



**Pacific Northwest**  
NATIONAL LABORATORY

*Proudly Operated by Battelle Since 1965*

# Validation of Ultrasonic Nondestructive Examination (NDE) Computational Models – Phase 1

**March 2017**

G Dib  
SL Crawford  
RE Jacob

MS Prowant  
AA Diaz



Prepared for the U.S. Nuclear Regulatory Commission  
under a Related Services Agreement with the U.S. Department of Energy  
CONTRACT DE-AC05-76RL01830

U.S. DEPARTMENT OF  
**ENERGY**

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY

*operated by*

BATTELLE

*for the*

UNITED STATES DEPARTMENT OF ENERGY

*under Contract DE-AC05-76RL01830*

Printed in the United States of America

Available to DOE and DOE contractors from the  
Office of Scientific and Technical Information,  
P.O. Box 62, Oak Ridge, TN 37831-0062;  
ph: (865) 576-8401  
fax: (865) 576-5728  
email: [reports@adonis.osti.gov](mailto:reports@adonis.osti.gov)

Available to the public from the National Technical Information Service  
5301 Shawnee Rd., Alexandria, VA 22312  
ph: (800) 553-NTIS (6847)  
email: [orders@ntis.gov](mailto:orders@ntis.gov) <<http://www.ntis.gov/about/form.aspx>>  
Online ordering: <http://www.ntis.gov>



This document was printed on recycled paper.

(8/2010)

# **Validation of Ultrasonic Nondestructive Examination (NDE) Computational Models – Phase 1**

G Dib  
SL Crawford  
RE Jacob

MS Prowant  
AA Diaz

March 2017

Prepared for  
the U.S. Nuclear Regulatory Commission  
under a Related Services Agreement  
with the U.S. Department of Energy  
Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory  
Richland, Washington 99352



# Summary

The Pacific Northwest National Laboratory (PNNL) is conducting confirmatory research for the U.S. Nuclear Regulatory Commission for the validation of computational models for ultrasonic testing (UT). The purpose of this research is to determine if UT computational models adequately represent reality, within their intended domain of application. This work attempts to assess the reliability of such models by directly comparing them to empirical measurements. We propose a framework that allows assessing the model by using both qualitative and quantitative validation indicators and metrics. This framework acknowledges the presence of errors and uncertainties in all models and experiments, and the quantitative validation stage uses a statistical approach that incorporates uncertainties to infer the adequacy of the model. This framework is used to assess the adequacy of CIVA-UT Version 11 to model conventional ultrasonic nondestructive examination inspections.

The validation study considers simple geometrical reflectors in isotropic, fine-grained, homogeneous materials with conventional ultrasonic transducers. In particular, wrought stainless steel plates with electro-discharge machined notch reflectors varying in length, depth, and orientation (tilt from the normal) are evaluated. Different types of conventional ultrasonic transducers are investigated, including different wave modes, beam angles, and frequencies.

In the qualitative analysis, the model prediction results were compared with empirical data in terms of the similarity of the relevant [visual] features of C-scan, B-scan, and A-scan images. Overall, the most prominently observed differences could be attributed to uncertainties in the validation process itself including uncertainties in model parameters, and measurement noise.

On the other hand, the quantitative analysis considered a metric that computes the effective amplitude of the flaw response under uncertainties in model parameters and measurement noise.

The validation results show that the models in CIVA were reasonably accurate in representing their empirical counterparts in a qualitative fashion. However, when comparing the results based on the quantitative metric employed in this study, there were large variances in accuracy and the associated empirical and simulation uncertainties were significantly large. Thus, we conclude that using models for applications that require quantitative reasoning (such as predicting the detectability of small flaws or model-assisted probability of detection studies) is rather unreliable and errors over 20 dB could occur.

Ultimately models need to be assessed on whether they can be used to predict if an inspector can find a flaw in a component. The quantitative analysis results show that using models to give definitive decisions about the detectability of flaws would be challenging due to the significant number of variables in the modeling process. This problem will be further addressed in the future by providing bounds on uncertainties and specifying guidelines for the modeling process. Those questions will be addressed as PNNL expands the scope to include realistic flaws in complex materials such as austenitic welds and inhomogeneous base material. This study focused on validating the CIVA model by comparing simulation results to empirical data and assessing the differences and limitations. Once the limitations are understood then the detection question can be more reliably addressed.



## Acknowledgments

The work reported here was sponsored by the U.S. Nuclear Regulatory Commission (NRC) under Pacific Northwest National Laboratory (PNNL) Project No. 62895; NRC JCN V6323. Ms. Carol Nove is the NRC Contracting Officer Representative and Program Monitor. PNNL would like to thank Ms. Nove for her guidance and technical direction over the course of this work and for providing a technical review of the contents of this Technical Letter Report (TLR).

PNNL appreciates the support of the Electric Power Research Institute (EPRI) Nondestructive Evaluation (NDE) Center, in Charlotte, North Carolina, for supporting activities and work associated with this effort as part of the NRC-EPRI Memorandum of Understanding. In addition, PNNL is grateful to EPRI for hosting PNNL at the NDE Center to conduct technical scoping meetings and status reviews, and for sharing data and mockups in support of facilitating PNNL's objectives and mission on this task. In particular, PNNL is grateful to Mr. Greg Selby, Mr. Mark Dennis, Mr. Thiago Seuaciuc-Osorio, and Mr. Myles Dunlap at the EPRI NDE Center for their technical assistance, hospitality, and support during PNNL's visits and on-going collaborative activities.

At PNNL, the authors are indebted to Mr. Royce Mathews for providing in-laboratory support in configuring the laboratory workspaces, moving mockups, packing and shipping necessary equipment for ongoing activities, and a variety of other support-related efforts. We are grateful to Mr. Samuel Sexton (PNNL summer intern) for his support of conventional ultrasonic testing data acquisition efforts. We would also like to extend our gratitude to Ms. Lori Bisping for all of her hard work in providing administrative and financial reporting support to this subtask. Finally, the authors would like to extend their thanks to Ms. Kay Hass for her ongoing support, attention to detail, and technical editing expertise in preparing and finalizing this TLR.

PNNL is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830.



## Acronyms and Abbreviations

AWGN	additive white Gaussian noise
AIAA	The American Institute of Aeronautics and Astronautics
ASME	The American Society of Mechanical Engineers
CASS	cast austenitic stainless steel
CEA	Commissariat à l’Energie Atomique
CMS	computer modeling and simulation
EDM	electro-discharge machined
FBH	flat-bottom hole
FDM	finite difference method
FDTD	finite difference time domain
FEM	finite element method
FFT	fast Fourier transform
GTD	geometric theory of diffraction
GUI	graphic user interface
HVS	Human Visual System
IEEE	Institute of Electrical and Electronics Engineers
ISI	inservice inspection
LRT	likelihood ratio test
NDE	nondestructive examination/evaluation
NPP	nuclear power plant
NRC	U.S. Nuclear Regulatory Commission
NRR	Office of Nuclear Reactor Regulation
PNNL	Pacific Northwest National Laboratory
RF	radio frequency
SCS	Society of Computing Simulation
SDH	side-drilled hole
SME	subject matter expert
SNR	signal-to-noise ratio
SOV	separation of variables
SRQ	System Response Quantity
TLR	technical letter report
TOFD	time-of-flight diffraction
TRL	transmit-receive longitudinal
UT	ultrasonic testing
V&V	verification and validation



# Contents

Summary .....	iii
Acknowledgments.....	v
Acronyms and Abbreviations .....	vii
1.0 Introduction .....	1.1
1.1 Scope of Work.....	1.2
1.2 Problem Statement and Research Objectives.....	1.4
1.3 Report Organization .....	1.4
2.0 Review of Verification and Validation.....	2.1
2.1 The Computer Modeling Process.....	2.1
2.2 Model Verification.....	2.2
2.3 Model Validation.....	2.2
2.4 Errors and Uncertainties.....	2.3
3.0 Review of Mathematical Models for UT.....	3.1
3.1 General Wave Physics.....	3.1
3.2 Asymptotic Approximations .....	3.2
3.3 Computational Models in CIVA .....	3.3
3.4 Performance of Kirchhoff Approximation.....	3.4
3.5 Comparison of CIVA and Wave3000+.....	3.6
4.0 Measurement and Modeling Procedures .....	4.1
4.1 Specimen and Probes .....	4.1
4.2 Empirical Data Collection.....	4.3
4.3 Setting up Models in CIVA.....	4.3
4.3.1 Specimen Specification .....	4.4
4.3.2 Probe Specification .....	4.5
4.3.3 Flaws Specification .....	4.8
4.3.4 Inspection Specification .....	4.8
4.3.5 Computation Specification.....	4.9
5.0 Model Validation Framework.....	5.1
5.1 Qualitative Analysis Methodology.....	5.2
5.1.1 C-scan Views.....	5.3
5.1.2 B-scan Views.....	5.4
5.1.3 A-scan Views .....	5.6
5.2 Probabilistic Framework.....	5.6
5.2.1 Derivation of the Validation Metric and Empirical Uncertainty.....	5.7
5.2.2 Uncertainty in Simulations.....	5.10
6.0 Analysis Results and Discussion .....	6.1

6.1	Qualitative Analysis .....	6.1
6.1.1	C-scans .....	6.2
6.1.2	B-scans .....	6.3
6.1.3	A-scans .....	6.5
6.1.4	Summary of Qualitative Analysis .....	6.7
6.2	Quantitative Probabilistic Analysis .....	6.7
6.2.1	Simulation Uncertainty: Parametric Study .....	6.7
6.2.2	Empirical Uncertainty .....	6.11
6.2.3	Effective Metric .....	6.11
7.0	Conclusions and Future Work .....	7.1
8.0	References .....	8.1
	Appendix A – Summary of Specimen and Probes Dimensions .....	A.1
	Appendix B – Composite Views of Ultrasonic Results .....	B.1
	Appendix C – Validation Metric Results .....	C.1

# Figures

1.1	Example of UT Modeling Scenario under Consideration .....	1.4
2.1	Phases in CMS and Its Relation with V&V .....	2.1
3.1	(a) Comparison between the Exact SOV Method and Kirchhoff Approximation. (b) Amplitude difference between separation of variables and Kirchhoff approximation.....	3.6
3.2	Geometry Used for the Models in CIVA and Wave3000+ .....	3.7
3.3	Comparison of A-scan Results for Two Different Simulation Packages: Wave3000+ and CIVA.....	3.7
4.1	Validation Tree Describing the Probes and Scans Directions Used for Each Specimen .....	4.1
4.2	Stainless Steel Plates with Reflectors Used for the Validation Study.....	4.2
4.3	Defining the Notch Reflector Tilt and Depth within the Global Coordinate System of the Specimen.....	4.2
4.4	Experimental Setup for Obtaining the UT Raster Scans.....	4.3
4.5	General Workflow to Define a Model of Homogeneous Isotropic Plates with Notches under Examination Using Conventional Transducers.....	4.4
4.6	<i>Left:</i> Model of a Probe with a Circular Piezoelectric Element and Angle Wedge. <i>Right:</i> CIVA's Implementation of a Piezoelectric Element, where it is Replaced by the Effective Pressure $s(t)$ at the Wedge's Surface.....	4.6
4.7	Block Diagram of the Pulse-Echo Ultrasound System, Represented as a Linear System in the frequency Domain.....	4.6
4.8	An Example of the Gaussian Tone Burst Signal Defined in CIVA for Probe 1 as Compared to the Measured Edge Response Signal .....	4.7
4.9	Dimensions Required for Specifying the Probe Wedges .....	4.8
5.1	B-scan View Describing the Ultrasonic Response Echoes from a Regularly Shaped Crack-Like Reflector .....	5.3
5.2	C-scan Views (top view of raster data) of Responses from Two Different Reflectors.....	5.4
5.3	B-scan Image of a Response from an EDM Reflector when Mode Conversion is Present .....	5.5
6.1	An Example Showing the Effect of the Median Filter on the B-scan Image .....	6.2
6.2	Example of the C-scans from Specimen 3, Flaw M, Skew 0°, Probe 3 .....	6.2
6.3	Example of the C-scans from Specimen 1, Flaw J, Skew 0°, Probe 2 .....	6.3
6.4	Example of the B-scans from Specimen 1, Flaw I, Skew 0°, Probe 2 .....	6.4
6.5	B-scan Images for Shallow Flaw, A and Deep Flaw, I.....	6.5
6.6	A-scans for a Shallow Flaw, A and Deep Flaw, I.....	6.6
6.7	Parametric Studies Error for the 19 Parameters Investigated.....	6.10
6.8	Results of the Metric Computation for Both Observation and Prediction from Scans in Specimen 1, Probe 4, Skew Angle 0° .....	6.12
6.9	Computed Metric in the Observation-Prediction Space.....	6.13
6.10	Computed Metric in the Observation-Prediction Space.....	6.14
6.11	Computed Metric in the Observation-Prediction Space for the Reflectors in Specimen 1, Probe 3, Skew 0 .....	6.15



# 1.0 Introduction

Based on a User Need Request from the U.S. Nuclear Regulatory Commission (NRC), Office of Nuclear Reactor Regulation (NRR), the Office of Research (RES) developed a task to validate industry activities with respect to ultrasonic modeling. The industry is beginning to use ultrasonic modeling as a basis to demonstrate that the ultrasonic testing (UT) examination techniques being applied are acceptable. An assessment to confirm that a solid technical basis exists for conducting, interpreting, and applying the results of ultrasonic modeling must be performed. Recently, Pacific Northwest National Laboratory (PNNL) has used acoustic modeling to support NRR in evaluating nondestructive examination (NDE)-related events at operating nuclear power plants, and licensee requests for alternative examinations. RES leveraged these activities, as well as on-going work in the modeling cooperative under the NRC-Electric Power Research Institute (EPRI) Memorandum of Understanding (MOU), to establish a new task within JCN-V6323 to address the issues surrounding UT simulations as bases for regulatory decision-making.

PNNL was scoped with employing NDE techniques and theoretical modeling to evaluate ultrasonic flaw signal responses impacted by varied microstructures and component geometry, and will validate and employ the results to identify issues to the NRC that should be addressed to enhance simulations and ensure examination reliability. This work will provide a better understanding and a technical basis to inform NRR regarding the confidence and accuracy of simulated results, when licensees employ models to support presumptions such as crack detection and volumetric coverage of materials. PNNL is conducting work to evaluate specific modeling and simulation tools currently used for UT. In particular, the focus is to evaluate semi-analytical modeling methods with CIVA as the software tool of choice for addressing these issues.

Computer modeling and simulation (CMS) is used for predicting a system's response in scenarios that are either impossible or costly to be implemented experimentally. In addition, CMS provides insights for theoretically explaining situations that are otherwise difficult to visualize through empirical efforts. In the field of UT, CMS allows personnel to assess the behavior of ultrasonic probes under various conditions, visualize and characterize generated sound fields in the material, study wave propagation characteristics and scattered energy from reflectors, and predict the probability of detecting different flaw types and orientations. Moreover, it is especially useful for making predictions under complex geometries and materials for rapidly evaluating coverage issues, impingement angles, insonification effectiveness of targeted volumes of material, signal-to-noise ratio (SNR), etc.

As computer processing speed continues to evolve, CMS has matured and its benefits are becoming better understood and acknowledged by both the commercial nuclear industry and government regulatory agencies. However, CMS tools are only as effective as the algorithms and physics engines used for computations. In addition, accuracy of model inputs, the NDE experience level of the user, and other factors play a key role in how reliable and effective a modeling and simulation tool can be. Thus, it is critically important to validate modeling tools in order to determine if the simulations adequately represent acoustic responses of their real counterparts. Also, CMS users need to understand how much information and what type of information is required (input parameters) for a CMS platform to be utilized effectively and obtain acceptable results. The interface to a computer model is complex with many menus, boxes, etc. to consider and without proper training the user can inadvertently ask the model to run an unintended simulation, one that is far different from its experimental counterpart. Without a doubt, there are many uncertainties associated with modeling. When reality is uncertain, it cannot be represented by a single model. Because of this, CMS users need to be trained, not only on how to use a particular software tool, but also on how to handle such uncertainties in their modeling, and how to draw resultant conclusions from them. During the early initial work in Phase 1, the PNNL modeling task team (hereafter called Team) encountered challenges associated with both empirically derived results and employing the

most effective inputs to the CIVA software. Over time the Team became more proficient and knowledgeable in the use of the software tool and its sensitivities to key input parameters that have significant impact to the simulated results. This learning process was a valuable component of this research, and has created a new level of awareness within the Team as both semi-analytical and full physics-based numerical models are assessed for use in supporting NDE applications.

Human factors in some sense also play a role. There are individual differences and experience levels to consider in the personnel involved in the work as well as group or team dynamics. Is the same person acquiring empirical data as is running the computer model? If not, how good is the communication between the two? Is the experience level of the goal setter adequate so that only realistic scenarios are examined? For example, one would not want to model the performance of a 500 kHz probe on a thin, 12.7-mm (0.5-in.) plate as this frequency is too low for the application. It is critical that simple scenarios are assessed first and the results understood before moving on to more realistic components that would be found in nuclear applications.

With regard to the effectiveness of modeling tools to accurately simulate sound fields, ultrasonic signal responses, and the inspected material volume of interest, many parameters play a role in determining complete insonification of a defined volume within a component. These include component material, weld configuration, surface conditions, access limitations, potential degradation mechanisms, probe type and characteristics, etc. As industry begins to rely more heavily on commercially available modeling tools for determination of inspected volumes, it will become increasingly necessary to understand the effectiveness and reliability of these tools to provide realistic and accurate results. Of critical importance is how these tools handle beam position and impingement in components where beam redirection and partitioning are prevalent. As an example, we can use the challenging scenario of UT in cast austenitic stainless steel (CASS) components. Often in CASS components, the effects of the microstructure on sound field propagation and directionality can be severe, and this can lead to gaps in inspected weld volumes where areas are not effectively insonified.

The increased use of CMS to infer critical pieces of UT information across the commercial nuclear sector is a driving force to better understand relevant strengths and limitations regarding the use of CMS tools. For CMS to have any impact on decision and policy makers, its trustworthiness and validity should be proven. Model validation aims at delivering a quantified confidence in a particular model and simulation software results. By associating a level of confidence with a given simulation model, engineers or policy makers can make informed decisions based on the level of quantified uncertainty of the model.

## **1.1 Scope of Work**

In the absence of empirical data, effective ultrasonic modeling and simulation may allow NDE personnel to evaluate how a UT probe might behave under various conditions, how sound fields may propagate in different materials or at boundaries within a material, or how ultrasonic signal responses from flaws might vary. Simulations can complement experimental work to provide further insights into volumetric coverage issues, flaw SNRs, optimum impingement angles, and sound field characteristics needed for targeting specific volumes of material within a component. However, the effectiveness of modeling/simulation tools is directly related to the accuracy of the algorithms and physics engines used for computations. In addition, model input variables, the NDE experience level of the user, and other factors play key roles in how reliable and effective a model/simulation tool can be. It is critically important to assess the reliability of modeling tools to determine if they adequately simulate realistic acoustic results. Conversely, it is just as critical to understand the weaknesses of modeling tools and the associated scenarios where these tools may provide results that are incorrect or inaccurate. Model users

need to understand the hierarchy and depth of information (input parameters) required to obtain acceptable results.

The physics engines and computational algorithms currently employed in CIVA (and other modeling and simulation tools) come with specific strengths and inherent limitations regarding sound field propagation, noise and attenuation, how flaws are simulated, and how material inhomogeneity and anisotropy are represented. As an example, CIVA employs a variety of scattering models that include a geometrical theory of diffraction (GTD) approach, a Kirchhoff approximation approach, and a hybrid approach to modeling defect responses. The hybrid approach is known as the Physical Theory of Diffraction approach. Other methods exist, including the Uniform Theory of Diffraction and the Uniform Asymptotic Theory of Diffraction. Depending on the parameters involved, one or more of these approaches may be appropriate, or inappropriate, for a particular simulation.

The work that PNNL is conducting in order to address the NRC need is focused on the evaluation and assessment of how effective CIVA's semi-analytical engines and computational algorithms are for providing realistic outputs in the form of 1-D ultrasonic signal responses (A-scans) and resultant composite ultrasonic 2-D perspectives (B-, C- and D-scan representations). It is known that attenuation (primarily scattering), beam redirection, sound field partitioning, and a general degradation in sound field coherence have been observed and empirically measured in coarse-grained austenitic welds, and centrifugally and statically cast components. Additionally, in order for a model to yield realistic and accurate results in terms of ultrasonic flaw signal responses, the modeling tools must be capable of accurately representing crack morphology and the scattered signal responses emanating from these types of reflectors (e.g., corner, specular, and tip-diffracted responses). Within the scope of this work, PNNL will evaluate and document the strengths and limitations of CIVA by taking a multi-faceted approach to the problem. PNNL will conduct confirmatory research across various activities focused on:

- working with the software developers to better understand the algorithms
- working collaboratively with colleagues at the EPRI NDE Center through efforts defined in the NRC-EPRI MOU
- conducting parametric studies to evaluate the importance of input parameters on the accuracy of output results
- developing and implementing a set of validation metrics/criteria for reducing subjectivity in empirical and simulated data comparisons
- modeling various component-material-flaw configurations to simulate flaw signal responses and then conducting laboratory studies to compare and validate the model results
- identifying and documenting gaps where the models fail to provide effective and reliable results

With regard to the effectiveness of modeling tools to accurately simulate sound fields, ultrasonic signal responses, and the inspected material volume of interest, many parameters play a role in determining complete insonification of a defined volume within a component. These include component materials, weld configurations, surface conditions, access limitations, potential degradation mechanisms, probe types and characteristics, etc. As industry begins to rely more heavily on commercially available modeling tools for determination of inspected volumes, it will become increasingly necessary to understand the effectiveness and reliability of these tools to provide realistic and accurate results. Of critical importance is how these tools handle beam position and impingement in components where beam redirection and partitioning are prevalent. Often in CASS and dissimilar metal welds (DMWs) components, the effects of the microstructure on sound field propagation and directionality can be severe, and this can lead to gaps in inspected weld volumes where areas are not effectively insonified.

## 1.2 Problem Statement and Research Objectives

PNNL is conducting confirmatory research for the NRC for the validation of computational models for UT. The fundamental question that this report attempts to answer is: *Within their intended domain of application, do ultrasonic wave propagation models adequately represent reality?* To be able to answer this question, we should precisely know how do we define “adequate,” how much do we know about reality and how well can we observe it, and how do we specify the intended use of the model. Without concise definitions for those terms, answering this question is a Sisyphean task that will wear down even the most determined researcher, to no avail. The difficulty of such a task has led many researchers to refute the possibility of truly validating models in certain engineering fields (see, for example, Oreskes et al. 1994), and others to conclude that “All models are wrong, but some are useful” (Box and Draper 1987).

However, unlike other engineering fields, the mechanics of wave propagation in fluids and elastic solids are well understood and mathematical models have been well developed in this field. This work attempts to assess the reliability of such models by directly comparing them to empirical measurements. We propose a framework that allows assessing the model by using both qualitative and quantitative validation indicators and metrics. This framework acknowledges the presence of errors and uncertainties in all models and experiments, and the quantitative validation stage uses a statistical approach that incorporates uncertainties to infer the adequacy of the model. This framework is used to assess the adequacy of CIVA-UT version 11<sup>(a)</sup> (CEA 2014) (mentioned hereafter as CIVA) to model conventional ultrasonic NDE inspections.

In the first phase of this study, the subject of this report, simple geometrical reflectors in isotropic, fine-grained, homogeneous materials with conventional ultrasonic transducers are considered, as shown in Figure 1.1. In particular, stainless steel plates with electro-discharge machined (EDM) notch reflectors varying in length, depth, and orientation (tilt from the normal) are considered. Different types of conventional ultrasonic transducers are investigated, including different wave modes, beam angles, and frequencies.

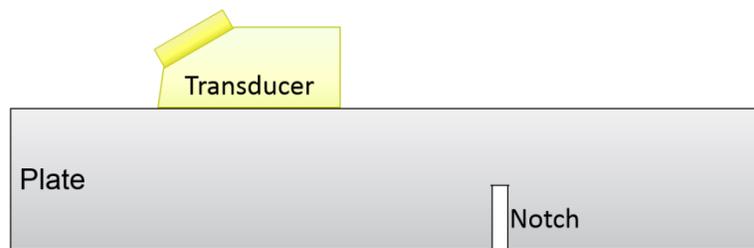


Figure 1.1. Example of UT Modeling Scenario under Consideration

## 1.3 Report Organization

The verification and validation (V&V) procedure conducted for this effort is described in this report as follows:

- A brief literature review of V&V and description of sources of errors and uncertainties in modeling is discussed in Section 2.0.

---

(a) CIVA-UT is a commercial semi-analytical simulation platform for UT developed by the French Commissariat à l’Energie Atomique (CEA).

