



HAL
open science

Assessing the reliability of local earthquake tomography for crustal imaging: 30 years of records in the Western Alps as a case study

Jean Virieux, Anne Paul, Mickaël Langlais, Gaël Janex, Philippe Guéguen, Agnès Helmstetter, Laurent Stehly

► To cite this version:

Jean Virieux, Anne Paul, Mickaël Langlais, Gaël Janex, Philippe Guéguen, et al.. Assessing the reliability of local earthquake tomography for crustal imaging: 30 years of records in the Western Alps as a case study. *Geophysical Journal International*, In press. hal-04218265v1

HAL Id: hal-04218265

<https://hal.science/hal-04218265v1>

Submitted on 26 Sep 2023 (v1), last revised 28 Sep 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted 2023 September 20. Received 2023 September 13; in original form 2023 March 28

Assessing the reliability of local earthquake tomography for crustal imaging: 30 years of records in the Western Alps as a case study

J. Virieux¹, A. Paul¹, M. Langlais¹, G. Janex¹, P. Guéguen¹, A. Helmstetter¹, and L. Stehly¹

¹ *Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, UGE, ISTerre, Grenoble, France*

E-mail: Jean.Virieux@univ-grenoble-alpes.fr

SUMMARY

Local earthquake tomography (LET) is a popular method for inverting arrival time picks of local-regional earthquakes for P - and S -wave velocity and hypocenter parameters in seismically active regions. This popularity is due to some robust and well-documented open-source codes that are sometimes used as black boxes. The availability of a very complete time-pick database on the Western Alps gives us a chance to thoroughly investigate the influence of the numerous processes and parameters involved when applying LET to the Western Alps or similar targets. From a subset of high-quality manual picks (1989-2014), we compute preliminary P and S velocity models that are used to predict arrival times for later events and allow the selected fusion of picks downloaded from different seismological agencies for a consistent 33-year database (1989-2021). Different model reconstructions are proposed by selecting different subsets of the arrival time dataset. Aside data input into the LET, influence of initial stratified or three-dimensional velocity models and related initial earthquake locations is investigated together with grid discretization, Laplacian

2 *Local earthquake tomography in the Western Alps*

smoothing and damping parameters in the standard penalty approach commonly used by LET codes. These numerical parameters account for the expected limited resolution of seismic waves due to their finite-frequency content. Parameter selection is handled by the user, whereas frequency impact is only implicit in the onset picks. The earthquake distribution allows a reconstruction down to 40-km depth over an area of $\sim 500 \times 500$ km². Robust features such as the high-velocity Ivrea body anomaly, and a deep low-velocity anomaly associated with crustal thickening underneath the mountain belt survive whatever the tomography strategy and parameters. Finally, a comparison with previous LET reconstructions suggests that finite-frequency content be explicitly incorporated through wave equation tomography to improve spatial resolution. This would fully exploit observables collected from seismograms, albeit with a significant increase in computer costs.

Key words: Seismic tomography; Crustal imaging; Body waves; Western Alps; Europe

1 INTRODUCTION

Local earthquake tomography (LET) is a widely used optimization technique for the joint reconstruction of P and S -wave velocities and hypocenter parameters (e.g., Thurber 1983; Eberhart-Phillips 1986). Several robust algorithms for LET schemes have been developed by different groups that are distributed freely. The SIMULPS code is very popular thanks to the publication of a detailed description (Evans et al. 1994) and clear instructions for use. Numerous other codes have been proposed that differ essentially by the velocity description (continuous or discontinuous) and the inversion strategies (e.g., Zhao et al. 1992; Roecker et al. 1993; Koulakov 2009). The availability of these codes, the reasonable size of input files and the rather low computation costs make LET an easy-to-use seismic tomography method in regions prone to local earthquakes. Unlike other methods such as ambient-noise tomography for example, both P and S -wave velocities (or V_p/V_s ratio) can be imaged provided that P and S -wave picks are available. The joint inversion for seismic velocity and hypocenter parameters is a unique feature of LET that makes it not only useful for structural investigations but also for seismotectonic studies and seismic hazard assessment. In the Western Alps, our case study,

Local earthquake tomography in the Western Alps 3

LET has been used extensively to probe crustal structures and improve earthquake locations (e.g., Solarino et al. 1997; Paul et al. 2001; Diehl et al. 2009a,b; Solarino et al. 2018).

This easy-to-use tool must however be handled with care. It is based on ray theory, resulting in an infinitely discrete illumination by rays of the continuous Earth block to be imaged. The trade-off between wave velocity and hypocenter parameters must be given close consideration. The non-linear inverse problem is solved through iterative linearized steps with smoothing constraints starting from an initial model with given velocity structure and hypocenter locations. Even for a given discretization of the velocity model, a wide range of output models would explain the data equally well regardless of input data quality.

The purpose of this paper, targeted for practitioners, is to show how to avoid these pitfalls, and why and how the LET is model-driven while honoring data fitting. The focus will be on possible standard strategies for reconstructing the velocity and hypocenter parameters, while assessing the influence of the selection of arrival time picks, the initial choice of parameters, the discrete velocity description and the optimization approach.

The Western Alps (France, Italy, Switzerland) is our case study. The seismicity level being low to moderate, body wave arrival time tomography requires that earthquakes be monitored for a long enough time period. Our arrival time database covers more than 30 years, from 1989 to 2021. The drawback of the long recording time is the heterogeneity of the dataset, derived from constantly evolving seismic networks, and various picking methods, manual, automatic or mixed. In the earlier time period 1989-2014, we take advantage of the high-quality dataset compiled by Potin (2016) that includes nearly a million of P and S times carefully picked by visual inspection of a single experienced operator. This high-quality dataset is used to investigate the effect of several inversion parameters and to design an optimal workflow that is later applied on a more exhaustive, but far more heterogeneous dataset spanning the 33-year time period. Applying this workflow developed on a high-quality dataset to a lower-quality dataset can be viewed as a fusion strategy for heterogeneous datasets. For the more heterogeneous datasets, we select initial model and hypocenter parameters from the outputs of the high-quality dataset. This dedicated fusion strategy therefore mitigates the tradeoff issue between hypocenter parameters and seismic velocities, in addition to removing outliers.

The Alpine belt is a prime target for seismic tomography because it is an emblematic mountain range where a number of important geological concepts have been proposed such as nappe theory (Argand et al. 1911), or continental subduction. Indeed, Chopin (1984) provided the first mineralogical, direct evidence of continental subduction with the discovery of coesite in eclogite-bearing rocks of the Dora Maira massif (NW Italy). In addition, the three-

4 *Local earthquake tomography in the Western Alps*

dimensional shape of the arc of the Western Alps poses a particular challenge for seismic tomography. A number of geophysical investigations have been carried out since the 1980's to probe the crustal structure of the Western Alps and constrain geological models. In addition to LET studies cited above, the Western Alps crust has been probed by controlled-source seismology (e.g., ECORS-CROP DSS Group 1989; Nicolas et al. 1990; Thouvenot et al. 2007), ambient-noise tomography (e.g., Stehly et al. 2009; Lu et al. 2018; Zhao et al. 2020; Nouibat et al. 2022), receiver-function studies (e.g., Zhao et al. 2015; Paul et al. 2022), and teleseismic full-waveform inversion (Beller et al. 2018). The review article by Malusà et al. (2021) presents a synthesis of recent seismic tomography works and their geological interpretations. Our LET applied to the most comprehensive arrival time database available to-date gives a new 3-D crustal P - and S -wave velocity model of the Western Alps. A short comparison with the most recent tomographic models will be done, as an in-depth discussion of its contribution to crustal structure is out of scope for this paper.

The second section of this article is devoted to a description of the three subsets of picked times considered in our reconstruction of velocity models and hypocenter locations. In the third section, we review the steps of local earthquake tomography and discuss parameters that matter in the inversion process: (i) representation of the continuous P and S velocity models by Cartesian grids with smoothing constraints, (ii) forward approach for computing source-receiver rays, arrival times and sensitivity parameters based on a widely used Eikonal solver (Podvin & Lecomte 1991), (iii) inversion procedure and practical implementation using a number of hyperparameters. The tuning of the parameters is illustrated in the fourth section through application to the high-quality arrival time dataset, which also provides some flavour of the variability of reconstructed models. We conclude by suggesting a few new approaches to improve local earthquake time tomography.

2 CASE STUDY: SEISMICITY DATABASES IN THE WESTERN ALPS

A crucial step for a high-quality reconstruction of LET parameters is the picking of accurate arrival times (Diehl et al. 2009b). Our first dataset is derived from the careful manual analysis of seismic waveforms for P and S time picks performed by Potin (2016), with data from five permanent seismic networks, SISmalp (<https://sismalp.osug.fr>), CEA-DASE (<http://www-dase.cea.fr>), RéNaSS (<https://renass.unistra.fr>), RSNI (<http://www.distav.unige.it/rsni/>), and SED (<http://www.seismo.ethz.ch>), complemented by a few temporary networks.

The target zone defined by Potin (2016) is $[4.5^{\circ}\text{E} - 10.5^{\circ}\text{E} ; 43^{\circ}\text{N} - 47.5^{\circ}\text{N}]$ (Fig. 1). The maximum depth below sea level is 90 km to include deep events noticed by Cattaneo et al.

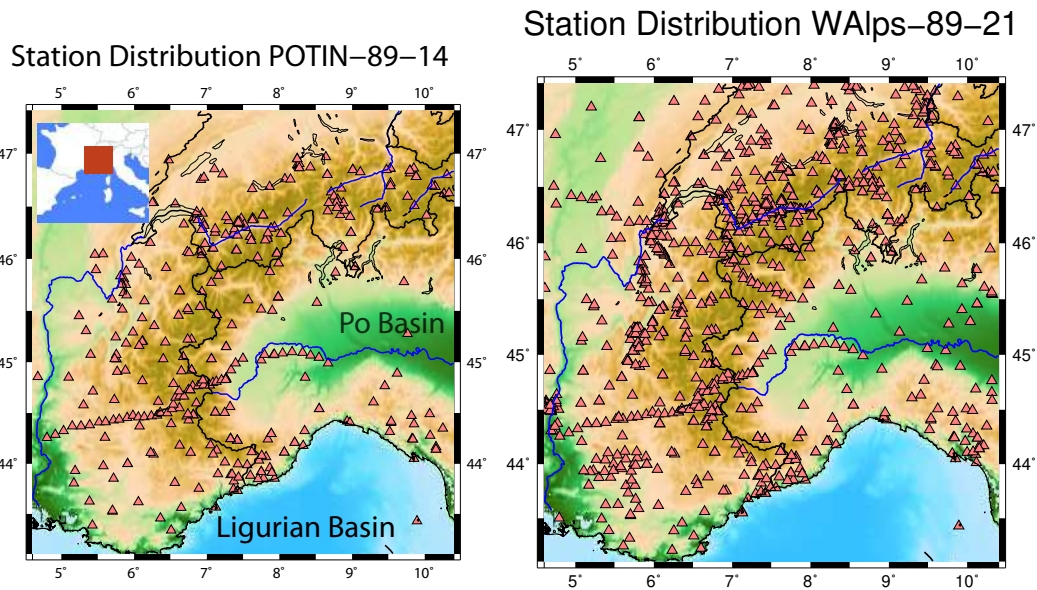


Figure 1. Map of the target zone in the Western Alps with station distribution for the Potin (2016) database on the left (1989-2014) and the new, full database on the right (1989-2021).

(1999) and Eva et al. (2015). The database is characterized by a dense station distribution (spacing 30-40 km), a completeness magnitude below 2 with $\sim 36,000$ well-located earthquakes (Potin 2016). The number of manually picked arrival times is almost a million over the time period 1989-2014. Earthquake foci are rather shallow with 50% above 10-km depth and 85% above 15 km. We will show that earthquake distribution allows for velocity reconstruction to ~ 40 -km depth in spite of shallow hypocenters.

The Potin (2016)'s dataset, named *POTIN-89-14*, is split into two different databases: a high-quality one named *HQ-89-14* with at least 12 *P* and 6 *S* picks per event (regardless of possible gaps in azimuthal distribution), and a standard database for tomographical purpose, named *SQ-89-14*, with at least 6 (*P* or *S*) picks per event (Table 1).

In Europe, the past two decades have witnessed a dramatic growth in seismic station coverage with increasing open access to seismic waveforms and products, thanks to ORFEUS (<http://www.orfeus-eu.org>) and EPOS (<https://www.epos-eu.org>). We could therefore complement the *POTIN-89-14* dataset with selected time picks downloaded from networks SISmalp, RéNaSS, RSNI and SED for the time period 2015-2021, and create a new database named *WALps-89-21* spanning the whole 1989-2021 period (Fig. 1 right). Events with at least 6 picks have been selected from the *WALps-89-21* dataset to obtain the *WA-89-21* database for tomography (Table 1).

Fusion of time-picks databases must be done with care, as data quality can be highly

6 *Local earthquake tomography in the Western Alps*

Database	Selection	Events	Stations	P picks	S picks
POTIN-89-14	All picks	54 409	373	542 818	460 129
HQ-89-14	> 12 P & 6 S	13 022	367	309 228	263 498
SQ-89-14	> 6 picks	50 331	373	533 499	451 517
WAlps-89-21	All picks	82 088	1043	952 317	670 786
WA-89-21	> 6 picks	75 538	1043	936 977	661 522

Table 1. Number of events, stations, P and S picks for the original Potin (2016)'s database *POTIN-89-14*, its derivatives *HQ-89-14* with at least 12 P picks and 6 S picks, and *SQ-89-14* with at least 6 picks, the newly created Western-Alps database *WAlps-89-21* and the selected database *WA-89-21* with at least 6 picks.

variable due to differences in acquisition systems and picking strategies, whether manual or automatic (Maupin 2020). We will show that strict screening of the new database is useless, because the 3-D velocity models obtained from database *HQ-89-14* are accurate enough for a fair prediction of synthetic times to be compared to picked times. We assume that initial hypocenter locations (HYPO71) are reliable enough to allow data selection based on the comparison of observed and synthetic times. In this way, incorrect picks are eliminated. We will only disregard observed times with an initial arrival time residual above 10 s when considering initial model and hypocenter parameters, leading to the specified time picks (Table 1). Using LET is thus an integral part of our fusion strategy. We believe that this strategy makes database *WAlps-89-21* of similar quality as database *POTIN-89-14*, but with a much larger number of picks from a greater number of stations improving spatial coverage (Tab. 1 and Fig. 1).

3 LET METHODOLOGY : A REVIEW

In this section, we decompose the LET process to clarify the contribution of parameters left to user's choice.

3.1 Discrete model space: velocity and hypocenter parameters

In the joint reconstruction of P and S wave velocities and hypocenter parameters by LET, special attention should be paid to the spatial velocity description and discrete hypocenter representation. Indeed, the various software codes differ in the velocity description, either continuous or discrete. We consider here a continuous description over a regular Cartesian grid, often referred to as nodal approach, built from a transverse Mercator projection which

1
2
3
4
5
6 122 approximatively honours distances over the 6°E longitude range in the target zone [4.5°E -
7 123 10.5°E]. Between grid nodes, we assume trilinear interpolation of the slowness, the inverse of
8
9 124 velocity. We consider regular sampling of 5 km for horizontal axes and 2 km for the vertical
10
11 125 axis. Such spacings are smaller than the expected spatial resolution of the reconstruction, but
12
13 126 independent of the discretisation for computing times and rays. Some hyperparameters of the
14
15 127 inversion procedure will be used for controlling the spatial content of the reconstructed image
16
17 128 in connection with the expected resolution. Doing so will remove any impact of the internal
18
19 129 linear interpolation of the slowness.

20
21 130 As our grid is made of $\sim 100 \times 120 \times 50$ nodes, the number of unknown velocity values
22
23 131 is 1,200,000, while the number of unknown hypocenter parameters ranges from 60,000 to
24
25 132 280,000 depending on the selected dataset. This rough estimate shows that the number of
26
27 133 unknowns with this discretisation is greater than the number of data, making the control of
28
29 134 the spatial content fairly crucial. A coarser grid is often used in LET although suggestions
30
31 135 have been made for relaxing the grid constraint (e.g., Thurber & Eberhart-Phillips 1999; Chiao
32
33 136 & Kuo 2001; Delost et al. 2008), or for a dynamically driven gridding strategy, illustrating
34
35 137 the importance of velocity discretisation (Fang et al. 2019). In this work, we rely on velocity
36
37 138 smoothing constraints proposed by nearly all used tomographic codes.

3.2 Forward problem and sensitivity assessment

38
39 140 Considering the current P or S velocity inversion grid, slowness values are re-sampled over a
40
41 141 fine regular 2-km grid by linear interpolation in the three directions. This step is required to
42
43 142 compute a travel time grid accurate enough for ray tracing using an efficient Eikonal solver
44
45 143 with receivers as initial sources (Podvin & Lecomte 1991). Inside the computed travel time
46
47 144 grid, ray tracing with a specific ray sampling of 0.5 km goes downhill on the time landscape
48
49 145 from any source back to the receiver, which is the only point with zero initial travel time (Le
50
51 146 Meur et al. 1997). Due to discrete time interpolation, very few rays can be missed in this
52
53 147 back-tracing strategy, which are often related to strong velocity gradients. The missing rays
54
55 148 do not impact the inversion as they are very few in number. Once the rays have been traced,
56
57 149 the forward grid is left behind and we end up with discrete segmented rays, discrete velocity
58
59 150 parameters and discrete hypocenter parameters.

60
61 151 Rather than using times computed by the Eikonal solver, we compute time along the
62
63 152 computed ray using the Simpson rule of the integral of the slowness assuming again a trilinear
64
65 153 interpolation of slowness over the grid. This improves accuracy of time estimation (e.g., Latorre
66
67 154 2004; Monteiller 2005; Monteiller et al. 2005). Simultaneously, Fréchet derivatives of travel

8 *Local earthquake tomography in the Western Alps*

times with respect to the cell nodes crossed by the ray are estimated using the same trilinear interpolation of slowness. The sensitivity matrix with respect to velocity parameters is quite sparse and it is stored through a compressed row storage (CRS) strategy. Let us emphasize that such a projection on grid nodes is the first discrete impact on the spatial spreading of such a sensitivity velocity matrix. The spatial resolution limit of the reconstructed image is not explicit and must be estimated during the optimization by the user. We will show that ray tomography can indeed build high-wavenumber images. Most users of ray tomography expect it to provide images with low-wavenumber content; this misconception is based on the common use of smoothing operators triggered by the user over the hard-coded grid spreading. The activation of such a triggering comes from the Occam's razor principle and not from the assumed ray approximation. Moreover, the known limited frequency range of seismic waves promotes such a low-wavenumber content when considering first-arrival times and the related transmission regime of wave-medium interaction.

We complement the velocity sensitivity matrix by derivatives with respect to hypocenter parameters, which is minus the slowness vector at the hypocenter for Cartesian coordinates and one for the origin time. The slowness vector at the hypocenter is deduced from the discrete ray by a finite-difference estimation at the hypocenter position. The CRS strategy is still active when adding these Fréchet derivatives to the global sensitivity matrix. However, such sensitivity matrix with respect to hypocenter parameters presents a very different pattern than P and S velocity sensitivity matrices with respect to data. Balancing influences of these four sensitivity matrices (one for P velocity values, one for S velocity values, one for Cartesian hypocenter coordinates, one for hypocenter origin times) is case-specific and a quantitative illustration of the potential impact of scaling these different sensitivity matrices is given by Le Meur et al. (1997) through a singular value decomposition (SVD) analysis. This challenge of multi-parameter inversion has not been fully overcome (Roecker et al. 2006). Most computer codes rely on a specific calibration between slowness/velocity sensitivity and hypocenter coordinates/origine-time sensitivity which is left up to the user through synthetic tests. We face the same issue in our tomographic application to the Western Alps target, even if we consider normalized quantities.

3.3 Inverse problem

Seismic tomography involves inverting discrete and thus incomplete information (collected at points on the Earth's surface), with uncertainties. For practical use, the imaged Earth block must be discretized. As a result, the inverse problem is ill-posed. In the LET case, hypocenter

188 parameters are naturally discretized, whilst the user must specify the discretisation of the
 189 velocity models. Numerous velocity models and hypocenter parameters fit data equally well,
 190 so that geologic interpretation of images must be made with caution. The search for the so-
 191 called “best” model combines a fit criterion to data with an operator assessing model roughness
 192 as seen by the physics of seismic wave propagation. This operator can be tuned to geological
 193 constraints.

