 5.17–5.21.

Folkner, W. M., V. Dehant, S. Le Maistre, M. Yseboodt, A. Rivoldini, T. Van Hoolst, S. W. Asmar, and M. P. Golombek (2018). The Rotation and Interior Structure Experiment on the InSight Mission to Mars, *Space Sci. Rev.* **214.5**.

Gudkova, T. V., P. Lognonné, and J. Gagnepain-Beyneix (2011). Large impacts detected by the Apollo seismometers: Impactor mass and source cutoff frequency estimations, *Icarus* **211.2** 1049–1065.

Hammer, C., M. Ohrnberger, and D. Fäh (2013). Classifying seismic waveforms from scratch: A case study in the alpine environment, *Geophys. J. Int.* **192.1** 425–439.

Hammer, C., M. Beyreuther, and M. Ohrnberger (2012). A seismic-event spotting system for volcano fast-response systems, *Bull. Seismol. Soc. Am.* **102.3** 948–960.

Jurkevics, A. (1988). Polarization analysis of three-component array data, *Bull. Seism. Soc. Am* **78.5** 1725–1743.

Kenda, B., P. Lognonné, A. Spiga, T. Kawamura, S. Kedar, W. B. Banerdt, R. Lorenz, D. Banfield, and M. Golombek (2017). Modeling of Ground Deformation and Shallow Surface Waves Generated by Martian Dust Devils and Perspectives for Near-Surface Structure Inversion, *Space Sci. Rev.* **211.1-4** 501–524.

Kennett, B. L. and T. Furumura (2013). High-frequency Po/So guided waves in the oceanic lithosphere: I-long-distance propagation, *Geophys. J. Int.* **195.3** 1862–1877.

Khan, A., M. van Driel, M. Böse, D. Giardini, S. Ceylan, J. Yan, J. F. Clinton, F. Euchner, P. Lognonné, N. Murdoch, D. Mimoun, M. Panning, M. Knapmeyer, and W. B. Banerdt (2016). Single-station and single-event marsquake location and inversion for structure using synthetic Martian waveforms, *Phys. Earth Planet. Inter.* **258** 28–42.

Knapmeyer, M., J. Oberst, E. Hauber, M. Wählisch, C. Deuchler, and R. Wagner (2006). Working models for spatial distribution and level of Mars' seismicity, *J. Geophys. Res.* **111.E11** E11006.

Knapmeyer-Endrun, B. and C. Hammer (2015). Identification of new events in Apollo 16 lunar seismic data by Hidden Markov Model-based event detection and classification, *J. Geophys. Res. Planets* **120.10** 1620–1645.

Krischer, L., T. Megies, R. Barsch, M. Beyreuther, T. Lecocq, C. Caudron, and J. Wassermann (2015). ObsPy: a bridge for seismology into the scientific Python ecosystem, *Comput. Sci. Discov.* **8.1** 014003.

Lin, F. C., V. C. Tsai, and B. Schmandt (2014). 3-D crustal structure of the western United States: Application of Rayleigh-wave ellipticity extracted from noise cross-correlations, *Geophys. J. Int.* **198.2** 656–670.

Lognonné, P. et al. (2019). SEIS: Insight's Seismic Experiment for Internal Structure of Mars, *Space Sci. Rev.* **215.1**.

Mimoun, D., N. Murdoch, P. Lognonné, K. Hurst, W. T. Pike, J. Hurley, T. Nébut, and W. B. Banerdt (2017). The Noise Model of the SEIS Seismometer of the InSight Mission to Mars, *Space Sci. Rev.* **211.1-4** 383–428.

Murdoch, N., B. Kenda, T. Kawamura, A. Spiga, P. Lognonné, D. Mimoun, and W. B. Banerdt (2017a). Estimations of the Seismic Pressure Noise on Mars Determined from Large Eddy Simulations and Demonstration of Pressure Decorrelation Techniques for the Insight Mission, *Space Sci. Rev.* **211.1-4** 457–483.

Murdoch, N., D. Mimoun, R. F. Garcia, W. Rapin, T. Kawamura, P. Lognonné, D. Banfield, and W. B. Banerdt (2017b). Evaluating the Wind-Induced Mechanical Noise on the InSight Seismometers, *Space Sci. Rev.* **211.1-4** 429–455.

Nissen-Meyer, T., M. van Driel, S. C. Stähler, K. Hosseini, S. Hempel, L. Auer, A. Colombi, and A. Fournier (2014). AxiSEM: broadband 3-D seismic wavefields in axisymmetric media, *Solid Earth* **5.1** 425–445.

Panning, M., É. Beucler, M. Drilleau, A. Mocquet, P. Lognonné, and W. B. Banerdt (2015). Verifying single-station seismic approaches using Earth-based data: Preparation for data return from the InSight mission to Mars, *Icarus* **248** 230–242.

Panning, M., P. Lognonné, W. Bruce Banerdt, R. Garcia, M. Golombek, S. Kedar, B. Knapmeyer-Endrun, A. Mocquet, N. A. Teanby, J. Tromp, R. Weber, É. Beucler, J.-F. Blanchette-Guertin, E. Bozdağ, M. Drilleau, T. Gudkova, S. Hempel, A. Khan, V. Lekić, N. Murdoch, A.-C. Plesa, A. Rivoldini, N. Schmerr, Y. Ruan, O. Verhoeven, C. Gao, U. Christensen, J. F. Clinton, V. Dehant, D. Giardini, D. Mimoun, W. Thomas Pike, S. Smrekar, M. Wieczorek, M. Knapmeyer, and J. Wookey (2017). Planned Products of the Mars Structure Service for the InSight Mission to Mars, *Space Sci. Rev.* **211.1-4** 611–650.

Plesa, A. C., M. Knapmeyer, M. Golombek, D. Breuer, M. Grott, T. Kawamura, P. Lognonné, N. Tosi, and R. C. Weber (2018). Present-Day Mars' Seismicity Predicted From 3-D Thermal Evolution Models of Interior Dynamics, *Geophys. Res. Lett.* **45.6** 2580–2589.

Rivoldini, A., T. Van Hoolst, O. Verhoeven, A. Mocquet, and V. Dehant (2011). Geodesy constraints on the interior structure and composition of Mars, *Icarus* **213.2** 451–472.

Ross, Z. E. and Y. Ben-Zion (2014). Automatic picking of direct P, S seismic phases and fault zone head waves, *Geophys. J. Int.* **199.1** 368–381.

Schimmel, M. (1999). Phase cross-correlations: Design, comparisons, and applications, *Bull. Seismol. Soc. Am.* **89.5** 1366–1378.

Schimmel, M., E. Stutzmann, F. Ardhuin, and J. Gallart (2011a). Polarized Earth's ambient microseismic noise, *Geochem. Geophys. Geosys.* **12.7**.

Schimmel, M., E. Stutzmann, and J. Gallart (2011b). Using instantaneous phase coherence for signal extraction from ambient noise data at a local to a global scale, *Geophys. J. Int.* **184.1** 494–506.