- The computational methods most commonly used for simulating ultrasonic signal responses are discussed in Section 3.0. The reasons for selecting CIVA as the software of choice in this V&V effort are also described in Section 3.0, along with a review of the literature for experimental validation of CIVA.
- Three different stainless steel plates with a total of 29 reflectors of various lengths, widths, depths, and orientations are considered. Each flaw was scanned using six different types of transducers and from two different directions relative to the flaw orientation. This resulted in a total of 348 ultrasonic raster scans. Section 4.0 describes this validation tree, the experimental procedures, the specimens, and probes used for validation. Also, the procedures for building a model in CIVA are described, along with detailed descriptions of the methods used to estimate the input parameters for the model.
- We introduce the model validation framework in Section 5.0. We give detailed description of the assessment criteria in the qualitative analysis. Then, a probabilistic framework is proposed to quantify the error in the model. A metric is derived based on a requirement for optimizing the probability of detection in the empirical data, given coherent and non-coherent sources of noise. To compute the model uncertainty associated with the derived metric, a parametric study is conducted on all the major input parameters to the model. Then the variation in the metric as a function of the model's inputs is used to determine the simulation uncertainty.
- In Section 6.0, we apply the validation framework proposed in Section 5.0 to all the raster scans considered in this validation study. We provide detailed discussions and reasoning about the results and outcomes.
- We provide conclusions and tasks for the next phase of the study in Section 7.0.
- References cited in this report are included in Section 8.0.

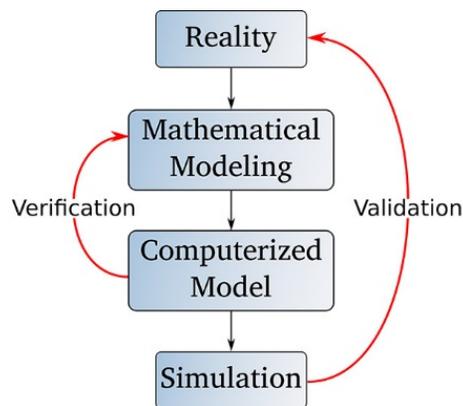


## 2.0 Review of Verification and Validation

This section briefly reviews the computer modeling process and how V&V fit in it. Because the main topic of this report is model validation, review of validation methodologies is emphasized. Particularly, we discuss sources of uncertainties that affect our judgment about model inadequacy, and review how such uncertainties are handled by the validation frameworks in the literature. The modeling process is described in Section 2.1. Section 2.2 discusses model verification while Section 2.3 explains model validation. Lastly, errors and uncertainties in the V&V process are presented in Section 2.4.

### 2.1 The Computer Modeling Process

Concepts in V&V have been defined by multiple engineering societies including the Society of Computing Simulation (SCS), the Institute of Electrical and Electronics Engineers (IEEE 2010), The American Institute of Aeronautics and Astronautics (AIAA 1998), and The American Society of Mechanical Engineers (ASME 2006). One of the earliest definitions of V&V was by the SCS where they described the relations between V&V and the computer modeling processes. This relation is demonstrated in Figure 2.1.



**Figure 2.1.** Phases in CMS and Its Relation with V&V

During a computer modeling process, a *mathematical model* is derived based on a known reality. Usually this conceptual model involves simplification of reality so that it could be conveniently represented by mathematical equations. For example, in ultrasonics, it can be the wave equation based on continuum mechanics, or an asymptotic model based on the ray theory. Most mathematical models of interest in the real world cannot be solved analytically, and require a computer to find a solution. Thus, once a *mathematical model* is developed, numerical methods are used to implement the conceptual model in a computer, and the *computerized model* is developed. Based on this CMS paradigm, verification and validation are defined by SCS (Schlesinger 1979) as:

- *Model Verification*: “substantiation that a computerized model represents a conceptual model within specified limits of accuracy.”
- *Model Validation*: “substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model.”

As was summarized by Oberkampf and Trucano (2008), an effective V&V procedure that makes a candidate software tool reliable and trustworthy should include: (1) estimation of numerical accuracy; (2) assessment of the quality of the physics models used; (3) quantification of parameter and experimental

uncertainties; (4) evaluation of the training and expertise of the users of the software; and (5) effective communication of the maturity and reliability of each of these elements. It is thus the job of V&V to determine which models are useful, and the usefulness of the same model will vary between different application areas.

## 2.2 Model Verification

Model verification is concerned with the computer implementation of an intended conceptual model to ensure its correctness. Verification is not concerned with comparing simulations with results from experiments. Model verification is further divided into two types (Oberkampf and Trucano 2002)—code verification and solution verification. Code verification deals with verifying that the written computer code gives the correct solution as intended by the mathematical conceptual model. This goes into the category of quality assurance of computer codes to ensure that the code is bug-free. However, it should be noted at this point that no code is ever bug-free, and there will always be instances where a computer code will fail. Identifying coding bugs can prove to be particularly elusive when a computer program does not fail nor show an error due to a coding bug. A study about inconsistencies and bugs in scientific computing codes written in C and FORTRAN was conducted by Hatton (1997). Solution verification deals with quantifying the errors due to the underlying discretization algorithm, errors due to numerical integration, errors due to the resolution of the geometry mesh, and convergence of the solution. In general, solution verification is conducted by comparing the model solution to other highly accurate benchmark solutions, or to known correct analytical solutions. If such solutions are not available, which is commonly the case for models of highly complex systems, the method of manufactured solutions has been proposed (Salari and Knupp 2000). This relies on the idea that to verify a computer code, the solution does not necessarily need to have any physical meaning, and hence enables the design of simple “manufactured” solutions for testing.

## 2.3 Model Validation

On the other hand, model validation is concerned with comparing the model results with empirical data to infer their accuracy. In the V&V process, the model code and solution should first be verified before embarking into the validation procedures. **A key concept in model validation is that empirical data will provide accumulated evidence to support confidence in the model output; however, empirical data cannot prove that the model is correct.** In the following, we break down the validation process into three main activities, as was described in Oberkampf and Trucano (2008):

- *Using experimental measurement to quantify the accuracy of a computational model:* To quantify the accuracy of a numerical model, some metric should be used for the comparison with empirical measurements. However, an informal analysis of the model’s output is important before computing the quantitative metric in a formal process (Bayarri et al. 2007b). The informal analysis allows rapid examination of the model output by a subject matter expert (SME) to ensure that it follows expected patterns. In the formal analysis, a validation metric will be based on the model’s output of interest, as required by its intended use. This metric will be used to compute the difference (or distance) between the model’s prediction and empirical observations. Due to the presence of multiple sources of uncertainties in both the model’s prediction and the empirical measurements, the metric is not deterministic, and it is represented by some range (or for example, as a probability density function). In Section 2.4, we will discuss sources of uncertainties and methods for handling them.
- *Determine if the estimated accuracy of the computational model satisfies the accuracy requirements specified.* Comparing a model’s prediction and empirical observations will inevitably result in an error measure. It is then up to the SME to select an appropriate threshold below which the model is

deemed unacceptable. Also, multiple validation metrics might be used for comparison, each resulting in a different error measure. Then, it would be required to combine all the metrics into a single quantity to be able to make an informed decision about the model accuracy. **At this point, it is very important to note that this metric cannot identify whether the simulation or the experiment is at fault. It can only tell the *difference* between them.** Thus, if the metric is deemed unacceptable, then there are two possible courses of action, which are up to the experience and discretion of the SME—(1) go back to the model and modify the input parameters until acceptable accuracy is obtained, or if that is not sufficient, then the conceptual model might be insufficient for modeling the physics of interest; (2) go back and review the experimental results, minimizing experimental uncertainties, in order to obtain more optimal responses. For example, this can be done by ensuring that the transducers are operating according to specification, detecting faulty equipment, or any other oversights that were not considered in the previous experimental measurements.

- *Interpolation and extrapolation of the validation space to include conditions corresponding to the intended use of the model.* In a validation effort, it is not possible to conduct experiments and simulation over the entire spectrum of domain conditions where the model is intended to be used. It is thus imperative to understand and possibly develop methods on how the validation results extrapolate to conditions outside the considered validation space, or interpolate to situations within the validation space, but not considered specifically in the validation.

## 2.4 Errors and Uncertainties

As we have touched upon previously in this section, there are multiple reasons for why a computerized model would not adequately represent reality. Sometimes it is not solely due to the model inadequacy, but also due to uncertainties within the V&V process itself, such as inherent variations in the physical properties of the system. Sources of errors and uncertainties are classified into five different categories; see, for example, Kennedy and O'Hagan (2001) and Oberkampf and Roy (2010):

- *Parameter uncertainty:* Parameters used as inputs for the model are not always known. Many times, these parameters are based on educated guess by the modeler due to unavailability of specifications. Moreover, even if it is possible to measure a parameter, it might vary between different experiments, and the parameter measurement itself has an uncertainty associated with it. In order to account for uncertainties in input parameters, stochastic methods such as the Monte-Carlo method and Latin Hypercube Sampling method can be used to propagate this uncertainty through the model to compute the probability distribution function of the model response (Oberkampf et al. 2002; Oberkampf and Barone 2006; Oberkampf and Trucano 2008; Sankararaman and Mahadevan 2015). Sometimes, the probability distribution of the input parameters is not known. In such cases, theories such as the possibility theory and fuzzy sets can be applied as methods for propagating input uncertainties through the computational model (Oberkampf et al. 2004).
- *Experimental uncertainties:* The quality of the experimental measurements can vary. Measurements have an inherent uncertainty due to various noise sources from instrumentation, or due to other sources of unacknowledged uncertainties such as variation of transducer coupling between multiple experiments. Although such variations are known to exist, they might be difficult to quantify in practice.
- *Model inadequacy:* Semi-analytical methods are based on assumptions that result in a mathematically convenient approximation of the wave equation. If those assumptions are not satisfied by the underlying physical problems, then inaccuracies will result. In general, even when the input parameters are known with absolute certainty, the model will exhibit values that deviate from the true results.

- *Residual variability*: The model inputs are not sufficiently detailed to represent the problem at hand. For example, heterogeneous materials are generally modeled as material with varying material properties, but this might not be an accurate representation of the true material properties. Another example is modeling a fatigue crack with complex branching morphology that cannot be represented exactly in the model.
- *Coding and numerical errors*: This can be from specific bugs in the programs that generate false results, without the program explicitly failing. In addition, numerical methods used to solve mathematical models in a computer always have a level of error associated with them.

Many validation frameworks have been proposed to handle these sources of errors and uncertainties. Rebba and Mahadevan (2008) investigated the use of hypothesis testing for accepting or rejecting a model. Hypothesis testing is a statistical method, which takes sources of uncertainty into account by representing the validation metric as a probability density function. Making a decision using a hypothesis test requires the specification of some level of significance (or a confidence threshold). Oberkampf and Barone (2006) proposed a methodology for defining confidence in the estimated mean error between model prediction and empirical observations. This method was based on the assumption that multiple empirical observations are available and a single model simulation. Thus, parameter uncertainty is not handled in this method.

A complete validation framework that could handle and identify sources of errors and uncertainties was proposed by Bayarri et al. (2007b). Uncertainties are handled by representing both the model predictions and empirical observations as Gaussian processes and using Bayesian methods to infer the uncertainty in the validation metric. This method has been extended also to handle multi-dimensional metrics (Bayarri et al. 2007a; Higdon et al. 2008).

## 3.0 Review of Mathematical Models for UT

Verifying and validating computer models requires a firm understanding of the underlying computational models and numerical methods. Thus, before delving into the procedures of V&V, it is worthwhile to take a look at the relevant models for ultrasound wave propagation and scattering. Describing and analyzing the intrinsic limitations, assumptions, and approximations of those theoretical models allows us to define their domain of applicability, and the scenarios where they are not expected to give the correct results. Also, this provides some confidence in our choice of the models and computer simulation software to be validated in this study.

Models in scientific computing must balance between two contending factors—accuracy and speed. Prioritizing one over the other is a function of the application and the amount of computer power available. If speed is not an issue, a model can be made arbitrarily complex to make it as close to reality as possible. Also, the opposite is true—if accuracy is not an issue, a model can be made arbitrarily fast! The accuracy and speed of a computer program depends on the complexity of the conceptual model and also on the numerical method used to discretize the conceptual model.

In this report, we are interested in the computational models used for wave propagation and scattering in nuclear power plant (NPP) components. NPP components of interest are often complex, involving welded regions and materials such as CASS, which exhibits both anisotropic and heterogeneous behavior, depending on grain size and orientation. In Section 3.1, we briefly introduce the elastodynamic equation and the numerical methods used to implement it on a computer. Then, we describe approximations of the elastodynamic equation in Section 3.2. Those approximations sacrifice accuracy for speed, and we argue that for UT applications in NPPs, direct solutions of the elastodynamic equation might be unfeasible, and it is necessary to use such approximations. In Section 3.3, we discuss the computational methods used by CIVA and review the literature related to the validation of CIVA. In Section 3.4, we compare the Kirchhoff approximate model with an exact solution of the wave equation. In Section 3.5, we compare CIVA with Wave3000, which is a finite difference time domain (FDTD) software, in terms of results and speed.

### 3.1 General Wave Physics

There is extensive literature (Auld 1973; Graff 1975; Červený 2001; Rose 2004) describing wave propagation in solids. This section summarizes only the aspect relevant to our goal. Consider a perfectly elastic anisotropic and heterogeneous solid in a Cartesian coordinate system. The wave equations in terms of particle displacement in three-dimensional space are described as:

$$(c_{ijkl}u_{k,l})_{,j} + f_i = \rho\ddot{u}_i, \quad i = 1,2,3 \quad (3.1)$$

where,

- $f_i$  = the body force in the  $i^{\text{th}}$  direction,
- $u_i$  = the particle displacement in the  $i^{\text{th}}$  direction,
- $\ddot{u}_i$  = the second time derivate, or particle acceleration,
- $\rho$  = the material density, and
- $c_{ijkl}$  = the stiffness tensor.

Equation (3.1) represents a system of three coupled partial differential equations of the second order for the three Cartesian components of the displacement vector  $\mathbf{u}$ . The standard tensor notation is used throughout: A letter subscript after a comma indicates the derivative with respect to space, and Einstein summation convention is implied (Auld 1973). This equation takes into account modeling heterogeneous materials by allowing the stiffness tensor  $c_{ijkl}$  to change as a function of position. Moreover, viscoelastic/attenuative materials can be represented by allowing the stiffness tensor  $c_{ijkl}$  to have complex quantities.

Analytical solutions to Eq. (3.1) can be found only for simple forcing functions  $f_i$  and in simple coordinate systems. For any geometries/transducers of significant interest, standard numerical methods such as the finite element method (FEM) or finite difference method (FDM) are required to solve this equation. For modeling heterogeneous materials whose mechanical properties change rapidly on the microscopic scale, the finite element heterogeneous multiscale model has been proposed (Stohrer 2013). Although such numerical methods can model very complicated geometries and materials, they are computationally intensive. This is because such methods require one to discretize (or mesh) the geometry into small parts called elements. Selecting the proper mesh size and time step size is important for an accurate solution, or even for converging to a solution. The mesh resolution requirements generally depends on the type of elements being used, but a general rule of thumb is 20 elements per wavelength. Thus as the wavelength becomes smaller (frequency increases), the model requires more elements, which increases the required computation time. For a typical size of an NPP component and UT inspection frequencies typically above 500 kHz, such methods might become unfeasible or prohibitively intensive.

## 3.2 Asymptotic Approximations

To overcome problems inherent with such numerical methods, the ultrasound NDE community has leaned on approximate methods for their modeling needs. Approximate methods can be classified into two types—(1) methods for solving ultrasound wave propagation from a given source and (2) methods for solving ultrasound scatter from a discontinuity due to an incident wave. CMS software can combine models from both categories to model a large range of UT inspection scenarios. This is unlike FEM/FDM methods that directly solve the wave equation on the given mesh, and do not distinguish between wave propagation, scatters, or wave modes (although it does solve for all of them).

Some examples of common approximate methods for modeling wave propagation include:

- *Ray theory*: This is based on the high-frequency approximation of the wave equation, which yields the Eikonal equation for controlling the travel-time (phase) function and the Transport equation that controls the amplitude function (Červený 2001).
- *Paraxial approximation*: This is based on approximating the Rayleigh-Sommerfeld equation, which solves for the wave propagation in a material given a transducer pressure source (Schmerr Jr. 1998). The observation point is assumed far enough from the transducers so that the angle between the center axis of the transducer and its edge is small (Schmerr Jr. 1998).
- *Multi-Gaussian beam*: This is based on concepts from optics, where the wave is approximated as multiple super-imposed Gaussian-shaped rays.

For modeling wave scatter from discontinuities, some of the models used in the literature include:

- *Kirchhoff approximation*: By assuming a plane wave incident on the discontinuity surface, the discontinuity surface is split into “lit” and “unlit” surfaces. Then, the wave scattering equation is only computed on the “lit” surface, which faces the direction of incidence. This approximation is used to

compute scatter from volumetric flaws, and it has been extended to compute scatter from cracks (non-volumetric/ plane discontinuity) (Schmerr Jr. and Song 2007). The approximation accuracy decreases as the smallest dimension of the discontinuity becomes smaller than the wavelength. A second limitation of the Kirchhoff approximation is that both the transmitter and receiver should be on the same half space defined by the axis along the length of the discontinuity. The model becomes less accurate for angles of observation different from the specular direction. A third limitation of the Kirchhoff model is that the amplitudes of tip-diffraction echoes from crack-like discontinuities cannot be predicted accurately for diffraction directions different from the specular direction.

- *Geometric theory of diffraction*: The GTD model is also a high-frequency approximation that computes tip-diffraction echoes from crack-like discontinuities, but cannot compute specular echoes. The GTD model also becomes less accurate as the smallest dimension of the reflector becomes smaller than the wavelength.
- *Born approximation*: This is a low-frequency, weak-scattering approximation, which requires that the material within the discontinuity be nearly the same as the base material.

### 3.3 Computational Models in CIVA

CIVA (Calmon et al. 2006) is a semi-analytical simulation platform for UT developed by the French Commissariat à l’Energie Atomique (CEA). CIVA is a widely used simulation software packages used in the commercial nuclear NDE sector, particularly for simulating inservice inspection (ISI) scenarios in NPPs. The main appeal of CIVA is that its user interface is designed specifically for applications in UT, providing predefined geometries, transducers, and scanning options commonly applied in UT for ISI of nuclear components. CIVA is unlike other ultrasound simulation tools whose applications are more general, and thus modeling standard geometries and transducers for UT might be more involved. Another advantage of CIVA is its fast computational speed compared with FEM and FDTD software packages, at the expense of accuracy. CIVA will be the main platform used in this report to assess computational methods.

CIVA incorporates algorithms to model both the transducer radiation and scatter from flaws (Darmon and Chatillon 2013). For modeling transducer radiation, CIVA uses an implementation of the ray theory (Deschamps 1972). For computing scatter from flaws, it includes a multitude of models for covering different scenarios. For example, different models are used to calculate the responses from inclusions than those used to calculate responses from cracks. In this report, we will focus on the case of a homogeneous and isotropic material, with the wave incident on a crack-like reflector embedded in the material, as shown in Figure 1.1. Because Kirchhoff and GTD complement each other in terms of their modeled reflected wave modality (specular and tip-diffracted), CIVA implemented a hybrid method to compute both specular and tip-diffracted echoes from crack-like reflectors.

The developers of CIVA have conducted an extensive amount of validation on different geometries, reflectors, and transducer configurations (CIVA 2015) (Meyer et al. 2012). CEA has created a laboratory specifically for the purpose of validating simulation results with experimental measurements. Several validation studies have been published involving different types of reflectors: side-drilled holes (SDH), flat-bottom holes (FBH), and corner echoes from notches. For these flaw types, various types of probes were used for testing using both longitudinal and shear wave modes, different refraction angles, transducer diameters, frequencies, water path lengths, and flaw depths. Also some studies were made for pitch/catch configurations and phased-array probes. Validations compared the response amplitudes from reflectors with varying depths and areas.

For the modeling of SDHs and FBHs, the results showed that simulations were consistent with experiment, with less than 2 dB variations in amplitudes. Simulating such reflectors in CIVA generally uses a full analytical method, separation of variables (SOV) (Schmerr Jr. 1998), to compute the wave scatter. When simulating notch reflectors, which uses the Kirchhoff approximation method to compute reflector scatter, more significant differences were found between experimental and simulation results, especially for small reflectors with size comparable to the wavelength. These simulations only considered machined reflectors (notches) in isotropic and homogeneous material.

Many other researchers have contributed to the validation of the algorithms used in CIVA. For example, to verify the validity of the Kirchhoff model in CIVA to compute specular echoes, Raillon et al. (2012) modeled a steel mockup plate containing back-wall notches and compared the simulation results with experiments.

In Raillon et al. (2013), the GTD model was verified using a time-of-flight diffraction (TOFD) setup with 45° and 60° longitudinal wave probes. Notch reflectors were evaluated and the amplitudes were compared with experiment for varying distances between the receiving and transmitting probes. They also investigated the effects of modeling the notch aperture (notch width) on the results. It was found that for smaller distances between transmitter and receiver, the models not accounting for reflector aperture had a large deviation from the experiment. In Darmon et al. (2014), the TOFD was simulated in CIVA and verified for various types of reflectors and wave reflections.

The simulation of more realistic reflector geometries and inhomogeneous materials was also investigated. In Dupond et al. (2009), the simulation of multi-faceted and branched reflectors was compared with experimental results using B-scans and signal amplitudes using three different modeling software packages—CIVA, Athena 2D (an FEM software), and a hybrid between CIVA and Athena 2D. Mahaut et al. (2006) conducted CIVA validations for the simulation of notches using corner echoes and immersion probes. They also considered realistic fatigue crack morphologies. The modeling of backscattered noise for reflector detection in DMWs was also investigated. The B-scan images were primarily used to compare experimental and simulation results. In Roth et al. (2009) modeling an aircraft skin-to-flange welded component was investigated. A 32-element phased-array probe was used to scan the component and detect a SDH in the weld, which simulated imperfections in the welding area. Experimental measurements were compared to the modeling results using B-scan images. In Gardahaut et al. (2014), DMWs were modeled as anisotropic and inhomogeneous media for simulating reflections from notches within the weld area. The results were compared to experiments. In Chassignole et al. (2011) austenitic stainless steel welds were simulated using both Athena2D and CIVA semi-analytical models. To model the inhomogeneous properties of the weld, the weld area was modeled as a finite number of anisotropic but homogeneous zones. Scattering from a corner echo within the weld was simulated and compared to experimental results.

On a final note, although there exists a reasonable amount of literature attempting to compare the results of simulations from CIVA with experimentally obtained results, the conclusions from such studies are rather vague, especially as the geometry becomes more complex. This only proves the need of a systematic and well-defined method for V&V in UT. Moreover, none of these studies attempted to quantify the numerical errors of CIVA, uncertainties in the input parameters or empirical uncertainty.

### **3.4 Performance of Kirchhoff Approximation**

In our verification and validation work, we are concerned mostly with crack-like flaws, thus the Kirchhoff approximation is the most relevant model for specular echo scattering. It is known that Kirchhoff is based

on high-frequency approximation of the plane wave equation. In this section, we demonstrate how accurate the Kirchhoff model is when compared with the exact solution of the wave equation.

An exact solution for the wave equation can be obtained only for canonical simple volumetric discontinuities, and thus here we will compare the responses from a SDH.

Consider a pulse-echo setup for inspecting a cylindrical SDH of radius  $b$  and length  $L$ . We are interested in the case when the incident wave has the opposite direction as the reflected wave due to normal incidence. Also, the wave direction is in a plane perpendicular to the axis of the cylinder.