The non-linear inverse problem proceeds in iterative linearized steps starting from an initial model with given velocity models and hypocenters. The linear system at a given iteration can be written as $\Delta t = \mathcal{J}\Delta m$ where Δt is the time difference between observed and synthetic arrival times, Δm is the model increment, and the sensitivity matrix \mathcal{J} is computed using rays (e.g., Um & Thurber 1987; Rawlinson & Sambridge 2004). The model increment Δm can be defined in a compact way by:

$$\arg \min_{\Delta m} \|\mathcal{J}\Delta m - \Delta t\|^2 + \lambda\|\mathcal{L}\Delta m\|^2 + \epsilon\|\Delta m\|^2, \quad (1)$$

where \mathcal{L} is a smoothing operator required for getting a solution of minimal norm using two hyperparameters, the scalar hyperparameter λ and the damping hyperparameter ϵ . This optimization is adapted to a mixed-determined problem where some parameters are sampled by many data while others may not be. The selected hyperparameters narrow the model description by favouring the smoothest models, according to the so-called principle of parsimony (also named Occam’s razor principle: the simplest explanation is usually the best one) as suggested by Constable et al. (1987) among others. We refer the reader to the seminal articles Tarantola & Valette (1982), Greenhalgh et al. (2006), and to textbooks on the general formulation of the inverse problem (e.g., Tarantola 2005; Menke 2012). The numerical solution of equation (1) may be performed in various ways, as hyperparameters operate in different ways. The usual way to solve this nonlinear inverse problem relies on the penalty approach where an augmented linear system must be solved (Menke 2012). This augmented linear system is described in Supplementary text S1 and could be written in the following way:

$$\begin{bmatrix} W_D \frac{\partial t}{\partial s_k} & W_D \frac{\partial t}{\partial h_k} \\ \lambda D & 0 \\ \epsilon I & 0 \\ 0 & \epsilon I \end{bmatrix} \begin{bmatrix} \Delta s_k \\ \Delta h_k \end{bmatrix} = \begin{bmatrix} W_D(t_{obs} - t(m_k)) \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (2)$$

194 This linear system highlights the difference between slowness/velocity parameters for contin-
 195 uous model description and hypocenter parameters which are discrete by definition, at least
 196 in the ray formulation. The model to be reconstructed is split into two sub-models $m = (s, h)$

10 *Local earthquake tomography in the Western Alps*

1
2
3
4
5
6 197 and increments $\Delta m = (\Delta s, \Delta h)$, where h denote the hypocenter parameter vector and s the
7
8 198 slowness vector. The initial velocity model (in fact slowness model) is often a stratified model
9
10 199 (*i.e.* depth-varying velocity). Initial hypocenter locations are computed with HYPO71, which
11
12 200 is the standard code for locating earthquakes in a stratified model (Lee & Lahr 1972). Arrival
13
14 201 times are computed in this initial model $m_0 = (s_0, h_0)$ and the inversion aims at computing
15
16 202 a better model (with a better fit to observed arrival times), assuming the earthquake origin
17
18 203 times are known. At iteration k , model m_k is perturbed by $\Delta m_k = (\Delta s_k, \Delta h_k)$, such that the
19
20 204 predicted arrival time $t(m_k) + \partial t / \partial m(m_k) \Delta m_k$ is closer to the observed arrival time t_{obs} . A
21
22 205 weighting matrix W_D is added for removing outliers from the arrival time dataset. In all our
23
24 206 applications, data with residuals larger than 4 s are removed, and a linear weight is applied
25
26 207 that decreases from 1 for 3 s residuals to 0 for 4 s residuals.

27
28 208 Because different physical quantities are considered in this linear system, we scale each
29
30 209 parameter of the four classes ($V_p, V_s, (x, y, z), t_0$) by the maximum norm of columns of the
31
32 210 sensitivity matrix for the corresponding class. This adimensionalization procedure is not used
33
34 211 in most LET computer codes, which rather apply empirical scaling of partial derivatives
35
36 212 through numerical tests as mentioned earlier. While it is rarely proposed in an explicit way,
37
38 213 such generic scaling appears to be working in most cases. We are using the LSQR algorithm
39
40 214 which requires only matrix-vector products for getting the damped least-squares solution
41
42 215 (Paige & Saunders 1982).

43
44 216 The smoothing required by the mixed-determined structure of seismic tomography is
45
46 217 an often-used Laplacian seven-points finite-difference smoothing operator \mathcal{L} . The spatial-
47
48 218 derivative operator D is deduced for each node of the inversion grid. More sophisticated
49
50 219 smoothing strategies essentially based on model covariance matrices can be considered as
51
52 220 long as spatial derivatives can be estimated (Tarantola 2005). The Laplacian-smoothing hy-
53
54 221 perparameter λ will limit variations of the spatial second derivative of the slowness model,
55
56 222 leading to smooth velocity models while still trying to fit the observations. A more physical
57
58 223 interpretation of this hyperparameter is the reduction of possible diffraction patterns in the
59
60 224 computed arrival times, since the diffraction curvature depends on spatial second derivatives
225
226 225 of slowness (Li & Fomel 2013). However, because of the acquisition geometry, we essentially
227
228 226 split the Laplace operator leading to two separate terms $\lambda_x D_x + \lambda_y D_y$ and $\lambda_z D_z$ where the
229
230 227 first one includes only horizontal derivatives and the second one vertical derivative. Even when
231
232 228 selecting $\lambda_x = \lambda_y = \lambda_z$, the spatial influence is different because of different horizontal and
233
234 229 vertical inversion grid spacings and because of the Laplacian split.

235
236 230 The damping hyperparameter ϵ helps regularize ill-conditioned systems. The LSQR algo-

231 rithm will consider this damping influence, still in the framework of penalty strategy, that
 232 prevents model perturbations from being too large at each iteration. The choice of this pa-
 233 rameter can be driven by the shape of the L-curve (Hansen 1992). Detecting the kick point in
 234 the L-curve may however be difficult (Potin 2016). In case the hyperparameter ϵ happens to
 235 be too small, model perturbations are limited by hard constraints, which amount to 800 m/s
 236 for V_p , 600 m/s for V_s , 1500 m for horizontal perturbation of hypocenter location, 500 m for
 237 vertical perturbation and 1.5 s for origin-time perturbation. These hard constraints are not
 238 critical and they may increase the number of iterations for reaching the expected values.

By solving this adimensional scaled system in the least-squares sense, we get a new model m_{k+1} with better fit to observed data and a data misfit function given by:

$$\mathcal{C}_D(\mathbf{s}, \mathbf{h}) = \frac{1}{2} \sum_{nobs} W_D(t_{obs} - t(m))^2, \quad (3)$$

a weighted root-mean-square time residual given by:

$$RMS_w = \frac{1}{nobs} \sqrt{\sum_{nobs} W_D(t_{obs} - t(m))^2}, \quad (4)$$

and a total root-mean-square time residual given by:

$$RMS = \frac{1}{nobs} \sqrt{\sum_{nobs} (t_{obs} - t(m))^2} \quad (5)$$

239 The reduction of both RMS_w and RMS is an expression of the database quality, and a
 240 way to feel confident that the least-squares strategy constrained by velocity model smoothing
 241 description is still mainly controlled by the data fit.

242 Aside these hyperparameters involved in the least-squares formulation, other inversion
 243 strategies can be considered mainly driven by prior information regarding the expected model.
 244 One way is by adding an extra term in the misfit definition where the distance to a prior
 245 model is controlled by a new hyperparameter: building such a prior model is more difficult at
 246 the lithospheric scale than for crustal reservoir targets with in-situ wells. A softer alternative
 247 comes by smoothing the slowness perturbation through a dedicated operator promoting ex-
 248 pected structural features. Limited illustrations of these smoothing strategies will be given in
 249 Supplementary text S2 (Suppl. Fig. S1), while building prior model requires a more in-depth
 250 analysis of the geodynamic context. However, in this work, we only consider the widely used
 251 Laplacian-smoothing strategy of LET codes.

252 Testing the influence of hyperparameters is an endless search. We advise the LET user to
 253 focus on parameters that most influence the outcome, *i.e.* the initial model and the spatial
 254 discretisation hampered by the Laplacian smoothing component controlled by the smoothing

12 *Local earthquake tomography in the Western Alps*

hyperparameter λ . High values should be chosen for the damping hyperparameter ϵ to ensure that the non-linear inverse problem performs correctly in linear steps (Kissling et al. 1994).

Once a minimum misfit is reached for a given initial model, a local sensitivity analysis can be achieved either by a checkerboard test or a spike test (Lévêque et al. 1993). In this paper, we will consider a spike test which benefits from the modular structure of the tomographic code that we are using (Latorre et al. 2004). However, LET users should be warned that such local sensitivity analysis is in no respect a test of the solution's unicity. The sensitivity analysis of the expected resolution only tells us the local topography of the data misfit landscape at the selected minimum solution. Even transdimensional approaches (Piana Agostinetti et al. 2015) cannot overcome this issue of solution unicity. We do not consider the related discrete-velocity-model optimization in this article devoted to classical LET methodology.

4 TESTING LET PARAMETERS ON THE HIGH-QUALITY *HQ-89-14* DATABASE

As mentioned in the introduction, data quality and consistency are essential for a reliable LET. In this section, we take benefit of the *HQ-89-14* database (Table 1), which meets these quality criteria, in order to test several inversion strategies and parameters and define the most effective one for application to lower quality databases with higher spatial coverage. We use the stratified initial model proposed by Potin (2016) for P velocity (shown in Supplementary Fig. S2) and the HYPO71 hypocenter locations performed in the reference SISmalp four-layer model used routinely for locating earthquakes by the French RéNaSS observatory (red dots in Fig. 2). The S velocity model is deduced by dividing P velocities by the V_p/V_s ratio of 1.6933 obtained from the Wadati diagram (Potin 2016).

All tested inversion scenarios lead to a reduction of the root-mean-square (RMS) time residual from 0.78 s to 0.5-0.6 s depending on the selected strategy and hyperparameters. Supplementary Fig. S4 shows the weighted ($\sim 36\%$) and the unweighted ($\sim 30\%$) residual reductions; their similarity illustrates the time picks quality. A damping parameter $\epsilon = 0.1$ ensures small enough model perturbations at each iteration. The Laplacian-smoothing strategy with a weight $\lambda_{x,y} = 5$ over the horizontal distance and a similar weight $\lambda_z = 5$ over the vertical distance provides a rather smooth model with a significant data misfit reduction.

Supplementary Fig. S5 shows that the residual distribution has more positive values when considering the initial layered model, while the residuals distribution driven by the least-squares strategy balancing positive and negative residuals after the inversion is more symmetric as expected by this misfit definition.

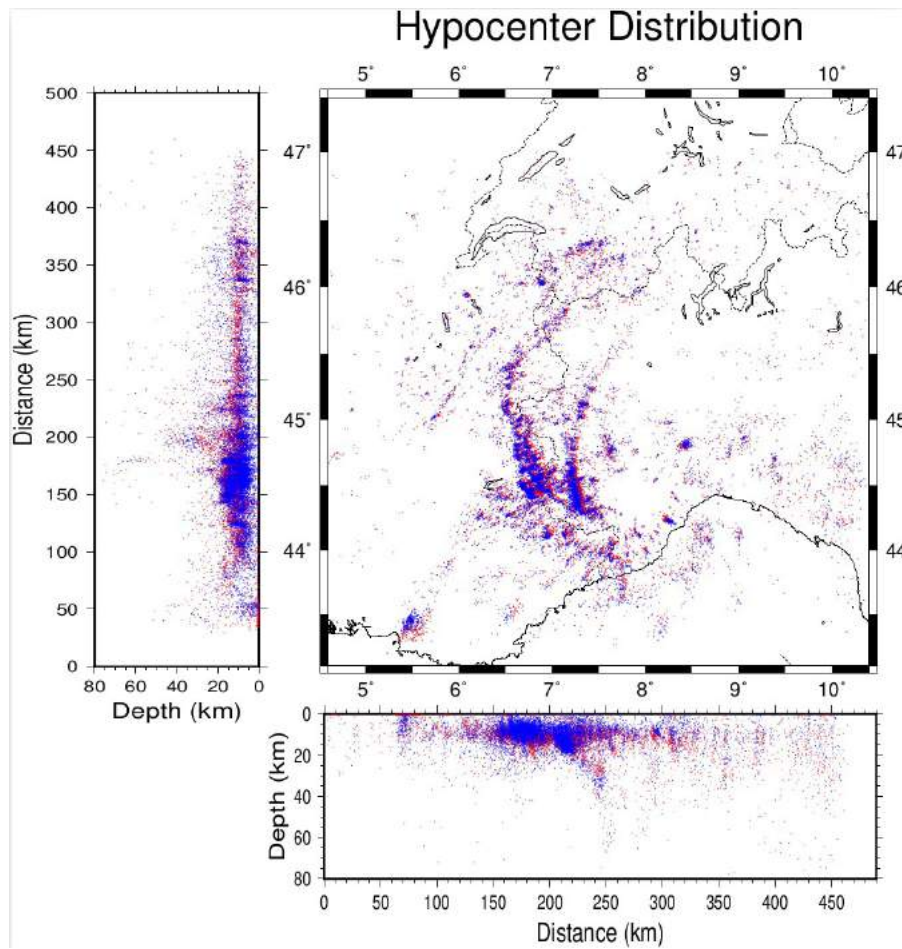


Figure 2. Distribution of hypocenters (*HQ-89-14* database). Red dots: initial locations; Blue dots: final locations. After inversion, hypocenters tend to move up at shallow depth (0-20 km) while they tend to move down at larger depths (30-50 km) as shown in Supplementary Fig. S3. Deep events (>40 km) are observed in the Italian (Eastern) side.

288 Reconstructed velocity models at iteration 21 are shown in Figure 3 at two depths, while
 289 the final hypocenter locations are displayed as blue dots in Figure 2. Initial hypocenter loca-
 290 tions are derived from the routine HYPO71 location process of the SISmalp network. Indeed,
 291 we did not carry out a first location step in the initial model to mitigate the trade-off be-
 292 tween velocity and hypocenter parameters, which is a well-known issue of such joint inversion.
 293 Velocities at 10-km depth reflect sampling of the upper crust by seismological stations, as il-
 294 lustrated by the clear correspondence between station locations in the Po basin and the strong
 295 low-velocity anomaly centred at [8.2°E, 45°N]. Conversely, velocity slices at 20-km depth dis-
 296 play rather stable features regardless of discretisation and hyperparameter values, such as the
 297 low-velocity zone (LVZ) at \sim [6.8°E, 45 - 46.3°N] and the fast-velocity anomaly at \sim [7.3 - 8°E;

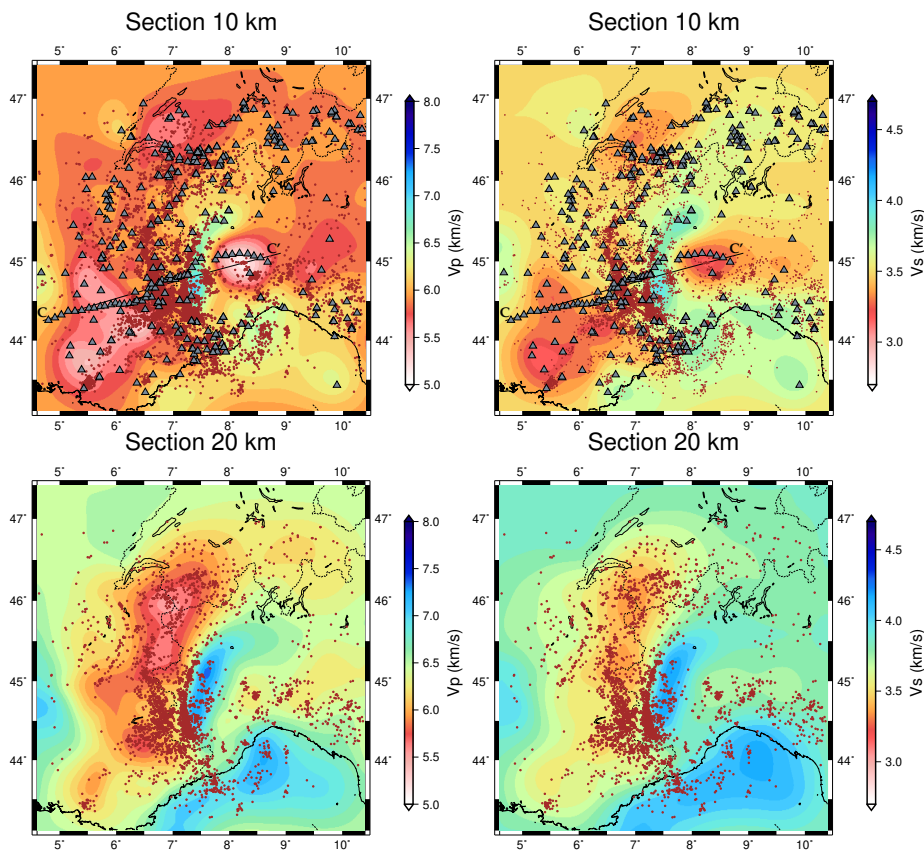
14 *Local earthquake tomography in the Western Alps*

Figure 3. Results of the inversion of the *HQ-89-14* dataset. The top panels show horizontal sections at 10 km depth, with V_p model on the left, and V_s model on the right. The bottom panels are sections at 20 km depth. Epicenters with foci at less than 10-km vertical distance are plotted as brown dots. Stations are shown as grey triangles. The black line labeled *CC'* is the CIFALPS profile.

44.3 - 45.5°N] that is well-known as the Ivrea body seismic anomaly (*e.g.* Paul et al. 2001). We will however see that the spatial extension and the amplitude of the velocity anomalies may change with changing inversion parameters. The hypocenter distributions are roughly similar for the initial model and final models, thanks to the large number of picks (> 12 P and 6 S) controlling locations (Fig. 2). Shallow earthquakes tend to move up during the inversion while some deeper events are moving down (supplementary Fig. S3) Depth sections in the V_p and V_s models along the CIFALPS reference profile are shown in Supplementary Fig. S6; they will be used later in comparisons with other final models.

4.1 Influence of hyperparameters

In this section, we investigate the sensitivity of reconstructed velocity models to hyperparameter values. The variability, or stability of velocity anomalies is valuable information for

309 preventing over-interpretation. Since there are so many possible hyperparameters, a limited
310 number of scenarios are considered for illustration purposes.

311 4.1.1 Damping hyperparameter

312 For a given discretisation of the velocity subspace of 5 km in horizontal and 2 km in vertical,
313 we may consider only variations of the ϵ damping hyperparameter applied in the LSQR solver
314 while keeping fixed the λ smoothing hyperparameter. We have tested values of ϵ from 100 to
315 0.01 while applying previously given values of the hyperparameter λ . The selected ϵ value of
316 0.1 allows an efficient decrease of the misfit when applying the penalty approach. Conversely,
317 a damping of 10 leads to a slow misfit decrease that may require hundreds of iterations. In
318 order to better separate the influences of the smoothing and damping hyperparameters, we
319 set $\lambda = 0$ (no laplacian smoothing) and $\epsilon = 10$ (strong damping). The 20-km depth slice in
320 the resulting V_p model is displayed in Figure 4. It shows that, due to the ray approach, the
321 size of velocity heterogeneities is only controlled by the grid discretisation since no spatial
322 smoothing is applied. The velocity image exhibits a high wavenumber content that mirrors
323 the infinitely thin ray paths. Indeed, the limited frequency content of seismic waves is ignored,
324 and time variations at receivers can be expressed as diffraction effects at the high-frequency
325 limit with no healing effects on wavefronts (Nolet & Dahlen 2000). The V_p depth section
326 along the CIFALPS profile CC' is shown in the bottom panel of Fig. 4. By contrast with
327 Supplementary Fig. S7, it illustrates the strong influence of the smoothing hyperparameter
328 λ , set to 0 in this example. Although arrival time fit is as good for this model as for the
329 previous one (Fig. 3 and Suppl. Fig. S7), such model reconstruction ignores the intrinsic
330 seismic frequency limitation, and therefore is not amenable to geological interpretation. This
331 velocity model will be considered as initial model for following inversions with the *SQ-14-*
332 *89* and *WA-89-21* databases, to help us pinpoint robust velocity anomalies versus anomalies
333 driven primarily by the discretisation and hyperparameter choices.

334 4.1.2 Laplacian-smoothing hyperparameter

335 The damping is now set to $\epsilon = 0.1$ and we keep the same discretisation (5 km horizontally,
336 2 km vertically). The inversion is firstly performed using a smoothing weight $\lambda = 5$, and
337 equal horizontal and vertical constraints on second-order spatial derivatives. The resulting
338 V_p and V_s models and new hypocenter locations after 21 iterations are then taken as input
339 parameters for another inversion with a smoothing weight $\lambda_{x,y} = 3$ for horizontal axes and
340 $\lambda_z = 2$ for the vertical axis. After 21 new iterations, we repeat the procedure with two smaller

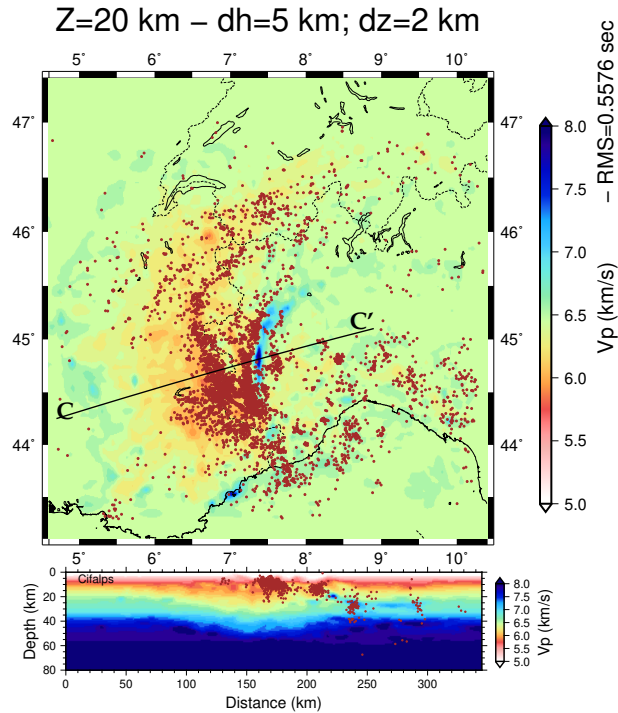
16 *Local earthquake tomography in the Western Alps*

Figure 4. Influence of the hyperparameter ϵ for a given spatial discretisation without any applied smoothing. (Top) Depth slice at 20-km depth; (Bottom) Depth section along profile CC'. The RMS is quite low while the size of velocity heterogeneities is only controlled by grid spacing and ray paths, due to the high-frequency assumption that mitigates healing effects on diffraction patterns at receivers.

smoothing weights $\lambda_{x,y} = 2$ and $\lambda_z = 1$. As shown by Supplementary Fig. S7, this procedure introduces progressively higher-wavenumber content in the resulting velocity model. In order to check that the low-velocity anomaly ($V_p \sim 5$ km/s) does not result from a possible trade-off between velocity and hypocenter parameters, we use the same workflow while keeping HYPO71 locations as initial hypocenter parameters. Indeed, velocity anomalies are slightly weaker in Suppl. Fig. S6 (bottom panels), but still present, which suggests that leakage is still present between model parameters. In order to avoid these low-velocity anomalies, we select the smoothing hyperparameter values $\lambda_{x,y} = \lambda_z = 5$ for our next inversions.