Selby, N. D. (2001). Association of Rayleigh Waves Using Backazimuth Measurements : Application to Test Ban Verification, *Bull. Seismol. Soc. Am.* **1.3** 580–593.

Spiga, A., D. Banfield, N. A. Teanby, F. Forget, A. Lucas, B. Kenda, J. A. Rodriguez Manfredi, R. Widmer-Schnidrig, N. Murdoch, M. T. Lemmon, R. F. Garcia, L. Martire, Ö. Karatekin, S. Le Maistre, B. Van Hove, V. Dehant, P. Lognonné, N. Mueller, R. Lorenz, D. Mimoun, S. Rodriguez, É. Beucler, I. Daubar, M. P. Golombek, T. Bertrand, Y. Nishikawa, E. Millour, L. Rolland, Q. Brissaud, T. Kawamura, A. Mocquet, R. Martin, J. Clinton, É. Stutzmann, T. Spohn, S. Smrekar, and W. B. Banerdt (2018). Atmospheric Science with InSight, *Space Sci. Rev.* **214.7**.

Spohn, T., M. Grott, S. E. Smrekar, J. Knollenberg, T. L. Hudson, C. Krause, N. Müller, J. Jänchen, A. Börner, T. Wippermann, O. Krömer, R. Lichtenheldt, L. Wisniewski, J. Grygorczuk, M. Fittock, S. Rheershemius, T. Spröwitz,

E. Kopp, I. Walter, A. C. Plesa, D. Breuer, P. Morgan, and W. B. Banerdt (2018). The Heat Flow and Physical Properties Package (HP3) for the InSight Mission, *Space Sci. Rev.* **214.5**.

Stähler, S. C. and K. Sigloch (2014). Fully probabilistic seismic source inversion – Part 1: Efficient parameterisation, *Solid Earth* **5.2** 1055–1069.

Teanby, N. (2015). Predicted detection rates of regional-scale meteorite impacts on Mars with the InSight short-period seismometer, *Icarus* **256** 49–62.

Van Driel, M., L. Krischer, S. C. Stähler, K. Hosseini, and T. Nissen-Meyer (2015). Instaseis: instant global seismograms based on a broadband waveform database, *Solid Earth* **6.2** 701–717.

Vidale, J. E. (1986). Complex polarization analysis of particle motion, *Bull. Seismol. Soc. Am.* **76.5** 1393–1405.

Wieczorek, M. A. and M. Zuber (2004). Thickness of the Martian crust: Improved constraints from geoid-to-topography ratios, *J. Geophys. Res.* **109.E1** E01009.

Zharkov, V. N. and T. V. Gudkova (2005). Construction of Martian interior model, *Sol. Syst. Res.* **39.5** 343–373.

Zheng, Y. and H. Hu (2017). Nonlinear Signal Comparison and High-Resolution Measurement of Surface-Wave Dispersion, *Bull. Seismol. Soc. Am.* **107.3** 1551–1556.

Zheng, Y., F. Nimmo, and T. Lay (2015). Seismological implications of a lithospheric low seismic velocity zone in Mars, *Phys. Earth Planet. Inter.* **240** 132–141.

List of Figures

Table 1: Contributions to the blind test data set

contribution	respsonsible co-authors (alphabetically ordered by last names)
marsquake catalogue	Savas Ceylan, John Clinton, Martin van Driel
impact catalogue	Ingrid Daubar, Matthew P. Golombek
synthetic seismograms	Martin van Driel, Melanie Drilleau
synthetic noise and pressure	Melanie Drilleau, Raphael Garcia, Balthasar Kenda, Philippe Lognonné, David Mimoun, Naomi Murdoch, Ludovic Perrin, Aymeric Spiga
compilation of 1D models	Amir Khan, Mark P. Panning
compilation of the data set and webservices	Savas Ceylan, Martin van Driel, Fabian Euchner
final choice of 1D model and catalogues	Bruce Banerdt, Martin van Driel
test conception and initiation	Domenico Giardini, Philippe Lognonné

Table 2: Participating teams and their members

group name	team members (alphabetically ordered by last names)
Bochum	Marc S. Boxberg, Manuel Ditz, Andre Lamert, Thomas Möller, Marcel Paffrath
Colorado	Shane Zhang
Geoazur	Hector Alemany, David Ambrois, Julien Balestra, Jérôme Chèze, Anne Deschamps, Diego Mercerat, Fabrice Peix, Lucie Rolland, Cédric Twardzik
SEISonMars@school	French Seismological Educational Network (SISMOS à l'Ecole) coordinated by Julien Balestra
Hamburg	Dirk Becker, Titus Casademont, Fabian Dethof, David Essing, Katharina Grunert, Celine Hadziioannou, Isabell Hochfeld, Tabea Kilchling, Sarah Mader, Lorenz Marten, Franziska Mehrkens, Paul Neumann, Robert Neurath, Christoph Schröer, René Steinmann, Noah Trumpik, Philipp Werdenbach-Jarklowski
Houston	Hao Hu, Jiaxuan Li, Yingcai Zheng
IPGP	Martin Schimmel, Eleonore Stutzmann
Max Planck	Conny Hammer, Brigitte Knapmeyer-Endrun
MQS	Maren Böse, Nienke Brinkman, Savas Ceylan, John Francis Clinton, Fabian Euchner, Domenico Giardini, Sharon Kedar, Amir Khan, Simon Christian Stähler
Oxford	Benjamin Fernando, Thomas Garth, Harriet Godwin, Claudia Haindl, Kasra Hosseini, Alexandre Szenicer, Maria Tsekhmistrenko
Utah	Amir Allam