The wave complex amplitude solution using Kirchhoff approximation can be found to be (Schmerr Jr. and Song 2007):

$$A_{Kirchhoff}(k) = \frac{kbL}{2} [J_1(2kb) - iS_1(2kb)] + \frac{ikbL}{\pi} \quad (3.2)$$

where

- $A_{Kirchhoff}(\cdot)$  = the scattering coefficient of the SDH
- $k$  = the wavenumber of the wave-mode under consideration
- $J_1$  = the Bessel function of order one
- $S_1$  = Struve function of order one
- $i$  = the imaginary number  $\sqrt{-1}$

On the other hand, the exact scattering function can be obtained from the wave equation using SOV. The following computed the scattering amplitude for longitudinal waves:

$$A_{SOV}(k) = \frac{iL}{2\pi} \sum_{n=0}^{\infty} (2 - \delta_{0n}) (-1)^n F_n \quad (3.3)$$

where,

$$F_n = 1 + \frac{C_n^{(2)}(k_p b) C_n^{(1)}(k_s b) - D_n^{(2)}(k_p b) D_n^{(1)}(k_s b)}{C_n^{(1)}(k_p b) C_n^{(1)}(k_s b) - D_n^{(1)}(k_p b) D_n^{(1)}(k_s b)}$$

$$C_n^{(i)}(x) = (n^2 + n - (k_s / b)^2 / 2) H_n^{(i)}(x) - (2n H_n^{(i)}(x) - x H_{n+1}^{(i)}(x))$$

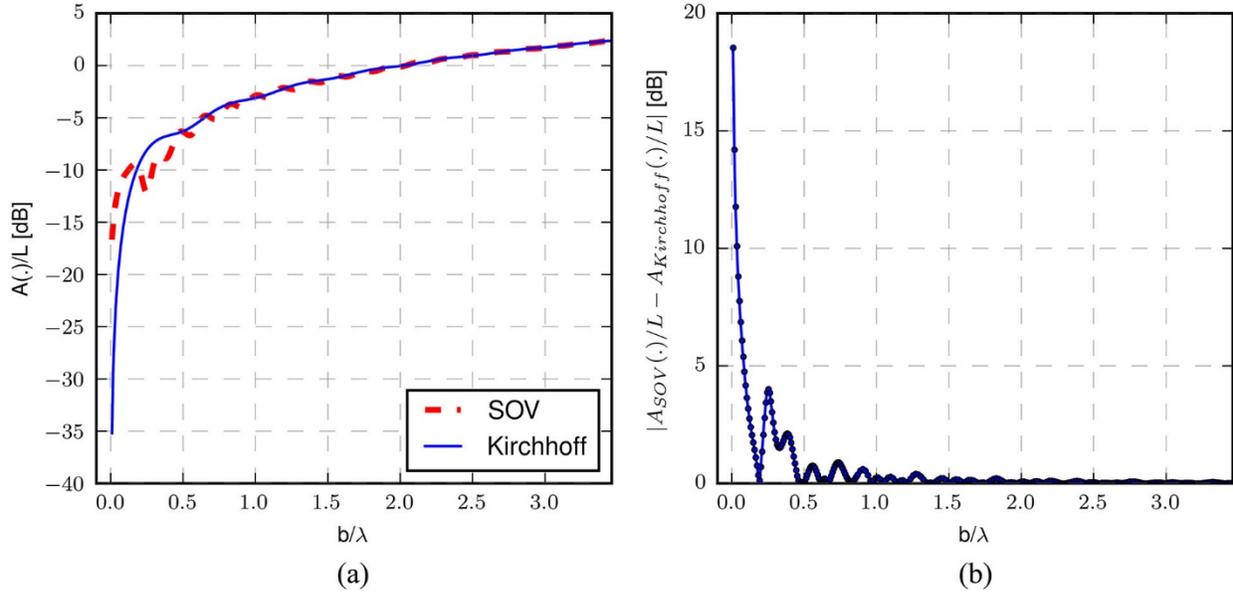
$$D_n^{(i)}(x) = n(n+1) H_n^{(i)}(x) - n(2n H_n^{(i)}(x) - x H_{n+1}^{(i)}(x))$$

and,

- $k_p$  = the wave number for the incident longitudinal wave
- $k_s$  = the wave number of the shear waves in the base material
- $H_n^{(i)}$  = the Hankel function of the  $i^{th}$  order
- $\delta_{0n}$  = the Kronecker delta function

The solutions for Eqs. (3.2) and (3.3) are plotted in Figure 3.1 for a 10 mm (0.39 in.) radius SDH as a function of the radius normalized by the wavelength. The amplitudes shown are normalized by the length  $L$  of the SDH, and we assume that  $L \gg b$ . Figure 3.1(b) shows the difference between the two methods

in decibels (dB). It can be seen that as the wavelength becomes larger than the radius (i.e.,  $b/\lambda < 1$ ), the error increases, with differences up to 18 dB. As the wavelength decreases (or frequency increases), the error oscillates up to where  $b/\lambda = 2$ . Then, at higher frequencies, the two methods give very close results, with difference near 0.



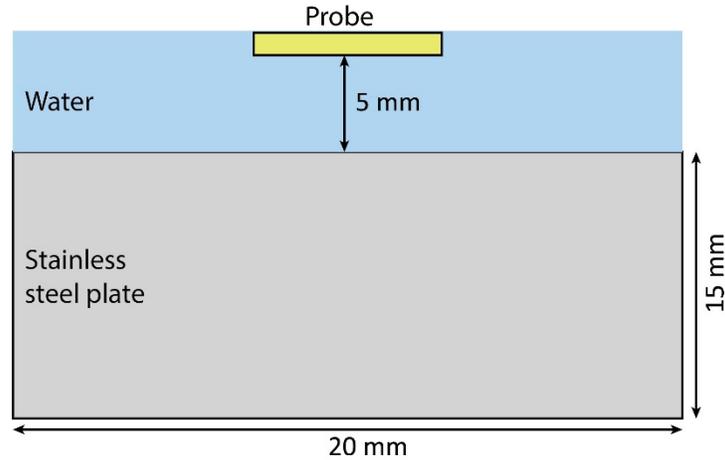
**Figure 3.1.** (a) Comparison between the Exact SOV Method and Kirchhoff Approximation. This is the amplitude response as a function of frequency for a 10 mm (0.39 in.) radius SDH. (b) Amplitude difference between separation of variables and Kirchhoff approximation.

Although this example is specifically for a side-drilled hole, the errors will be similar when modeling crack-like flaws. In such a case, the relevant dimension that determines accuracy will be the height of the flaw relative to the wavelength, assuming that the length is much larger than the height. Moreover, Eqs. (3.2) and (3.3) only represent the scatter coefficient of the flaw. Also, approximation errors will occur in the model for the ultrasonic beam propagation. The analysis for this beam propagation errors is not shown here, and the interested reader is referred to Červený (2001) for detailed analysis.

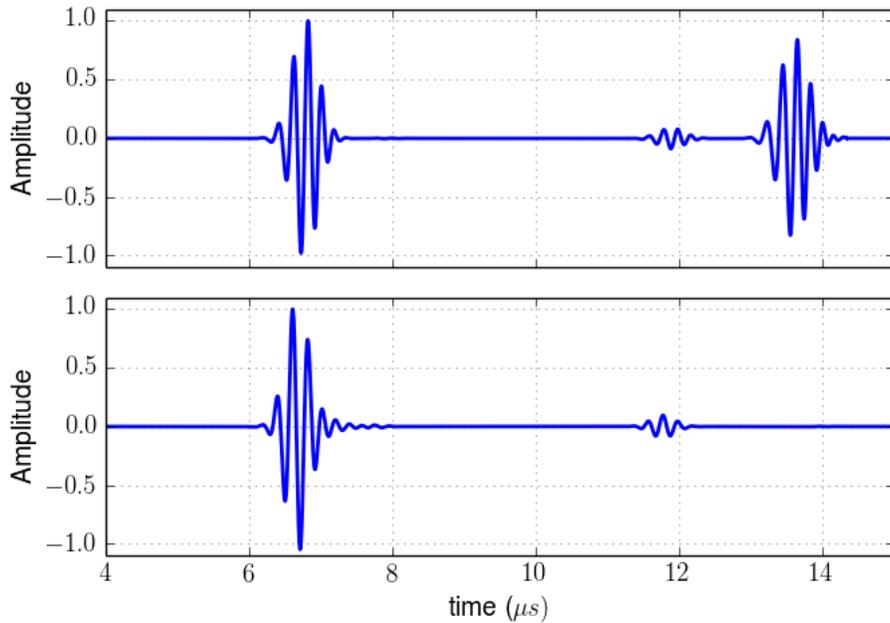
### 3.5 Comparison of CIVA and Wave3000+

To understand the superiority of CIVA in terms of computation speed, a comparison was made with another software package—Wave3000+, developed by Cyberlogic Inc. Wave3000+ is an ultrasound modeling software tool based on the FDTD method. The geometry modeled using both software packages is shown in Figure 3.2. All simulation parameters were identical for both modeling software packages. The plate material was specified to be stainless steel. A single pulse-echo immersion probe was modeled with a 6 mm (0.24 in.) diameter and a 5 mm (0.2 in.) water path. The probe’s surface was oriented parallel to the plate surface and generated a normal beam longitudinal wave. The excitation signal was a Gaussian modulated pulse with a 5.0 MHz center frequency and 50% bandwidth at  $-6$  dB. For the CIVA simulation, it was specified that the signal responses (echoes) be computed from both top and bottom surfaces of the plate. In Wave3000+, this step is not required because it directly solves the wave equation, and hence all interfaces and surfaces are accounted for. The computed responses from both software packages are shown in Figure 3.3. It can be seen that the responses are similar. Both software packages computed the first echo, which corresponds to the reflection from the top surface of the plate and the second echo corresponding to the reflection from the bottom surface of the plate. However, CIVA did not

compute the third echo, which is the second reflection from the top surface of the plate, because it was not instructed to do so. This echo automatically shows up in the Wave3000+ simulation. Also note that although the exact same excitation signal was provided as the source for both software packages, there are slight differences in the echo shapes. This is due to the differences between the two software packages and how the excitation signal is applied to the geometry boundary.



**Figure 3.2.** Geometry Used for the Models in CIVA and Wave3000+. The actual geometry is a 3D geometry and the width of the plate (not shown in figure) is 20 mm (0.79 in.).



**Figure 3.3.** Comparison of A-scan Results for Two Different Simulation Packages: Wave3000+ (*top*) and CIVA (*bottom*)

Although it has been shown that both software packages are able to produce similar results, the main difference is the execution time:

- *Wave3000+*: Required 1435 minutes (nearly 24 hours) using 12 cores and 47 GB of memory to compute the results. It was run on dual Intel Xeon X5680 processors and with 96 GB of memory.
- *CIVA*: Required 0.3 minutes (18 seconds) and less than 1 GB of memory to compute the same result! Also, it was run on a slightly slower processor, dual Intel Xeon E6-2630 processors with 64 GB of memory.

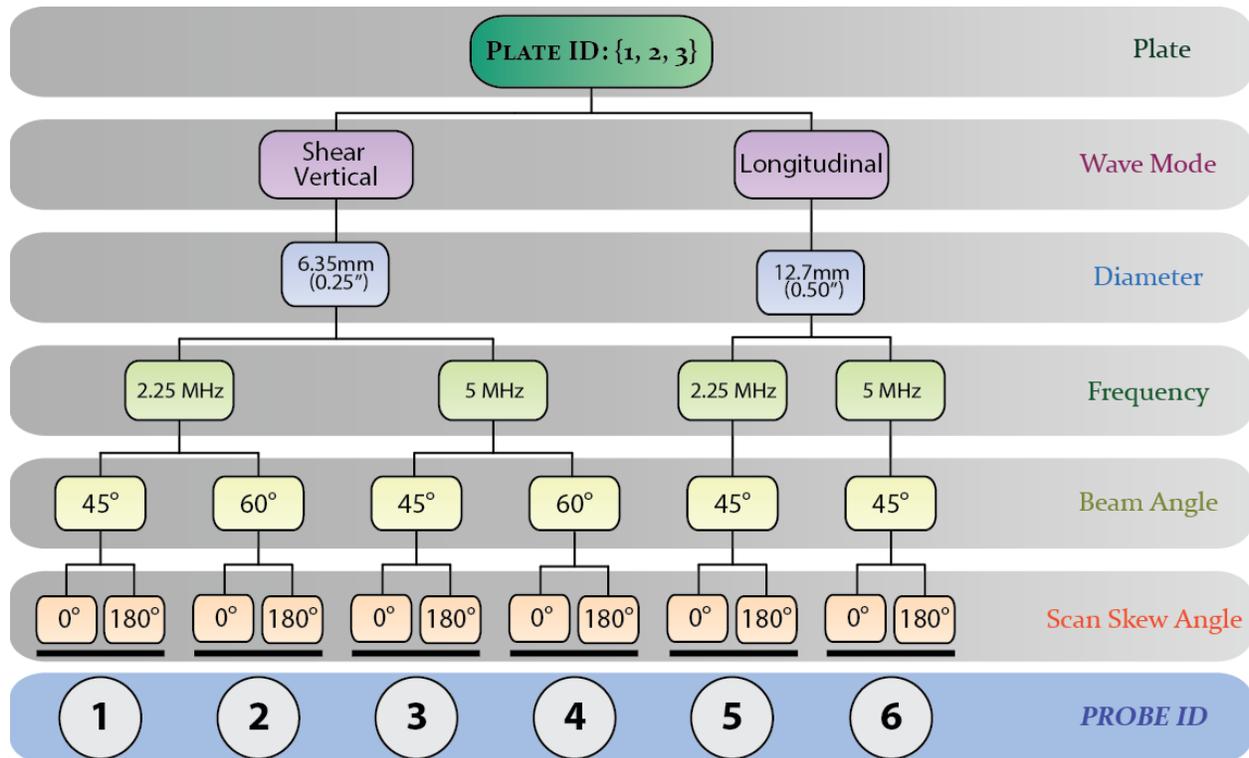
Thus, the Wave3000+ software required 4783 times more computation time than CIVA. It is important to note here that the running times largely depend on the mesh size and time step for the Wave3000+, and on the accuracy level parameter for CIVA (see Section 4.3.5). In any case, Wave3000+ requires a much longer computation time, which makes it impractical to use for simulating UT problems typically encountered in commercial nuclear ISI scenarios. Particularly in UT, it is of interest to model C-scans using a raster scan data acquisition protocol. For example, consider a 50×50 point raster scan. Using the same model described in Figure 3.2, Wave3000+ would require nearly 7 years to compute the C-scan, compared with 12.5 hours using CIVA. For these reasons, this work focuses on validating the results from CIVA using experimental measurements.

## 4.0 Measurement and Modeling Procedures

The UT scenario that we consider for validation has been briefly described in Section 1.0. In Section 4.1, we describe the validation tree containing all the materials and probes considered in this validation study. The procedures used for collecting the empirical data are given in Section 4.2. Then, in Section 4.3 we introduce how to specify a model in CIVA, and all the relevant inputs required to define such models. For the inputs that require knowledge of their empirical counterparts, we provide detailed description of the methods used for measuring them. Also, as we have established in Section 2.0, uncertainties in specifying model inputs will contribute to uncertainties in its output. Minimizing uncertainty in the model inputs will increase the confidence in the model output. To quantify such uncertainty, we associate the tolerance level for each of the measured input parameters.

### 4.1 Specimen and Probes

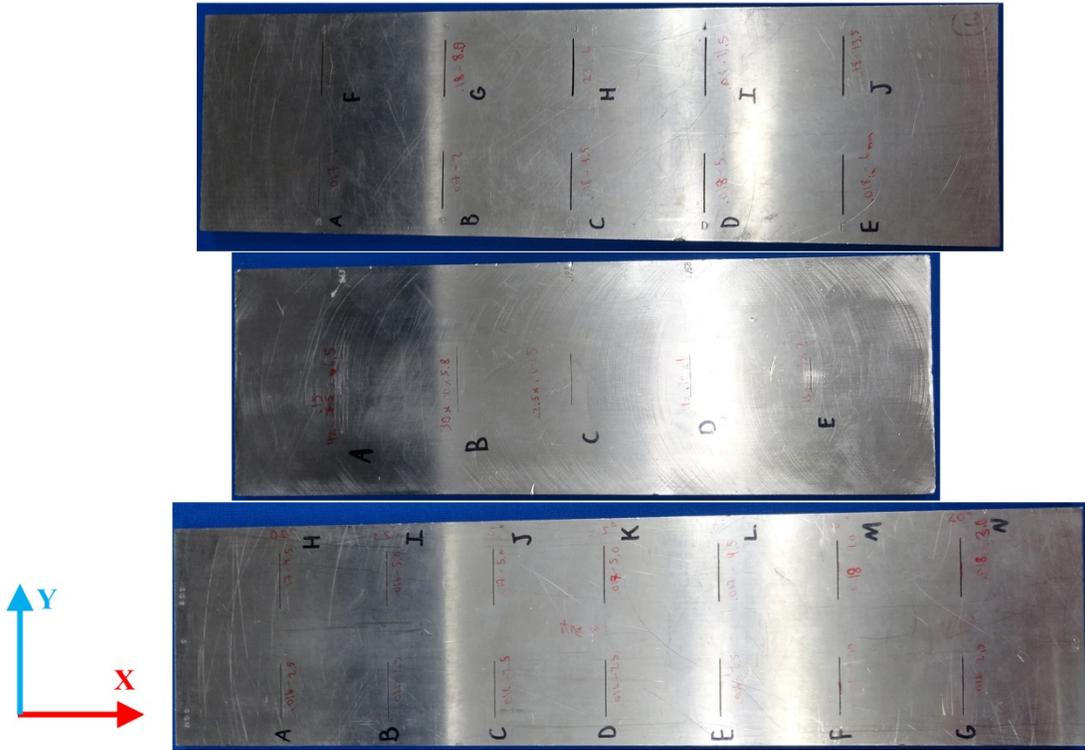
The validation tree describing the specimen, probes, and scans is given in Figure 4.1.



**Figure 4.1.** Validation Tree Describing the Probes and Scans Directions Used for Each Specimen

The specimens considered are three rectangular 304 grade stainless steel plates, as shown in Figure 4.2, each with a different number of crack-like reflectors varying in length, depth, and tilt as follows:

- *Plate 1:* Contains 10 rectangular back-wall EDM reflectors varying in depth with 0° tilt.
- *Plate 2:* Contains 5 semi-elliptical back-wall saw-cut reflectors varying in depth and length with 0° tilt.
- *Plate 3:* Contains 14 rectangular back-wall EDM reflectors varying in both depth and tilt.



**Figure 4.2.** Stainless Steel Plates with Reflectors Used for the Validation Study. Specimen 1 (*top*), specimen 2 (*middle*), and specimen 3 (*bottom*).

The length of the reflectors in all specimens is along the Y-direction as defined in Figure 4.2. The reflector tilt describes its angle relative to the Z-direction, as defined in Figure 4.3. The convention for defining flaw depth for tilted flaws is also shown in Figure 4.3. Note that because all specimens have crack-like reflectors, their widths are much smaller than their other dimensions, and the effect of width was not considered in the validation study.

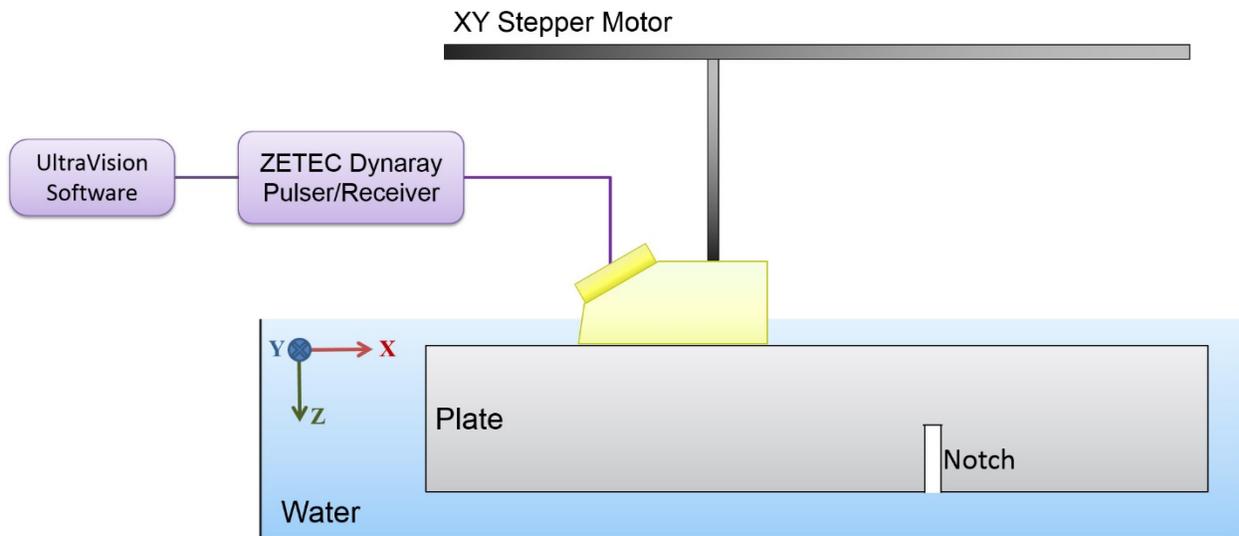


**Figure 4.3.** Defining the Notch Reflector Tilt ( $\theta$ ) and Depth within the Global Coordinate System of the Specimen

Each specimen was scanned by six different conventional circular probes along two different directions, as described by the “scan skew angle” field in Figure 4.1. We selected six probes from combinations of two frequencies, two beam angles, two diameters, and two wave modes. The probes consist of both the transducer containing the piezoelectric element and the wedge for introducing an angled beam (refracted) through the specimen. All the probes have single piezoelectric element transducers are used in a pulse-echo mode. The validation tree in Figure 4.1 provides the manufacturer’s specifications for the diameter, frequency, refracted beam angle, and wave mode. However, the actual probe properties are likely to vary from the given specifications due to typical variations in manufacturing tolerances.

## 4.2 Empirical Data Collection

A simplified schematic of the setup that we used to collect empirical data is shown in Figure 4.4. We collected data as 2-D raster scans of the plate surface (along the X-Y plane) and saved full radio-frequency (RF) waveforms at each scan point. All the raster scans were acquired in the pulse-echo inspection mode. The probe was water coupled to the specimen, providing a smooth and consistent coupling medium. We used a scan step size (along X-axis) of 0.5 mm (0.02 in.) and index step size (along Y-axis) of 1 mm (0.04 in.). Each raster scan included the responses from all the reflectors in the specimen, as well as the specimen's forward edge in the X-direction. The response from the specimen edge is required so that we can normalize the voltage amplitudes of the flaws with respect to the edge response. For each specimen and probe combination, we collected scans with the probe oriented along both the positive X-direction (denoted as a 0° scan skew angle) and also along the negative X-direction (180° scan skew angle). This allows us to insonify both surfaces of the reflectors and is particularly important for tilted notches.



**Figure 4.4.** Experimental Setup for Obtaining the UT Raster Scans

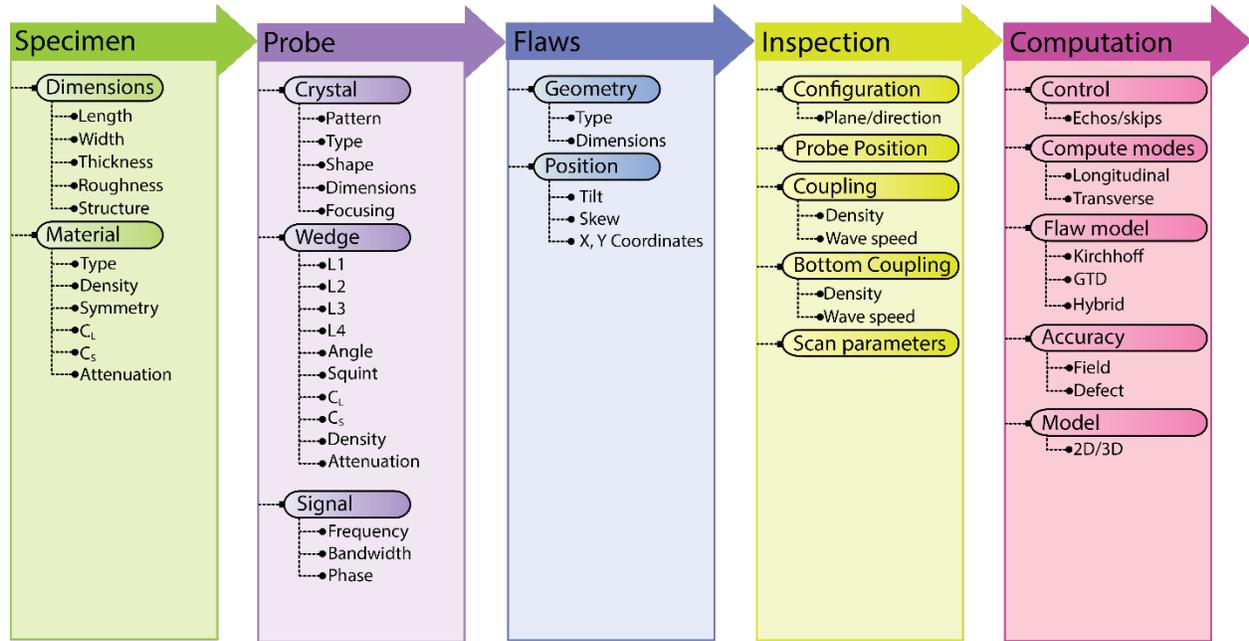
As for the pulser/receiver and scanning equipment, we used the Zetec DYNARAY data collection system interfaced via a Zetec Motor Control Drive Unit, MCDU-02 (or ZMC2), to an ATCO GPS-1000 scanner. A computer running UltraVision software (version 3.6R5) controlled the motor, pulser/receiver, and data acquisition. We modified the pulser excitation pulse width depending on the nominal frequency of the probe such that the width was half the period [ $1 / (\text{frequency} \times 2)$ ]. All data were digitized using a 16-bit data acquisition card at a sampling frequency of 100 MHz.

## 4.3 Setting up Models in CIVA

All the simulations in this validation study are based on models in CIVA. For all the raster scans that we collected empirically (henceforth called **observations**), we attempt to replicate them using simulations (henceforth called **predictions**). To achieve this, all the initial/boundary conditions and parameters from the observations should be known in order to obtain accurate predictions.

CIVA's graphical user interface is designed with an intuitive workflow for modeling NDE inspections. The general workflow in CIVA for simulating an isotropic homogeneous plate with a back-wall EDM

notch reflector is shown in Figure 4.5. The model requires defining the specimen, probe, reflectors, inspection parameters, and computation parameters.