4.1.3 Grid discretisation

The inversion grid choice is critical and it should be tied to the resolution that depends on the wave frequency content. However the selection of the discretisation is left to the user since the LET does not explicitly depend on the frequency content. Because the station network lies mostly on the free surface, the horizontal discretisation should be different from the vertical one. We consider here a vertical discretisation of 2 km, and two strategies for progressively

refining the grid in the horizontal directions: one from 10 km to 5 km and another one with two steps from 20 km to 10 km, then from 10 km to 5 km. The results of these inversions are shown at 20-km depth in Supplementary Fig. S8, and compared to the initial inversion with 5-km grid. The three inversions result in similar RMS values around 0.5 s. The grid refinement strategies induce spreading of the velocity anomalies and strong edge effects. The projection of ray trajectories onto nodes of the inversion grid through trilinear interpolation induces low-wavenumber artifacts, in particular along edges, that survive the grid multiscale procedure (Suppl. Fig. S7). Improved multiscale strategies are possible using wavelets for velocity model description; those wavelets better control the wavenumber content injected in the inversion (Chiao & Kuo 2001; Delost et al. 2008).

However, the joint inversion for velocity and hypocenter locations is always challenging because hypocenter parameters are intrinsically discrete. The LET user must therefore pay attention to potential leakages of hypocenter/velocity parameters that may be induced by specific discrete velocity description. Synthetic tests in a configuration which should be representative of the final solution are strongly advised. These tests should be repeated for newly found solutions.

4.1.4 *Summary on hyperparameter influence*

These investigations of velocity grid discretisation and hyperparameter influence emphasize that ray-based tomography probes velocity structures along ray paths only, disregarding nearby velocity values (Fig. 4). This is related to the frequency-independent assumption of the LET (high-frequency) approach. Gap filling between ray paths primarily depends on the grid discretisation. However, smooth velocity changes may be expected in relation to the frequency content of seismic waves. In the classical LET approach considered here, this smoothing depends on the user's choice from the often-used Laplacian-smoothing strategy. Attempts to define a more physics-driven strategy where the source frequency content is taken into account have been proposed by different authors (Dahlen et al. 2000; Nolet 2008; Zelt & Chen 2016). To our knowledge, these methodological advances have not yet been extensively applied to real datasets at the lithospheric scale for local earthquake tomography.

Our tests show that using the penalty approach combining smoothing and damping strategies with a fine enough grid yields the optimal compromise between data fit and model smoothness. This strategy is indeed adopted by most LET applications. Performances of such a strategy depend on the user's control of the limited frequency content of seismic waves. Such physical limitation is not explicit in the picked P and S data.

4.2 Initial model design

Trade-off between hypocenter and velocity parameters is a well-known issue of local earthquake tomography. In this section, we investigate the influence of initial hypocenter locations when the initial velocity model is 3-D. We show that the trade-off is not amplified when performing the LET with initial earthquake locations in a stratified model (HYPO71).

This analysis is carried out on the *HQ-89-14* dataset, with expected well-constrained hypocenter locations, using the smoothing strategy and a small damping parameter. In section 4, we used the stratified model proposed by Potin (2016) as initial velocity model with initial HYPO71 locations in the SISmalp four-layer model. We will refer to this initial model as *Z-HQ* in the following. A second initial model is constructed by combining the stratified velocity model with a 3-D model from ambient noise tomography (ANT, Nouibat et al. 2022); It assumes that the shallow layers are better sampled by surface-wave tomography than by LET, which is limited by the sparsity of the station array. The ANT V_s model is used at shallow depth (< 10 km) and its influence is progressively damped with depth back to the stratified V_s model at 30-km depth. V_p is computed from V_s using the constant P/S velocity ratio 1.6933. Supplementary Fig. S9 shows the 10- and 20-km depth slices in this *3D-ANT* initial model. HYPO71 locations are still considered as initial locations in a first test using this initial 3D model. In a second step, hypocenters are relocated using the NLLOC software (Lomax et al. 2000) in the 3D model before starting the inversion. We therefore test three different initial model configurations: the stratified model with HYPO71 locations (*Z-HQ*), the *3D-ANT* model with HYPO71 locations and the *3D-ANT* model with NLLOC locations.

Supplementary Fig. S10 shows that the initial distribution of residuals is sharper for HYPO71 locations in a stratified model than for NLLOC locations in the *3D-ANT* model. The NLLOC location process yields a lot of negative residuals before the inversion, which have their origin in the location strategy of NLLOC. It starts with an initial spatial grid search before deducing the origin time as the average of time delays left at the location point. The final residual distribution is much sharper for the *Z-HQ* initial model than for the two *3D-ANT* models. The *3D-ANT* initial velocity model has a low-wavenumber (high-velocity) content which is not fully attenuated by inversion regardless of initial hypocentral locations. The data misfit curves of Suppl. Fig. S11 illustrate the influence of initial locations of hypocenters. HYPO71 locations give a smaller initial misfit for the stratified model than for the 3-D model. Initial data misfit is not significantly improved with NLLOC locations in the initial 3-D model. Moreover, very similar misfits after 16 iterations suggest that final locations are mostly similar for HYPO71 and NLLOC initial locations.

Fig. 5 shows the 10 and 20-km depth slices in the reconstructed V_p models for the three tested initial models. At 10-km depth, similar velocity anomalies observed in all three models are data-driven since they do not exist in the initial stratified model, while other patterns are remnants of the stratified or 3-D initial velocity models. For example, the small-scale low-velocity anomaly beneath the easternmost stations of the CIFALPS profile in the left panel (initial stratified model) still exists but with a larger size in the center and right panels (initial 3-D model). Indeed, the 3-D initial model includes a broad low-velocity anomaly at 10-km depth below stations at the eastern end of the CIFALPS profile and further east (Supplementary Fig. S9). This suggests that the low-velocity anomaly is probably real, but with different size and shape. This larger-size low-velocity anomaly exists at a depth of 10 km in the initial model *3D-ANT*, below stations at the eastern end of the CIFALPS profile and further east (Supplementary Fig. S9). We will check in section 6 if the newly built *WA-89-21* dataset better captures this anomaly due to the larger station number. At 20-km depth, the three velocity maps are quite similar albeit velocity contrasts are smaller with the initial 3-D model. This again illustrates the trade-off between velocity and hypocenter parameters, and also the dependence on the choice of the starting model. Supplementary Fig. S12 compares final hypocenter distributions for the three different initial models. Epicenter locations differ close to box boundaries, in particular along the coast and in the Ligurian Sea, while they are very similar in the Alps thanks to the high number of P and S picks. Depth sections show that hypocenter locations tend to be shallower for the initial *3D-ANT* model and HYPO71 locations.

Locating hypocenters in an initial 3-D structure does not seem to be crucial at such regional scale, even with a 3-D initial velocity model. The LET sets hypocenters at nearly the same final locations when starting from initial HYPO71 locations. Therefore, such HYPO71 locations will be used in inversions with the full database.

An alternative strategy for assessing the influence of the initial velocity model consists in creating a large set of stratified random models about a reference model. Supplementary text S3 and Figs. S13-16 present the results of such tests for a series of 1024 random initial stratified models around the Potin (2016)'s stratified model used in previous inversions. The ratio between the RMS of the final P velocity models and the RMS of the initial velocity models (shown in Suppl. Fig. S14) may be considered as a proxy of the initial-model influence on the reconstructed P velocity models. As noticed earlier, the influence of the initial model is strong at 10-km depth, in particular along the boundaries of the box due to poor station

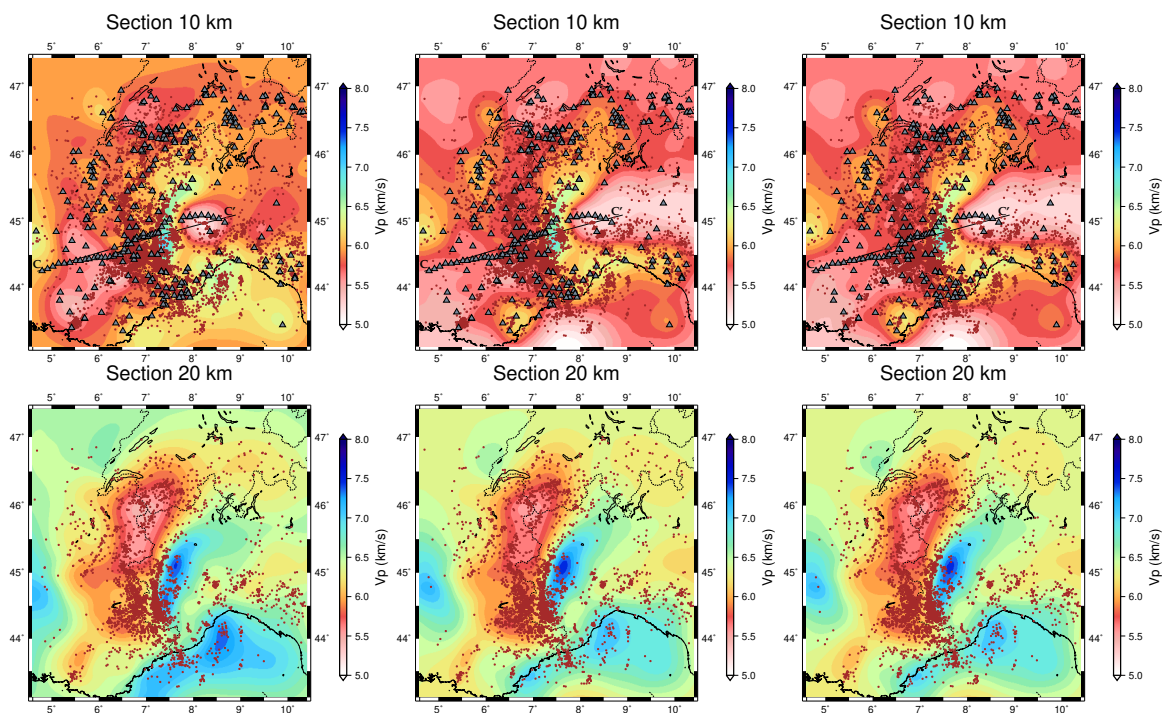
20 *Local earthquake tomography in the Western Alps*

Figure 5. Results of the inversion for three initial models. Left: initial stratified velocity model *Z-HQ* with initial HYPO71 hypocenter locations (in a stratified model); Center: initial 3-D velocity model *3D-ANT* with HYPO71 locations; Right: initial *3D-ANT* model with NLLOC locations in the initial 3-D model. Top: depth slices at 10-km depth, with all epicenters (red dots) and stations (black triangles); Cifalps profile shown as a black line; Bottom: depth slices at 20-km depth with epicenters in the 10 – 30 km depth range.

coverage. It is much weaker at 20-km depth, particularly in regions with numerous earthquake foci.

5 ASSESSING THE VARIABILITY OF FINAL MODELS WITH THE ORIGINAL *SQ-89-14* DATABASE

As explained in the section 2, the *HQ-89-14* arrival time database used so far is derived from a severe selection (min. 12 *P* and 6 *S*) in the initial *POTIN-89-14* database. In section 4, the *HQ-89-14* database was used to set up the optimal inversion strategy and hyperparameter values. We also showed that the initial model strongly influences the inversion results in areas with low ray coverage. In this section, we further investigate the initial-model influence by using the *SQ-89-14* arrival time base, which includes a much larger number of picks (+72%, see Table 1), hence a better ray coverage, but still high-quality picks by a single operator, and at least 6 picks per event. By analysing variability in final models computed for different

initial models, we can identify stable, hence reliable features of the final models, as well as features that are not so stable but remain valuable for interpretations.

We use the strategy with least impact on the velocity model structure, that is the penalty approach with hyperparameter $\lambda = 5$ in all directions, combined with a small damping hyperparameter $\epsilon = 0.1$. The selected horizontal penalty term based on a horizontal finite-difference Laplacian term is different from the vertical penalty term based on a vertical finite-difference second derivative. Velocity discretisation is 5 km horizontally and 2 km vertically. Therefore, the Laplacian-smoothing constraint will be different in the horizontal and vertical directions. We proceed with initial HYPO71 locations in all workflows.

5.1 Testing four initial velocity models

The first inversion starts from the layered model named *Z-HQ* proposed by Potin (2016) and shown in Fig. S1. The second initial model named *3D-HQ* is the 3-D model reconstructed from the *HQ-89-14* database with the initial stratified model (Fig. 3). The third one named *3D-ANT* is the 3-D model reconstructed from the *HQ-89-14* database, with the initial 3-D model obtained by combining the stratified velocity model with a 3-D model from ambient noise tomography (center panels in Fig. 5). The fourth one named *3D-HW* is the reconstructed model from the *HQ-89-14* database, the initial stratified model and a damping hyperparameter $\epsilon = 10$ with no Laplacian smoothing (Fig. 4). This model has a high-wavenumber content related to ray and grid sampling.

All reconstructed models have almost the same data misfit (~ 0.43 s for the weighted misfit and ~ 0.72 s for the unweighted misfit). The 20-km depth slices in the four reconstructed *P* velocity models show broad similarities, with second-order differences capturing the influence of initial models (Fig. 6). Lateral velocity contrasts are stronger when starting from the stratified initial model (top left panel in Fig. 6), with lower velocities in the broad low-velocity anomaly, and higher velocities along the northern margin of the Ligurian basin in the southern part of the model. The final model built from the high-wavenumber *3D-HW* initial model (bottom right panel in Fig. 6) gives the lowest data misfit, but it cannot be regarded as best model because its sharp localized velocity variations are not easily interpretable. Inverted hypocenter parameters compensate velocity differences between models to give similar final data misfits.

Differences between final models are best seen on the CIFALPS (CC') vertical section that crosses the two major anomalies (Fig. 7). Strong differences in the 100 – 200-km distance range between the *3D-ANT* V_p model (third row in Fig. 7) and the other ones might reflect

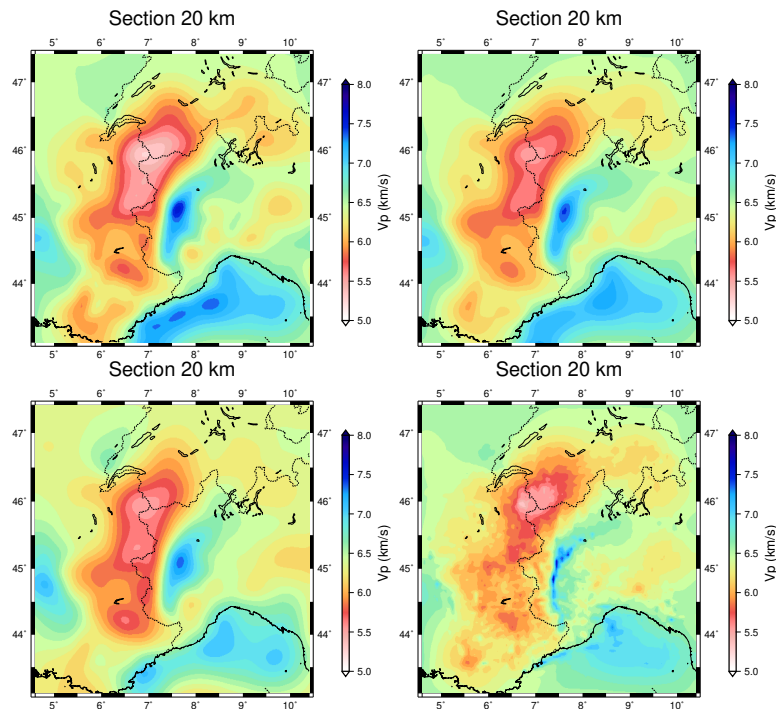
22 *Local earthquake tomography in the Western Alps*

Figure 6. Depth slices at 20-km depth in the reconstructed models from *SQ-89-14* database using four different initial velocity models. Top left: initial stratified model (*Z-HQ*); Top right: initial *3D-HQ* model; Bottom left: initial *3D-ANT* model; Bottom right: initial *3D-HW* model.

500 limited ray sampling of the shallow subsurface. Velocity anomalies at shallow depth are still
 501 too close to the initial *3D-ANT* model, as documented by a comparison with cross-sections
 502 in the initial model displayed in Supplementary Fig. S9. On the other hand, the V_s model is
 503 closer to other ones than the V_p model, perhaps due to the stronger sensitivity of Rayleigh
 504 waves used in ANT to S -wave velocities. This observation may mean that the V_s *3D-ANT*
 505 model derived from ANT is a better initial model for LET than the corresponding V_p model.
 506 We now disregard the *3D-ANT* model, and notice that some vertical V_p changes in the
 507 100 – 200-km low-velocity anomaly of the *3D-HW* final model (bottom left in Fig. 7) can
 508 be detected in the *3D-HQ* and *Z-HQ* results. Similarly, the undulating western boundary of
 509 the Ivrea Body high-velocity anomaly (at ~ 200 -km distance) appears in the three *Z-HQ*,
 510 *3D-HQ* and *3D-HW* models. Such similarities are indications of the reliability of small-scale
 511 velocity anomalies in the reconstructed *Z-HQ* and *3D-HQ* V_p models. The four V_s models in
 512 the right column of Fig. 7 also display similar patterns, as for example the small low-velocity
 513 anomaly at $z = 20\text{--}35$ km and $x = 150\text{--}200$ km, which can therefore be considered as reliable.

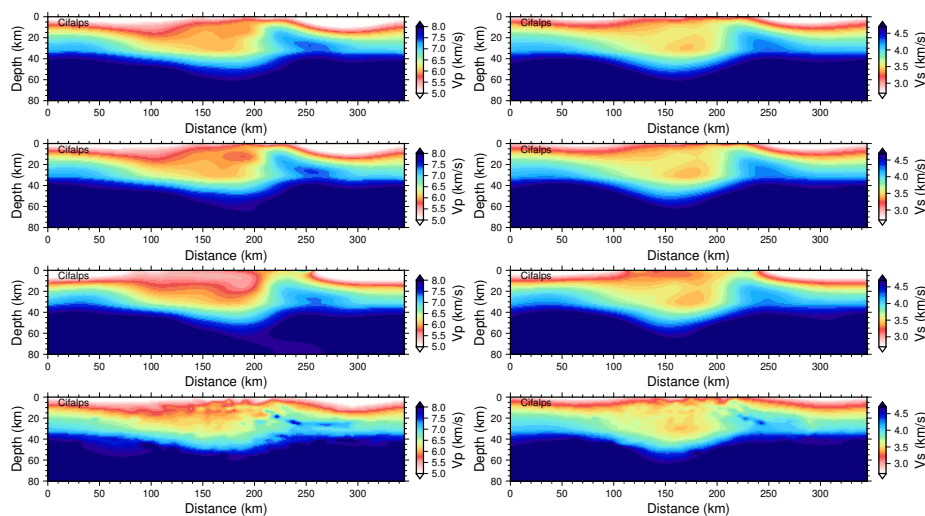


Figure 7. Depth sections along the Cifalps profile (CC') in the V_p (left) and V_s (right) velocity models reconstructed from *SQ-89-14* database with four different initial velocity models. Top: initial stratified model (*Z-HQ*); Second row: initial *3D-HQ* model; Third row: initial *3D-ANT* model; Bottom: initial *3D-HW* model.

5.2 Testing random Gaussian initial models nearby the reconstructed model

We now further investigate the variability of recovered velocity models by testing a set of initial models obtained by random perturbations around a given 3-D model. Sampling the misfit landscape by such a perturbation strategy is almost impossible considering the number of degrees of freedom. However, the limited frequency content of seismic waves allows the exploration of initial random models to be limited to smooth perturbations. The 3-D model reconstructed from the *3D-HQ* initial model in the section 5.1 is considered as original model for the random perturbations (slice at 20-km depth in the top panel of Fig. 8). We add to the original model random Gaussian velocity perturbations scaled by the local velocity (to ensure smoothness), creating 1024 random initial models. The mean and the RMS of the ensemble of P velocity models are shown in the second row of Fig. 8. Predicted arrival times from HYPO71 initial locations do not fit observed ones (red dots in the third row of Fig. 8). Inversions are performed with each random model as initial model, and the RMS values decrease from 0.80 - 0.84 s to a narrow interval around 0.51 s after inversion (blue dots in the third row of Fig. 8). We select the 20 best models (with lowest RMS) and compute the mean and RMS of the reconstructed P velocity models shown in the bottom row of Fig. 8. The mean P velocity model is rather close to the original model (top panel of Fig. 8), while the RMS is low (< 0.02 km/s) and displays no organized feature, which means that the final models are quite similar.