Table 3: Overview of participating teams and methods employed

group name		methods
Bochum	**detection**:	STA/LTA triggering and manual review;
	location:	three probabilistic polarization analysis methods for azimuth (Eisermann et al. 2015; Selby 2001); probabilistic body wave and Rayleigh group traveltimes for distance (Panning et al. 2015; Böse et al. 2016).
Colorado	**detection**:	manual event detection on bandpass filtered traces;
	location:	probabilistic polarization analysis for azimuth (Böse et al. 2016); probabilistic body wave and Rayleigh group traveltimes for distance (Panning et al. 2015);
	magnitudes:	Clinton et al. 2017;
	other efforts:	attempt of pressure decorrelation (Murdoch et al. 2017a); verification of the methods on synthetics (van Driel et al. 2015; Ceylan et al. 2017).
Geoazur	**detection**:	automated event detection using different STA/LTA triggers, manual classification;
	location:	distance based on relative P-S traveltime, azimuth based on P and Rayleigh polarization (Jurkevics 1988; Bayer et al. 2012; Panning et al. 2015; Khan et al. 2016);
	other efforts:	correct model chosen based on surface wave dispersion.
SEISonMars@school	**detection**:	visual inspection of the data, manual event detection.
Hamburg	**detection**:	visual (data and spectrograms) and automated event detection (STA/LTA triggers with variable parameter settings, spectrogram detector);
	location:	visual azimuth determination using hodograms; distance based on relative P, S, R1 and multiple orbit surface waves;
	other efforts:	correct model chosen based on traveltimes and dispersion curves; automated pressure event classification.
Houston	**location**:	surface wave polarization for azimuth (Vidale 1986); relative surface wave traveltimes for distance (including minor arc only);
	other efforts:	high resolution dispersion analysis of multi-orbit surface waves to determine phase velocity and the correct model (Zheng et al. 2015; Zheng and Hu 2017); depth based on depth phases.
IPGP	**key efforts**:	autocorrelation to detect crustal discontinuities (Schimmel 1999; Schimmel et al. 2011b); degree of polarization Rayleigh wave detection and azimuth (Schimmel et al. 2011a); no catalogue submitted.
Max Planck	**key efforts**:	automated event detection and classification using Hidden Markov Models (Hammer et al. 2012; Hammer et al. 2013; Knapmeyer-Endrun and Hammer 2015); no catalogue submitted.
Marsquake Service	**detection**:	event detection by visual screening of spectrograms;
	location:	four probabilistic methods for distance and azimuth for body and surface waves (Böse et al. 2016); new model set for probabilistic methods based on the largest events; distances refined by visual alignment of waveforms vs. distance for all events; multiple iterations in relocation to detect outliers;
	magnitudes:	Böse et al. 2018;
	other efforts:	event classification based on quality of location (Clinton et al. 2018); correct model chosen; by comparing event waveforms at similar distances, depths were indicated and one event was correctly identified as an impact.
Oxford	**detection**:	visual event detection on bandpass filtered traces;
	location:	differential traveltimes and surface wave dispersion for distance; particle motion and polarization for azimuth (three different methods); detailed description in Fernando et al. (2018);
	other efforts:	three models suggested, including the correct one.
Utah	**detection**:	manual event detection assisted by STA/LTA using multiple filter bands and polarization (Jurkevics 1988; Allam et al. 2014; Ross and Ben-Zion 2014);
	location:	azimuth based on P and Rayleigh polarization; distance based on relative P and S traveltimes;
	other efforts:	model wrongly detected based on H/V ratio (Lin et al. 2014) and receiver functions (Allam et al. 2017); event classification based on radial/transverse amplitude ratio.

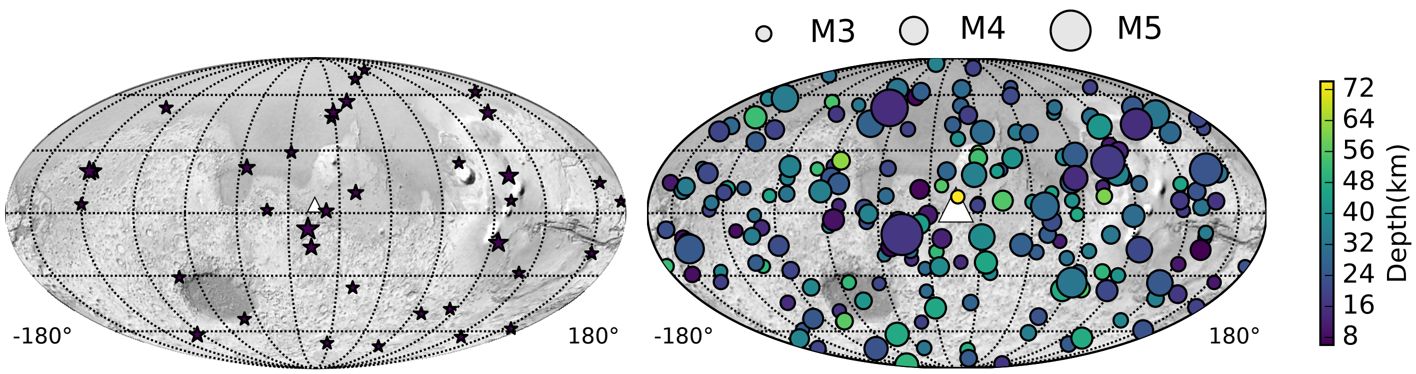


Figure 1: Catalogue summary maps: distribution of impacts (left) and marsquakes (right) in the true catalogue, both randomly distributed over the sphere. The maps are centered on the InSight landing site (white triangle). Only a fraction of these events were detectable above the noise level.

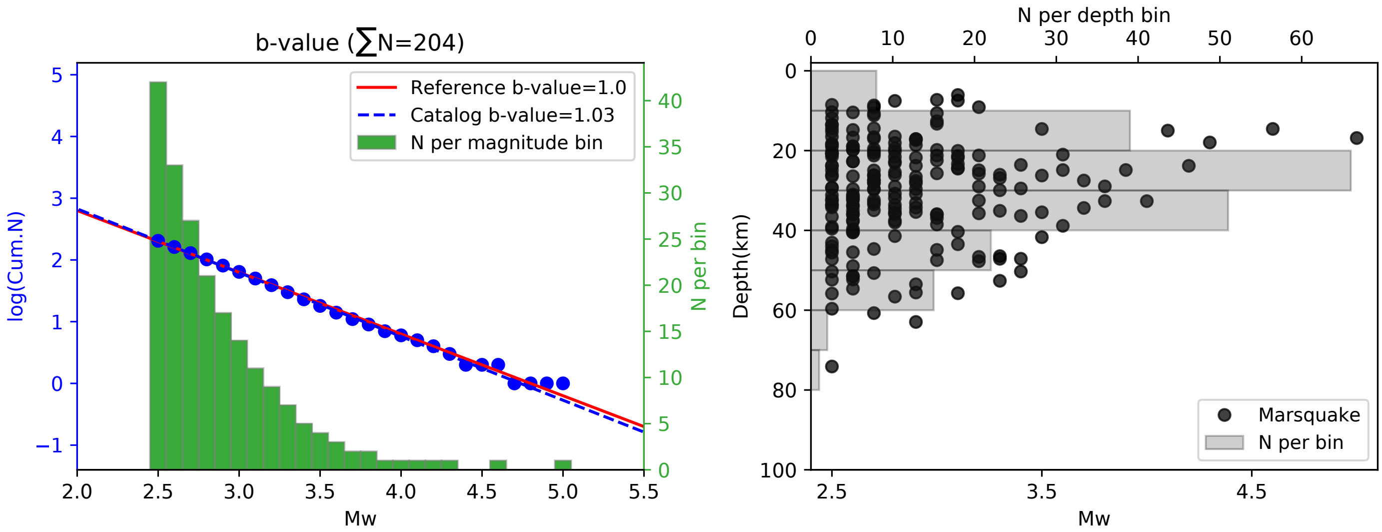


Figure 2: Statistics for marsquakes in the true catalogue: (left) The magnitude-frequency distribution approximates a Gutenberg-Richter distribution with b-value 1.0. The largest event in the catalogue has a magnitude $M_w = 5.0$. (right) The magnitude-depth distribution of the marsquakes in the true catalogue is a skewed Gaussian with a maximum event number around 20 km and maximum allowed event depth of 80 km.