**Figure 4.5.** General Workflow to Define a Model of Homogeneous Isotropic Plates with Notches under Examination Using Conventional Transducers

### 4.3.1 Specimen Specification

The specimens are defined by specifying their dimensions and material mechanical properties. All the specimens that we consider in this study are rectangular plates, and thus their dimensions are defined by the plate length, width, and thickness. We measured those dimensions for each of the three specimens considered in this study using a Vernier caliper. For each specimen, we made thickness measurements at six randomly selected positions to ensure that there are no variations in thickness within the same plate. The measured values for the specimen dimensions are given in Table A.1. All the dimension measurements were within 0.2 mm (0.01 in.), which is the estimated accuracy of the Vernier caliper.

Moreover, CIVA can handle rough specimen surfaces at the interface with the probe by defining a roughness parameter. For all simulations, we assumed that the specimen surface is smooth, and thus the roughness parameter was set to 0. The *structure* parameter (see Figure 4.5) allows specifying multilayered specimens. In this study, all specimens are specified as a single-layered homogeneous structure.

At the ultrasound frequencies of interest in UT (less than 10 MHz), we can safely assume the stainless steel specimens to be isotropic, homogeneous, and have negligible material attenuation. Thus, we can fully specify the mechanical properties of the specimens by their density, and ultrasonic longitudinal and shear wave velocities. We obtained the values for density from the literature, and determined it to be  $8 \text{ g/cm}^3$  with  $\pm 0.2 \text{ g/cm}^3$  uncertainty.

We measured the ultrasonic velocities in each of the specimens using two 5.0 MHz normal beam longitudinal and shear wave transducers. The transducers were operated in pulse-echo mode, and directly coupled to the plate surface using conventional ultrasonic gel for the longitudinal wave transducer and a high viscosity couplant for the shear wave transducer. The transducers were excited by a square wave

pulse, and their responses were recorded on a digital oscilloscope. We obtained the through-thickness time of flight using a cross-correlation method that computes the delay between consecutive echoes in the recorded signal response (Marioli et al. 1992). Then, we used the measured specimen thickness dimension to compute the wave velocity. For each specimen, we made the normal beam measurements at six randomly selected points in order to ensure that there is no significant material properties variation within the same specimen. A summary of the measured specimen dimensions and velocities are given in Table A.1 (Appendix A).

To obtain an estimate about the uncertainty in the measured ultrasonic velocities, we consider our estimated velocity to be:

$$\hat{c} = \frac{2\hat{d}}{\hat{t}} \quad (4.1)$$

where  $\hat{c}$  is the estimated velocity,  $\hat{d}$  is the estimated specimen thickness as described above, and  $\hat{t}$  is the estimated time of flight computed using the cross-correlation method. On the other hand, we can express the true velocity value as:

$$c = \frac{2(\hat{d} + \Delta d)}{\hat{t} + \Delta t} \quad (4.2)$$

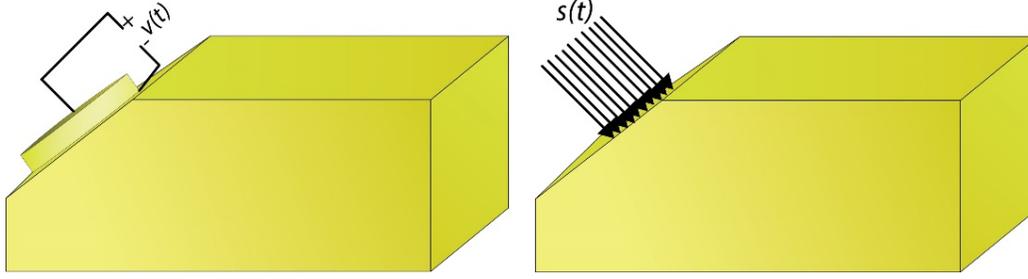
where  $\Delta d$  and  $\Delta t$  are the errors in the specimen thickness and time-of-flight measurements, respectively. Thus, the error in the velocity computation is:

$$\begin{aligned} |\hat{c} - c| &= \left| \frac{2\hat{d}}{\hat{t}} - \frac{2(\hat{d} + \Delta d)}{\hat{t} + \Delta t} \right| \\ &= \left| \frac{\hat{c}\Delta t}{\hat{t} + \Delta t} - \frac{2\Delta d}{\hat{t} + \Delta t} \right| \end{aligned} \quad (4.3)$$

Now the error in the time-of-flight computation ( $\Delta t$ ) is on the order of the sampling interval of the acquired signals, which is 1 nanosecond. The error in the thickness computation is 0.2 mm (0.01 in.), as measured earlier in this section. Thus, we know all the parameters required to determine Eq. (4.3). A summary of the uncertainties in the specimen-measured parameters is given in Table A.2, Appendix A.

### 4.3.2 Probe Specification

Specifying the probes in CIVA requires defining the piezoelectric element properties, excitation signal, and wedge properties. CIVA does not model the complete dynamics of the piezoelectric element. Rather, the piezoelectric element is replaced by an effective pressure at the surface of the wedge, as depicted in Figure 4.6. If a pulser applies a voltage  $v(t)$  to the piezoelectric transducer, CIVA models the effective pressure  $s(t)$  applied on the wedge's surface by the transducer. Because all piezoelectric transducers have a limited bandwidth, the signal  $s(t)$  will have different shape than  $v(t)$ .

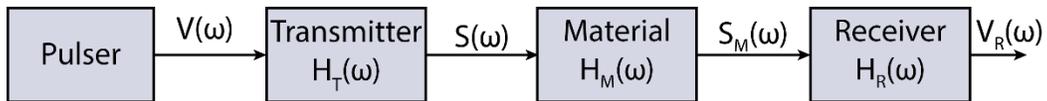


**Figure 4.6.** *Left:* Model of a Probe with a Circular Piezoelectric Element and Angle Wedge. The voltage signal  $v(t)$  is the applied pulse excitation. *Right:* CIVA's Implementation of a Piezoelectric Element, where it is Replaced by the Effective Pressure  $s(t)$  at the Wedge's Surface.

Based on this simplified transducer model in CIVA, defining a piezoelectric element requires knowledge only of its effective surface area. All the probes used in this study have circular piezoelectric elements and we used the manufacturer's specifications to obtain the diameter. We could not directly measure the element's diameter because it resides inside the probe casing, making it impossible to access without damaging the probe. Also, we assumed that the piezoelectric element surface is parallel to the wedge surface. Another required input parameter is the probe's squint angle. This is the angle between the surfaces of the piezoelectric element and wedge. If this angle is  $0^\circ$ , this represents perfect coupling between the element and wedge. This parameter was set to  $0^\circ$  for all probes, with a  $0.5^\circ$  uncertainty.

To define the excitation signal shape, we need a method to measure the signal  $s(t)$  described in Figure 4.6. Note that the signal  $v(t)$  can be easily obtained by measuring the excitation voltage from the pulser. However, measuring  $s(t)$  is more involved. To obtain  $s(t)$ , we consider a pulse-echo corner response measurement from a specimen's edge. In the following, we explain how the corner response measurements can give us information about the signal of interest,  $s(t)$ .

We model this scenario as a linear system whose block diagram is shown in Figure 4.7. To simplify matters, we represent the signals in the frequency domain using their Fourier transform. The transmitter represents the probe we are trying to characterize, as well as the power amplifier, diplexer, and cables. The transfer function of the transmitter side is given by its frequency response  $H_T(\omega)$ . The material transfer function  $H_M(\omega)$  represents the propagation of the ultrasonic beam within the material and its reflection from the specimen edge. The receiver transfer function  $H_R(\omega)$  represents the frequency response of the probe, amplifier, and cables.



**Figure 4.7.** Block Diagram of the Pulse-Echo Ultrasound System, Represented as a Linear System in the frequency Domain

Then, the received voltage  $V_R(\omega)$  signal due to an excitation pulse with a Fourier transform  $V(\omega)$  is:

$$V_R(\omega) = H_T(\omega)H_M(\omega)H_R(\omega)V(\omega) \quad (4.4)$$

To simplify Eq. (4.4), we assume that the material edge response is flat within the frequency band of interest, and thus  $H_M(\omega) = 1$ . Moreover, we assume the power amplifier and the receiver amplifier have a

flat frequency response, and thus the transfer functions  $H_T(\omega)$  and  $H_R(\omega)$  are solely due to the frequency response of the transducer. Due to the reciprocity of the piezoelectric transducer, we have  $H_T(\omega) = H_R(\omega) = H(\omega)$ . Thus, Eq. (4.4) now becomes:

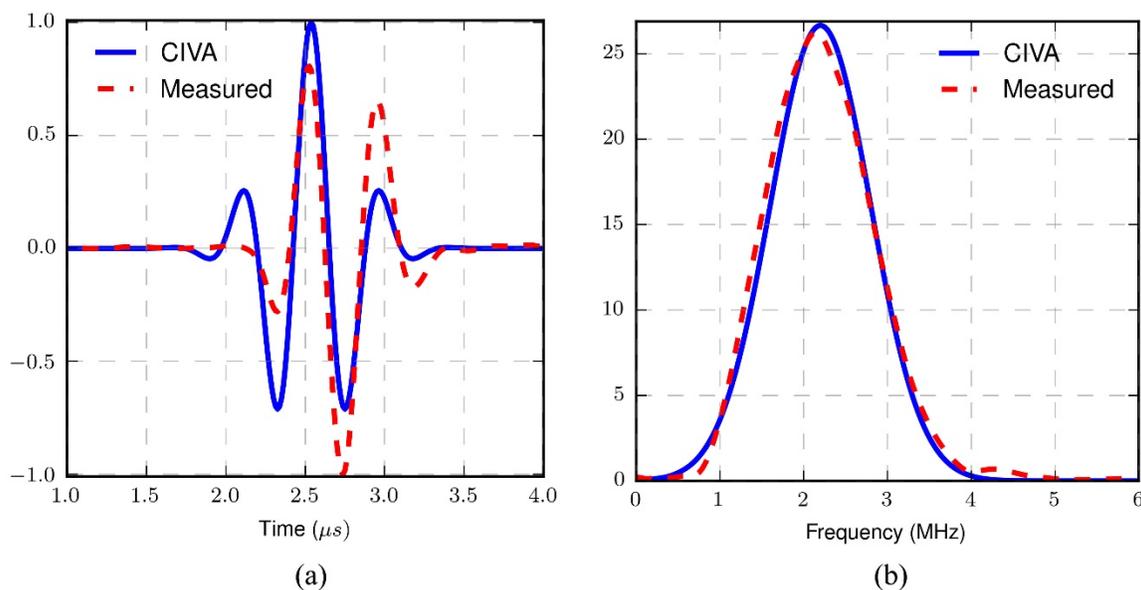
$$V_R(\omega) = H^2(\omega)V(\omega) \quad (4.5)$$

The signal that we are interested in estimating is  $S(\omega) = H(\omega)V(\omega)$ . Thus, we have:

$$V_R(\omega) = H(\omega)S(\omega) \quad (4.6)$$

Thus, the received corner echo signal  $V_R(\omega)$  do not give us exactly the signal of interest  $S(\omega)$ .

However,  $S(\omega)$  and  $V_R(\omega)$  should have the same center frequency, and approximately similar bandwidths. To define the excitation signal, we compute the center frequency and bandwidth of the corner echo response, and define the excitation signal in CIVA as a Gaussian modulated tone burst with the estimated center frequency and bandwidth (Figure 4.8).

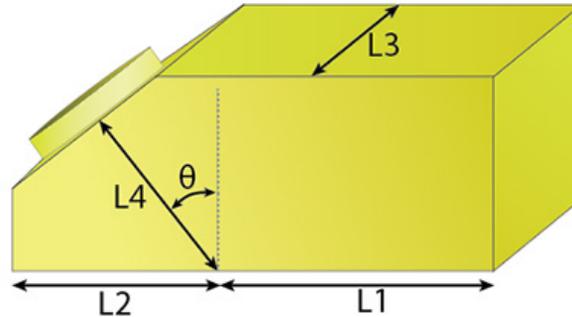


**Figure 4.8.** An Example of the Gaussian Tone Burst Signal Defined in CIVA for Probe 1 as Compared to the Measured Edge Response Signal; (a) time domain, (b) frequency spectrum amplitude

Note that in general this type of frequency response measurement should be repeated every time a different configuration of the pulser/receiver instrumentation is used. However, in this study, this procedure was conducted only once for each probe. It is expected that if the instrumentation configuration is changed between different raster scan measurements, the effective signal bandwidth will be changed to some extent, but this should not result in any significant change in the excitation center frequency. The uncertainties in measures of the probe's center frequency and bandwidth were computed as the standard deviation in the values from different specimens.

To complete the specification of the probe, we still need to define the wedge properties. Because all the wedges considered in this study are based on a pulse-echo angle beam inspection, the wedge is fully specified by the dimensions shown in Figure 4.9. The length dimensions ( $L_1$ ,  $L_2$ ,  $L_3$ , and  $L_4$ ) were obtained by physically measuring them using a Vernier caliper. The wedge angle ( $\theta$ ) was measured using a digital protractor. According to the manufacturer of the wedges, they are made out of Plexiglas material.

The mechanical properties of the wedges were not measured, and literature values for the density, longitudinal wave velocity, and shear wave velocity of Plexiglas were used. A summary of all the probes' properties and their associated uncertainties are provided in Appendix A, Tables A.6 and A.7, respectively, based on the measurement methodology describe above.



**Figure 4.9.** Dimensions Required for Specifying the Probe Wedges

### 4.3.3 Flaws Specification

All three specimens considered in this study have crack-like reflectors. Their lengths, widths, depths, and tilt were known based on previous documentations of the specimens. To verify the documented values, we used different techniques to characterize the reflectors. To verify the reflectors' length, depth, and profile, we imaged the specimens using X-ray radiography. All the reflectors in specimens 1 and 3 had EDM notches with a rectangular profile, while the reflectors in specimen 2 had a semi-elliptical profile. We measured the reflectors' widths using a gap measurement tool (feeler gauge). Although the reflectors' widths were measured, it is not required as input to the model. CIVA provides a planar profile for a flaw that allows us to model both rectangular notches and semi-elliptical notches. This is a reasonable simplification, because the width of the notch is much smaller than its other dimensions. For the reflectors' tilt, we used the gap measurement tool and a digital protractor. A summary of the measured reflectors' parameters is given in Tables A.3, A.4, and A.5 for the reflectors in specimens 1, 2, and 3, respectively. The estimated uncertainty in the measurement methodology is also provided.

### 4.3.4 Inspection Specification

Inspection specification includes defining the scan parameters and the coupling parameters. The scan and index step sizes in CIVA were defined to be the same as those used in the empirical data collection (Section 4.2). Raster scans were defined to obtain the responses from all the reflectors in the specimens. The probe scan direction was set to be in the positive X-direction for a scan skew angle of  $0^\circ$ , and negative X-direction for a scan skew angle of  $180^\circ$ . The  $180^\circ$  skew angle scans were not simulated for specimens 1 and 2, because those two specimens do not have tilted reflectors; thus, they will yield the same results at the  $0^\circ$  skew angle scans.

For Probes 5 and 6, only line scans were simulated, with the probe positioned at the lengthwise center of the reflectors. This is because those two probes excite the longitudinal wave mode, which will result in mode conversion at discontinuities. Accounting for mode conversion will significantly increase the computation time (3–6 days) so it was not feasible to obtain full raster scans. Because all the flaws considered here are either rectangular or semi-elliptical, they are symmetric around their center (lengthwise), and thus little information is lost by just looking at the line scans.

### 4.3.5 Computation Specification

The final step in defining a model as described by Figure 4.5 is specifying the computation parameters. One of the parameters allows specification of the number of echoes being reflected by the back-wall and front-wall before reaching the reflector. Increasing the number of computed echoes will significantly increase the computation time. For all the scenarios in this study, we are interested in echoes with a single reflection from the back-wall or the reflector or both. Thus, all computations were specified as half-skip. Also, all computations were specified to use the hybrid Kirchhoff and GTD method to compute the scatter from the reflectors. As for accounting for mode conversion, only probes 5 and 6, the longitudinal wave probes, were specified to account for mode conversion while all scans using Probes 1–4 did not account for mode conversion. This is because Probes 1–4 are designed to excite a refracted shear vertical wave in the specimen. To achieve this, the angle of incidence of the probes' wedges was below the critical angle to excite longitudinal waves in the specimens. Thus, enabling mode conversion computation was not required. Accounting for mode conversion significantly increases simulation computation time. Due to the large computation time, all the simulations involving probes 5 and 6 (which account for mode conversion) were run with only a single line scan running through the center of the reflectors. As a result, C-scan images could not be obtained for those probes (see Section 5 for analysis details). As for the simulations involving probes 1-4 (which do not account for mode conversion), full 2-D raster scans were defined and C-scan images were available for those probes.

There are two parameters that control the accuracy of the computation—*field accuracy*, which controls the mesh density at the probe excitation surface, and *defect accuracy*, which controls the reflector mesh density. The nominal value used for those two parameters was set to 1. Increasing these values increases the accuracy of the results at the expense of increasing the computational time.



## 5.0 Model Validation Framework

To understand what the “ingredients” of a model validation framework should be, we reiterate the definition of model validation by the SCS (Schlesinger 1979) that has been introduced in Section 2.0: “*Substantiation that a computerized model, within its domain of applicability, possesses a satisfactory range of accuracy consistent with the intended application of the model.*” To make more sense of this definition, we dissect it into its keywords:

- *Domain of applicability*: This is the domain where the model is known to be applicable. This has been discussed in Section 3.0, where we described the asymptotic ultrasound models that we are considering in this validation study. For example, we know that the ray theory beam propagation model and the Kirchhoff scattering model are based on high-frequency approximations. Thus, using this prior knowledge, we can quantify the error in those models by comparing them to known exact analytical solutions. Cases that do not have a satisfactory accuracy for the intended use are rejected and need not to be validated with empirical data.
- *Intended application of the model*: Computer models can be used for different purposes. For example, a model would be used to compute the probability of detection of small flaws in homogeneous material, or beam mapping and characterization in complex geometries, etc. A model might have acceptable accuracy for one intended application, but might not for a different application.
- *Substantiation of accuracy*: Generally a metric or a set of criteria that compute the difference or distance between the model’s prediction and the empirical observations will be sufficient to substantiate the accuracy of the model. If the model produces multidimensional output data, which is the case of UT inspection, then the metric should reduce the data output to a single value for computing the distance between the prediction and the observation. How this metric is selected will be based on the intended application of the model. However, depending on the intended use of the model, the data might be pre-processed by transforming the data, or by extracting particular features from the multidimensional data.
- *Range of accuracy*: Whenever we attempt to quantify the model’s accuracy, we must acknowledge that uncertainties will always exist in both the observations and predictions, as has been discussed in Section 2.0. Those uncertainties must be factored into the metric used for determining the model’s accuracy. Then, due to those uncertainties, the metric will have a range or an associated confidence interval.
- *Satisfactory [range of accuracy]*: Once the range of accuracy of the model has been determined, we should have criteria that allow us to decide if the calculated level of accuracy is satisfactory or not. Human decision is still required to determine such criteria, but once such criteria are established, there is no uncertainty to the rejection or acceptance of a given result.

Thus, a validation framework should include a set of metrics that are based on the output that will be used for the intended use of the model. Those metrics should take into account sources of uncertainties in both observations and predictions. The models used for ultrasound produce outputs that represent amplitude as well as spatial and temporal dimensions. In a strict sense, the metric would be some variation of a norm metric, which computes the net point difference in the output of prediction and observations, or by analyzing individual A-scan (time-series) data, or composite UT images (B-scan/C-scan/D-scan images). See Hellier (2003) for more information about UT inspection and UT image views. Often biases due to some input parameters will result in an “apparent” high error in the model. Those biases may not affect the decision that would be made based on the intended use of the model. Examples of such biases may be that the scan starting point relative to the flaw is different between the model and empirical measurements, or the excitation signal shape is not exactly the same. Moreover, it will be difficult to

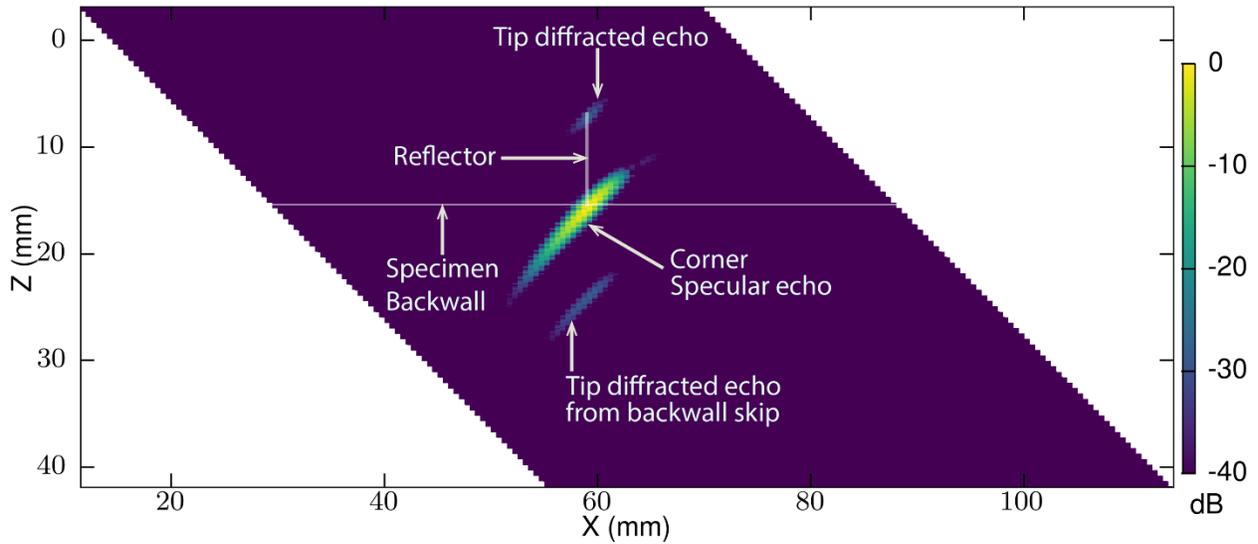
account for uncertainties in the empirical data output if there are no multiple measurements for the same scenario. In the following, we propose a two-step validation framework that both ensures the data output is qualitatively and quantitatively acceptable:

- *Qualitative Analysis:* When encoded UT data are collected, data analysis generally involves trained technicians who analyze individual A-scan (time-series) data, or composite UT images (B-scan/C-scan/D-scan images) for detection and characterization of reflectors. This generally is an acceptable technique because the human eye is very powerful and any automated algorithm attempting to automate reflector detection seeks to emulate human visual capabilities. This also applies for the comparison of two images from experiment and simulation. Thus, this task will focus on visually comparing and contrasting empirical results with simulation. Things to look for here include: (1) a presence of all the expected responses the simulation (e.g., both specular and tip-diffracted, or any expected mode-converted echoes); (2) echo shapes in the composite views are consistent within acceptable limits; and (3) echo locations are where they are expected to be (specular echoes reflected from corners, and tip-diffracted echoes coming from the flaw tip). We provide details of the qualitative analysis methodology in Section 5.1.
- *Probabilistic Analysis:* Although the human eye is quick to detect any differences and similarities between different images, the decision about the similarity of the two images made by different people will differ, and so might their confidence in such a decision. This problem can be eliminated by using quantitative metrics that would instantly give a similarity index between two images. However, human decision is still required to determine whether the value of this similarity index is acceptable or not, but once such criteria are established, there is no uncertainty to the rejection or acceptance of a given result. In this analysis, we obtain metrics depending on a specific intended use for the model. In this study, we derive an amplitude-based metric, which would be of interest for probability of detection applications. This metric uses RF signals from a line scan and reduces them to a single value. This metric also takes into account sources of uncertainties. The purpose of the probabilistic assessment is to separate the effects of parameter uncertainty, experimental uncertainty, and residual variability from model inadequacy and coding and numerical errors. Details of this probabilistic analysis will be given in Section 5.2.

## 5.1 Qualitative Analysis Methodology

In this validation study, we consider isotropic homogeneous materials and regularly shaped crack-like planar reflectors. Damage detection is based on corner trap responses from the back-wall reflectors. This simple scenario will result in well-defined ultrasonic responses due to specular echoes from corner reflections, and diffracted echoes from the tip of the reflector. Perhaps, a B-scan view is best to describe those responses, as shown in Figure 5.1. The figure shows the response due to an EDM notch reflector. We can see that there are three uniquely identifiable responses from such a notch.

Based on a qualitative analysis, a model prediction can be labeled as acceptable if both the specular and diffracted responses are present. However, often one can find discrepancies between the model prediction and empirical observations. In the following, we describe possible discrepancies and the possible factors contributing to such discrepancies. The assessment criteria will be based on three views of the raster-scan data: C-scan, B-scan, and A-scan. Because all those views represent the same data, any factor contributing to discrepancies between observation and prediction will result in discrepancies in all three views.