1
2
3 24 *Local earthquake tomography in the Western Alps*
4

5
6 533 At 20-km depth, the strong high-velocity anomaly of the Ivrea Body is recovered whatever
7 534 the initial model, as well as the low-velocity zones in the west.

8
9 535 Checkerboard or individual-spike tests are alternate strategies for identifying data-constrained
10 536 velocity anomalies. We will perform a spike inversion for the extensive arrival time database
11
12 537 in section 6.3.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Local earthquake tomography in the Western Alps 25

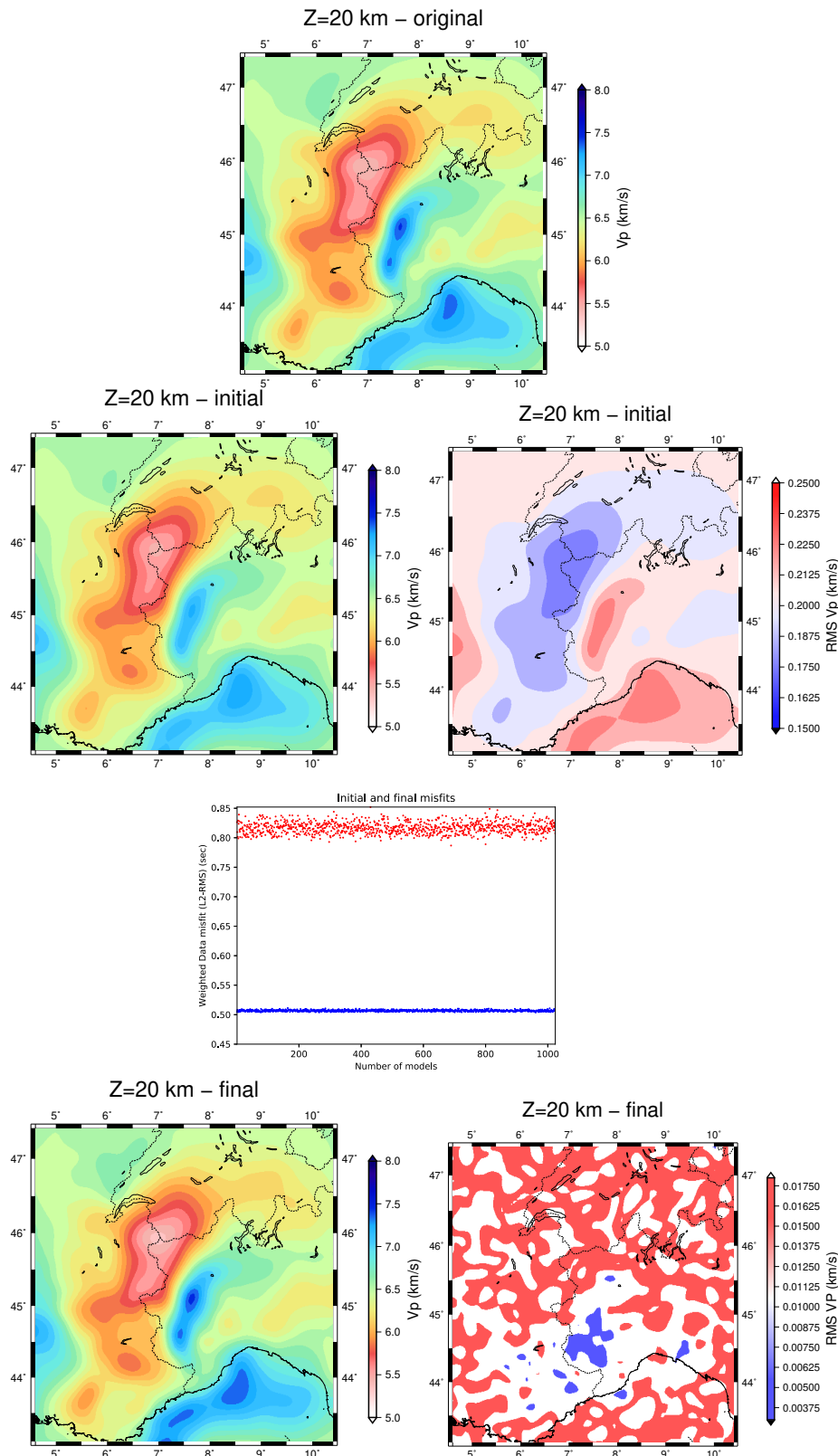


Figure 8. Results of the inversion test of *SQ-89-14* arrival time base with random 3-D initial models. Top: 20-km depth slice in the original model; Second row: mean (left) and RMS (right) over 1024 randomly selected P initial velocity models about the original model; Third row: data misfit in the 1024 models before (red dots) and after (blue dots) inversion; Bottom: final mean and RMS P velocity models after inversion. The absence of organized pattern in the RMS velocity slice shows that a local minimum of the data misfit landscape is reached.

6 FINAL TOMOGRAPHY USING THE EXTENSIVE 1989-2021 DATABASE

After intensive testing of inversion workflows and hyperparameter influence on the high-quality *HQ-89-14* and *SQ-89-14* arrival time datasets, we are now ready to use LET to merge the set of high-quality picks with a heterogeneous set collected from different observatories. We will also compute an optimal reconstructed velocity and hypocenter location model using more than 30 years of local earthquake arrival time data in the Western Alps.

6.1 Database and workflow

We now consider the extensive database collected for the time period 1989-2021 from the fusion of Potin (2016)'s database referred to as *POTIN-89-14* in Table 1, and the arrival time set collected from agencies with different picking strategies and different station networks. As explained in section 2, our fusion strategy is to select time picks with residuals < 10 s in the tomographic model derived from the inversion workflow set up using *HQ-89-14* database. This selection process results in the *WAlps-89-21* database that involves 1043 seismic stations in the target region (Fig. 1). In spite of the low-to-moderate seismicity level of the Western Alps, 33 years of seismological observation provide data for more than 75,000 local earthquakes with at least 6 picks, and more than 1,590,000 picks in dataset *WA-14-21* (Table 1).

The selected optimization workflow is the smoothing-damping strategy with hyperparameters values used in section 5. We proceed in a recursive way for designing the input velocity models for the inversion of database *WA-89-21*. Output models for the *HQ-89-14* database were used as input models for inversion of the *SQ-89-14* database (section 5.1), providing output models that are in turn used as input models for the *WA-89-21* database. We keep the same labels as in section 5.1 for the four workflows. The first inversion is labelled *Z-HQ* because its initial model is the final model of the inversion of database *SQ-89-14* with initial model *Z-HQ*. In the same way, the second workflow is labelled *3D-HQ*, the third one *3D-ANT* and the fourth one *3D-HW*.

6.2 Results

Even though it uses 60 % more arrival time picks, the inversion of *WA-89-21* database yields similar results to the inversion of *SQ-89-14* database. Depth sections in the reconstructed V_p and V_s models with the four different initial models are shown in Fig. 9. It can be compared to Fig. 7 and overall similar conclusions can be drawn about the reliability of velocity anomalies. The low-velocity P anomaly at 100-200-km distance has stronger amplitudes with the

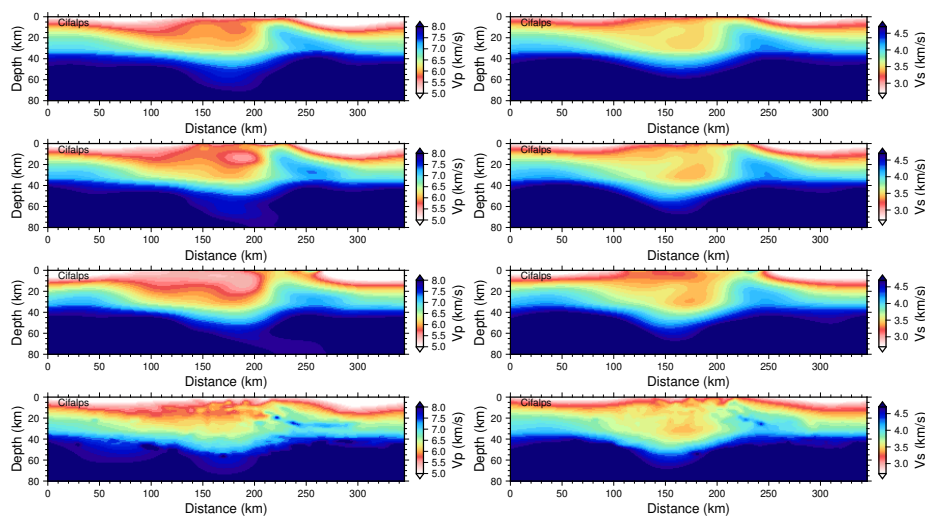


Figure 9. Results of the inversion of *WA-14-21* full database: depth sections in the V_p (left) and V_s (right) models along the CIFALPS (CC') profile, for four different initial models. Top: *Z-HQ* model; Second row: *3D-HQ* model; Third row: *3D-ANT* model; Bottom: *3D-HW* model.

569 extended database and *Z-HQ* and *3D-HQ* initial models than with the *SQ-89-14* database
 570 (top rows in Figs. 9 and 7). As in Fig. 7, the V_s sections of Fig. 9 show less variability with
 571 the initial model than V_p .

572 The final V_p and V_s models share the same global features, whatever the inversion workflow.
 573 We therefore consider as the best final model results of the most standard smoothing-damping
 574 approach with *3D-HQ* as initial model (Fig. 10).

575 6.3 Spike tests for assessing the sensitivity at the convergence

576 The reliability of our reconstructed models has already been discussed in detail, namely by
 577 comparing the inversion results with different initial models and different hyperparameters
 578 (sections 5 and 6.2). Since the spike and/or checkerboard tests are the most classical methods
 579 for assessing the quality of LET results, we present the results of a spike test in Supplementary
 580 text S4 and Figs. S17-18. We follow the strategy of Latorre et al. (2004), which is more CPU
 581 intensive than checkerboard or simultaneous-spike tests (Rawlinson & Sambridge 2004), but
 582 still doable.

583 Our test with four Gaussian synthetic velocity anomalies of characteristic lengths 15 and 5
 584 km in the horizontal and vertical directions, located at 20-depth along the CIFALPS profile
 585 show that our reconstruction is of overall good quality (Suppl. Figs. S17 and S18). Additional
 586 spike tests (not displayed here) show that the reconstruction cannot be achieved at depths
 587 larger than 40 km, due to the weak ray coverage.

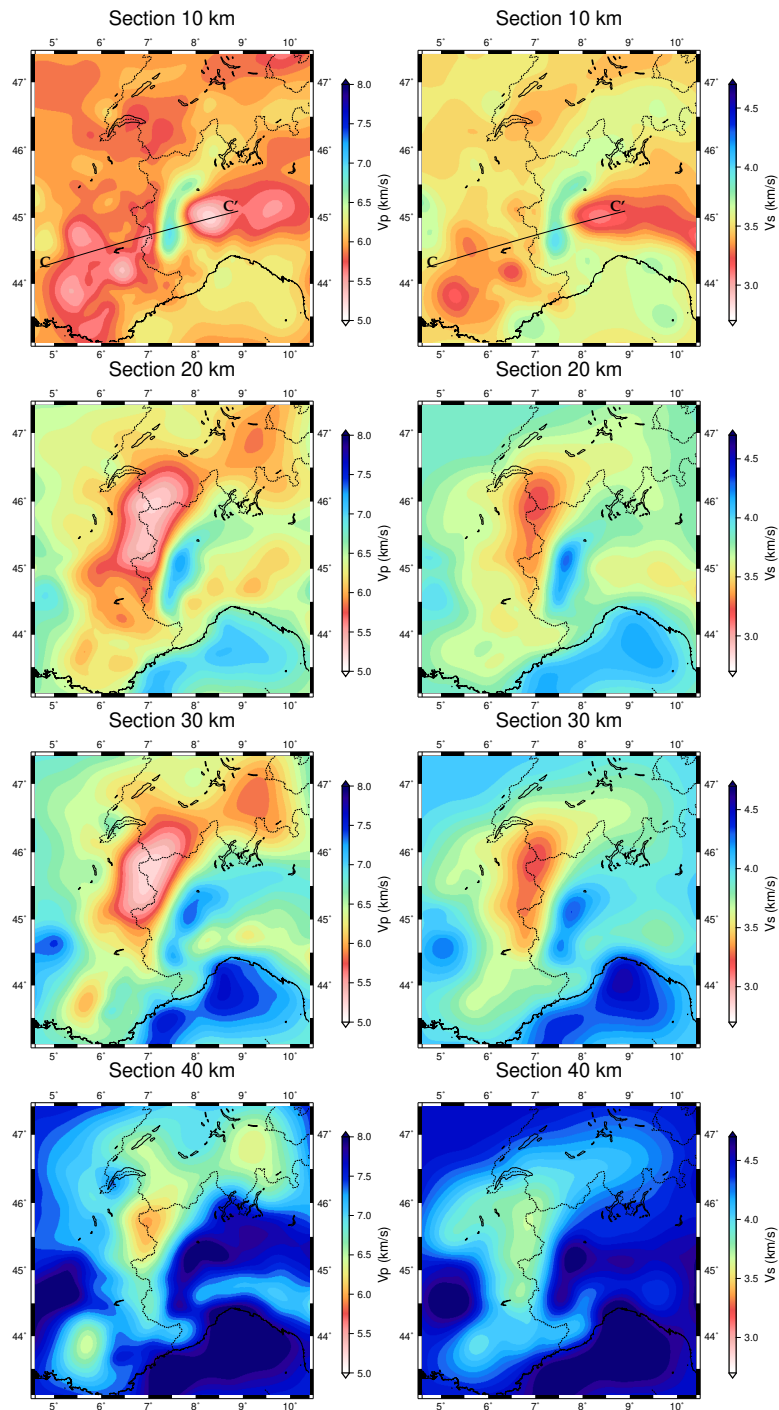
28 *Local earthquake tomography in the Western Alps*

Figure 10. Depth slices in the inversion result of database *WA-89-21* using the standard smoothing-damping approach, considered as our best final model. Left: *P*-velocity slices; Right: *S*-velocity slices. The initial model (labelled *3D-HQ*) is built in a hierarchical strategy since it results from the inversion of *SQ-89-14* database using as initial model the *3D-HQ* result of the inversion of *HQ-89-14* database.

1
2
3
4
5
6 588 This kind of sensitivity analysis can be performed at specific locations that are relevant
7 589 to geological interpretation of velocity models. Yet, we should be aware that such sensitivity
8
9 590 analysis is strongly tied to the prior information injected in the inversion process by the
10
11 591 Laplacian-smoothing approach mitigated by the grid discretisation. The results of the spike
12
13 592 tests should be interpreted with caution because LET relies on ray theory, which carries
14
15 593 no information on the frequency content of seismic waves controlling the Fresnel diffraction
16
17 594 resolution. Spike test results are primarily driven by source and receiver locations, and by
18
19 595 ray paths in the heterogeneous velocity model. Influence of the frequency content can be
20
21 596 determined by wave equation tomography, which is more CPU intensive, thus limiting the
22
23 597 spike sensitivity analysis.

23 598 **6.4 Discussion: Comparison with published 3-D V_p and V_s crustal models of the** 24 25 599 **Western Alps**

26
27 600 We here briefly compare our final V_p and V_s models (Fig. 10) with the most recent published
28
29 601 local earthquake tomography V_p models of Diehl et al. (2009a) and Solarino et al. (2018),
30
31 602 and the ambient-noise tomography V_s model of Nouibat et al. (2022). Both LET models
32
33 603 were computed using the SIMULPS code (e.g., Thurber 1983; Eberhart-Phillips 1986) and
34
35 604 carefully selected P (and S) arrival times. Diehl et al. (2009a)'s dataset covers most of the
36
37 605 Alps and includes more than 13,000 P first arrival picks from 552 earthquakes recorded at 391
38
39 606 stations from 13 national and regional networks. The data quality and consistency is ensured
40
41 607 by an automated re-picking of manual picks that includes quality weight estimates (Diehl et al.
42
43 608 2009b). The LET by Solarino et al. (2018) is centered on a smaller region around the CIFALPS
44
45 609 temporary seismic transect and it uses carefully hand-picked P and S arrivals recorded by
46
47 610 the CIFALPS experiment and surrounding permanent Italian and French permanent stations,
48
49 611 complemented by picks of intermediate-depth events beneath the westernmost Po plain. The
50
51 612 number of arrival time picks used by Solarino et al. (2018) is not reported, but it is certainly
52
53 613 much lower than ours (Table 1). The ambient-noise tomography model of Nouibat et al. (2022)
54
55 614 is the highest-quality published V_s model to-date for the broad Alpine region. It uses the
56
57 615 most comprehensive noise correlation dataset computed for ~ 1440 permanent and temporary
58
59 616 seismic stations in Europe and four years of continuous vertical-component records.

60
61 617 Figure 11 shows horizontal slices in our final V_p model compared to slices at similar depths
62
63 618 in the V_p models of Diehl et al. (2009a) and Solarino et al. (2018), referred to as Diehl's and
64
65 619 Solarino's in the following. Our model is obviously smoother than others in all slices. We
66
67 620 have chosen a Laplacian-smoothing influence such that the shallow velocity reconstruction

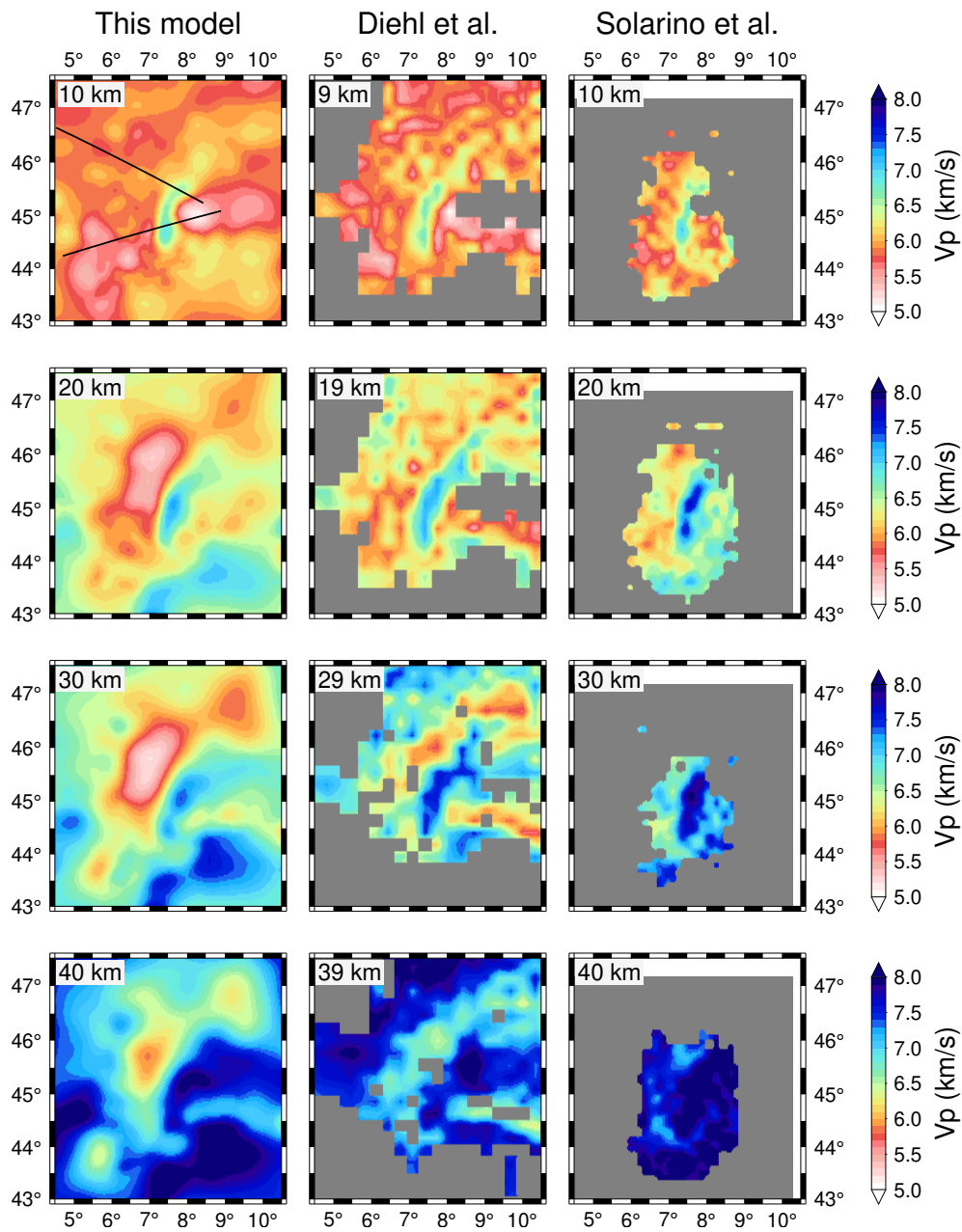
30 *Local earthquake tomography in the Western Alps*

Figure 11. Horizontal slices in the P -wave velocity models of: (left) this paper, (center) Diehl et al. (2009a), (right) Solarino et al. (2018). Slice depth is indicated in the top-left corner of each map. Depths are not exactly the same because maps are shown at depths provided by the authors. Areas with low resolution (diagonal elements of the resolution matrix < 0.1 or 0.15) are masked in Diehl's and Solarino's models. The black lines in the top left map show locations of the CIFALPS2 (north) and CIFALPS (south) cross-sections.