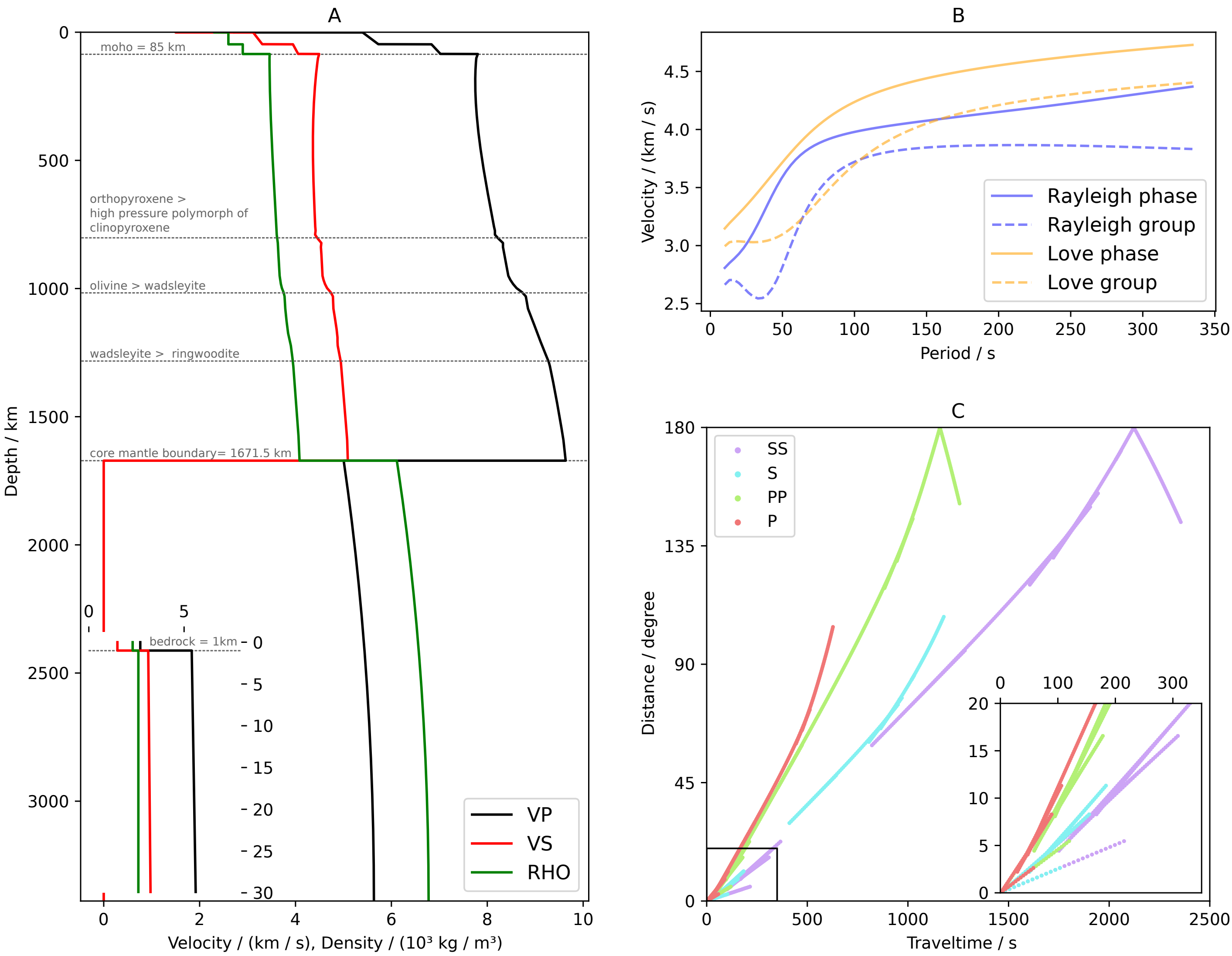


Figure 3: Summary of the model EH45TcoldCrust1b that was used in the blind test. Vertical profile of (A) seismic velocities and density, (B) dispersion curves, and (C) travel times. This model includes a low-velocity zone (LVZ, a region with a negative velocity gradient for either or both P and S). The LVZ leads to broad shadow zones for direct-arriving S-phases as indicated by gaps in the travel time curves in (C).