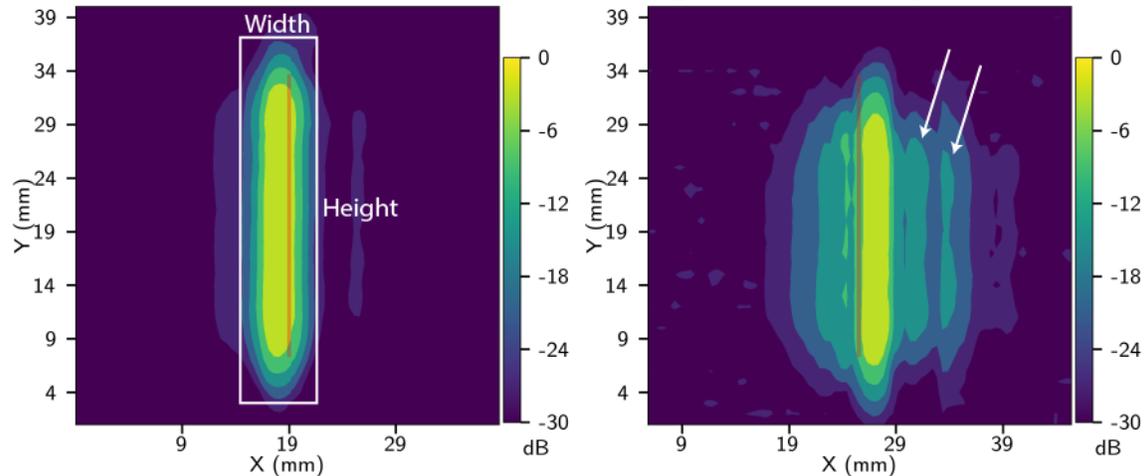


**Figure 5.1.** B-scan View Describing the Ultrasonic Response Echoes from a Regularly Shaped Crack-Like Reflector

### 5.1.1 C-scan Views

We show an example of a typical C-scan image from two different rectangular EDM reflectors in Figure 5.2. C-scan images are a top view of the raster scan data, which provides a view of the reflector's length profile, while excluding information about its depth. Thus, the response height (see Figure 5.2) is directly related to the length profile of the flaw. On the other hand, the response width is related to beam angle and the beam spread of the ultrasonic probe. Because this is a qualitative comparison, there is no hard threshold value for determining the width and length of the response; however, based on the scale used in Figure 5.2, this value can be estimated to be around  $-20$  dB. In some scenarios, multiple echoes are present in the response due to mode conversion or multiple reflections between the specimen backwall and the reflector. These multiple echoes typically separate in the B-scan view discussed in the next section. With multiple echoes, the C-scan view will have a more complicated response pattern as shown in the right of Figure 5.2. Hence, the criteria for comparing model predictions and empirical C-scans are defined as follows:

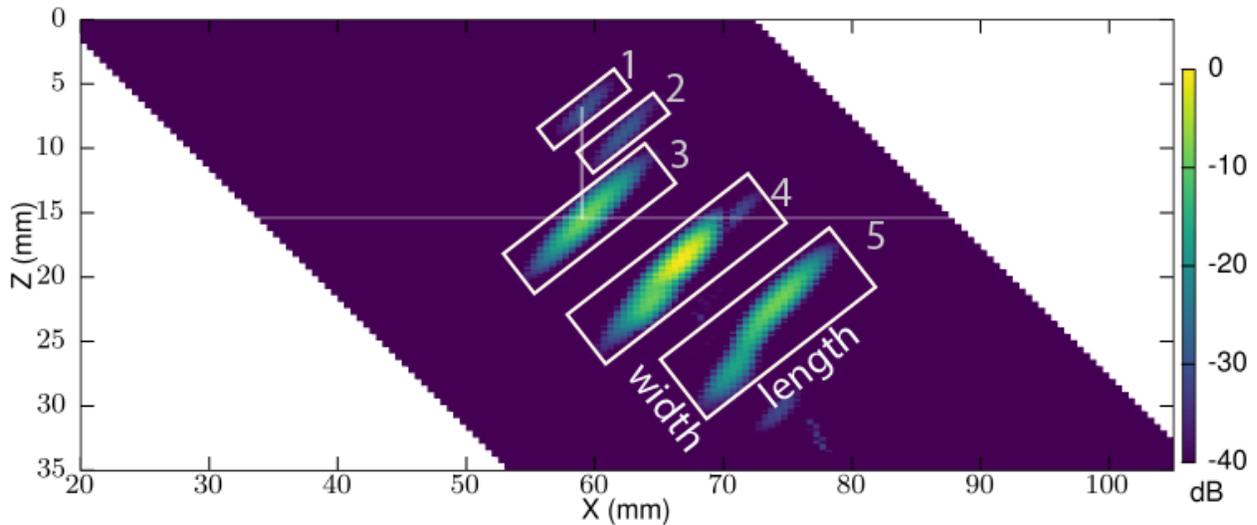
- *Number of resolvable responses:* If the number of resolvable echoes in predictions is not the same as the observation, then either the model does not accurately model the scatter from the reflector, or the observation contains echoes from scatter other than the modeled reflector.
- *Width of resolvable responses:* Discrepancies in the width of the response indicate that the model does not accurately model the beam angle and the effective beam spread of the probe. This inaccuracy might be a result of inaccurate input parameters such as the material properties or probe diameter.
- *Height of resolvable responses:* As we have mentioned, the height of the responses are related to the reflector length. Thus, any discrepancies in the response height can be attributed to either excessive noise in the experiment, the reflector length given as input to the model is inaccurate, or the model does not adequately model the scatter from the reflector. Additionally response height (a measure of flaw length) is also affected by the probe's beam width or spot size in the lateral direction. Differences between experimental and modeled data can come from inaccurately modeled beam spread in this direction.



**Figure 5.2.** C-scan Views (top view of raster data) of Responses from Two Different Reflectors. The length profile (as determined by radiography) of the reflector is indicated as a red line at the center of the response. *Left:* The approximate width and height of the response are shown. *Right:* C-scan view for responses with multiple echoes to the right, noted by arrows.

### 5.1.2 B-scan Views

We have shown a typical B-scan image from a rectangular EDM reflector in Figure 5.1. B-scans provide cross-section view that give depth (through-thickness) information and exclude information about flaw length. The response shown in Figure 5.1 is the simplest possible response from a corner-trap response. Multiple echoes might also occur as was explained in Section 5.1.1. An example of a B-scan that contains multiple echoes from mode conversion is shown in Figure 5.3. This shows a response due to the excitation by a longitudinal wave transducer. When a probe is used to generate a refracted longitudinal wave, mode conversion will occur at the plate surface and at the reflector interface. This is because the incidence angle required for exciting a refracted longitudinal wave is typically below the shear wave critical angle; however, the opposite is not true. See Hellier (2003) for more details. Therefore, with shear probes, mode-converted signals are generally not present. In the case of mode conversion, there are five uniquely identifiable echoes, as labeled in Figure 5.3: (1) echo 1 represents longitudinal wave tip-diffracted signals; (2) echo 2 is tip-diffracted after mode conversion (either at the specimen-wedge interface or at the reflector interface); (3) echo 3 is specular echo due to incident and reflected (pure) longitudinal wave; (4) echo 4 is specular echo to combinations of incident and reflected shear/longitudinal wave modes; and (5) echo 5 is specular echo due to a pure shear wave.



**Figure 5.3.** B-scan Image of a Response from an EDM Reflector when Mode Conversion is Present

Now we can identify possible sources of discrepancies between model predictions and empirical observations in B-scan views:

- *Number of resolvable echoes:* Differences in the number of echoes indicate that either the observations had scattering sources that were not possible to account for in the model (unacknowledged uncertainties), or the model does not adequately model the scatter from the reflector.
- *Width of echo:* The definition of the echo width is as labeled for echo 5 in Figure 5.3. The echo width depends on the excitation pulse width, or equivalently the signal bandwidth. Thus, any noticeable discrepancies between predictions and observations will most likely be due to inaccuracies in specifying the excitation signal within the model. Based on the ultrasonic models discussed in Section 3.0, it is unlikely that such discrepancies are related to model inadequacy.
- *Length of echo:* This is proportional to the C-scan response width, as defined in Figure 5.2. Thus, we can attribute any discrepancies in this feature to the same reasons related to discrepancies in C-scan width, as described in Section 5.1.1.
- *Location of echo:* As we have discussed earlier, the center of the specular echo will appear at the corner between the specimen back-wall and reflector. The center of the tip-diffracted echo will be at the tip of the crack-like reflector. The locations of any other echoes depend on their scattering sources, or the type of mode conversion (longitudinal to shear, or shear to longitudinal). Discrepancies along the X-axis direction are due to errors in location encoding during the experiment. Discrepancies along the Z-axis are attributed to inaccurate specifications of the wave speed within the specimen, or inaccurate specification of the wedge delay.
- *Amplitude of echo:* The relative amplitudes between echoes can be approximated from the B-scans. Thus, we can evaluate any discrepancies in the relative echo amplitudes. If discrepancies exist in the relative echo amplitudes, there are multiple sources that would contribute to such differences. This includes sources from the empirical data such as noise and irregularity in the edges of the flaws. Sources from the model that contribute to such discrepancies include inadequacy of the underlying model. Uncertainty in the model input parameter might also have some contribution.

### 5.1.3 A-scan Views

A-scan views show a time-series signal at a single scan point from the raster scan. The types of echoes expected to be present in an A-scan depend on the scan point location. Generally, the echoes expected to be seen within an A-scan are the same as was described for the B-scans in Section 5.1.2. We summarize the possible discrepancies between model predictions and empirical observations as follows:

- *Pulse shape*: When comparing time-series, the pulse shape will be affected by its phase, frequency, bandwidth, and window function. All those parameters are influenced by the specification of the excitation signal in the model, and by the transducer's response and electrical equipment in the experiments. Such uncertainties are difficult to control, and they do not represent any model inadequacy. Although the A-scans might appear to have significant discrepancies due to uncertainties in such parameters, they do not have significant effects on decisions made based on the intended use of the model.
- *Pulse arrival time*: There are two main sources contributing to the differences in pulse arrival time. First, if the specified excitation signal phase angle in the model is different than the empirical counterpart, it will affect the apparent pulse arrival time. Second, uncertainties due to irregularities in the reflector's surface, specimen height, wedge delays, and material properties contribute to differences in the pulse arrival time. Those uncertainties and biases will have major effects on the pulse arrival time, which will overshadow any model inadequacy contributing to errors in the pulse arrival time.
- *Amplitudes*: Differences in the relative amplitudes between the prediction and observation are due to different factors. Empirically, high levels of noise, or variations in the coupling conditions and electrical equipment, might significantly affect the A-scan amplitudes. Such factors are not accounted for in the model. On the other hand, inadequacy of the model will also contribute to differences in the amplitudes. Because amplitudes are critical for making decisions about the reliability of NDE methods, in Section 5.2 we will provide a framework that incorporates those uncertainties into a quantitative validation metric.

## 5.2 Probabilistic Framework

Although many discrepancies are possible between observations and predictions, those discrepancies alone cannot tell us much information about the adequacy of the model. Also, as described in the previous section, the sources of such discrepancies can be minimized by careful calibration of the model, and by carefully selecting the excitation signal shape. **However, the intent of model validation is not the assessment of how precisely it can replicate a given observation, it is rather a measurement of how adequate a given model would be for its intended use, knowing that uncertainties exist.**

The qualitative analysis is based on simulation predictions that were given nominal values for the input parameters. However, as has been discussed in Section 2.0, it is important that we account for sources of uncertainties in our analysis before reaching a decision about the validity of a model. In UT, two main applications of modeling are assessing the probability of flaw detection and the ability to characterize flaws. Thus, knowing that uncertainties exist in practice, how well can information from a prediction tell us about the reliability of our inspection technique? The objective here is: Given simulated data and empirical data based on the same input, how can we decide if the simulation adequately represents the empirical data. To do that, we will represent the signals mathematically, and account for all sources of uncertainties.

### 5.2.1 Derivation of the Validation Metric and Empirical Uncertainty

In this section, we derive an amplitude-based metric by mathematically modeling the RF signals in a line scan or a raster scan. A generally accepted approach in UT is to select the peak amplitude of the scan as the indicator of the effective signal value from the flaw. However, picking a single value is highly susceptible to noise, and also discards a lot of information present in a raster scan. From a model validation point of view, it is anticipated that a metric that involves the raster scan as a whole will better represent the adequacy of the model.

To compute this metric, we start by posing the flaw detection problem as the requirement to decide between two hypotheses:

$$\begin{aligned}\mathcal{H}_0 : & \quad \tilde{\mathbf{r}} = \tilde{\mathbf{w}} \\ \mathcal{H}_1 : & \quad \tilde{\mathbf{r}} = \tilde{\mathbf{s}} + \tilde{\mathbf{w}}\end{aligned}\tag{5.1}$$

$\mathcal{H}_0$  is the hypothesis that no defect is present. In this case, the received signal is composed of only an additive white Gaussian noise (AWGN)  $\tilde{\mathbf{w}}$  due to the thermal noise of the receiver hardware. On the other hand,  $\mathcal{H}_1$  is the hypothesis that a flaw is present. The received signal is composed of the scatter from the defect  $\tilde{\mathbf{s}}$  and the AWGN term.

To proceed further, we need to find a model for both the defect signal and the AWGN. We start by considering a raster scan or a line scan with a total of  $L$  scan points. At each scan point, we apply a narrow-band excitation signal to the transmitting transducer. We represent this excitation signal in the continuous time domain as follows:

$$\tilde{u}(t) = \begin{cases} \tilde{A}g(t)\exp[j2\pi f_c t] & 0 \leq t \leq T_p \\ 0 & \text{otherwise} \end{cases}\tag{5.2}$$

where  $g(t)$  is the pulse shaping window due to the transducer's narrow-band frequency response,  $f_c$  is the center frequency of the excitation pulse, and  $T_p$  is the pulse width, and  $\tilde{A}$  is the complex valued amplitude. Note that in the following, all complex valued quantities will be denoted by a tilde, and all vector quantities are shown in bold face. If a defect is present, the received signal at the  $l^{\text{th}}$  scan point is:

$$\tilde{r}_l(t) = \tilde{s}_l(t) + \tilde{w}_l(t)\tag{5.3}$$

where the term  $\tilde{s}_l(t)$  is the received signal at the  $l^{\text{th}}$  scan point due to scatter from the defect, and  $\tilde{w}_l(t)$  is the AWGN term. The scatter from the defect is composed of the superposition of multiple echoes, which are delayed, phase shifted and attenuated versions of  $\tilde{u}(t)$ , as follows:

$$\tilde{s}_l(t) = \sum_{k=0}^K (\tilde{A}_{lk}^D + \tilde{A}_{lk}^I)g(t - \tau_{lk})\exp[j2\pi f_c(t - \tau_{lk})] \quad 0 \leq t \leq T\tag{5.4}$$

where  $K$  is the number of reflected echoes, and  $T$  is the total data acquisition time. Each echo  $k$  in the signal at the  $l^{\text{th}}$  scan point is represented by its complex amplitude  $\tilde{A}_{lk}^D$ , its propagation time delay  $\tau_{lk}$ , and an interference term with a complex amplitude  $\tilde{A}_{lk}^I$ . The interference term represents the variations in

amplitude and phase of the received echo between different measurements due to experimental uncertainties.

Because the pulse-shaping window  $g(t - \tau_k)$  does not have a role in detection, for simplicity we assume that  $g(t) = 1$  for  $0 \leq t \leq T_p$ . Also, we divide the time signal into time bins of equal length. After digitizing the received signal, we can represent it in vector form as:

$$\tilde{\mathbf{r}}_l = [\tilde{\mathbf{r}}_{l0}^T \quad \tilde{\mathbf{r}}_{l1}^T \quad \dots \quad \tilde{\mathbf{r}}_{lm}^T \quad \dots \quad \tilde{\mathbf{r}}_{l(M-1)}^T]^T \quad (5.5)$$

where  $M$  is the total number of time bins (which can be different than the total number of echoes  $K$ ),  $\tilde{\mathbf{r}}_{lm}$  is the received signal within the  $m^{\text{th}}$  time bin of the signal from the  $l^{\text{th}}$  scan point, and  $T$  is the transpose operator. Now, if there are multiple echoes within each bin, they will be non-resolvable, as long as the time bin width is smaller than  $T_p/2$  (half the excitation pulse width defined in Eq. (5.2)). Then, in the case when a defect is present (hypothesis  $\mathcal{H}_1$ ), the received signal within each time bin has the following form:

$$\begin{aligned} \tilde{\mathbf{r}}_{lm} &= \tilde{\mathbf{s}}_{lm}^D + \tilde{\mathbf{s}}_{lm}^I + \tilde{\mathbf{w}}_{lm} \\ &= \tilde{A}_{lm}^D \tilde{\mathbf{v}} + \tilde{A}_{lm}^I \tilde{\mathbf{v}} + \tilde{\mathbf{w}}_{lm} \end{aligned} \quad (5.6)$$

and

$$\tilde{\mathbf{v}} = [1 \quad \exp[j2\pi f_c] \quad \dots \quad \exp[j2\pi f_c(N-1)]]^T \quad (5.7)$$

where  $N$  is the number of digitized samples within each time bin. The complex scalar  $\tilde{A}_{lm}^D$  is the unknown complex amplitude of the defect scatter, representing the effective sum of all wave echoes within the  $m^{\text{th}}$  time bin and the  $l^{\text{th}}$  scan point. The complex interference amplitude term  $\tilde{A}_{lm}^I$  has a circular Gaussian distribution  $\mathcal{CN}(0, \sigma_I^2)$ . The variance  $\sigma_I^2$  of the interference term is determined by the level of uncertainty in the experimental measurements. This variance is equal for all signals in the raster scan. The vector  $\tilde{\mathbf{v}}$  has length  $N$  with known values as defined in Eq. (5.7). Thus, the defect scatter includes an unknown deterministic signal and a coherent noise due to experimental uncertainties.

The AWGN vector  $\tilde{\mathbf{w}}_{lm}$  is also of length  $N$ , and its components are identically and independently distributed according to the circular Gaussian distribution  $\mathcal{CN}(0, \sigma^2)$ . Finally, we arrange all the received signals within the raster scan into a single vector as follows:

$$\begin{aligned} \tilde{\mathbf{r}} &= \tilde{\mathbf{s}}^D + \tilde{\mathbf{s}}^I + \tilde{\mathbf{w}} \\ &= [\tilde{\mathbf{r}}_0^T \quad \tilde{\mathbf{r}}_1^T \quad \dots \quad \tilde{\mathbf{r}}_l^T \quad \dots \quad \tilde{\mathbf{r}}_{(L-1)}^T]^T \end{aligned} \quad (5.8)$$

where the components of the vector  $\tilde{\mathbf{r}}$  have been defined in Eqs. (5.5) and (5.6), and it has a total length of  $L \times M \times N$ .

Now that we have a model for both the defect scatter and the AWGN, we can compute the likelihood probability function of the received signals. Under hypothesis  $\mathcal{H}_0$ , the likelihood is given by:

$$\begin{aligned}
P(\tilde{\mathbf{r}} | \mathcal{H}_0) &= \frac{1}{(2\pi\sigma^2)^{(LMN)/2}} \exp\left[-\frac{1}{2\sigma^2} \tilde{\mathbf{r}}^\dagger \tilde{\mathbf{r}}\right] \\
&= \prod_{l=0}^{L-1} \prod_{m=0}^{M-1} \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \tilde{\mathbf{r}}_{lm}^\dagger \tilde{\mathbf{r}}_{lm}\right]
\end{aligned} \tag{5.9}$$

where the symbol  $\dagger$  is the Hermitian transpose of the vector.

Under hypothesis  $\mathcal{H}_1$ , we assume that that interference amplitudes between different time bins and between signals at different scan points are independent. Then, the likelihood function is:

$$P(\tilde{\mathbf{r}} | \mathcal{H}_1) = \prod_{l=0}^{L-1} \prod_{m=0}^{M-1} \frac{1}{(2\pi)^{N/2} |\mathbf{C}|^{1/2}} \exp\left[-\frac{1}{2} (\tilde{\mathbf{r}}_{lm} - \tilde{\mathbf{s}}_{lm}^{\mathbf{D}})^\dagger \mathbf{C}^{-1} (\tilde{\mathbf{r}}_{lm} - \tilde{\mathbf{s}}_{lm}^{\mathbf{D}})\right] \tag{5.10}$$

where  $\mathbf{C}$  is the covariance matrix of the noise and interference terms within each time bin at a single scan point,

$$\begin{aligned}
\mathbf{C} &= \text{COV}\left[\tilde{\mathbf{s}}_{lm}^{\mathbf{I}} + \tilde{\mathbf{w}}_{lm}\right] \\
&= \sigma_I^2 \tilde{\mathbf{v}} \tilde{\mathbf{v}}^\dagger + \sigma^2 \mathbf{I}
\end{aligned} \tag{5.11}$$

where  $\mathbf{I}$  is the identity matrix. Because the interference noise is independent and identical between different time bins, then the covariance matrix is the same for all time bins.

The value of the flaw scatter signal  $\tilde{\mathbf{s}}_{lm}^{\mathbf{D}}$  is unknown, and so we estimate it using the maximum likelihood estimate (MLE) (see Kay 1994), which is given by:

$$\hat{\tilde{\mathbf{s}}}_{lm}^{\mathbf{D}} = \left(\tilde{\mathbf{v}}^\dagger \mathbf{C}^{-1} \tilde{\mathbf{v}}\right)^{-1} \tilde{\mathbf{v}}^\dagger \mathbf{C}^{-1} \tilde{\mathbf{r}}_{lm} \tag{5.12}$$

To compute Eq. (5.12), we need to find the inverse of the covariance matrix, which is obtained by an Eigen-decomposition formulation:

$$\mathbf{C}^{-1} = \frac{1}{\sigma^2} \left[ \mathbf{I} - \frac{\sigma_I^2}{\sigma^2 + N\sigma_I^2} \tilde{\mathbf{v}} \tilde{\mathbf{v}}^\dagger \right] \tag{5.13}$$

Now, by substituting Eqs. (5.7) and (5.13) into Eq. (5.12), we obtain the MLE of the flaw signal within each time bin:

$$\begin{aligned}
\hat{\tilde{\mathbf{s}}}_{lm}^{\mathbf{D}} &= \frac{1}{N} \tilde{\mathbf{v}}^\dagger \tilde{\mathbf{r}}_{lm} \\
&= \frac{1}{N} \mathcal{F}(\tilde{\mathbf{r}}_{lm})
\end{aligned} \tag{5.14}$$

Where, by definition, based on the form of  $\tilde{\mathbf{v}}$  defined in eq. (5.7),  $\mathcal{F}(\tilde{\mathbf{r}}_{lm})$  is the discrete Fourier transform of the  $m^{\text{th}}$  time bin of the  $l^{\text{th}}$  scan signal, at the frequency  $f_c$ . Note that the dependency of the Fourier transform on the center frequency  $f_c$  is not shown in the notation for brevity.

To obtain the optimal feature (or sufficient statistic) for detection, we use the log likelihood ratio test (LRT) (Kay 1998; Van Trees et al. 2013):

$$T(\tilde{\mathbf{r}}) = \ln \frac{P(\tilde{\mathbf{r}} | \mathcal{H}_1)}{P(\tilde{\mathbf{r}} | \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \gamma \quad (5.15)$$

where  $T(\tilde{\mathbf{r}})$  is the sufficient statistic and  $\gamma$  is some given threshold, which can be computed based on a Bayesian cost function or a Neyman-Pearson formulation. Equation (5.15) states that we decide a defect is present (hypothesis  $\mathcal{H}_1$ ) if  $T(\tilde{\mathbf{r}})$  is greater than the threshold, and decide no defect is present (hypothesis  $\mathcal{H}_0$ ) if  $T(\tilde{\mathbf{r}})$  is less than the threshold.

We now compute the LRT by substituting Eq. (5.14) into Eq. (5.10) and then Eqs. (5.9) and (5.10) into Eq. (5.15), and moving all the terms that do not depend on the received signal to the right-hand side. We obtain the LRT:

$$T(\tilde{\mathbf{r}}) = \frac{1}{N} \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} |\mathcal{F}(\tilde{\mathbf{r}}_{lm})|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \gamma' \quad (5.16)$$

Thus, we have now derived the metric that will be used for validation and it is given by Eq. (5.16).

Because in the validation study, we are only interested in the received signal when a defect is present, we need to compute the probability density function of  $T(\tilde{\mathbf{r}})$  in Eq. (5.16) under hypothesis  $\mathcal{H}_1$ . Based on the computed probability density function, we can estimate the experimental uncertainty in the metric. According to Dib and Udpa (2016) the probability density function of the metric in eq. (5.16) is given by the non-central  $\chi^2$  density function with  $2LM$  degrees of freedom, and a non-centrality parameter:

$$\lambda = \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} \frac{1}{N\sigma^2} \frac{|\tilde{A}_{lm}^D|^2}{1 + N\sigma_l^2 / \sigma^2} \quad (5.17)$$

To quantify the uncertainty associated with the metric in Eq. (5.16), we use the standard deviation of the distribution, defined as:

$$Uncertainty_{empirical} = \sqrt{2(2LM + 2\lambda)} \quad (5.18)$$

where  $\lambda$  is defined in Eq. (5.17), and  $L$  and  $M$  are the number of scan points and time bins, respectively.