1
2
3
4
5
6 621 has no imprint from the station distribution. The Ivrea body narrow fast-velocity anomaly is
7 622 detected by all models at 10 - 30-km depth, with rather similar shapes and higher velocities in
8
9 623 Solarino's model at 20 and 30 km. In all three models, it is bordered by low-velocity anomalies
10
11 624 on its eastern side at 10-km depth, and on its western side at all depths. The eastern-side
12 625 10-km depth anomaly is typical of the Po sedimentary basin. A major difference appears
13
14 626 between our model and others in the western-side anomalies at 20 km and deeper. While
15 627 we find a broad and strong low-velocity anomaly in the 20 and 30 km slices (45 - 46.5°N,
16 628 6.5 - 8°E), the corresponding low-velocity anomalies of Diehl's model are weaker and rather
17
18 629 discontinuous, similar to Solarino's model where it is resolved. This anomaly attenuates at
19
20 630 40-km depth, but it remains stronger in our model than in Diehl's. Its shape that follows
21
22 631 the arc of the western Alps and its location to the west of the Ivrea body indicates that this
23
24 632 low-velocity anomaly corresponds to the European crust being underthrust beneath the
25 633 Ivrea body mantle wedge of Adriatic affinity (*e.g.* Malusà et al. 2021). A depth slice in our V_p
26 634 model along the CIFALPS2 section (northwestern Alps, location in Fig. 11 top) shows that
27
28 635 this strong low-velocity anomaly covers the depth range 15 - 40 km in the European crust
29
30 636 beneath the internal zones of the northwestern Alps (Suppl. Fig. S17).

31
32 637 We compare our LET V_s model with the ANT model of Nouibat et al. (2022) in Fig.
33
34 638 12. Unlike the V_p models of Fig. 11, the two V_s models display heterogeneities of similar
35 639 characteristic lengths. As stated above, the spatial smoothing strategy is under the control of
36
37 640 the user. The anomaly distributions are roughly similar in the two models, but with substantial
38
39 641 discrepancies when looking in detail. At 10-km depth, the low- V_s anomaly of the Po basin is
40 642 much wider in the ANT model, confirming that LET results at shallow depth are strongly
41
42 643 controlled by station distribution (see section 4, Fig. 3). As in V_p , the mid- and lower crust
43 644 of the internal zones west of the Ivrea body (high-velocity) anomaly have rather low S -wave
44
45 645 velocity in both models, but with strong discrepancies in locations and amplitudes. In the
46 646 northwestern Alps, our model shows a low- V_s anomaly in the depth range 15-35 km which is
47
48 647 weaker and spread laterally in the European lower crust of the ANT model (see CIFALPS2
49
50 648 depth section in Suppl. Fig. S17). Further south in the CIFALPS depth section (Suppl. Fig.
51 649 S18), the ANT model displays a strong low-velocity anomaly at 25 - 30-km depth and 70 -
52
53 650 130-km distance while our LET model has a weaker anomaly at larger depth, closer to the
54
55 651 Ivrea body. Along the CIFALPS and CIFALPS2 seismic profiles, Nouibat et al. (2022) and
56 652 Paul et al. (2022) have documented the striking coherency between the Moho depth defined
57
58 653 from receiver functions and the depth of the $V_s = 4.3$ km/s contour in the ANT model.
59 654 Supplementary Figs. S17 and S18 show that the $V_s = 4.3$ km/s contour in our LET model
60

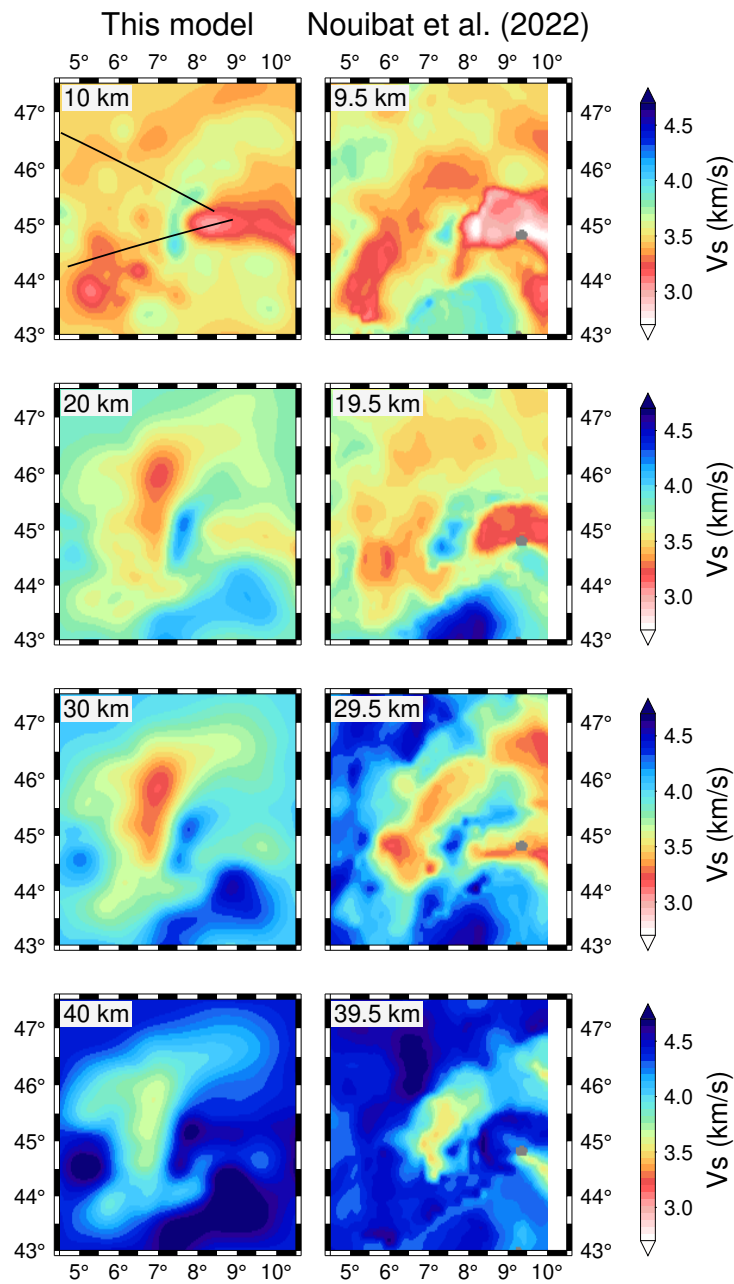
32 *Local earthquake tomography in the Western Alps*

Figure 12. Horizontal slices in the S -wave velocity models of this paper (left) and Nouibat et al. (2022) (right). Slice depth is indicated in the top-left corner of each map. Depths are not exactly the same because maps are shown at depths provided by the authors. The black lines in the top left map show the locations of the CIFALPS2 (north) and CIFALPS (south) cross-sections.

655 is very different from the contour in the ANT model. These mismatches highlight the poor
 656 resolution of the LET models at depth larger than 40 km, which precludes them from imaging
 657 the deepening of the European Moho in the Western Alps subduction.

7 CONCLUSION: LESSONS LEARNED AND PROSPECTS

Seismic time tomography is a widespread tool for velocity-model reconstruction using earthquakes at different scales from crust to lithosphere-asthenosphere and global Earth targets. At the regional scale, LET output models are very useful as initial models for full waveform inversion (FWI), in which initial model quality is of primary importance. Here, we have used a high quality dataset of local earthquake arrival times in the Western Alps to compute crustal V_p and V_s models as a means of illustrating the influences of various input parameters in the tomography, besides earthquake and station distributions and raw onset picked times. We have shown how several possible interpretations of the interaction between seismic waves and the actual medium can lead to various models that fit the observed data equally well. Such a study helps to avoid a possible over-interpretation of the tomographic images, which explain the first arrival times, but do not take into account the frequency content of the seismic waves. This intrinsic limitation can be mitigated by proper selection of input parameters. Such proper choice is particularly crucial when considering the joint inversion of velocity and hypocenter parameters, as first-arrival times have very different sensitivities to these parameters. The bandwidth of seismic waves still has to be considered as an external constraint for the model reconstruction. By contrast, the single point assumption for hypocenters should be replaced by a more generic one based on the probability density function, as done for the location problem (Lomax et al. 2000).

In a time of continuous densification of seismic networks, namely in the study area (e.g., Hetényi et al. 2018), data coverage is no longer a limiting factor for the quality of time tomographies, assuming that data are open. The 33-year database used in our tomography of the Western Alps was assembled thanks to the open data policy of several seismic monitoring agencies. Indeed, the low-to-moderate seismicity of a large part of Europe must be balanced by a data sharing policy and a close collaboration between national agencies, which is clearly under way. Moreover, the detection level and the picking quality are rapidly improving thanks to new technologies based on machine-learning tools (e.g., Zhu & Beroza 2018; Beaucé et al. 2019).

Finding methods that reduce variability in LET images thus becomes an important goal, particularly in areas of low-to-moderate seismicity. A first issue is that ray-based tomography assumes an infinite frequency ignoring any diffraction effect. Other possible asymptotic approaches are based on the Eikonal equation that includes diffraction at the same computational cost. Nevertheless, ray-based or Eikonal-based methods do not take the frequency content of seismic waves into account. Tomography based on wave equation does, but at the

34 *Local earthquake tomography in the Western Alps*

cost of significantly increased computational needs. Since computational power is usually no longer an issue, wave-equation tomography allows the frequency content of collected data to be assessed through numerical simulations of wave propagation and velocity sensitivity kernels with respect to phase (in the sense of time). However, a joint inversion method for velocity and hypocenter parameters using wave-equation-based tomography has not been designed yet. Such a methodological effort should be carried out to fully exploit recent dense seismic networks. Quantifying the uncertainties of these new images will be of utmost importance, but it will require new approaches with lower computational costs. Such approaches can be based on ray or Eikonal approximations, once the optimal image based on the wave equation has been constructed (Zhang et al. 2020). This model including uncertainties will prevent the misinterpretation of diverging images resulting from the assumed simplification of the interactions between seismic waves and real Earth. Another potential way to improve LET, not excluding previously proposed ones, is to develop model-driven approaches. In such approaches, output models must be compatible with selected input geologic constraints besides fitting arrival time data.

In a time of dramatically increasing density of arrival time observations, we believe that local earthquake tomography is still a useful tool for regional studies at crustal scale. A valuable advance is to combine LET with other powerful and complementary imaging methods, notably ambient-noise tomography and wave-equation tomography. We expect that new, better resolved velocity models with uncertainty estimates will soon yield important constraints on geological models of the complex collision zone of the broad Mediterranean region, including the Alps.

8 ACKNOWLEDGEMENTS

The first SISmalp seismic network (1987-2014) was created and operated in Grenoble by our former colleagues François Thouvenot, Julien Fréchet and Liliane Jenatton, whom we thank for their pioneering work. The careful work of Bertrand Potin, Liliane Jenatton, François Thouvenot and Bernard Valette has produced the impressive dataset of high-quality manual picks (time period: 1989-2014), which plays an important part in the present paper. We thank the operators of several permanent and temporary seismic networks that provided data, although tracking each contribution is challenging. AP warmly thanks her team-mates of the CIFALPS working group for letting her use some embargoed data from the CIFALPS2 temporary network (Coralie Aubert, Elena Eva, Marco Malusà, Silvia Pondrelli, Simone Salimbeni, Stefano Solarino and Liang Zhao; doi:10.15778/RESIF.XT2018). The open-mindedness of var-

ious national agency contacts should be highlighted. We are pleased to thank in particular Daniele Spallarossa of RSNI/DipTeris (Genova) for his fast feedback, as well as John Clinton and Tobias Diehl of SED (ETH Zürich) for the tracking of metadata related to seismic stations. We also thank Marc Grunberg (RéNaSS, Strasbourg) for helping us with webservice handling and Christophe Maron (Geoazur, Nice) for station information. Furthermore, we thank the ISC for its tremendous effort in collecting arrival times in the ISC Bulletin. Most of the plots are generated using the Generic Mapping Tool by Wessel & Smith (1995).

We thank Steve Roecker for a careful reading. His comments have improved significantly the content of this paper.

This study was granted access to the HPC resources of the Dahu platform of the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes region (GRANT CPER07 13 CIRA), the OSUG@2020 LabEx (reference ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche. AP's work on the Alpine region is funded by Agence Nationale de la Recherche under grants ANR-15-CE31-0015 (AlpArray-FR) and ANR-20-CE49-0007 (LiSAIps).

9 DATA AVAILABILITY

The data underlying this article is available at the public repository:

https://github.com/jeanvirieux/Western_Alps.

REFERENCES

- Argand, E., Bloesch, E., Heim, A., & Heim, A., 1911. *I. Les nappes de recouvrement des Alpes pennines et leurs prolongements structuraux*, no. 31-33, In Kommission bei A. Francke (vorm. Schmid & Francke).
- Beaucé, E., Frank, W. B., Paul, A., Campillo, M., & van der Hilst, R. D., 2019. Systematic detection of clustered seismicity beneath the Southwestern Alps, *Journal of Geophysical Research: Solid Earth*, **124**(11), 11531–11548.
- Beller, S., Monteiller, V., Operto, S., Nolet, G., Paul, A., & Zhao, L., 2018. Lithospheric architecture of the South-Western Alps revealed by multi-parameter teleseismic full-waveform inversion, *Geophysical Journal International*, **212**(2), 1369–1388.
- Cattaneo, M., Augliera, P., Parolai, S., & Spallarossa, D., 1999. Anomalous deep earthquakes in northwestern Italy, *Journal of Seismology*, **3**, 421–435.

36 *Local earthquake tomography in the Western Alps*

- 756 Chiao, L.-Y. & Kuo, B.-Y., 2001. Multiscale seismic tomography, *Geophysical Journal International*,
757 **145**(2), 517–527.
- 758 Chopin, C., 1984. Coesite and pure pyrope in high-grade blueschists of the Western Alps: a first
759 record and some consequences, *Contributions to Mineralogy and Petrology*, **86**, 107–118.
- 760 Constable, S. C., Parker, R. L., & Constable, C. G., 1987. Occam's inversion : a practical algorithm
761 for generating smooth models from electromagnetic sounding data, *Geophysics*, **52**, 289–300.
- 762 Dahlen, F. A., Hung, S.-H., & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes - I.
763 theory, *Geophysical Journal International*, **141**, 157–174.
- 764 Delost, M., Virieux, J., & Operto, S., 2008. First-arrival travelttime tomography using second gener-
765 ation wavelets, *Geophysical Prospecting*, **56**, 505–526.
- 766 Diehl, T., Husen, S., Kissling, E., & Deichmann, N., 2009a. High-resolution 3-D P-wave model of the
767 Alpine crust, *Geophysical Journal International*, **179**(2), 1133–1147.
- 768 Diehl, T., Kissling, E., Husen, S., & Aldersons, F., 2009b. Consistent phase picking for regional
769 tomography models: application to the greater Alpine region, *Geophysical Journal International*,
770 **176**(2), 542–554.
- 771 Eberhart-Phillips, D., 1986. Three-dimensional velocity structure in northern California Coast Ranges
772 from inversion of local earthquake arrival times, *Bulletin of the Seismological Society of America*,
773 **76**(4), 1025–1052.
- 774 ECORS-CROP DSS Group, 1989. A new picture of the Moho under the western Alps, *Nature*, **337**,
775 249–251.
- 776 Eva, E., Malusà, M., & Solarino, S., 2015. A seismotectonic picture of the inner southern western
777 Alps based on the analysis of anomalously deep earthquakes, *Tectonophysics*, **661**, 190–199.
- 778 Evans, J. R., Eberhart-Phillips, D., & Thurber, C. H., 1994. User's manual for SIMULPS12 for
779 imaging Vp and Vp/Vs; a derivative of the "Thurber" tomographic inversion SIMUL3 for local
780 earthquakes and explosions, Tech. rep., Survey, U. S. Geological.
- 781 Fang, H., van der Hilst, R. D., de Hoop, M. V., Kothari, K., Gupta, S., & Dokmanić, I., 2019.
782 Parsimonious seismic tomography with Poisson Voronoi projections: Methodology and validation,
783 *Seismological Research Letters*, **91**(1), 343–355.
- 784 Greenhalgh, S., Zhou, B., & Green, A., 2006. Solutions, algorithms and inter-relations for local
785 minimization search geophysical inversion, *Journal of Geophysics and Engineering*, **3**(2), 101–113.
- 786 Hansen, P. C., 1992. Analysis of discrete ill-posed problems by means of the L-curve, *SIAM Review*,
787 **34**(4), 561–580.
- 788 Hetényi, G., Molinari, I., Clinton, J., Bokelmann, G., Bondár, I., Crawford, W. C., Dessa, J.-X.,
789 Doubre, C., Friederich, W., Fuchs, F., Giardini, D., Gráczner, Z., Handy, M. R., Herak, M., Jia, Y.,
790 Kissling, E., Kopp, H., Korn, M., Margheriti, L., Meier, T., Mucciarelli, M., Paul, A., Pesaresi,
791 D., Piromallo, C., Plenefisch, T., Plomerová, J., Ritter, J., Rümpker, G., Šipka, V., Spallarossa,
792 D., Thomas, C., Tilmann, F., Wassermann, J., Weber, M., Wéber, Z., Wesztergom, V., Živčić, M.,

Local earthquake tomography in the Western Alps 37

- 793 Abreu, R., Allegretti, I., Apoloner, M.-T., Aubert, C., Besançon, S., Bès de Berc, M., Brunel, D.,
794 Capello, M., Čarman, M., Cavaliere, A., Chèze, J., Chiarabba, C., Cougoulat, G., Cristiano, L.,
795 Czifra, T., D'Alema, E., Danesi, S., Daniel, R., Dannowski, A., Dasović, I., Deschamps, A., Egdorf,
796 S., Fiket, T., Fischer, K., Funke, S., Govoni, A., Gröschl, G., Heimers, S., Heit, B., Herak, D., Huber,
797 J., Jarić, D., Jedlička, P., Jund, H., Klingen, S., Klotz, B., Kolínský, P., Kotek, J., Kühne, L., Kuk,
798 K., Lange, D., Loos, J., Lovati, S., Malengros, D., Maron, C., Martin, X., Massa, M., Mazzarini, F.,
799 Métral, L., Moretti, M., Munzarová, H., Nardi, A., Pahor, J., Péquegnat, C., Petersen, F., Piccinini,
800 D., Pondrelli, S., Prevolnik, S., Racine, R., Régnier, M., Reiss, M., Salimbeni, S., Santulin, M.,
801 Scherer, W., Schippkus, S., Schulte-Kortnack, D., Solarino, S., Spieker, K., Stipčević, J., Strollo, A.,
802 Süle, B., Szanyi, G., Szűcs, E., Thorwart, M., Ueding, S., Vallocchia, M., Vecsey, L., Voigt, R., Weidle,
803 C., Weyland, G., Wiemer, S., Wolf, F., Wolyniec, D., Zieke, T., AlpArray Seismic Network Team,
804 ETHZ-SED Electronics Lab, AlpArray OBS Cruise Crew, & AlpArray Working Group, 2018. The
805 AlpArray seismic network: A large-scale European experiment to image the Alpine orogen, *Surveys*
806 *in Geophysics*, **39**(5), 1009–1033.
- 807 Kissling, E., Ellsworth, W. L., Eberhart-Phillips, D., & Kradolfer, U., 1994. Initial reference models in
808 local earthquake tomography, *Journal of Geophysical Research: Solid Earth*, **99**(B10), 19635–19646.
- 809 Koulakov, I., 2009. LOTOS code for local earthquake tomographic inversion: Benchmarks for testing
810 tomographic algorithms, *Bulletin of the Seismological Society of America*, **99**(1), 194–214.
- 811 Latorre, D., 2004. *Imagerie sismique du milieu de propagation à partir des ondes directes et converties*
812 *: application à la région d'Aigion (Golfe de Corinthe, Grèce)*, PhD thesis in French, Université Nice
813 Sophia-Antipolis.
- 814 Latorre, D., Virieux, J., Monfret, T., Monteiller, V., Vanorio, T., Got, J. L., & Lyon-Caen, H., 2004. A
815 new seismic tomography of Aigion area (Gulf of Corinth, Greece) from the 1991 data set, *Geophysical*
816 *Journal International*, **159**, 1013–1031.
- 817 Le Meur, H., Virieux, J., & Podvin, P., 1997. Seismic tomography of the Gulf of Corinth: a comparison
818 of methods, *Annali di Geofisica*, **XL**, 1–25.
- 819 Lee, W. & Lahr, J., 1972. HYPO71: a computer program for determining hypocenter, magnitude,
820 and first motion pattern of local earthquakes, Tech. rep., U.S. Geological Survey.
- 821 Lévêque, J.-J., Rivera, L., & Wittlinger, G., 1993. On the use of the checker-board test to assess the
822 resolution of tomographic inversions, *Geophysical Journal International*, **115**(1), 313–318.
- 823 Li, S. & Fomel, S., 2013. Kirchhoff migration using eikonal-based computation of traveltimes source-
824 derivatives, *Geophysics*, **78**(4), S211–S219.
- 825 Lomax, A., Virieux, J., Volant, P., & Berge, C., 2000. Probabilistic earthquake location in 3D and
826 layered models: Introduction of a Metropolis-Gibbs method and comparison with linear locations,
827 in *Advances in Seismic Event Location*, vol. 281, pp. 101–134, eds Thurber, C. H. & Rabinowitz,
828 N., Kluwer Academic Pub., Amsterdam.
- 829 Lu, Y., Stehly, L., Paul, A., & AlpArray Working Group, 2018. High-resolution surface wave tomog-