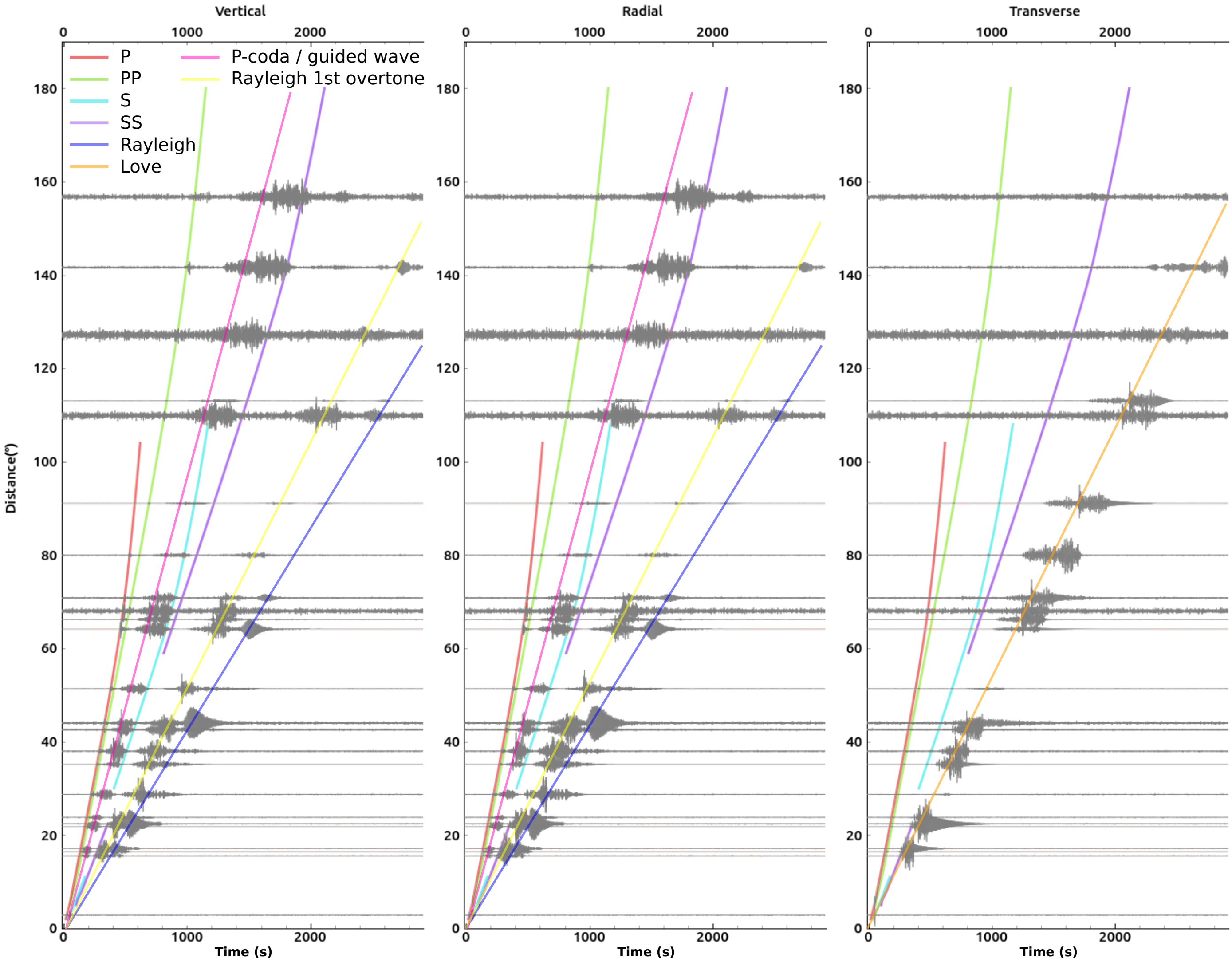


Figure 4: The most visible events in the data set, plotted as a function of distance from the station. Travel time curves for the most prominent phases are shown in the legend. The waveforms are bandpass filtered between 1.5 and 10 s.

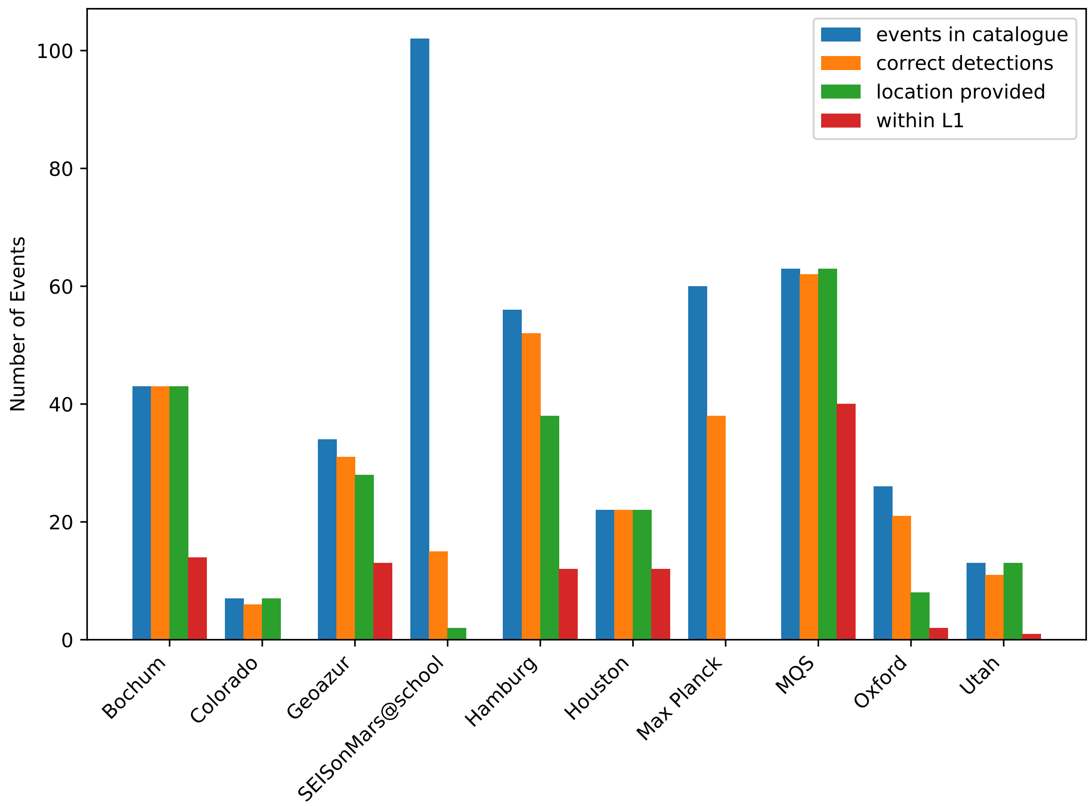


Figure 5: Summary of the team performances: total number of detected events in the submitted catalogues (blue); detected events that can be associated with an event in the true catalogue (orange); detected events in the submitted catalogues with full locations provided (green); and number of these events that lie within L1 mission requirements (red). Note the difference between orange and blue indicates false detections.