## 5.2.2 Uncertainty in Simulations

As we have mentioned previously, one of the factors that contributes to the uncertainty in the simulations is the uncertainty in the input parameters, which is propagated to the output. Thus, this uncertainty must be taken into consideration when computing the validation metric. In this validation study, we consider that all the relevant model inputs can be measured, as discussed in Section 4.0. We denote the vector of model inputs by  $\boldsymbol{\theta}$ . Then, the model is a function that maps input vector into a scalar output:

$$y = f(\boldsymbol{\theta}), \quad \boldsymbol{\theta} = [\theta_0 \quad \theta_1 \quad \dots \quad \theta_{p-1}] \quad (5.19)$$

where  $\boldsymbol{\theta}$  is a vector of size  $P$ , which is the total number of input parameters. Note that the output  $y$  is our validation metric which, as we have established in Section 5.2.1, can be computed using Eq. (5.16). Computing the uncertainty in the output requires a multi-dimensional Monte-Carlo simulation to compute the probability density function of the output  $y$ . Due to the high dimensionality of the problem, this will require millions of simulations, which is an unfeasible task. To make the computation of the uncertainty feasible, we need to make some simplification assumptions:

1. The simulation problem is reduced from a multi-dimensional problem to a series of one-dimensional problems by using the nominal values of the inputs  $\boldsymbol{\theta}$ , and conducting a one-at-a-time variation of each  $\theta_i$  in the vector  $\boldsymbol{\theta}$ . Now, we have a series of functions

$$y_i = f_i(\theta_i) \quad i = \{0, 1, \dots, P-1\} \quad (5.20)$$

2. Each input parameter  $\theta_i$  is assumed to have a Normal distribution  $\theta_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ . We have given the methods used to obtain estimates for the hyper-parameters of the input parameters in Section 4.0, and summarized in Appendix A. The mean  $\mu_i$  will be the measured nominal value, and the standard deviation  $\sigma_i$  is the uncertainty in the measured value.
3. The model function  $f(\cdot)$  is a nonlinear function of the input parameters. However, if care has been taken to obtain accurate estimates of the input parameter, the uncertainty will be small. Then, we can approximate the functions  $f_i(\cdot)$  using linear regression, where we have:

$$y_i = \alpha_i \theta_i + \beta_i \quad (5.21)$$

where  $\alpha_i$  and  $\beta_i$  are the parameters of the linear regression.

4. Because the output  $y_i$  is a linear function of a Normal distribution,  $y_i$  itself will have a Normal distribution with  $y_i \sim \mathcal{N}(\alpha_i \mu_i + \beta_i, \alpha_i^2 \sigma_i^2)$ .

Now we have a series of probability density functions for the model output as a function of each input parameter  $\theta_i$ . We can examine the effect of each parameter on the validation metric by computing the error as:

$$\%error = \frac{\alpha_i \sigma_i}{\alpha_i \mu_i + \beta_i} \times 100 \quad (5.22)$$

which is the ratio to standard deviation to the mean. Also, we associate the uncertainty in the output to be the largest standard deviation:

$$\sigma_{model} = \max_{i \in \{0 \dots P-1\}} \alpha_i \sigma_i \quad (5.23)$$

This model uncertainty measure is a conservative estimate of the total uncertainty in our model.

At this point, we have a metric that reduces the multi-dimensional data from an ultrasonic scan into a single value. Also, by deriving the metric from first principles, we are able to associate experimental uncertainty with the metric. Also, an approximation of the model uncertainty is computed based on a change one-at-a-time sensitivity study.



## 6.0 Analysis Results and Discussion

As we have described in Section 4.0, we have a total of three specimens and six probes that were considered in this study. The aim of this section is to use the validation framework described in Section 5.0 on the limited validation tree described in Section 4.0. Then, we will use the findings of this analysis to draw more general conclusions regarding the requirements and general recommended procedures for modeling and simulation.

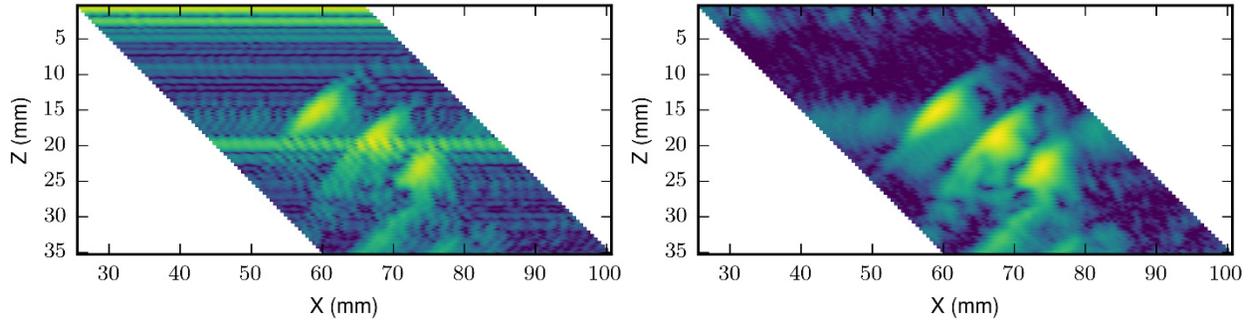
Before we can delve into our analysis, it is pertinent to define some terminology. Thus far, we have used model and simulation interchangeably. In this section, we differentiate between the two terms. The term *model* will refer to the underlying algorithms and physics engine used by the software. The term *simulation* will refer to an instance of solving a model using the modeling software, after providing all the proper input parameters. This distinction in terminology is important, because we will need to indicate if a discrepancy in a particular result is due to errors in the simulation process, or the underlying model itself.

### 6.1 Qualitative Analysis

We give the C-scan, B-scan, and A-scan images for all the specimen, probe, and skew combinations in Appendix B. The C-scan images show the observations (or empirical measurements) and the predictions (or simulations) side by side, where each image contains all the reflectors in the specimen. Each reflector's identification is labeled on the image. The dotted lines represent the Y-axis position where the B-scans have been extracted from the raster scan. Solid red lines represent the length profiles of the reflectors in the specimen, as determined from a radiographic examination. In the B-scan images, both the back-wall and the true-state depth profile of the reflectors are shown for reference as solid white lines. The dashed white lines indicate the location at which the A-scans were extracted and plotted. For each reflector, two A-scans were extracted. The first A-scan, labeled as "specular," corresponds to the location where the beam is directly incident to the corner of the reflector with the back-wall. The second A-scan, labeled as "tip," corresponds to the location where the beam is directly incident to the tip of the reflector. All the A-scans are normalized by the amplitude of the corner echo response from the same specimen and probe. This is done in order to have the same amplitude scale for both predicted and observed A-scans.

The inputs for the simulations used to generate these images were based on the nominal (or mean) measurement values as described in Section 4.0. For the empirical observation data, a filter was applied in the scan direction (X-axis) in order to reduce the effect of coherent noise due to the initial bang, as well as reflections from the plate surface and echoes within the probe wedge, which might interfere with the echoes from the reflectors. An example showing the effect of the filter in the B-scan image is shown in Figure 6.1.

In the following we summarize our qualitative analysis for each of the C-scans, B-scans, and A-scans based on the criteria discussed in Section 5.1.

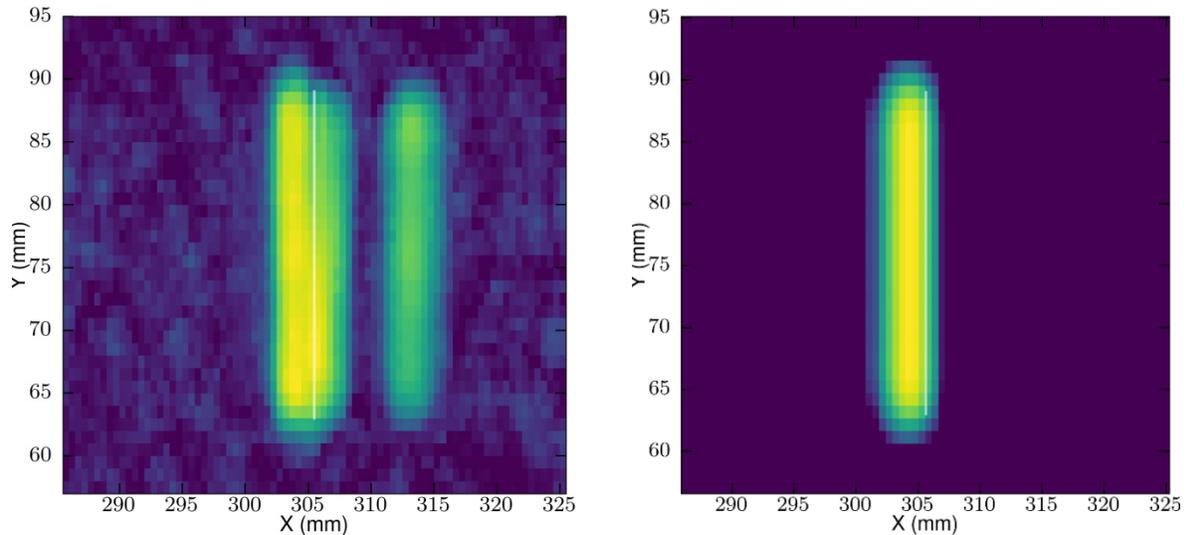


**Figure 6.1.** An Example Showing the Effect of the Median Filter on the B-scan Image. (*left*) before filtering; (*right*) after filtering.

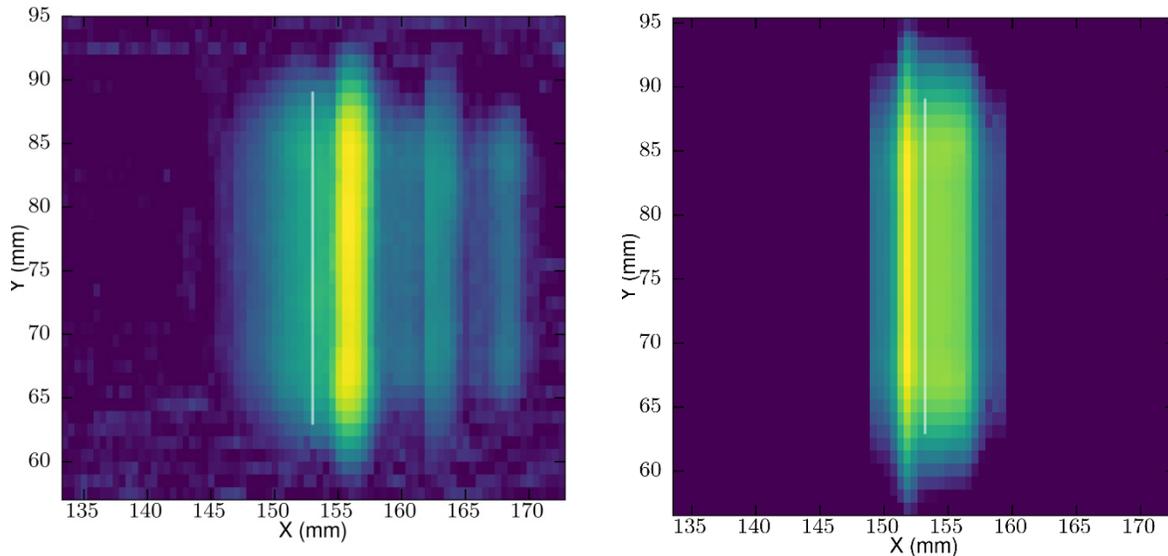
### 6.1.1 C-scans

Because only line scans were simulated for Probes 5 and 6, C-scan views are only available for Probes 1–4. The qualitative analysis for the C-scans based on the criteria given in Section 5.1.1 is as follows:

- *Number of resolvable responses:* In the observations, the shear wave probes with  $45^\circ$  beam angle (Probes 1 and 3) showed two resolvable echoes from reflectors with a tilt larger than  $10^\circ$  (specimen 3). However, the predicted responses showed only one resolvable echo. Example C-scans demonstrating this discrepancy are shown in Figure 6.2. Moreover, the shear wave probes with  $60^\circ$  beam angle (Probes 2 and 4) showed multiple resolvable echoes in the observations. The predictions showed one or two echoes, which have greater width and in closer proximity of each other than their observed counterparts. C-scan examples of such a scenario are shown in Figure 6.3.



**Figure 6.2.** Example of the C-scans from Specimen 3, Flaw M ( $15^\circ$  tilt), Skew  $0^\circ$ , Probe 3 (shear wave,  $45^\circ$  beam angle). This figure demonstrates the presence of multiple responses in the empirical observations (*left*), while only a single resolvable response is present in the model prediction (*right*).



**Figure 6.3.** Example of the C-scans from Specimen 1, Flaw J, Skew  $0^\circ$ , Probe 2 (shear wave,  $60^\circ$  beam angle). This figure demonstrates the presence of multiple scattered responses in the empirical observations (*left*), while the response from the model prediction (*right*) is more coherent.

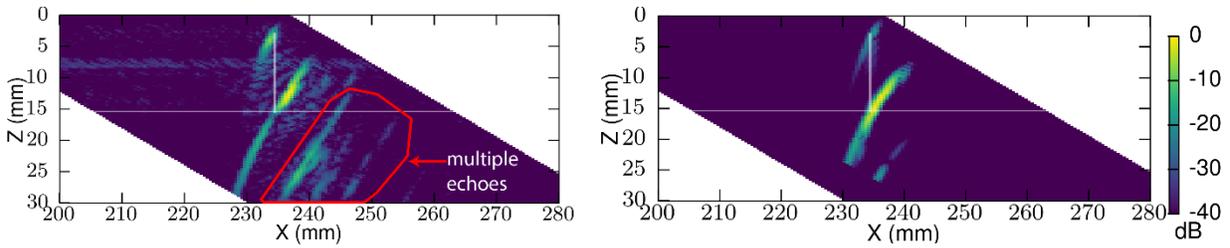
- *Width of resolvable responses:* In all the C-scan image views considered in this study, the predicted responses had wider responses than their empirical observation counterparts.
- *Height of resolvable responses:* The response height (as defined in Figure 5.2) for both the predictions and the observations were equivalent to the true-state reflector length, as expected.

In summary, the analysis of C-scans tells us that the simulations do not accurately represent the beam angle and beam spread of the probe. This inaccuracy is most likely due to inaccuracies in the model's input parameters that directly affect those properties of the ultrasonic beam. Those input parameters include the probe diameter, wedge angle, and specimen/wedge material properties. **Nevertheless, the discrepancies observed in the C-scans do not give any evidence about the inadequacy of the underlying computational/mathematical models.** However, depending on the intended use of the model, a comprehensive uncertainty analysis of relevant model inputs parameters would be required to provide sufficient evidence of the validity of the model.

### 6.1.2 B-scans

Based on the criteria described in the model validation framework in Section 5.1.2, the following summarizes the qualitative analysis of the B-scans:

- *Number of resolvable echoes:* In the observations, the shear wave Probes with  $60^\circ$  beam angle (Probes 2 and 4) have multiple echoes in addition to the specular and tip-diffracted echoes. However, the predictions do not have such echoes. We show an example of this discrepancy in Figure 6.4. This discrepancy reinforces the findings from the C-scan echoes, as mentioned in Sections 5.1 and 6.1.1. Moreover, the predictions show that the amplitude of the tip-diffracted echo from a back-wall skip (defined in Figure 5.1) is larger than the direct tip-diffracted echo. However, the observations do not appear to have a tip-diffracted echo from a back-wall skip but do have other multiple echoes.



**Figure 6.4.** Example of the B-scans from Specimen 1, Flaw I, Skew  $0^\circ$ , Probe 2 (shear wave,  $60^\circ$  beam angle). This figure demonstrates the presence of multiple scattered echoes below the mean specular echo in the empirical observations (*left*), while the response from the model prediction (*right*) is more coherent and does not show the multiple echoes.

- *Echo width:* Although there might be instances where the echoes width do not match, there were no major discrepancies related to the echoes width. For example, in Figure 6.4 we can see that the specular echo is slightly wider in the predictions. However, the general shape of the response is consistent between the predictions and observations.
- *Echo length:* There is a general bias in all the results where the length of individual echoes in the predictions is larger than that of the observations counterparts. This reinforces the discrepancies in the C-scan widths, as mentioned in Sections 5.1 and 6.1.1.
- *Echo relative location:* There were no major discernable discrepancies related to relative echo locations.
- *Echo relative amplitudes:* The peak amplitude of the tip-diffracted echo relative to the specular echo is smaller in the predictions than the observations. This difference is consistent in all the scans, and it can be seen in the example given in Figure 6.4. The response from the tip of the reflector is approximately 6 dB lower in the prediction than that of the observation. Note that because the reflectors in specimen 2 are semi-elliptical, the tip-diffracted echoes are weak, and are not identifiable from the plotted B-scans in both predictions and observations. Due to this consistent bias in the model's prediction of the tip-diffracted amplitudes, the source can be attributed to the model's inadequacy to handle such echoes, or due to the simplification in the reflector's modeling, which are represented as 2D planar discontinuities.

Note that with the shear wave probes with  $60^\circ$  beam angle (Probes 2 and 4), the specular echo appears to be cut off. This is related to the method in which the model computation was specified. A computation zone relative to the probe's location is specified in order to reduce the computation time. Any echoes from sources outside this zone were not computed, and hence resulted in an apparently cut off response for the  $60^\circ$  beam angle probes.

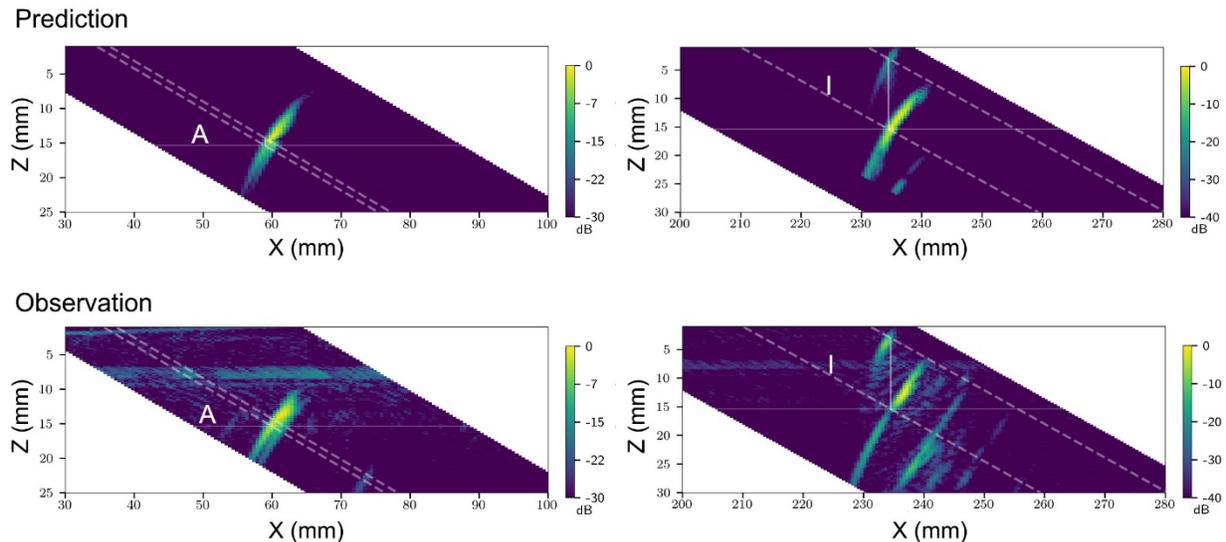
In summary, the B-scan image views reinforce the findings from the C-scans. In addition, they provided evidence that the models consistently underestimate the amplitude of the tip-diffracted signal, relative to the maximum amplitude of the specular echoes. This analysis cannot provide proof of the source of this bias (is it because of the simulation's definition of the flaw shape, or it is because of the inadequacy of the GTD model?). However, because the tip-diffracted echo amplitude is always much smaller than that of the specular echo, this bias has no implications on applications intended for assessing the probability of detection. However, this might be of importance for flaw sizing applications, and warrants further investigation.

### 6.1.3 A-scans

The A-scan images for all the validation tree scenarios are given in Appendix B.3. As we have described at the onset of this section, there are two A-scan plots for each reflector, corresponding to a line through the specular response and one through the tip response. The dotted lines in the B-scans in Appendix B.2 represent the locations of the A-scans that are shown in Appendix B.3. Extractions of two flaws from the appendices are depicted in Figure 6.5 and Figure 6.6 for the B-scans and A-scans, respectively. These data represent specimen 1, probe 2, and skew  $0^\circ$ . Reflector A is shallow with 1.1 mm or 0.04 in. flaw height. Reflector I is deep with 12.4 mm or 0.49 in. height. The units of the A-scan vertical axis are unitless, representing the normalized amplitudes, relative to a corner-echo amplitude. Note that the shallow flaw tip response is not resolved from its corner response while the tip response from a deep flaw clearly is separable from its corner response.

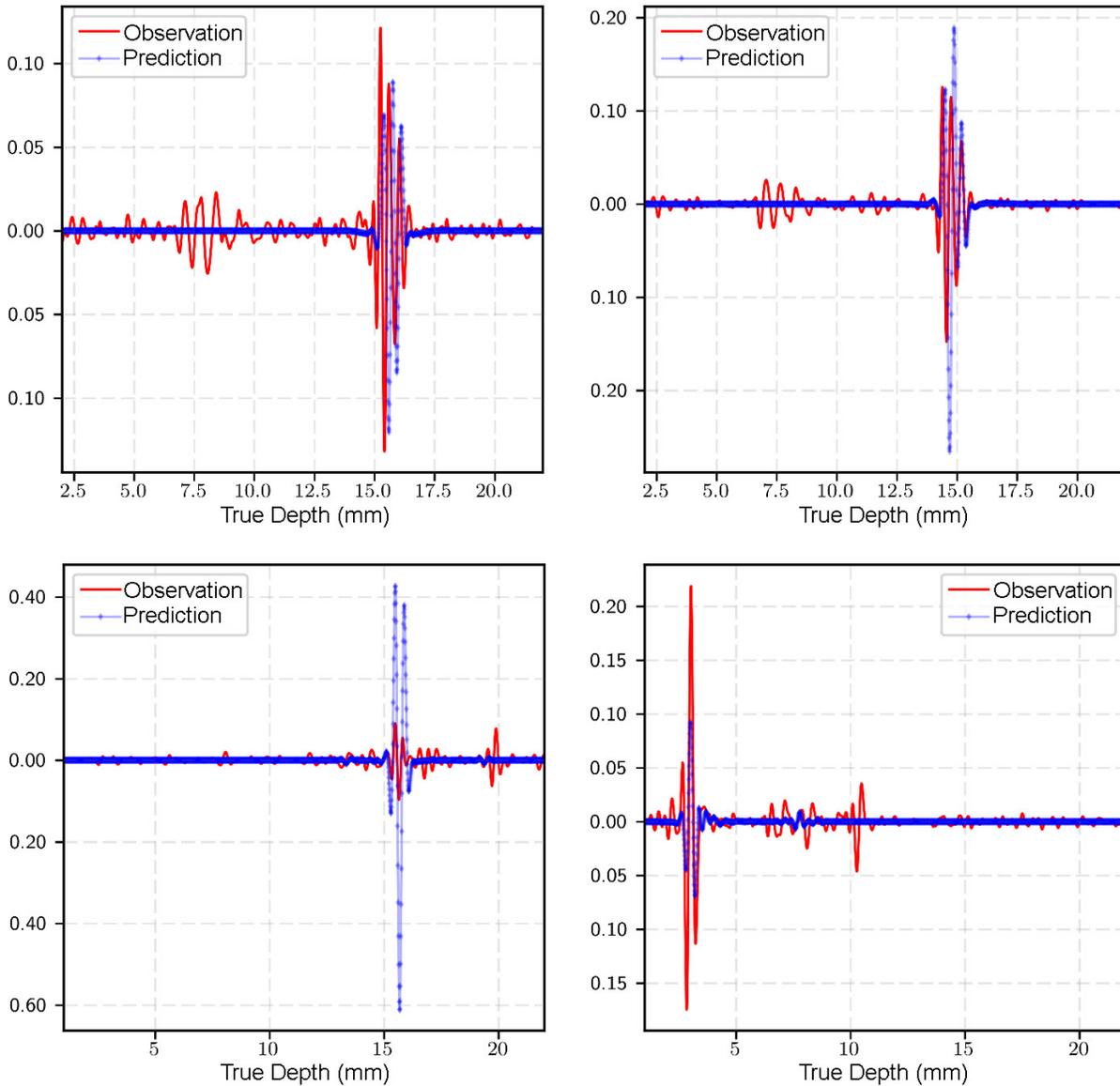
Based on the criteria described in the model validation framework in Section 5.1.3 (see also Section 4.3.2), the following summarizes the qualitative analysis of the A-scans:

- *Pulse shape:* We expected to see significant differences in the pulse shapes between predictions and observations due to the methodology we use to define the excitation signal. The A-scans largely depend on the defined excitation signal. If the intended use of the model requires precise representations of the RF signals within a specific experiment, then more sophisticated measures should be taken to account for the precise shape of the excitation signal. In most UT applications, taking such measures is unnecessary because precise RF signal shape is not required. Using only representative parameters such as bandwidth and frequency is enough.
- *Pulse arrival time:* Although there are few cases from the A-scans that have shown little to no difference in the pulse arrival times, most A-scans did have some difference in the pulse arrival time. Differences in the pulse arrival time were near or less than one pulse cycle. This gives strong evidence that the source of this discrepancy is the uncertainty in defining the excitation pulse phase angle in the simulation inputs.