38 *Local earthquake tomography in the Western Alps*

- raphy of the European crust and uppermost mantle from ambient seismic noise, *Geophysical Journal International*, **214**(2), 1136–1150.
- Malusà, M. G., Guillot, S., Zhao, L., Paul, A., Solarino, S., Dumont, T., Schwartz, S., Aubert, C., Baccheschi, P., Eva, E., et al., 2021. The deep structure of the Alps based on the CIFALPS seismic experiment: A synthesis, *Geochemistry, Geophysics, Geosystems*, **22**(3), e2020GC009466.
- Maupin, V., 2020. Combining asynchronous data sets in regional body-wave tomography, *Geophysical Journal International*, **224**, 401–415.
- Menke, W., 2012. *Geophysical data analysis: Discrete inverse theory (Third Edition)*, Academic press, Boston, USA, third edition edn.
- Monteiller, V., 2005. *Tomographie à l'aide de décalages temporels d'ondes sismiques P : développements méthodologiques et applications*, PhD thesis in French, Université Savoie Mont Blanc.
- Monteiller, V., Got, J.-L., Virieux, J., & Okubo, P., 2005. An efficient algorithm for double-difference tomography and location in heterogeneous media, with an application to the Kilauea volcano, *Journal of Geophysical Research*, **110**(B12306), doi:10.1029/2004JB00346.
- Nicolas, A., Hirn, A., Nicolich, R., Polino, R., & ECORS-CROP Working Group, 1990. Lithospheric wedging in the western Alps inferred from the ECORS-CROP traverse, *Geology*, **18**(7), 587–590.
- Nolet, G., 2008. *A Breviary of Seismic Tomography*, Cambridge University Press, Cambridge, UK.
- Nolet, G. & Dahlen, F., 2000. Wave front healing and the evolution of seismic delay times, *Journal of Geophysical Research*, **105**, 19 043–19 054.
- Nouibat, A., Stehly, L., Paul, A., Schwartz, S., Bodin, T., Dumont, T., Rolland, Y., Brossier, R., Cifalps Team, & AlpArray Working Group, 2022. Lithospheric transdimensional ambient-noise tomography of W-Europe: implications for crustal-scale geometry of the W-Alps, *Geophysical Journal International*, **229**(2), 862–879.
- Paige, C. C. & Saunders, M. A., 1982. ALGORITHM 583 LSQR : Sparse linear equations and least squares problems, *ACM Transactions on Mathematical Software*, **8**(2), 195–209.
- Paul, A., Cattaneo, M., Thouvenot, F., Spallarossa, D., Béthoux, N., & Fréchet, J., 2001. A three-dimensional crustal velocity model of the southwestern Alps from local earthquake tomography, *Journal of Geophysical Research*, **106**(19), 367–390.
- Paul, A., Malusà, M. G., Solarino, S., Salimbeni, S., Eva, E., Nouibat, A., Pondrelli, S., Aubert, C., Dumont, T., Guillot, S., et al., 2022. Along-strike variations in the fossil subduction zone of the Western Alps revealed by the CIFALPS seismic experiments and their implications for exhumation of (ultra-) high-pressure rocks, *Earth and Planetary Science Letters*, **598**, 117843.
- Piana Agostinetti, N., Giacomuzzi, G., & Malinverno, A., 2015. Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, *Geophysical Journal International*, **201**(3), 1598–1617.
- Podvin, P. & Lecomte, I., 1991. Finite difference computation of traveltimes in very contrasted velocity model : a massively parallel approach and its associated tools, *Geophysical Journal International*,

1
2
3
4
5
6 867 **105**, 271–284.

7 868 Potin, B., 2016. *Les Alpes occidentales : tomographie, localisation de séismes et topographie du Moho*,
8 869 PhD thesis, Université Grenoble Alpes.

9
10 870 Rawlinson, N. & Sambridge, M., 2004. Multiple reflection and transmission phases in complex layered
11 871 media using a multistage fast marching method, *Geophysics*, **69**(5), 1328–1350.

12
13 872 Roecker, S., Sabitova, T. M., Vinnik, L. P., Burmakov, Y. A., Golvanov, M. I., Mamatkanova, R.,
14 873 & Munirova, L., 1993. Three-dimensional elastic wave velocity structure of the western and central
15 874 Tien Shan, *Journal of Geophysical Research*, **98**, 15779–15795.

16
17 875 Roecker, S., Thurber, C., Roberts, K., & Powell, L., 2006. Refining the image of the San Andreas Fault
18 876 near Parkfield, California using a finite difference travel time computation technique, *Tectonophysics*,
19 877 **426**(1-2), 189–205.

20
21 878 Solarino, S., Kissling, E., Sellami, S., Smriglio, G., Thouvenot, F., Granet, M., Bonjer, K., & Slejko,
22 879 D., 1997. Compilation of a recent seismicity data base of the greater Alpine region from several
23 880 seismological networks and preliminary 3D tomographic results, *Ann. Geophys.*, **40**(1), 161–174.

24
25 881 Solarino, S., Malusà, M. G., Eva, E., Guillot, S., Paul, A., Schwartz, S., Zhao, L., Aubert, C., Dumont,
26 882 T., Pondrelli, S., et al., 2018. Mantle wedge exhumation beneath the Dora-Maira (U)HP dome
27 883 unravelled by local earthquake tomography (Western Alps), *Lithos*, **296**, 623–636.

28
29 884 Stehly, L., Fry, B., Campillo, M., Shapiro, N., Guilbert, J., Boschi, L., & Giardini, D., 2009. To-
30 885 mography of the Alpine region from observations of seismic ambient noise, *Geophysical Journal*
31 886 *International*, **178**, 338–350.

32
33 887 Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society
34 888 for Industrial and Applied Mathematics, Philadelphia.

35
36 889 Tarantola, A. & Valette, B., 1982. Inverse problems = quest for information, *Journal of Geophysics*,
37 890 **50**, 159–170.

38
39 891 Thouvenot, F., Paul, A., Fréchet, J., Béthoux, N., Jenatton, L., & Guiguet, R., 2007. Are there
40 892 really superposed Mohos in the southwestern Alps? New seismic data from fan-profiling reflections,
41 893 *Geophysical Journal International*, **170**(3), 1180–1194.

42
43 894 Thurber, C. & Eberhart-Phillips, D., 1999. Local earthquake tomography with flexible gridding,
44 895 *Computers & Geosciences*, **25**(7), 809–818.

45
46 896 Thurber, C. H., 1983. Earthquake locations and three-dimensional crustal structure in the Coyote
47 897 Lake Area, central California, *Journal of Geophysical Research: Solid Earth*, **88**(B10), 8226–8236.

48
49 898 Um, J. & Thurber, C., 1987. A fast algorithm for two-point seismic ray tracing, *Bulletin of the*
50 899 *Seismological Society of America*, **77**, 972–986.

51
52 900 Zelt, C. A. & Chen, J., 2016. Frequency-dependent traveltimes tomography for near-surface seismic
53 901 refraction data, *Geophysical Journal International*, **207**, 72–88.

54
55 902 Zhang, X., Roy, C., Curtis, A., Nowacki, A., & Baptie, B., 2020. Imaging the subsurface using induced
56 903 seismicity and ambient noise: 3-D tomographic Monte Carlo joint inversion of earthquake body wave

57
58
59
60

40 *Local earthquake tomography in the Western Alps*

904 traveltimes and surface wave dispersion, *Geophysical Journal International*, **222**(3), 1639–1655.

905 Zhao, D., Hasegawa, A., & Horiuchi, S., 1992. Tomographic imaging of P and S wave velocity structure
906 beneath northeastern Japan, *Journal of Geophysical Research: Solid Earth*, **97**(B13), 19909–19928.

907 Zhao, L., Paul, A., Guillot, S., Solarino, S., Malusà, M. G., Zheng, T., Aubert, C., Salimbeni, S., Du-
908 mont, T., Schwartz, S., Zhu, R., & Wang, Q., 2015. First seismic evidence for continental subduction
909 beneath the Western Alps, *Geology*, **43**(9), 815–818.

910 Zhao, L., Malusà, M. G., Yuan, H., Paul, A., Guillot, S., Lu, Y., Stehly, L., Solarino, S., Eva, E., Lu,
911 G., Bodin, T., Zhao, L., Malusà, M. G., Paul, A., Guillot, S., Solarino, S., Eva, E., Lu, G., Paul,
912 A., Solarino, S., CIFALPS Group, & AlpArray Working Group, 2020. Evidence for a serpentinized
913 plate interface favouring continental subduction, *Nature Communications*, **11**(1), 2171.

914 Zhu, W. & Beroza, G. C., 2018. PhaseNet: a deep-neural-network-based seismic arrival-time picking
915 method, *Geophysical Journal International*, **216**(1), 261–273.

916 f

SUPPLEMENTARY MATERIAL

ASSESSING THE RELIABILITY OF LOCAL EARTHQUAKE TOMOGRAPHY FOR CRUSTAL IMAGING: 30 YEARS OF RECORDS IN THE WESTERN ALPS AS A CASE STUDY

J. Virieux¹, A. Paul¹, M. Langlais¹, G. Janex¹, P. Guéguen¹, A. Helmstetter¹, and L. Stehly¹

¹ Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, UGE, ISTerre, 38000 Grenoble, France.

1 Local earthquake tomography as an ill-posed inverse problem

1.1 Text S1: The penalty approach as an augmented linear system

As mentioned in the section 3.3, the model increment Δm is defined in a compact way by

$$\arg \min_{\Delta m} \|\mathcal{J}\Delta m - \Delta t\|^2 + \lambda \|\mathcal{L}\Delta m\|^2 + \epsilon \|\Delta m\|^2. \quad (1)$$

The smoothing operator \mathcal{L} and the minimal-norm requirement depend on two hyperparameters: the scalar hyperparameter λ and the damping hyperparameter ϵ .

The first step is the choice of an initial model made up of initial V_p and V_s models (in fact slowness models) and initial hypocenter locations, described by two sub-models $m = (s, h)$ and increments $\Delta m = (\Delta s, \Delta h)$, where the hypocenter (resp. slowness) parameter vector is denoted by the symbol h (resp. s). Initial parameter values play an important role in the model search. The inversion aims at computing a better model (with a better fit to observed arrival times). At iteration k , model m_k is perturbed by $\Delta m_k = (\Delta s_k, \Delta h_k)$, such that the predicted arrival time $t(m_k) + \partial t / \partial m(m_k) \Delta m_k$ is closer to the observed arrival time t_{obs} . The linear system for the model increment Δm is:

$$W_D \mathcal{J}_k \Delta m_k = W_D \frac{\partial t}{\partial m}(m_k) \Delta m_k = W_D (t_{obs} - t(m_k)), \quad (2)$$

where a weighting matrix W_D is added for removing outliers from the arrival time data set.

Because different physical quantities are considered in this linear system, we scale each parameter of the four classes (V_p , V_s , (x, y, z) , t_0) by the maximum norm of columns of the sensitivity matrix for the corresponding class.

By solving this adimensional scaled system in the least-squares sense, we get a new model m_{k+1} with better fit to observed data and a data misfit function given by:

$$\mathcal{C}_D(\mathbf{s}, \mathbf{h}) = \frac{1}{2} \sum_{nobs} W_D (t_{obs} - t(m))^2, \quad (3)$$

Considering constraints in the velocity model is needed for this mixed-determined structure of seismic tomography. Instead of requiring exact constraints through Lagrangian multipliers, the simpler penalty approach is often preferred. This approach consists in adding new linear relations to the previous linear system both for the smoothing and damping components. We therefore consider the first-order seven-points finite-difference Laplace operator \mathcal{L} over the grid of velocity parameters. The spatial-derivation operator D of this operator for each node of the inversion grid which leads to the general system:

$$\begin{bmatrix} W_D \frac{\partial t}{\partial s_k} & W_D \frac{\partial t}{\partial h_k} \\ \lambda D & 0 \end{bmatrix} \begin{bmatrix} \Delta s_k \\ \Delta h_k \end{bmatrix} = \begin{bmatrix} W_D (t_{obs} - t(m_k)) \\ 0 \end{bmatrix}. \quad (4)$$

The hyperparameter λ will limit variations of the spatial second derivative of the slowness model, leading to smooth velocity models while still trying to fit the observations. Such smoothing operator could be split into Cartesian components. Often, differences are made between horizontal and vertical smoothing conditions complementing possible horizontal and vertical grid discretization. Using the same penalty strategy, the LSQR algorithm adds hyperparameter ϵ to this adimensional extended linear system for making it solvable, leading to a new extended linear system:

$$\begin{bmatrix} W_D \frac{\partial t}{\partial s_k} & W_D \frac{\partial t}{\partial h_k} \\ \lambda D & 0 \\ \epsilon I & 0 \\ 0 & \epsilon I \end{bmatrix} \begin{bmatrix} \Delta s_k \\ \Delta h_k \end{bmatrix} = \begin{bmatrix} W_D(t_{obs} - t(m_k)) \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (5)$$

This system highlights the difference between slowness/velocity parameters for continuous model description and hypocenter parameters which are discrete by definition, at least in the ray formulation. Ingredients of the inverse problem are the inversion grid discretization, the initial grid-related velocity and hypocenter values, and the two hyperparameters constraining the ill-posed problem of local earthquake tomography.

1.2 Text S2: How to introduce prior model information?

Target-oriented seismic reconstruction of upper-crust reservoirs could be constrained by information from well data (Asnaashari et al., 2012) or expected geological features (Guitton, 2012). At the lithospheric scale, geodynamic interpretation may help such a design of a prior model: such a model may attract somehow the running model of the inversion scheme through iterations. Another new hyperparameter will control the influence of this prior model in the global misfit estimation. Such a design is beyond the scope of this current work. However, two simpler inversion strategies can be considered by applying an operator once the slowness perturbation Δs has been estimated. The slowness perturbation could be transformed either by a smoothing kernel (Fomel and Claerbout, 2003) or by a Total-Variation (TV) kernel (Dahl et al., 2009) before updating the slowness model. These operators essentially filter out high-wavenumber components in the velocity model, essentially artificially filled by the ray geometry assumption. They are commonly used for full-waveform inversion.

The first strategy promotes specific features of the slowness perturbation. Because of its numerical efficiency, a tensorial Gaussian filter is often considered: it has another smoothing influence than the Laplacian-smoothing operator. Other alternative more elaborated filtering operators with specific boosting of expected geological features can be considered as well (Guitton, 2012; Wellington et al., 2017). Again, such an operator depends on prior information regarding the expected model to be specified in the case study.

The second strategy promotes more or less equal-value zones: a reasonable assumption of more or less uniform geological structures. It is derived from image denoising techniques. Such operations are not consistent with the minimization problem. They are only valid because the reduction of unwanted features does not affect the decrease of the data misfit function.

Limited illustrations of these smoothing strategies are given for making the reader aware of the numerous ways of updating the velocity model. For example, the Gaussian smoothing operator with a characteristic length of 20 km in all three directions (Fig. S1 left) gives a velocity model with nearly same data misfit than the model obtained with the seven-points Laplacian operator. Of course, the strong vertical smoothing illustrates the impact of such filtering strategy. Similarly, the TV operator can be applied to the slowness perturbation at each iteration. Such operator depend on various parameters not described in this work. They control the expected seismic velocity variations over characteristic zones. The reconstructed model (Fig. S1 right) displays less pronounced low-velocity structures at shallow depths. These low-velocity zones are very sensitive to the requested smoothness of the velocity reconstruction, mitigating possible interpretation of these quantitative values.

2 Testing LET parameters on the high-quality *HQ-89-14* database

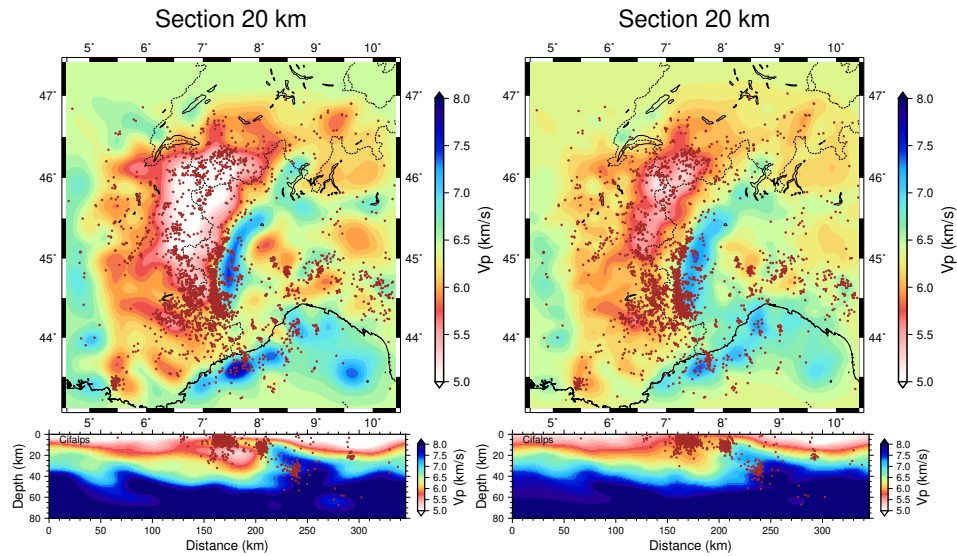


Figure S1: Influence of the smoothing (left) and the TV (right) operators applied to the slowness perturbation model. Top panels: depth slices in the P velocity model at 20-km depth; Bottom panels: P velocity depth sections along the CIFALPS profile.

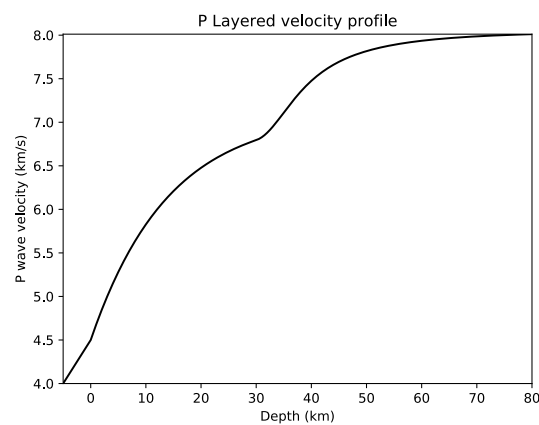


Figure S2: Initial stratified P velocity model proposed by Potin (2016). The initial S velocity is deduced by considering a constant V_p/V_s ratio of 1.6933.

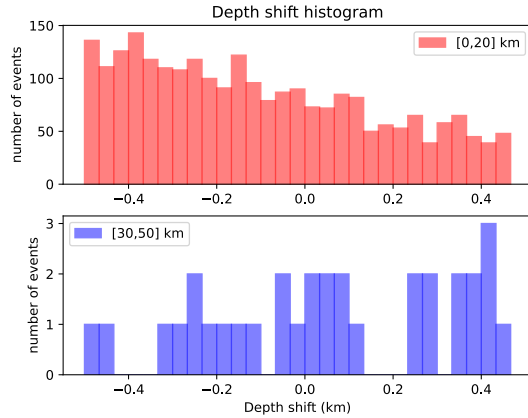


Figure S3: Histograms of hypocenter depth shifts between initial HYPO71 location and final inversion location for the 0-20 km layer (top) and the 30-50 km layer (bottom). Superficial events tend to move to shallower depths, while deeper events tend to move to greater depths.