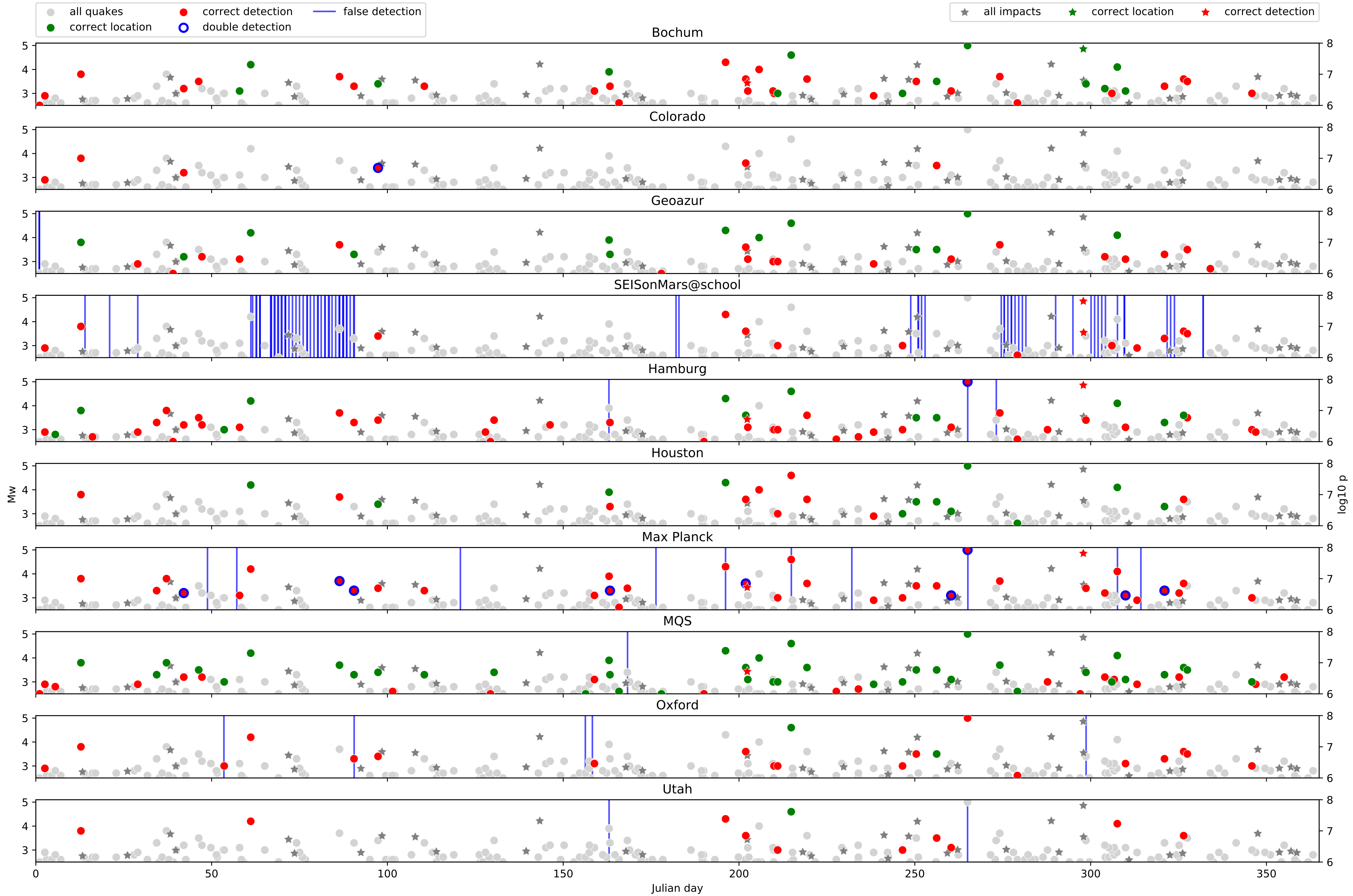
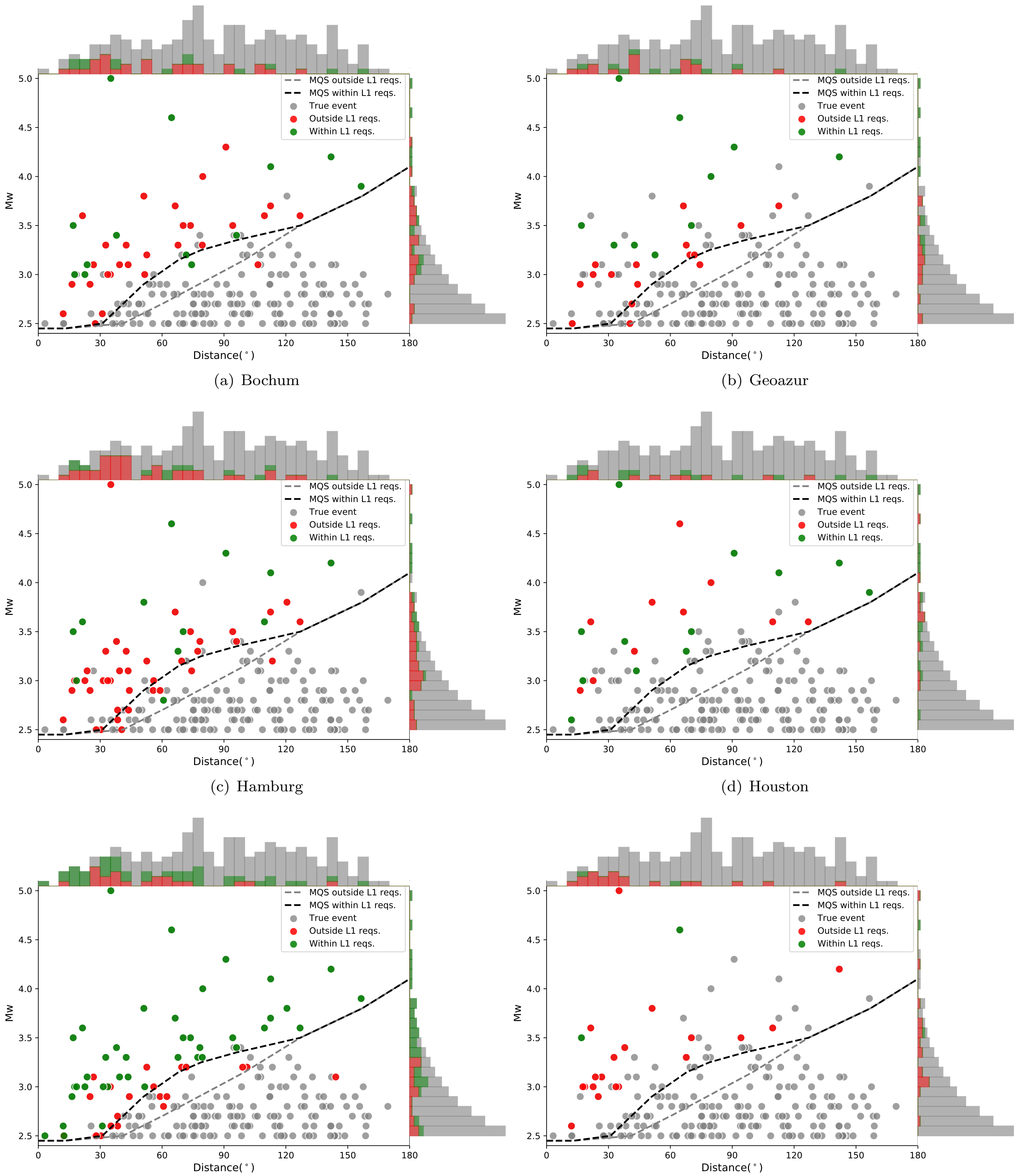
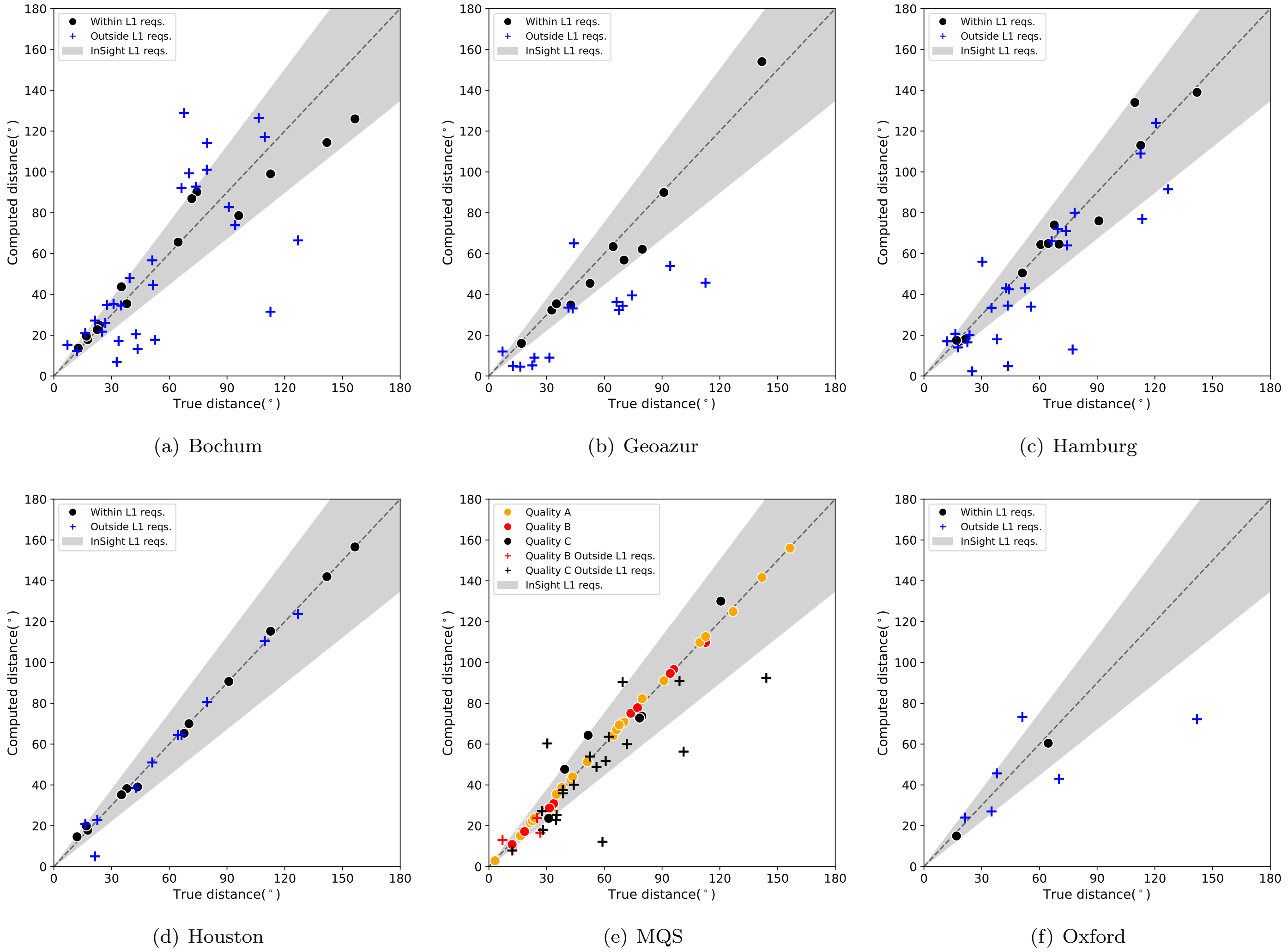