**Figure 6.5.** B-scan Images for Shallow Flaw, A (*left*) and Deep Flaw, I (*right*). Predictions are shown on the top and observations on the bottom. The dashed lines represent A-scan locations for the tip and corner signals in Figure 6.6.

- Amplitudes:** All the A-scan results show amplitude discrepancies with varying levels of significance. However, the amplitude of the tip-diffracted echo relative to specular echo can be seen to be consistently smaller in the predictions. As an example, we show in the bottom right of Figure 6.6 a case for a large reflector (reflector I). The echo from the reflector's tip as predicted by the simulation is less than half of the empirical observation. Note that for smaller reflectors, it is not possible to see this bias in the model. In the top right of Figure 6.6 (reflector A) shows the tip-reflected A-scan for a reflector with height 1.1 mm (0.04 in.). Due to the small height, the tip-reflected echo is not resolvable from the specular echo, and hence the A-scan tip signal is dominated by the specular echo response. As for the discrepancies in the specular echo amplitudes, we do not attempt to quantify those discrepancies here, as this will be the topic of the quantitative analysis in Section 6.2.



**Figure 6.6.** A-scans for a Shallow Flaw, A (top) and Deep Flaw, I (bottom). The specular responses are on the left and tip signals on the right.

#### 6.1.4 Summary of Qualitative Analysis

In Sections 6.1.1, 6.1.2, and 6.1.3, we provided an overall analysis of the C-scan, B-scan, and A-scan by directly inspecting results from model predictions and empirical observations. The aim of this analysis was to confirm that the simulation results are acceptable from the point of view of a SME, and also that they did not significantly deviate from reality. We constrained our analysis to a few criteria that we deemed sufficient for the qualitative analysis purposes while acknowledging that more specific criteria might be required, depending on the intended use of the model. Positive outcomes from such an analysis increase our confidence in any findings from subsequent quantitative validation. It is indeed the case that we have reached a positive outcome from the qualitative analysis. This analysis allowed us to confirm that within limits, the simulations results do accurately represent observations.

This qualitative analysis was based on a general comparison. For an analysis that is based on a specific intended use of the models, some of the discrepancies discussed in this analysis would lead us to conclude that there is no sufficient evidence about the accuracy of the models. In this case, we have two options—reject the models, or conduct a quantitative analysis based on the metrics that are of interest to our intended use. If we decide to reject the model, this does not provide any proof about the inadequacy of the model. However, based on the evidence we have provided in the discussions of this section, the most likely culprits (at least of the scenarios considered in these studies) are the uncertainties in the input parameters and/or the experimental variabilities, which could not be captured in a numerical model.

### 6.2 Quantitative Probabilistic Analysis

Apart from providing insight and physical intuition, CMS is predominantly used for estimating response amplitude from flaws, given a particular inspection method. Due to this reason, and the fact that amplitudes cannot be assessed appropriately in a qualitative analysis, our quantitative analysis uses an amplitude-based metric for assessing the simulations. The metric described by Eq. (5.16) in Section 5.2 will be used to quantify the difference between predictions and observations.

As we have discussed in Section 5.2, the metric considers acquired RF signals at multiple scan points. The analysis in this section will consider the scan point locations represented by the B-scan cross sections at the center (lengthwise) of the reflector. In order to be able to directly compare the results of predictions and observations, the RF B-scan data was normalized by the amplitude of an edge response. To be able to reach any useful conclusions about the validity of the models and simulations, we need to have an estimate of the metric's uncertainty from both predictions and observations.

In Section 6.2.1 we provide an estimate of the simulation uncertainty. This is based on a parametric study, where all the key input parameters of the simulation are varied around their nominal values. Thus, in addition to providing an estimate of simulation uncertainty, the parametric study allows us to find which of the input parameters has the largest effect on the simulation output. Then, in Section 6.2.2 we describe the methodology used to estimate the empirical uncertainty. Using those computed uncertainties, in Section 6.2.3 we compute the metric for both predictions and observations, and provide a discussion of the finding and their implications on the validity of the simulations.

#### 6.2.1 Simulation Uncertainty: Parametric Study

To quantify the effects of the uncertainty in the input parameters on the simulations output, we conducted a comprehensive parametric study. The parametric studies were conducted with six template models, one for each of the six probes. Then, for each template model, the values of 19 different parameters were varied. We recall from Section 5.2.2 that we approximate each of the input parameters ( $p_{ij}$ ) to be

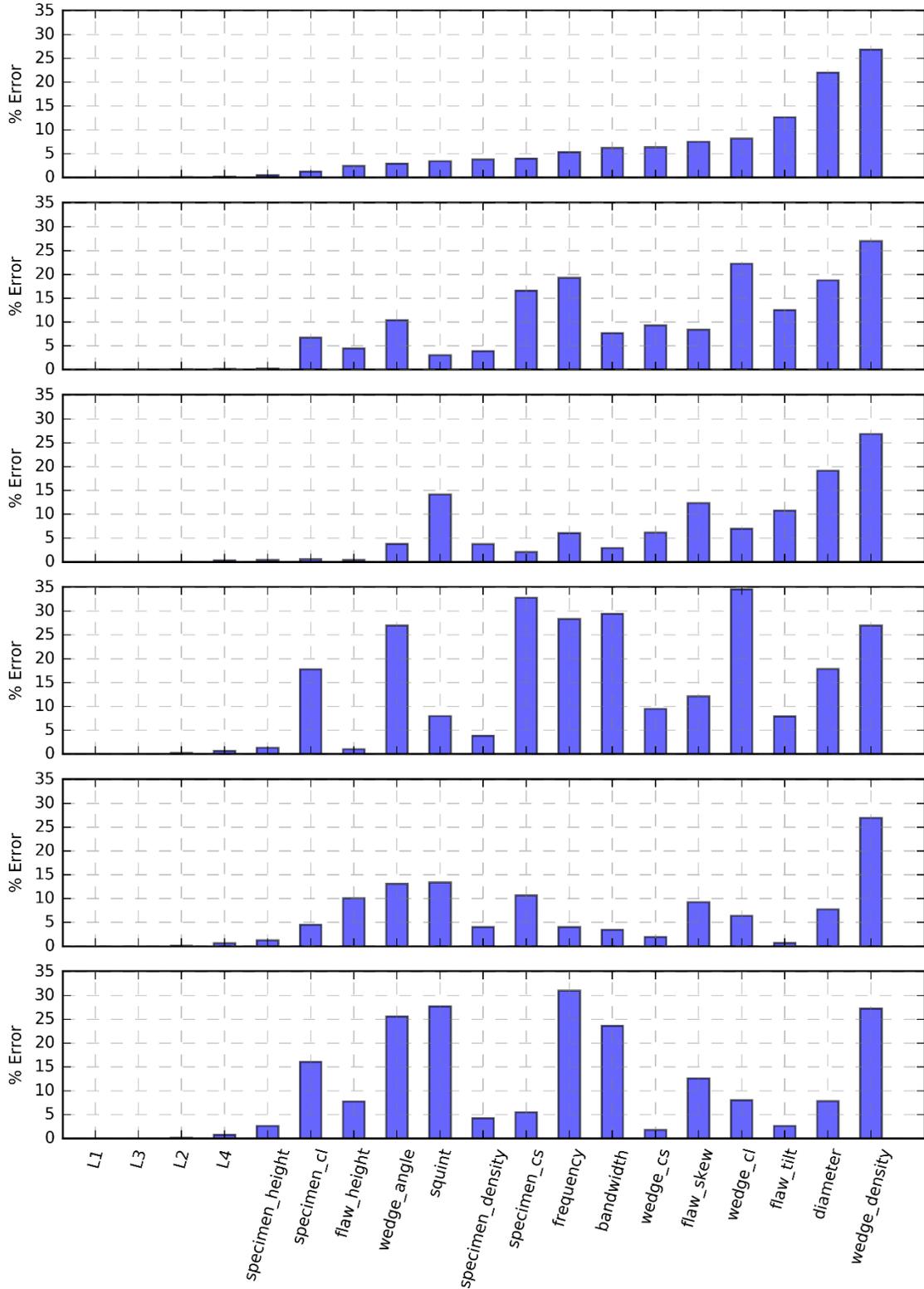
normally distributed with  $p_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij})$ , where  $\mu_{ij}$  is the mean given by the measured nominal value of the respective  $i^{\text{th}}$  parameter and  $j^{\text{th}}$  probe, and  $\sigma_{ij}$  is the associated uncertainty. In this case,  $i = 1, 2, \dots, 19$  corresponds to the 19 different parameters considered in this study, and  $j = 1, 2, \dots, 6$  corresponds to the 6 probes. Thus, for a given probe and input parameter, the values were varied by  $3\sigma_{ij}$  around the nominal value  $\mu_{ij}$ . After specifying all the nominal values and uncertainties for all the model parameters, the uncertainty in the metric from the simulation can be computed as described in Section 5.2.2. Detailed discussions of measuring nominal values and uncertainties for all relevant parameters were given previously in Section 4.3, and summarized in Tables A.2–A.7. The standard error in the output due to each input parameter is computed using Eq. (5.22), and shown in Figure 6.7 for the parametric studies of Probes 1 to 6. The bar plots in Figure 6.7 show the error in the output due to the uncertainty in each of the 19 input parameters. The bar plot for Probe 1 (top plot) is sorted in ascending order, and all other plots keep that same order on the horizontal axis to facilitate comparisons among the six probes. Overall, all errors were below 35%. The parametric studies show that there are certain parameters that have a larger effect on the output than others. In the following, we recall what each parameter represents, and describe their effect on the metric:

- *L1, L2, L3, L4*: The four dimensions that define the probe’s wedge. Those parameters had negligible effect on the metric, with errors less than 1%.
- *wedge\_angle*: The wedge angle of incidence. This parameter determines the refraction angle for the ultrasonic beam in the specimen. Thus, this metric is anticipated to have significant effect physically, especially for incidence angles near the critical angle where the wave ceases to exist. This explains why this parameter did not have a significant effect on Probes 1 and 3 compared to the other probes because the 45° shear probes are further away from the critical angle than the 60° shear probes.
- *wedge\_cl*: The longitudinal wave velocity in the wedge’s material. Similar to the wedge angle, this parameter determines the refraction angle of the ultrasonic beam in the specimen and has a similar effect as the *wedge\_angle*.
- *wedge\_cs, wedge\_density*: The shear wave velocity and density of the probe’s wedge material, respectively. The *wedge\_density* parameter has a significant effect on the metric, and it is one of the most significant parameters for all the probes, with errors more than 25%. This is because the wedge density determines the acoustic impedance of the wedge, and thus controls the energy that is transmitted through the wedge to the specimen. *wedge\_cs* also controls the apparent acoustic incidence of the wedge, but its effect is not as significant as the density.
- *squint*: Related to squint of the probe’s piezoelectric element. This is the angle between the surfaces of the piezoelectric element and wedge. According to the parametric study results, this parameter has a larger effect on both longitudinal probes (5 and 6) and on the 5 MHz, 45° Probe 3.
- *diameter*: Diameter of the probe’s piezoelectric element. It is apparent that this parameter has a significant effect on the metric, especially for the quarter-inch diameter probes (Probes 1–4), with errors over 17%. However, for the half-inch diameter probes (Probes 5 and 6), this parameter had less of an effect, with errors around 8%.
- *frequency, bandwidth*: Properties of the excitation signal for the probe. Those parameters do have significant effects in some of the probes, with errors over 30% for Probe 6. Probes 1, 3, and 5 are not as sensitive, with errors 5% or less.

- *specimen\_height*: The height of the specimen, which is set to a nominal value of 15.4 mm (0.61 in.) for all parametric studies of all the probes, and uncertainty of 0.2 mm (0.01 in.). This parameter has little effect on the metric. The largest error observed due to variation of this parameter is in Probe 6, with an error near 3%. This is a relatively small error when compared with the effect of the other input parameters.
- *specimen\_cs*: The shear wave velocity of the specimen material. The nominal value is set to 3150 m/s for all parametric studies with uncertainty of 42 m/s. This parameter determines the ultrasonic beam angle in the material for the shear wave probes, 1–4. It has a greater effect on the results with Probes 2 and 4, with errors of 16% and 32%, respectively. This is because the nominal refracted angle for those probes is 60°, which is near the critical angle for shear wave propagation. Probes 1 and 3, with nominal beam angles of 45°, had errors that were less than 11%.
- *specimen\_cl*: The longitudinal wave velocity of the specimen material. The nominal value is set to 5750 m/s for all parametric studies with uncertainty of 75 m/s. This parameter determines the ultrasonic beam angle in the material for longitudinal wave probes 5 and 6. It had an effect on the higher frequency 5 MHz Probe 6 (error over 15%) but not on the 2.25 MHz Probe 5 (error less than 5%). Unexpectedly, the longitudinal wave velocity affected the shear wave Probe 4 with error over 15%.
- *specimen\_density*: The density of specimen material. The nominal value is set to 8 g/cm<sup>3</sup> for all parametric studies with uncertainty of 0.2 g/cm<sup>3</sup>. Variation in this parameter had less than 5% error for all probes.
- *flaw\_height*: The nominal flaw height for all the parametric studies was set to 7.5 mm (0.30 in.) with uncertainty of 0.5 mm (0.20 in.). The longitudinal wave probes (Probes 5 and 6) showed more significant errors than the other probes, with 10% and 8% errors, respectively, versus errors less than 5%. Although not investigated in this study, it is anticipated that variations in flaws with a smaller nominal size will demonstrate much larger errors than the 7.5 mm (0.30 in.) flaw height considered in this study.
- *flaw\_skew*: The nominal flaw skew was set to 0° with uncertainty of 0.5°. Flaw skew does have a large impact on the scattered wave amplitudes. Because the uncertainty in the skew in this study is very low (0.5°), the sensitivity to its variation is not as significant; metric errors vary between 7%–13% for the different probes.
- *flat\_tilt*: The nominal flaw skew was set to 0° with uncertainty of 0.5°. This parameter has a small effect with the longitudinal wave probes (Probes 5 and 6) with errors between 1%–3%. Its effect on the shear wave probes is larger with errors between 8%–13%.

Now we have described the effects of all the relevant input parameters on the simulations output using the metric that we have derived in Section 5.2. Even though considerable efforts had been made to reduce uncertainties in the simulations input parameters (as described in Section 4.3), the results show that the input parameters still have considerable effects on the simulation output uncertainty (up to 35% variations). Thus, depending on the intended use of the model and the level accuracy of the model, it may be pertinent to conduct a parametric study or Monte-Carlo simulations to reflect the effects of uncertainties.

Another purpose of the parametric study is to estimate the effective simulation uncertainty, as given in Eq. (5.23). We will use the uncertainty computed from this parametric study as the uncertainty on all the nominal simulations results from the validation tree.



**Figure 6.7.** Parametric Studies Error for the 19 Parameters Investigated. The plots correspond to parametric studies related to Probes 1–6 from top to bottom, respectively.

## 6.2.2 Empirical Uncertainty

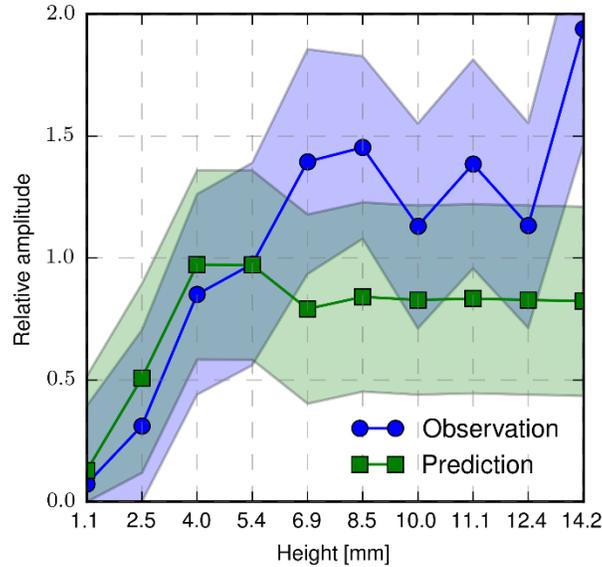
As we discussed in Section 5.2, the uncertainty associated with the experiment is represented by the thermal noise power  $\sigma^2$  and the interference term power  $\sigma_I^2$ . The values for the thermal noise are computed by taking the Fourier transform of each time-series in the B-scans images and computing the average power in the spectrum starting from 20 MHz up to the maximum frequency in the specimen, which is half of the sampling frequency (50 MHz). The reason we compute the average power above 20 MHz is to exclude the possibility of any coherent signal power in the lower frequency bands of the spectrum.

To compute the interference term power  $\sigma_I^2$ , the time-series signals due to the reflections from the specimen edge are extracted from each of the empirical measurements raster scans. Then, variance of the power of each time-series signal is computed. This variance gives us an indication of the uncertainties in the experimental procedures due to variations in coupling and equipment.

## 6.2.3 Effective Metric

The metric based on Eq. (5.16) is computed for each of the predictions and observations based on all the validation tree scenarios described in Section 4.0. The computed values are shown in Appendix C. Those values are considered the mean values, and the confidence in this computed mean value is shown as bands around it. The confidence is computed using Eq. (5.18) for the empirical data, and Eq. (5.23) for the simulations. The metric value is shown as a function of the reflector height for scans from specimen 1, as a function of reflector area for scans from specimen 2, and as a function of reflector tilt for scans from specimen 3.

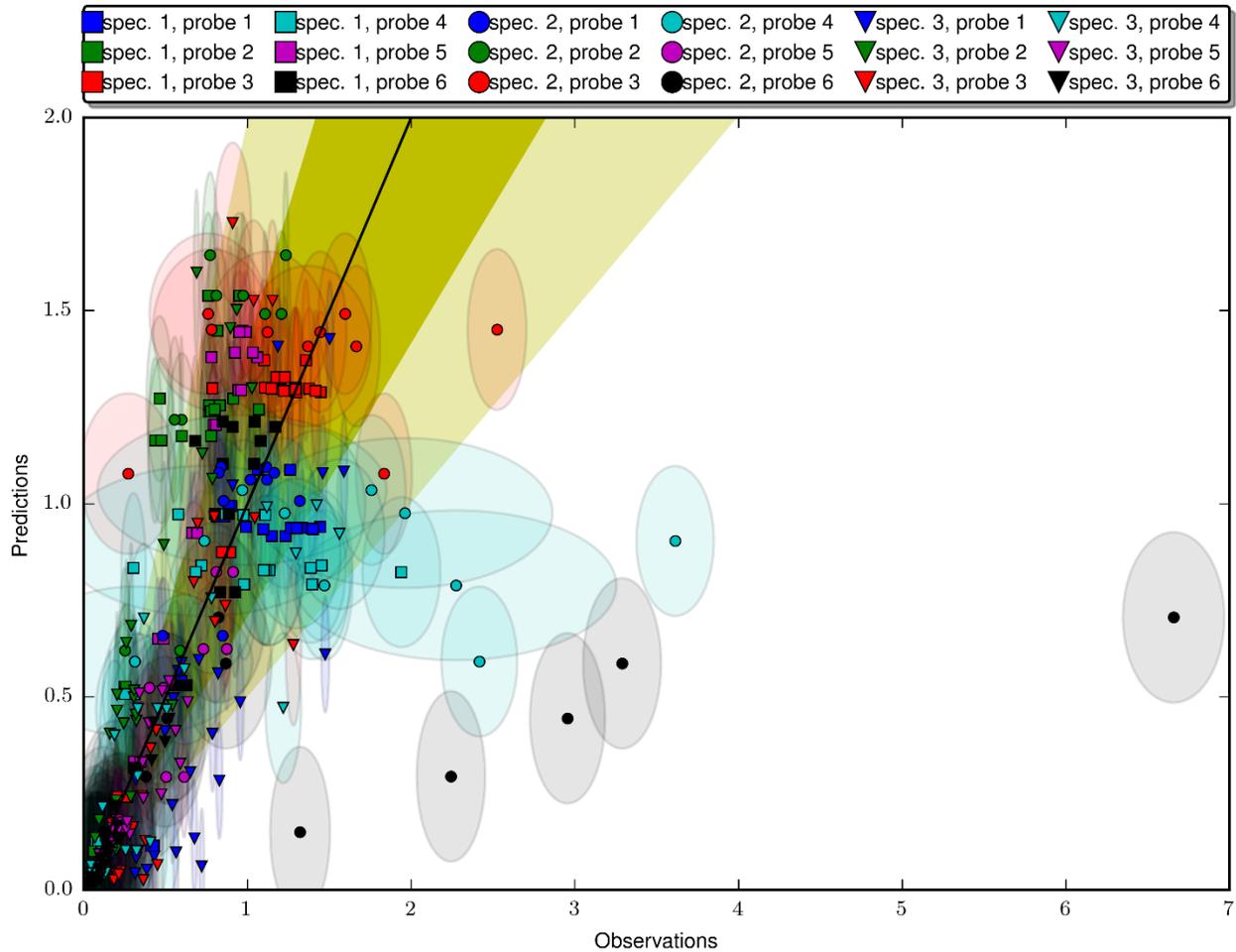
The computed metric has large variations in the differences between predicted and observed results, depending on the specimen, probe, reflector, and scan skew angle. Also, the associated confidence in the computed metric varies considerably for both the predicted and observed scans. For example, we consider a specific case, related to the raster scans made on specimen 1, Probe 4, and scan skew angle  $0^\circ$ . The results of the metric computation are shown in Figure 6.8, which is a replica of the figure in Appendix B for convenience. If the bands for predictions and observations are mostly overlapping, this is an indication that differences are mostly due to uncertainties considered in this study. If the bands of the predictions diverge away from those of the observations (or vice versa), then this is an indication on the presence of biases that have not been accounted for in this study. In Figure 6.8, for reflector heights less than 5.4 mm (0.21 in.), the difference between predictions and simulations is small compared to the values for reflector heights of more than 5.4 mm (0.21 in.). For a reflector height of 14.2 mm (0.56 in.), the observation value diverges from the predicted value, with the value more than two times larger than the prediction. In this case, based on the expected trend of signal amplitudes with reflector height in ultrasound, there is some bias that is introduced into the observations when the probe scans the 14.2 mm (0.56 in.) reflector. Although we can infer from this figure that the observations have some bias, at least for larger reflectors, there is nothing that can be said about the predictions. If the largest difference between observations and predictions, which in this case is about 7 dB, is larger than the level of accuracy required for our intended use, then we cannot accept the model.



**Figure 6.8.** Results of the Metric Computation for Both Observation and Prediction from Scans in Specimen 1, Probe 4, Skew Angle  $0^\circ$ . The bands around the line show the confidence intervals.

By investigating the metric computation results for all the specimens, probes, and skew angles, it is difficult to reach a conclusion about which particular cases of the model can be rejected, and for which cases it should be accepted. To provide a combined view of the results from all the scans, we show the metrics (from 6 probes, 29 flaws, and 2 skews, or 348 calculations) as points in the Observation-Prediction space. Figure 6.9 shows the amplitude of the observation metric on the X-axis and the prediction metric on the Y-axis. The coordinates of each point are represented as (prediction value, observation value) for the same specimen, probe, skew, and reflector. Around each point, the confidence interval is shown as an ellipse, where the focal length in the X-direction represents the observation uncertainty, and the focal length in the Y-direction represents the prediction uncertainty. If the observation is exactly the same as the prediction, all the points should lie on the line with slope 1, and is shown in Figure 6.9 as a solid black line. Around the solid black line, there are two yellow bands. The inner yellow band (darker color) shows the region where the predictions and observations differ by less than 3 dB. The outer yellow band (lighter color) shows the region where the predictions and observations differ by less than 6 dB. The farther away the points are from those shaded regions, the larger is the error.

Figure 6.9 clearly shows two cases that are outliers, due to empirical bias. Those are the scans from (specimen 2, Probe 4, skew 0) and (specimen 2, Probe 6, skew 0). The five points of the scans from (specimen 2 and Probe 6, skew 0) representing the five reflectors in this specimen, lie quite far from the 6 dB error region. Further investigation of those scans found that this is due to bias in the response from the specimen's edge corner, which is used for normalization. This is further confirmed by the fact that the scans with skew 180 are all within the 6 dB region, although ideally should be the same as the skew 0 scans. Except for those two obvious cases, most of the points lie within the 6 dB region, with few outliers.