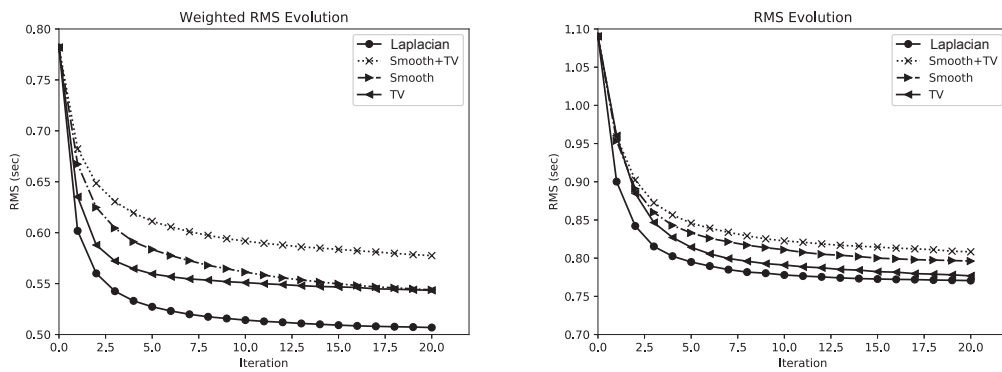


Figure S4: Weighted (left) and raw (right) data misfit reduction curves through iterations for four inversion strategies controlling the model roughness applied to *HQ-89-14* database

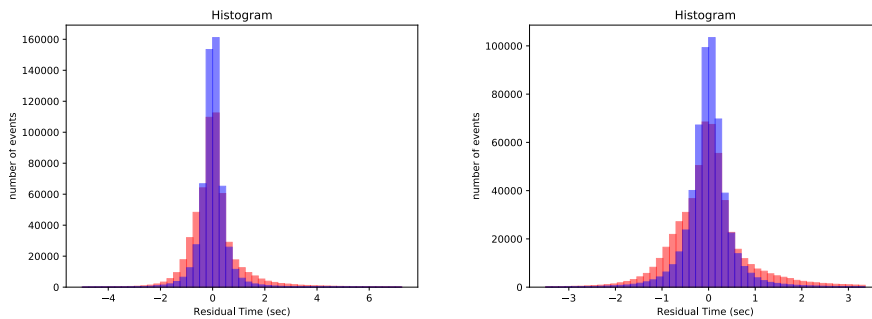


Figure S5: Arrival time residuals for database *HQ-89-14* for the initial model (red) and the recovered model (blue). Left: range $[-6 \text{ s}, 8 \text{ s}]$; Right: zoom in the range $[-4 \text{ s}, 4 \text{ s}]$ that will be used from now on. The histogram for the initial model (red) has a larger tail for positive values, meaning that observed arrival times are initially higher than predicted times, while the histogram is more symmetric for the recovered model (blue).

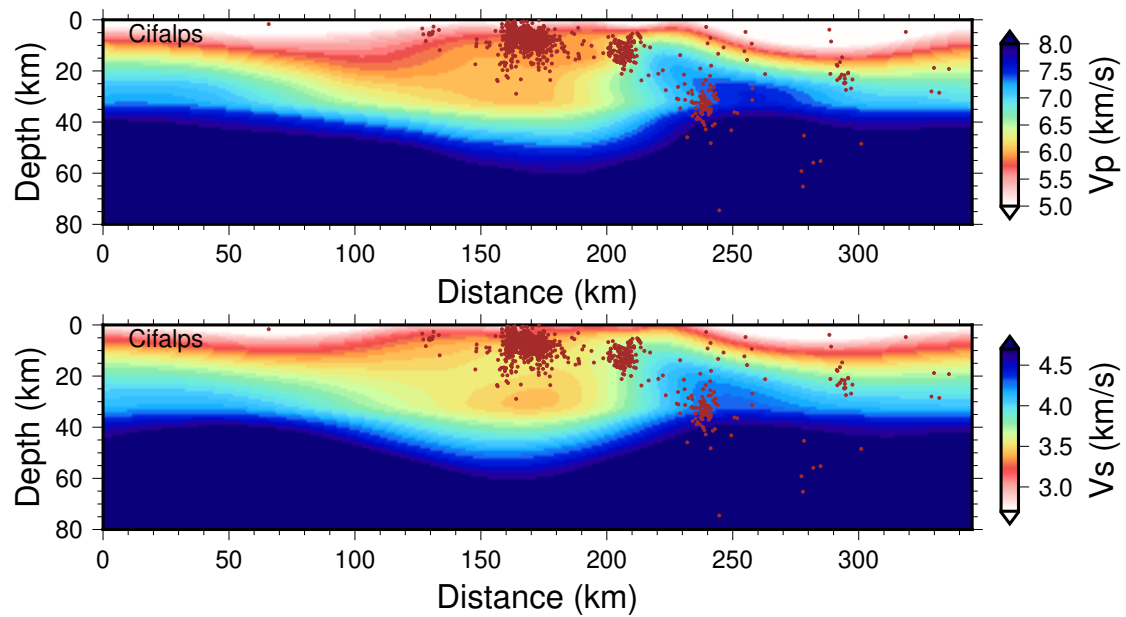


Figure S6: Vertical cross-sections along the CIFALPS profile CC' which locations in shown in Fig. 3 (top) of the main text. This reference profile, which follows the dense CIFALPS temporary experiment (Zhao et al., 2015) will be used in comparison to results of later investigations. Top: P velocity section; Bottom: S velocity section. Both models display a strong high-velocity anomaly at distances of 200 – 250 km rising to ~ 10 -km depth. This is the so-called Ivrea body anomaly. The S velocity structure displays a deep low-velocity zone at 150 - 200 km distance and 20 - 35 km depth. This deep LVZ can be guessed in the P velocity section.

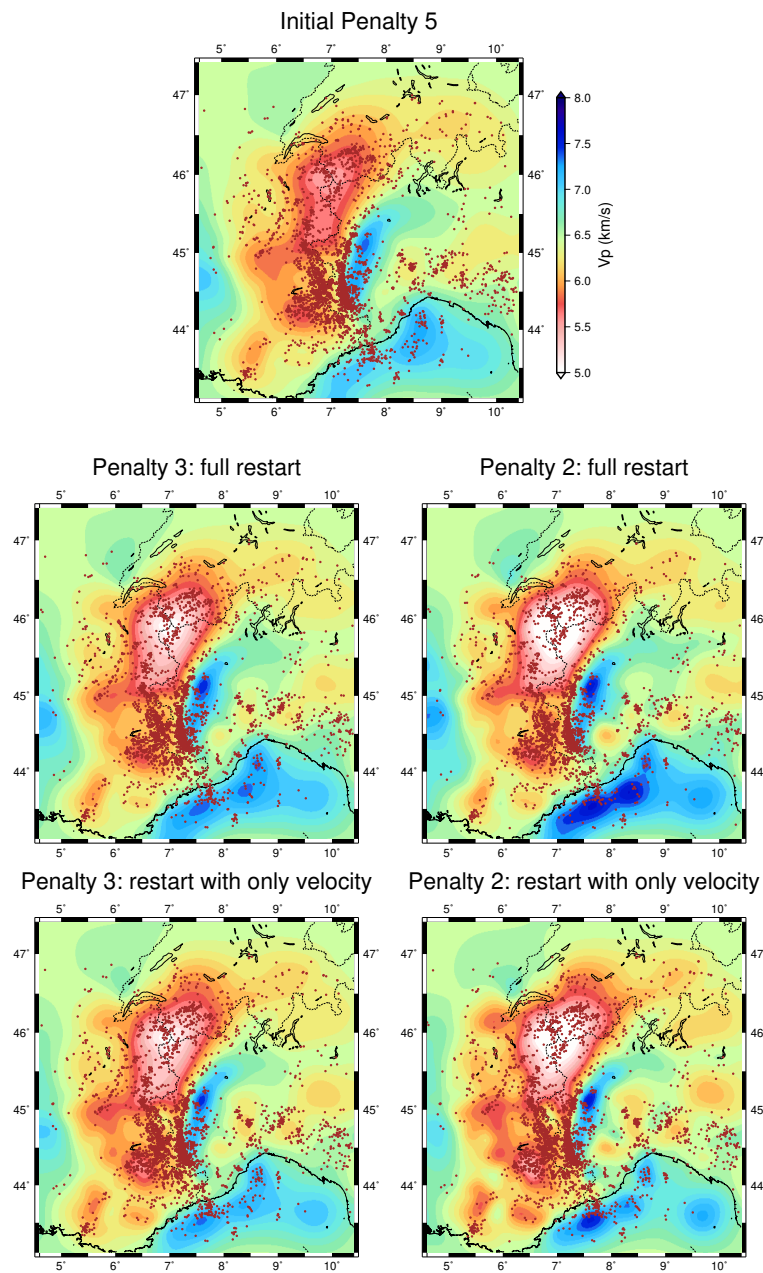


Figure S7: Results on the V_p slice at 20-km depth of a cascade of inversions with decreasing Laplacian hyperparameters. Top: result of the initial Laplacian workflow with weight $\lambda = 5$ in all directions. Middle: the resulting model (velocities and hypocenters) is then input in successive inversions with weights $\lambda_{x,y} = 3$, $\lambda_z = 2$ (left panel) and $\lambda_{x,y} = 2$, $\lambda_z = 1$ (right panel). Bottom: same as second row but with HYPO71 locations as initial model. Although velocity patterns are very similar, small differences illustrate the respective contributions of initial velocity and hypocenter parameters.

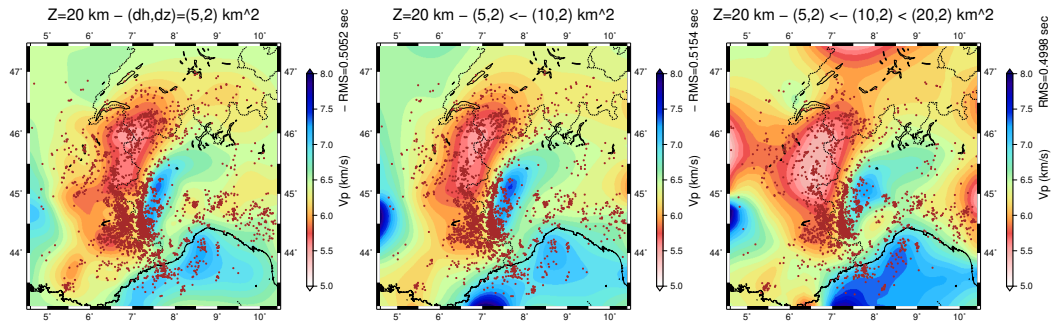


Figure S8: Influence of the grid discretisation on the 20-km depth V_p slice. Same values of hyperparameters as in the first tomography run. Left: velocity slice with a horizontal discretisation of 5 km. Center and right: velocity slices with two different grid refinement strategies (10 to 5 km, and 20 to 10 to 5 km). RMS values are almost identical for the three models. Initial coarse grid steps of 10 km or 20 km induce strong edge effects.

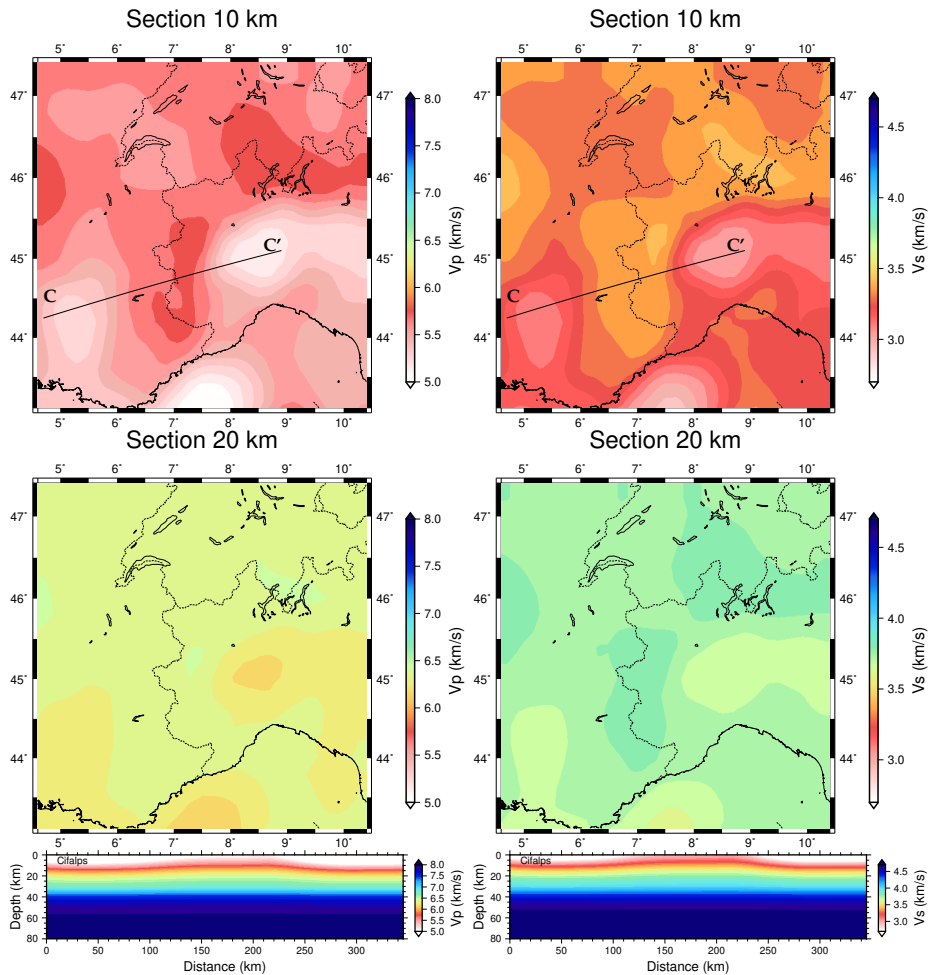


Figure S9: Initial $3D-ANT$ model combining an ambient-noise-tomography model and the reference layered model: a smooth change between the 2 models is performed in the range [10 km-30 km]. V_p in the left-hand side and V_s in the right-hand side. Top: horizontal sections at 10-km depth; Middle: horizontal sections at 20 km; Bottom: vertical sections along the Cifalps profile.

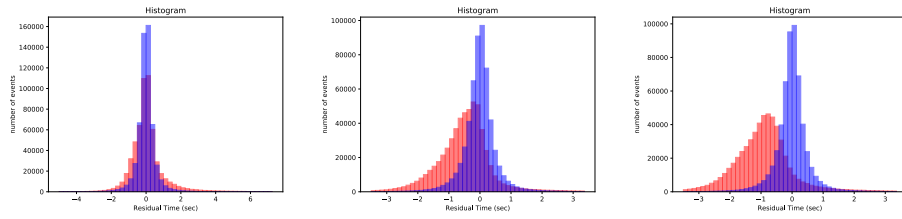


Figure S10: Arrival time residuals in the $[-4 \text{ s}, 4 \text{ s}]$ range for initial (red) and recovered (blue) models. Left: initial $Z\text{-}HQ$ model with HYP071 locations; Center: initial $3D\text{-}ANT$ model with HYP071 locations; Right: initial $3D\text{-}ANT$ model with NLLOC locations. More negative initial residuals appear with the NLLOC procedure while the final (blue) histogram is sharper when starting from a stratified model (right).

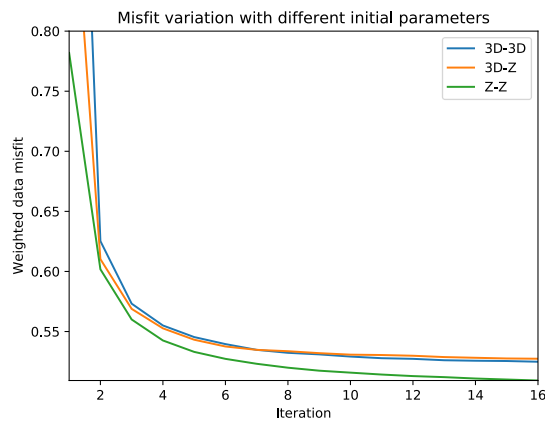


Figure S11: Date misfit evolution with iterations; Green line: initial layered model $Z\text{-}HQ$ with HYP071 locations; Red line: 3-D (it 3D-ANT) model with HYP071 locations; Blue line: 3-D model with NLLOC locations. Again, the stratified model looks like a better initial guess than a 3-D initial model too far from the target 3-D model.

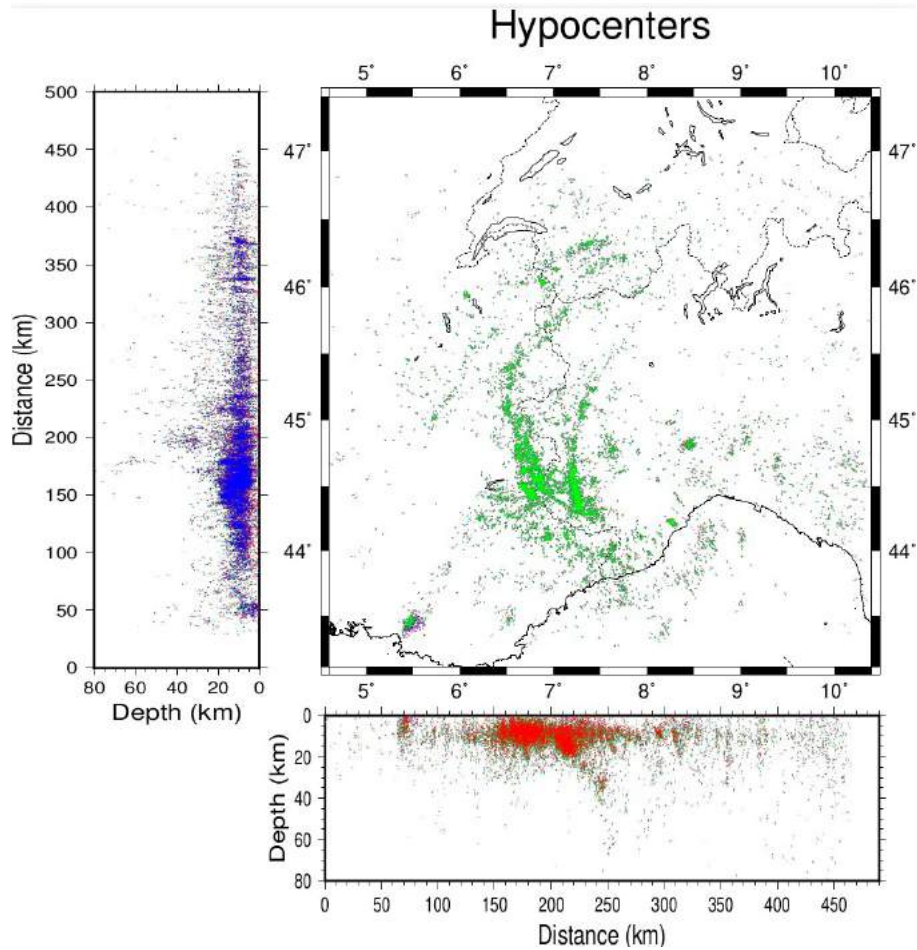


Figure S12: Final hypocenter locations after inversion with three different initial models. Blue: stratified initial model with HYPO71 locations (plotted in the foreground in the left panel); Red: 3-D initial model with HYPO71 locations (in the foreground in the bottom panel); Green: 3-D initial model with NLLOC locations (in the foreground in the central panel).

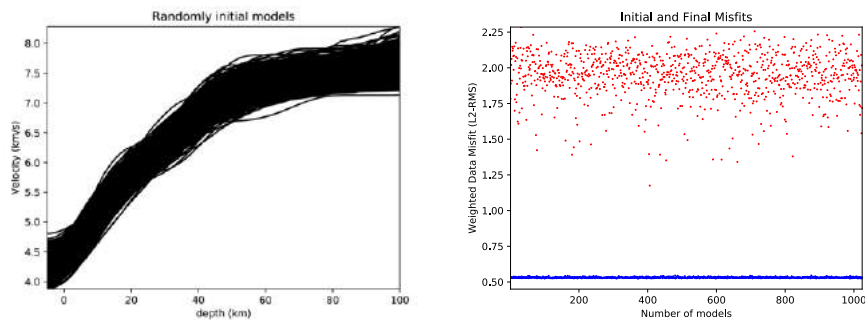


Figure S13: Left panel: Random initial stratified models computed around Potin (2016)'s initial model. Right panel: initial (red) and final (blue) data misfits for the 1024 inversions with initial random stratified models shown in the left-hand side. Final misfits concentrate between 0.52s and 0.53s, in the same range as with Potin (2016)'s initial stratified model.

69 2.1 Text S3 and Figs. S13-S16: Randomly-layered (stratified) initial models

70 To further assess the influence of the initial velocity model, we generated a set of 1024 stratified
 71 random models around Potin (2016)'s initial model of Fig. S2. They are shown in Fig. S13 (left). Then,
 72 we performed 1024 tomographies using each model as initial velocity model, the *HQ-89-14* database,
 73 and a Laplacian-smoothing inversion procedure with the same hyperparameters as in section 4 of the
 74 main text. The initial and final data misfits are shown in Fig. S13 (right). For further analyses of the
 75 inversion results, we selected final models with a data misfit below 0.525 s, that is $\sim 1\%$ of random
 76 models. The conclusions that we will draw do not strongly depend on this selection.

77 We now have two sets of velocity models, a set of initial (stratified) and a set of final (3-D) models.
 78 For each set, we compute an average velocity model and a standard deviation (RMS) model. At a
 79 given depth, areas sampled by seismic waves should have lower final than initial RMS values. Figure
 80 S14 shows the RMS ratio between the final and initial *P* velocity models. It may be considered as a
 81 proxy of the initial-model influence on the reconstructed *P* velocity models. A nearly similar pattern is
 82 observed for the reconstructed *S* velocity models. At 10-km depth, poorly sampled zones corresponding
 83 to high values of the RMS ratio are low-seismicity areas with poor station coverage. At 20-km depth,
 84 areas of low proxy values have a high density of hypocenters.

85 We may also evaluate the amplitude of velocity changes during the inversions. Figure S15 displays
 86 the average percentage of velocity perturbations (with respect to the initially constant velocity) at 10
 87 and 20-km depths. At 10 km, areas of small velocity perturbations concentrate along the borders of the
 88 tomographic box. The low-velocity anomaly at the end of the CIFALPS CC' profile is well identified.
 89 At 20 km depth, a strong (positive) velocity variation is required in the Ivrea body region.