Figure 6: Temporal overview of the submitted catalogues indicating correct detections and locations as well as double and false detections. All events in the true catalogue are shown, red and green correspond to correct detection and correct location, those in gray are missing in the submitted catalogue. Marsquakes are shown as circles, and impacts as stars. Note the scale based on linear momentum p for the impacts on the right hand side.



(a) Bochum
(b) Geoazur
(c) Hamburg
(d) Houston
(e) MQS
(f) Oxford

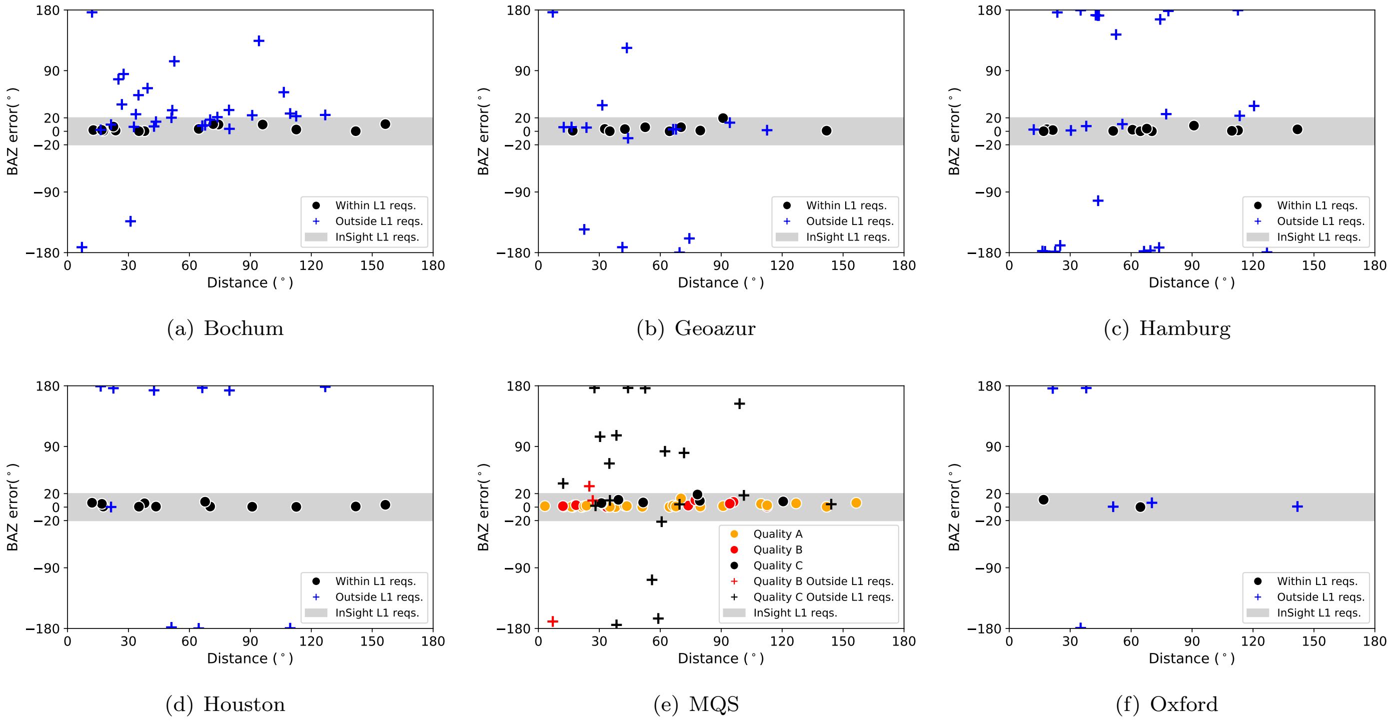
Figure 7: Distance-magnitude summary for the six most complete submitted catalogues. All events in the true catalogue are shown for each team, correctly detected in red, correctly located in green and missed events in gray. The dashed lines approximate the detection threshold (gray dashed line) and correct location threshold (black dashed line) for MQS. Histograms at the top and right side show the number of correctly detected (red), correctly located (green) and missed events (gray) for a number of distance and magnitude bins.



(a) Bochum (b) Geoazur (c) Hamburg

(d) Houston (e) MQS (f) Oxford

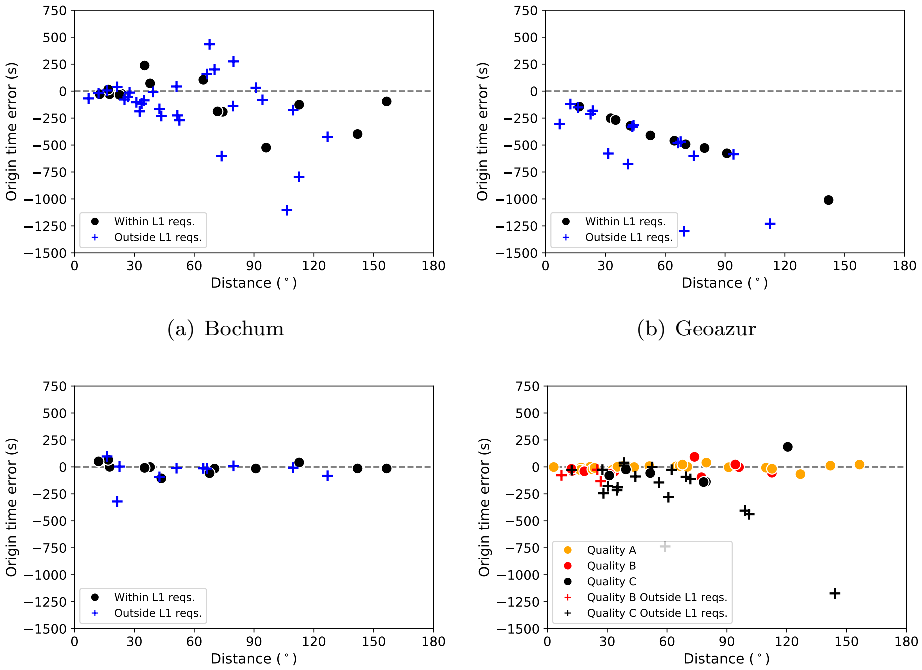
Figure 8: Distance performance - comparing the distances provided in the six most complete submitted catalogues with the true event distance. Gray area marks the L1 requirement. Note that for an event to be located within L1 we also required correct azimuth and origin time. For MQS, their data quality classification is indicated.



(a) Bochum (b) Geoazur (c) Hamburg

(d) Houston (e) MQS (f) Oxford

Figure 9: Back-azimuth performance for the six most complete submitted catalogues in terms of the back-aimuth estimation error as a function of distance. The gray area marks the mission L1 requirement. Note that for an event to be located within L1 we also required correct distance and origin time.



(a) Bochum (b) Geoazur (c) Hamburg

(d) Houston (e) MQS

Figure 10: Origin time performance for the five most complete submitted catalogues in terms of the timing error as a function of distance. Note that there is no L1 requirement, but for an event to be located within L1 we required correct azimuth and distance. Oxford's catalogue did not include origin times, but only arrival times; hence it is omitted here.

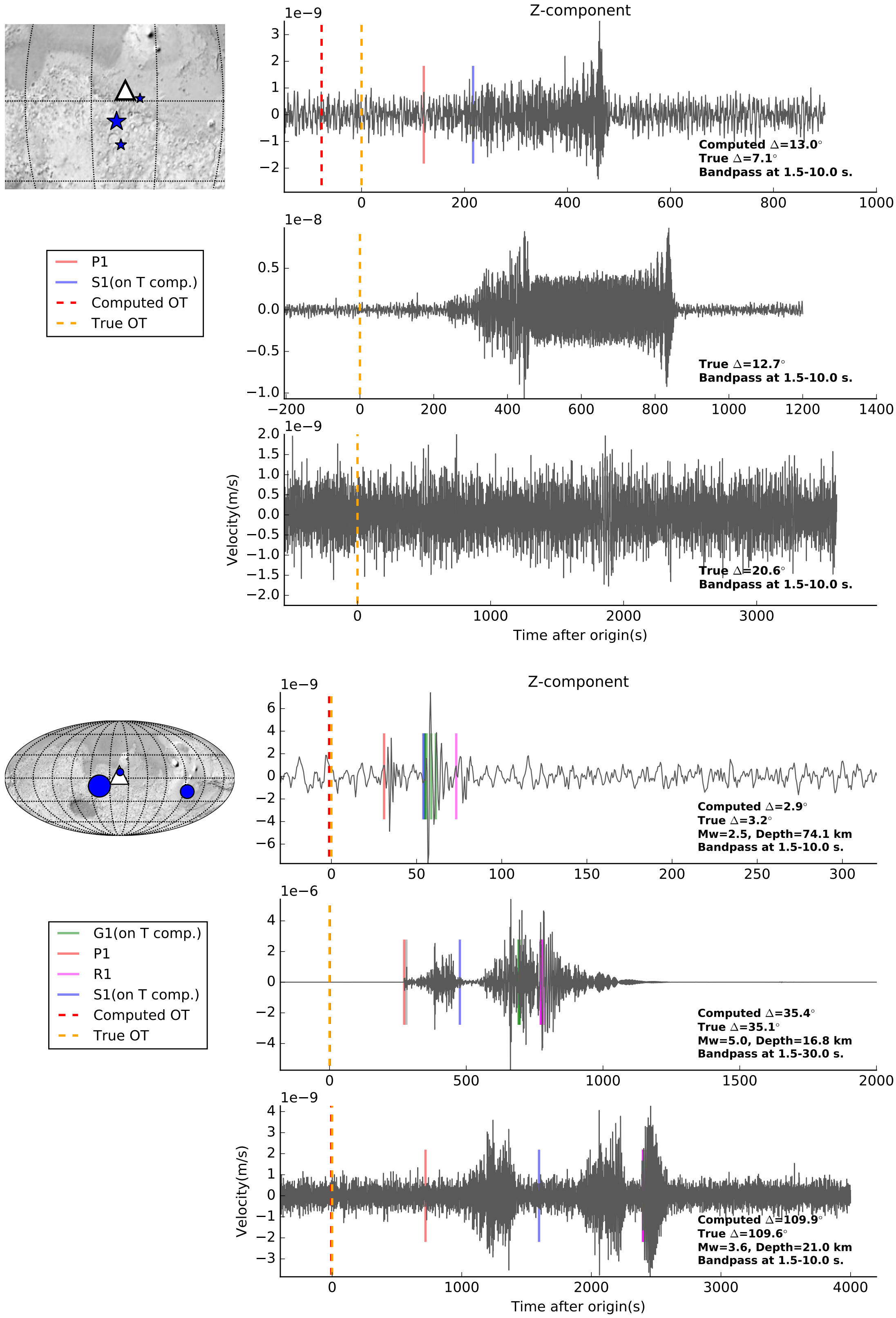


Figure 11: (top) Location and vertical component waveforms for the three strongest impact signals in the true catalogue. On the map, the impacts are indicated by stars (size proportional to the linear momentum), the station is marked with the triangle. The closest event was correctly identified as an impact by MQS. Though some other teams identified the largest event, no other team classified it as an impact in their catalogues. (bottom) Similar plot for three quakes for comparison. Seismic phases in both plots are annotated as: S1/P1 - first arriving S/P wave, where S was only visible on the tranverse component, G1/R1 - minor arc Love/Rayleigh waves, OT - source origin time.