90 The average final velocity model shown in Fig. S16 might be considered as a possible initial 3-D
 91 model in future inversions. However, poorly sampled zones may have velocity values depending strongly
 92 on the way the random distribution of initial velocities is computed. We disregard such possible 3-D
 93 initial models because of these unwanted low-wavenumber components.

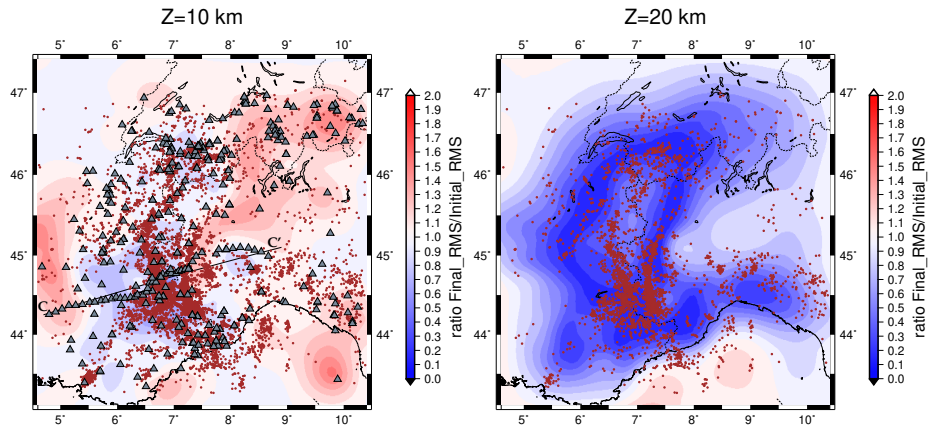


Figure S14: Maps at 10 and 20 km depth of the ratio between RMS values of final P velocity models and RMS values of initial P velocity models. Low values mean that different inversions with different initial stratified models provide similar results. Note the greater variability of final results at 10 than at 20-km depth, the influence of earthquake distribution at both depths, and the influence of station coverage at 10 km. This ratio is a good proxy of the influence of the initial model.

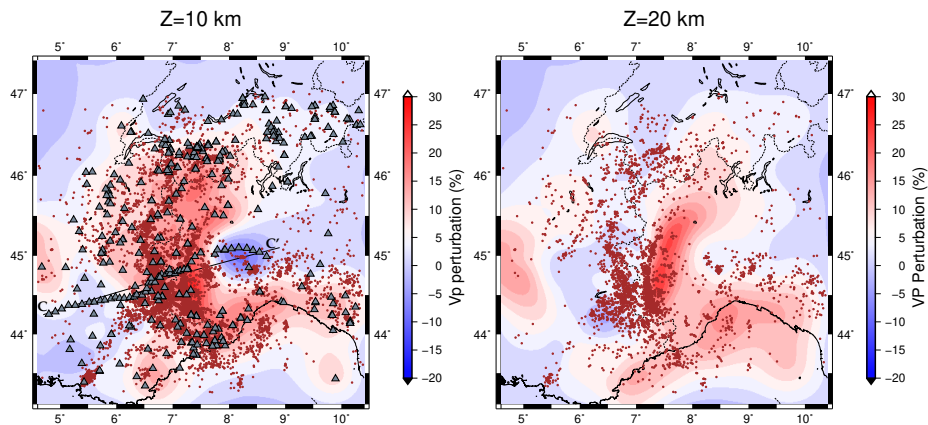


Figure S15: Maps at 10 and 20 km depth of the average velocity perturbations in the final models with respect to their initial models. The maps display areas where fitting arrival time data requires significant velocity changes.

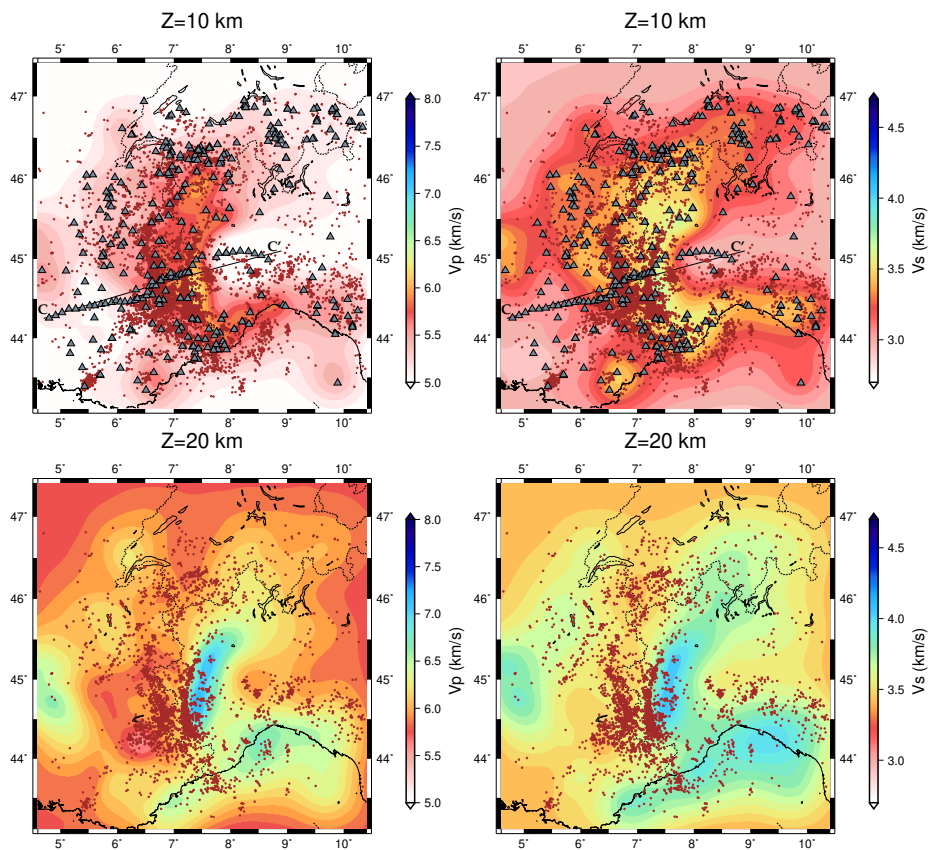


Figure S16: Depth slices at 10-km (top) and 20-km depths (bottom) in the average P (right) and S (left) velocity models reconstructed after inversions with initial random stratified models.

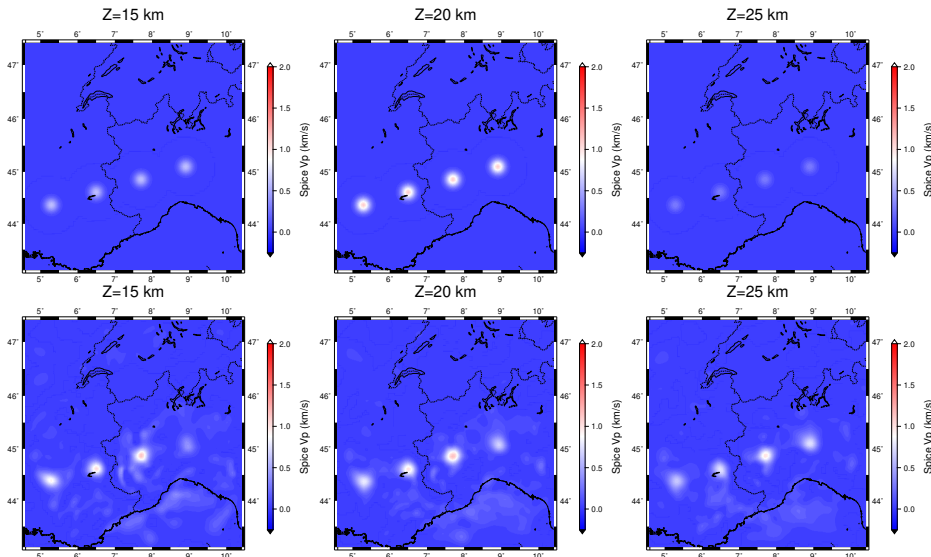


Figure S17: Horizontal sections at 15, 20 and 30-km depth for individual spike tests with a Gaussian V_p anomaly centred at 20 km-depth along the CIFALPS profile; Top: input synthetic model; Bottom: recovered model. Each spike is recovered independently, even if the four models are plotted together.

2.2 Text S4 and Figs. S16-S17: Spike tests

We follow the spike test strategy promoted by Spakman (1991) and add velocity perturbations on top of the final model m_{opt} . Since we are using a standard Laplacian-smoothing/damping approach, we consider a small velocity perturbation described by a Gaussian function centred at a given point with different characteristic lengths along the three directions. The perturbation should be small enough to avoid significant ray deviation.

Synthetic arrival times are computed in model m_{opt} perturbed by a single spike for the same earthquake-station couples as in the observed dataset. An inversion is performed starting from model m_{opt} with a few iterations. The difference between the newly recovered model and the final model is then compared to the input synthetic velocity anomaly.

As proposed by Rawlinson and Spakman (2016), we proceed with individual spikes even if they are combined for plotting. With this method, the inversion may reveal velocity anomalies at specific locations independently of spike locations, but tightly connected to station and earthquake distribution.

Input and output models are shown in Figs. S17 and S18 for an example with four independent positive anomalies at 20-km depth along the CIFALPS profile. The maximum perturbation is 1200 km/s and the characteristic lengths are 15 km in the horizontal directions and 5 km in the vertical one. Horizontal slices at 15, 20 and 25-km depths and depth sections illustrate amplitude changes during the reconstruction associated with slight spatial spreading (Figs. S17 and S18). The velocity reconstruction has an overall satisfactory quality.

A few anomalous patterns are observed whatever the location of the spike anomaly, which are highlighted by our tests with individual spikes combined in a second stage for plotting. This is the case for the anomaly at 90-km distance and 10-km depth in Fig. S18, as well as the anomalies spread in the south-eastern corner of the study region in Fig. S17. As outlined above, these spurious anomalies are connected to station and event distribution. They could lead to ambiguous interpretations of the results of tests performed with multiple spikes, even if they are sparsely distributed.

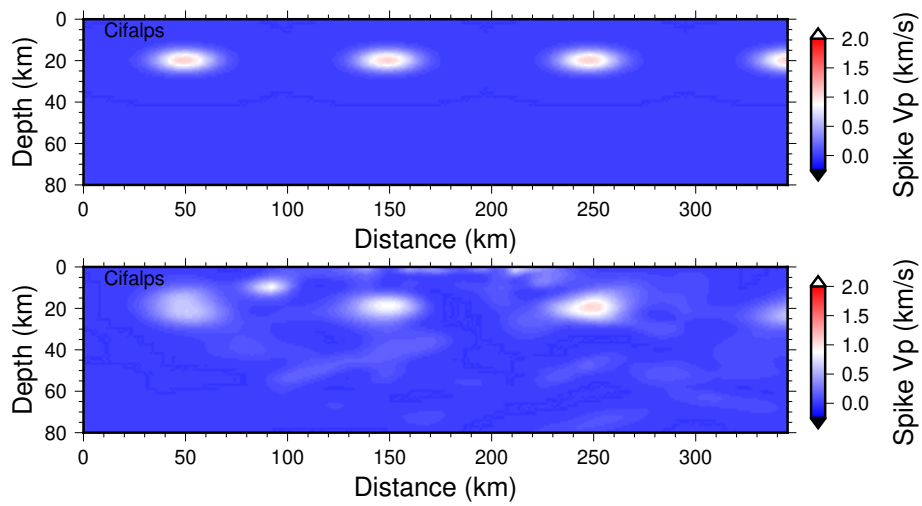


Figure S18: Depth sections along the CIFALPS profile in the same individual spike test model as in Fig. S17, with a Gaussian anomaly centred at 20 km-depth; Top: input synthetic model; Bottom: recovered model.

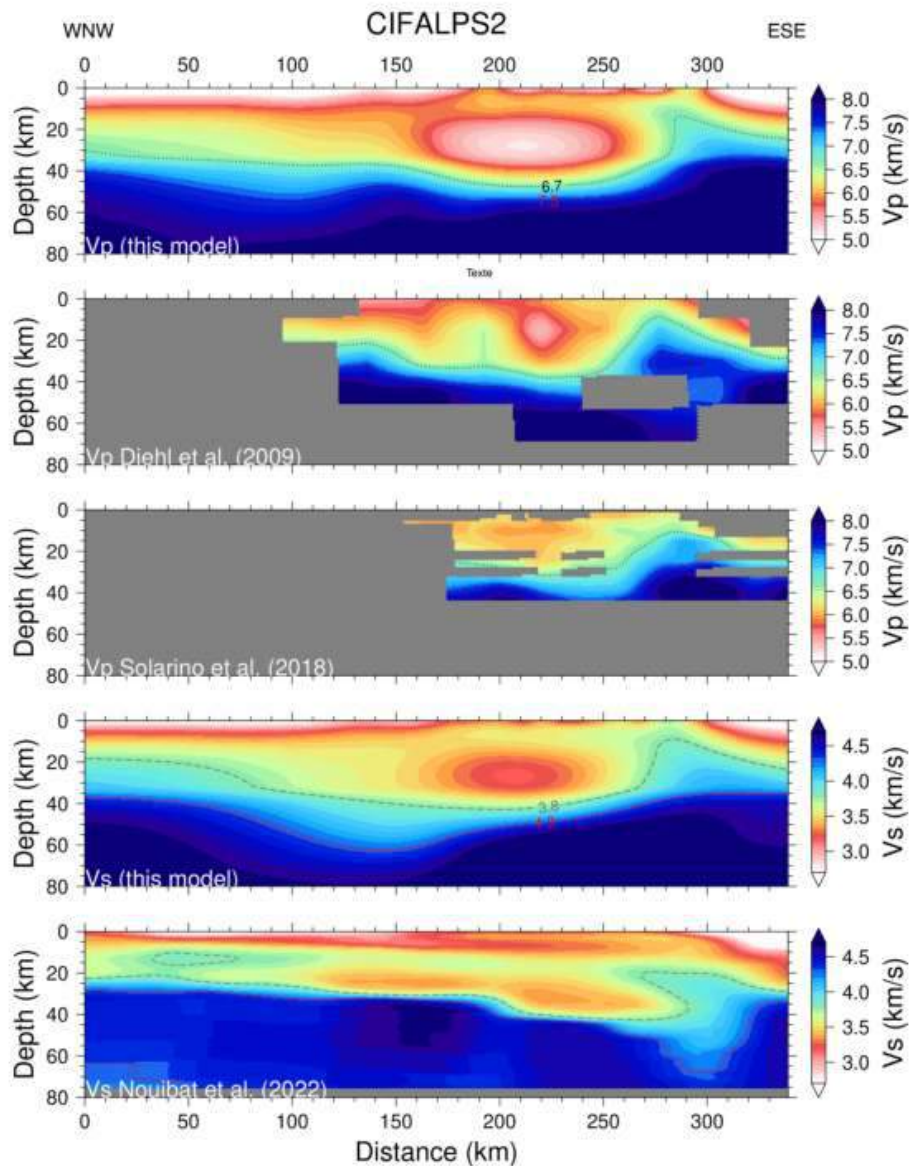
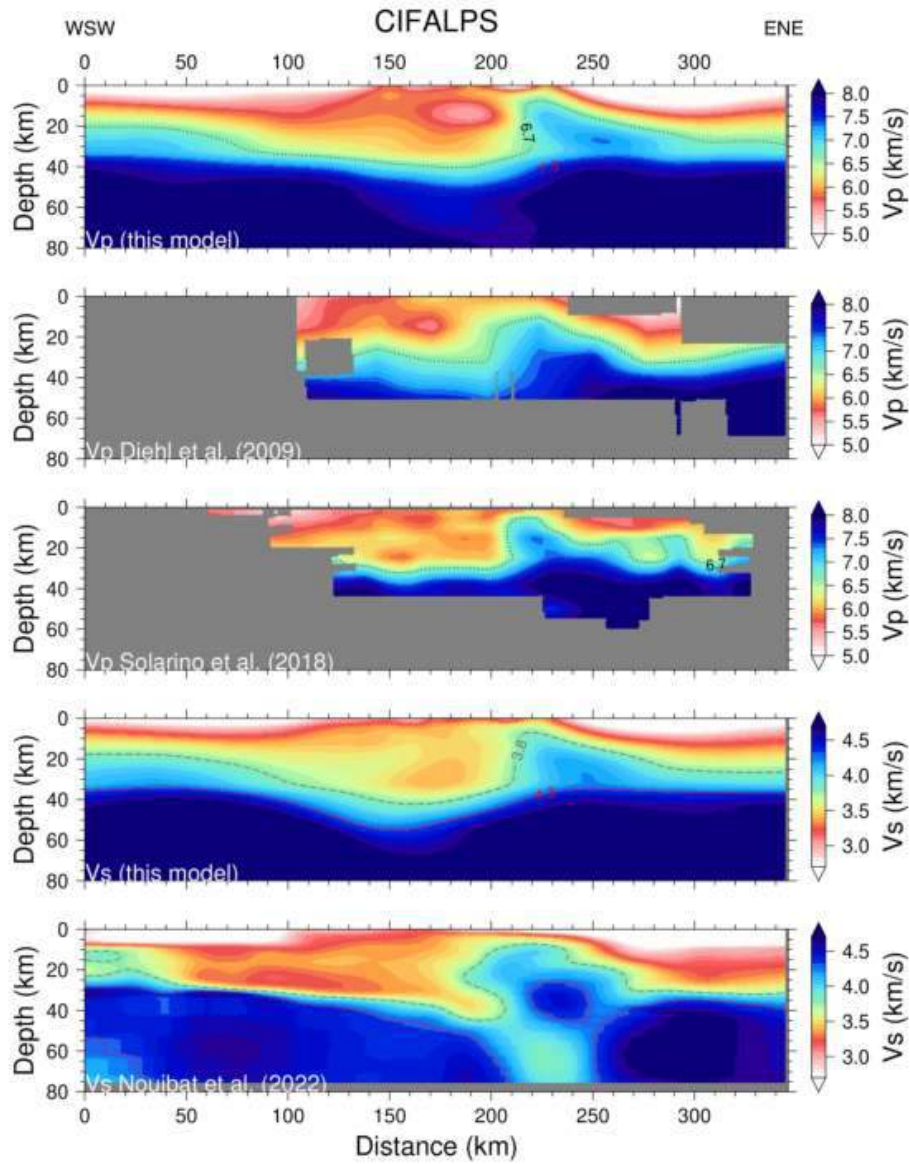


Figure S19: Depth slices in the V_p (top, second, and third rows) and V_s (fourth and bottom rows) models along the CIFALPS2 section across the northwestern Alps (location in the top left panel of Fig. 12 in the main text). The model's name is indicated in the lower left corner of each section. The 6.7 km/s (black) and 7.5 km/s (red) contours are shown as dotted lines in the three V_p sections, while the 3.8 km/s (black) and 4.3 km/s (red) contours are shown as dashed lines in the two V_s sections. The red contours may be considered as Moho proxies, while the black contours highlight the Ivrea body high-velocity shallow anomaly in the eastern side of the sections (distance range: 270 - 310 km).



53 Figure S20: Same legend as Fig. S19 for the CIFALPS section across the southwestern Alps (location
54 in the top left panel of Fig. 12 in the main text).
55
56
57
58
59
60

119 **References**

- 120 Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., and Virieux, J. (2012). Regu-
121 larized full waveform inversion including prior model information. In *EAGE, 74th Conference and*
122 *Exhibition, Expanded abstracts*, page W031.
- 123 Dahl, J., Hansen, P. C., Jensen, S. H., and Jensen, T. L. (2009). Algorithms and software for total
124 variation image reconstruction via first-order methods. *Numerical Algorithms*, 53:67–92.
- 125 Fomel, S. and Claerbout, J. F. (2003). Mutlidimensional recursive filter preconditioning in geophysical
126 estimation problems. *Geophysics*, 68(2):1–12.
- 127 Guitton, A. (2012). Blocky regularization schemes for full waveform inversion. *Geophysical Prospecting*,
128 60(5):870–884.
- 129 Potin, B. (2016). *Les Alpes occidentales : tomographie, localisation de séismes et topographie du Moho*.
130 PhD thesis, Université Grenoble Alpes.
- 131 Rawlinson, N. and Spakman, W. (2016). On the use of sensitivity tests in seismic tomography. *Geo-*
132 *physical Journal International*, 205:1221–1243.
- 133 Spakman, W. (1991). Delay-time tomography of the upper mantle below Europe, the Mediterranean,
134 and Asia Minor. *Geophysical Journal International*, 107:309–332.
- 135 Wellington, P., Brossier, R., Hamitou, O., Trinh, P., and Virieux, J. (2017). Efficient anisotropic dip
136 filtering via inverse correlation functions. *Geophysics*, 82(4):A31–A35.
- 137 Zhao, L., Paul, A., Guillot, S., Solarino, S., Malusà, M. G., Zheng, T., Aubert, C., Salimbeni, S.,
138 Dumont, T., Schwartz, S., Zhu, R., and Wang, Q. (2015). First seismic evidence for continental
139 subduction beneath the Western Alps. *Geology*, 43(9):815–818.