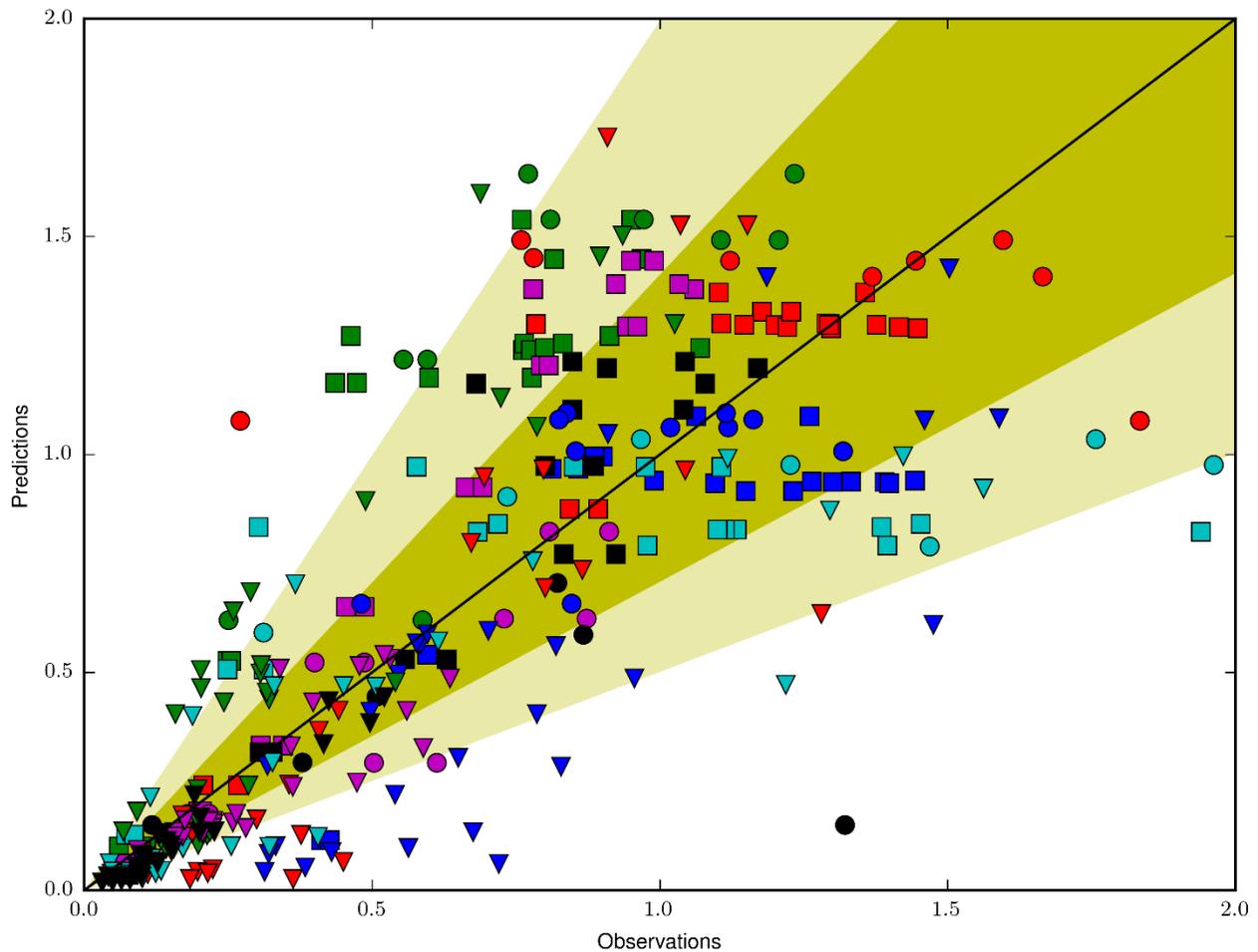
**Figure 6.9.** Computed Metric in the Observation-Prediction Space. Each point has an ellipse around it, indicating the uncertainty in its coordinates.

Figure 6.10 provides an enlarged view of the points near the origin. For clarity, the confidence areas are not shown here. As can be seen, most points lie within the 6 dB region, with some outliers. In total 71 points out of a total of 348 (20%) lie outside the 6 dB, and 169 points of the 348 (49%) lie outside the 3 dB region. However, this does not take into account the uncertainty in the coordinates, which might pull a point in or push it out of the 3/6 dB regions. For example, consider the case shown in Figure 6.11. This shows the metric for the reflectors only in specimen 1, probe 4, and skew 0. The uncertainty in the metric value could shift a point that is within the 6 dB into the 3 dB or into the 20 dB. It is obvious that points with lower amplitudes (small reflectors) are more affected by the uncertainty than points with higher amplitudes.

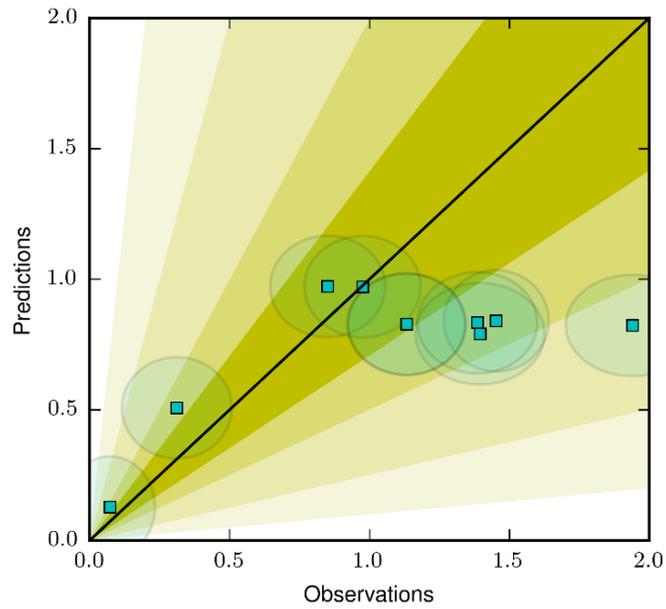
In Figure 6.10, points that are above the line with slope 1 have prediction values that are larger than the observations. On the other hand, points below the line have prediction values that are smaller than the observations. Out of the 71 outliers, only 20 were above the slope 1 line, while 51 points were below the lines. We have already shown that the points in specimen 2, Probe 6 have artificially large empirical values due to a bias in the specimen edge corner response. The fact that we have more points below the slope 1 line might be an indication that this bias is also occurring in other scans.

In summary, the scatter in the metric amplitudes varied from less than 3 dB, which is considered acceptable for most applications of UT, up to over 20 dB, which should be unacceptable to most UT applications. There is no specific trend in which specific combinations of (specimen, probe, skew) performed better. Taking sources of uncertainties into account, each validation point will have a domain that is represented by the ellipses in Figure 6.9. This domain provides us with a range of values, controlling the probability that the validation point will be within the 3 dB or 6 dB error regions.

If after accounting for the uncertainty, the error in the validation point is still unacceptable, then we attribute this to biases in our validation. It cannot be discerned whether the source of those biases are from the simulation or experiment, but this makes no difference, because we have shown that the simulation is not able to represent reality.



**Figure 6.10.** Computed Metric in the Observation-Prediction Space. This is the same as Figure 6.9, but provides a close up view of the points near the 6 dB error region.



**Figure 6.11.** Computed Metric in the Observation-Prediction Space for the Reflectors in Specimen 1, Probe 3, Skew 0. The four yellow bands represent the error regions corresponding to 3, 6, 12, and 20 dB.



## 7.0 Conclusions and Future Work

This report describes the validation of computational models used in modeling UT for commercial nuclear applications. The study investigated the response of machined notch reflectors in homogeneous stainless steel plates using different conventional ultrasonic probes. A database of a total of 348 simulated and experimental scans was generated and analyzed. This database includes notches of varying sizes and orientations, and six standard commercial ultrasonic probes using different frequencies, wave modes, diameters, and refraction angles.

We generated the simulations using CIVA, a semi-analytical modeling software that provides an efficient platform for simulating UT problems. CIVA was selected due to its popularity in the NDE industry for NPPs. We discussed computational models and physics engines used by CIVA, along with the approximation and assumptions used by those models. A numerical example was given based on the Kirchhoff approximate model, to provide an explanation about the nature of inaccuracies that are incurred by such approximate models.

We provided detailed descriptions of all the materials, probes, and scans that were considered in this study. We described the details related to the required inputs for defining simulations in CIVA, and we provided rigorous descriptions about the methods used to obtain or measure those input parameters, along with approximations of the uncertainties associated with the inputs parameters.

We proposed a two-step framework for validating the simulation results. The first step is based on qualitative criteria, and the second step is based on a quantitative probabilistic approach. The qualitative analysis is an informal procedure designed so that an SME could validate that the simulations adequately represent UT responses.

In the quantitative approach, different metrics could be considered, depending on the intended use of the model. In our proposed framework, we considered an amplitude-based metric that reduces a B-scan, represented by multidimensional spatio-temporal RF signals, into a scalar value. The reason why an amplitude-based metric was considered is because the use of amplitude is pervasive in UT modeling for quantifying the probability of detection. The quantitative approach accounts for sources of uncertainties in both measurements of model inputs. The findings using the proposed validation framework show that CIVA models were able to give the correct responses qualitatively, but there were substantial differences in the quantitative validation metric due to biases that cannot be represented by a model.

The results of the validation study allow us to reach the following conclusion:

*The procedures used for defining a model have significant effect on the model's output. Although models are useful for providing insight and visualizing wave fields, there is no evidence that models can be used to replace experiments in applications that require quantification of amplitudes. This is because there is a large uncertainty in the model's output associated with uncertain, and in some cases unknown, input parameters, and uncertainty and biases in experimental procedures.*

Another aspect we would like to touch upon is that it is essential for modeling software to be verified before conducting any validation study. Also, documentation of such verification is of utter importance. Although CIVA has an extensive validation database published, there is no publication of any verification benchmarks of the software. This is required to prove that the software has a correct numerical implementations of the underlying computational models.

As PNNL continues to assess the utility and limitations of these types of modeling tools, and as these efforts evolve through Phases 2 and 3, it is critical that the impact of these limitations on the resultant accuracy of the simulated results is defined and better understood. Additionally, future work should address the findings in terms of how the Regulator's confidence in CIVA results might be impacted by performance limitations identified in this study.

PNNL only considered homogeneous isotropic materials in this study, and left heterogeneous and more realistic cracks for future work. All the inputs for the cases considered thus far can be directly measured empirically, or are well known and documented. However, when considering complex cracks, different morphologies of cracks are present, and it is often expensive to fully characterize a crack. Even if it is determined, it is inefficient to implement such level of details and granularity in a model. The question is, how necessary is it? Additionally, heterogeneous materials, which result in backscatter of an ultrasonic beam, often require defining material properties that are difficult to measure empirically. These issues will be investigated in future work.

## 8.0 References

- AIAA. 1998. *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*. AIAA-G-077-1998, The American Institute of Aeronautics and Astronautics (AIAA), Reston, Virginia.
- ASME. 2006. *Guide for Verification and Validation in Computational Solid Mechanics*. ASME Standard V V 10 - 2006, The American Society of Mechanical Engineers (ASME), New York, New York.
- Auld BA. 1973. *Acoustic Fields and Waves in Solids, Volume 1*, John Wiley & Sons, New York.
- Bayarri MJ, JO Berger, J Cafeo, G Garcia-Donato, F Liu, J Palomo, RJ Parthasarathy, R Paulo, J Sacks and D Walsh. 2007a. "Computer Model Validation with Functional Output." *The Annals of Statistics* 35(5):1874-1906.
- Bayarri MJ, JO Berger, R Paulo, J Sacks, JA Cafeo, J Cavendish, C-H Lin and J Tu. 2007b. "A Framework for Validation of Computer Models." *Technometrics* 49(2):138-154. DOI: 10.1198/004017007000000092.
- Box GE and NR Draper. 1987. *Empirical Model-Building and Response Surfaces*, Wiley, New York.
- Calmon P, S Mahaut, S Chatillon and R Raillon. 2006. "CIVA: An Expertise Platform for Simulation and Processing NDT Data." *Ultrasonics* 44, Supplement:e975-e979. DOI: 10.1016/j.ultras.2006.05.218.
- CEA. 2014. *CIVA 11 User Manual*. Commissariat à l'Energie Atomique (CEA), Paris, France.
- Červený V. 2001. *Seismic Ray Theory*, Cambridge University Press, Cambridge, New York.
- Chassignole B, O Dupond, T Fouquet, AL Brun, J Moysan and P Benoist. 2011. "Application of Modelling for Enhanced Ultrasonic Inspection of Stainless Steel Welds." *Welding in the World* 55(7-8):75-82. DOI: 10.1007/bf03321310.
- Darmon M and S Chatillon. 2013. "Main Features of a Complete Ultrasonic Measurement Model: Formal Aspects of Modeling of Both Transducers Radiation and Ultrasonic Flaws Responses." *Open Journal of Acoustics* 3:43-53. DOI: 10.4236/oja.2013.33A008.
- Darmon M, A Ferrand, V Dorval, S Chatillon and S Lonne. 2014. "Recent Modelling Advances for Ultrasonic TOFD Inspections." In *41st Annual Review of Progress in Quantitative Nondestructive Evaluation*, pp. 1757-1765. July 20-25, 2014, Boise, Idaho. DOI 10.1063/1.4914799. American Institute of Physics.
- Deschamps GA. 1972. "Ray Techniques in Electromagnetics." *Proceedings of the IEEE* 60(9):1022-1035. DOI: 10.1109/PROC.1972.8850.
- Dib G and L Udpa. 2016. "Design and Performance of Optimal Detectors for Guided Wave Structural Health Monitoring." *Structural Health Monitoring* 15(1):21-37. DOI: 10.1177/1475921715620003.
- Dupond O, V Duwig and T Fouquet. 2009. "Influence of Stress Corrosion Crack Morphology on Ultrasonic Examination Performances." In *Proceedings of the 35th Annual Review of Progress in Quantitative Nondestructive Evaluation*, pp. 89-96. July 20-25, 2008, Chicago, Illinois. DOI 10.1063/1.3114353. American Institute of Physics, Melville, New York. Available at <http://scitation.aip.org/content/aip/proceeding/aipcp/10.1063/1.3114353>.

- Gardahaut A, H Lourme, F Jensen, S Lin and M Nagai. 2014. "Ultrasonic Wave Propagation in Dissimilar Metal Welds - Application of a Ray-Based Model and Comparison with Experimental Results." In *11th European Conference on Non-Destructive Testing (ECNDT 2014)*, pp. 302-311. October 6-10, 2014, Prague, Czech Republic. European Federation for Non-Destructive Testing.
- Graff KF. 1975. *Wave Motion in Elastic Solids*, Dover Publications, Inc., Mineola, New York. ISBN 0486667456.
- Hatton L. 1997. "The T Experiments: Errors in Scientific Software." *IEEE Computational Science & Engineering* 4(2):27-38. DOI: 10.1109/99.609829.
- Hellier CJ. 2003. *Handbook of Nondestructive Evaluation*, McGraw-Hill.
- Higdon D, J Gattiker, B Williams and M Rightley. 2008. "Computer Model Calibration Using High-Dimensional Output." *Journal of the American Statistical Association* 103(482):570-583. DOI: 10.1198/016214507000000888.
- IEEE. 2010. *Systems and Software Engineering -- Vocabulary*. ISO/IEC/IEEE-24765, Institute of Electrical and Electronics Engineers, Inc. (IEEE), New York, New York.
- Kay SM. 1994. *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall.
- Kay SM. 1998. *Fundamentals of Statistical Signal Processing Volume II: Detection Theory*, Prentice Hall.
- Kennedy MC and A O'Hagan. 2001. "Bayesian Calibration of Computer Models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(3):425-464. DOI: 10.1111/1467-9868.00294.
- Mahaut S, S Lonne, L de Roumilly and G Cattiaux. 2006. "Validation of CIVA Simulation Tools for Ultrasonic Inspection in Realistic Configuration." In *9th European Conference on NDT*. September 25-29, 2006, Berlin, Germany. Session We.1.4.4.
- Marioli D, C Narduzzi, C Offelli, D Petri, E Sardini and A Taroni. 1992. "Digital Time-of-Flight Measurement for Ultrasonic Sensors." *IEEE Transactions on Instrumentation and Measurement* 41(1):93-97.
- Meyer RM, LJ Bond and P Ramuhalli. 2012. "Online Condition Monitoring to Enable Extended Operation of Nuclear Power Plants." *International Journal of Nuclear Safety and Simulation* 3(1):31-50.
- Oberkampf WL and MF Barone. 2006. "Measures of Agreement between Computation and Experiment: Validation Metrics." *Journal of Computational Physics* 217(1):5-36. DOI: 10.1016/j.jcp.2006.03.037.
- Oberkampf WL, SM DeLand, BM Rutherford, KV Diegert and KF Alvin. 2002. "Error and Uncertainty in Modeling and Simulation." *Reliability Engineering & System Safety* 75(3):333-357. DOI: 10.1016/S0951-8320(01)00120-X.
- Oberkampf WL, JC Helton, CA Joslyn, SF Wojtkiewicz and S Ferson. 2004. "Challenge Problems: Uncertainty in System Response Given Uncertain Parameters." *Reliability Engineering & System Safety* 85(1-3):11-19. DOI: 10.1016/j.res.2004.03.002.

- Oberkampf WL and CJ Roy. 2010. *Verification and Validation in Scientific Computing*, Cambridge University Press, United Kingdom.
- Oberkampf WL and TG Trucano. 2002. *Verification and Validation in Computational Fluid Dynamics*. SAND2002-0529, Sandia National Laboratories, Albuquerque, New Mexico.
- Oberkampf WL and TG Trucano. 2008. "Verification and Validation Benchmarks." *Nuclear Engineering and Design* 238(3):716-743. DOI: 10.1016/j.nucengdes.2007.02.032.
- Oreskes N, K Shrader-Frechette and K Belitz. 1994. "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences." *Science* 263(5147):641-646.
- Raillon R, G Toullelan, M Darmon, P Calmon and S Lonne. 2012. "Validation of CIVA Ultrasonic Simulation in Canonical Configurations." In *18th World Conference on Non-destructive Testing*, pp. 1013-1022. April 16-20, 2012, Durban, South Africa. British Institute of Non-Destructive Testing, Northampton, United Kingdom. Available at [http://www.ndt.net/article/wcndt2012/papers/308\\_wcndtfinal00308.pdf](http://www.ndt.net/article/wcndt2012/papers/308_wcndtfinal00308.pdf).
- Raillon R, G Toullelan, M Darmon and S Lonne. 2013. "Experimental Study for the Validation of CIVA Predictions in TOFD Inspections." In *10th International Conference on NDE in Relation to Structural Integrity for Nuclear and Pressurized Components*, pp. 660-667. October 1-3, 2013, Cannes, France. Available at [http://www.ndt.net/article/wcndt2012/papers/308\\_wcndtfinal00308.pdf](http://www.ndt.net/article/wcndt2012/papers/308_wcndtfinal00308.pdf).
- Rebba R and S Mahadevan. 2008. "Computational Methods for Model Reliability Assessment." *Reliability Engineering & System Safety* 93(8):1197-1207. DOI: 10.1016/j.res.2007.08.001.
- Rose JL. 2004. *Ultrasonic Waves in Solid Media*, Cambridge University Press, Cambridge, United Kingdom.
- Roth DJ, RP Tokars, RE Martin, RW Rauser and JC Aldrin. 2009. "Ultrasonic Phased Array Simulations of Welded Components at NASA." In *Proceedings of the 35th Annual Review of Progress in Quantitative Nondestructive Evaluation*, pp. 1190-1197. July 20-25, 2008, Chicago, Illinois. DOI 10.1063/1.3114090. American Institute of Physics, Melville, New York.
- Salari K and P Knupp. 2000. *Code Verification by the Method of Manufactured Solutions*. SAND2000-1444, Sandia National Laboratories, Albuquerque, New Mexico. DOI: 10.2172/759450.
- Sankararaman S and S Mahadevan. 2015. "Integration of Model Verification, Validation, and Calibration for Uncertainty Quantification in Engineering Systems." *Reliability Engineering & System Safety* 138:194-209. DOI: 10.1016/j.res.2015.01.023.
- Schlesinger S. 1979. "Terminology for Model Credibility." *SIMULATION* 32(3):103-104. DOI: 10.1177/003754977903200304.
- Schmerr Jr. LW. 1998. *Fundamentals of Ultrasonic Nondestructive Evaluation: A Modeling Approach*, Springer US, New York. ISBN 978-1-4899-0144-6.
- Schmerr Jr. LW and S-J Song. 2007. *Ultrasonic Nondestructive Evaluation Systems – Models and Measurements*, Springer, New York.

Stohrer C. 2013. *Finite Element Heterogeneous Multiscale Methods for the Wave Equation*. Ph.D. Thesis, University of Basel, Basel, Switzerland.

Van Trees HL, KL Bell and Z Tian. 2013. *Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory, 2nd Edition*, John Wiley & Sons, Inc. ISBN 978-0-470-54296-5.

## **Appendix A**

### **Summary of Specimen and Probes Dimensions**



# Appendix A

## Summary of Specimen and Probes Dimensions

The nominal values and their associated uncertainties for the plate specimen, probes, and reflectors are given in this appendix.

### A.1 Specimen and Flaws

**Table A.1.** Properties of the Plates Considered for the Validation Study. Note that  $c_L$  is the longitudinal wave velocity and  $c_S$  is the shear wave velocity in the plate.

Plate ID	Length, mm (in.)	Width, mm (in.)	Thickness, mm (in.)	Number of Reflectors	$c_L$ (m/s)	$c_S$ (m/s)	Density (g/cm <sup>3</sup> )
1	356 (14.02)	102 (4.02)	15.4 (0.61)	10	5743	3133	8.00
2	306 (12.05)	105 (4.13)	15.1 (0.59)	5	5760	3150	8.00
3	406 (15.98)	102 (4.02)	15.4 (0.61)	14	5756	3150	8.00

**Table A.2.** Uncertainty Associated with the Measured/Assigned Plates Properties

Plate ID	Thickness, mm (in.)	$c_L$ (m/s)	$c_S$ (m/s)	Density (g/cm <sup>3</sup> )
1	0.2 (0.0079)	75	42	0.20
2	0.2 (0.0079)	75	42	0.20
3	0.2 (0.0079)	75	42	0.20

**Table A.3.** Plate 1 Reflector Dimensions

Reflector ID	Length, mm (in.)	Width, mm (mil)	Depth, mm (in.)	Depth Uncertainty mm (in.)	Tilt (degrees)	Tilt Uncertainty (degrees)
1-A	26.0 (1.02)	0.43 (17)	1.1 (0.043)	0.5 (0.020)	0	0.5
1-B	26.0 (1.02)	0.43 (17)	2.5 (0.098)	0.5 (0.020)	0	0.5
1-C	26.0 (1.02)	0.46 (18)	4.0 (0.157)	0.5 (0.020)	0	0.5
1-D	26.0 (1.02)	0.46 (18)	5.4 (0.213)	0.5 (0.020)	0	0.5
1-E	26.0 (1.02)	0.46 (18)	6.9 (0.272)	0.5 (0.020)	0	0.5
1-F	26.0 (1.02)	0.46 (18)	8.5 (0.335)	0.5 (0.020)	0	0.5
1-G	26.0 (1.02)	0.46 (18)	10.0 (0.394)	0.5 (0.020)	0	0.5
1-H	26.0 (1.02)	0.58 (23)	11.1 (0.260)	0.5 (0.020)	0	0.5
1-I	26.0 (1.02)	0.48 (19)	12.4 (0.488)	0.5 (0.020)	0	0.5
1-J	26.0 (1.02)	0.46 (18)	14.2 (0.559)	0.5 (0.020)	0	0.5

**Table A.4.** Plate 2 Reflector Dimensions

Reflector ID	Length mm (in.)	Width mm (mil)	Depth mm (in.)	Depth Uncertainty mm (in.)	Tilt (degrees)	Tilt Uncertainty (degrees)
2-A	42.0 (1.653)	0.15 (5.91)	6.5 (0.256)	0.5 (0.020)	0	0.5
2-B	30.0 (1.181)	0.10 (3.94)	5.8 (0.228)	0.5 (0.020)	0	0.5
2-C	22.5 (0.886)	0.10 (3.94)	5.0 (0.197)	0.5 (0.020)	0	0.5
2-D	19.0 (0.748)	0.10 (3.94)	4.0 (0.157)	0.5 (0.020)	0	0.5
2-E	15.0 (0.590)	0.13 (5.12)	2.8 (0.110)	0.5 (0.020)	0	0.5

**Table A.5.** Plate 3 Reflector Dimensions

Reflector ID	Length mm (in.)	Width mm (mil)	Depth mm (in.)	Depth Uncertainty mm (in.)	Tilt (degrees)	Tilt Uncertainty (degrees)
3-A	26.0 (1.02)	0.41 (16.14)	2.5 (0.098)	0.5 (0.020)	0.0	0.5
3-B	26.0 (1.02)	0.41 (16.14)	2.5 (0.098)	0.5 (0.020)	2.5	0.5
3-C	26.0 (1.02)	0.41 (16.14)	2.5 (0.098)	0.5 (0.020)	5.0	0.5
3-D	26.0 (1.02)	0.41 (16.14)	2.5 (0.098)	0.5 (0.020)	7.5	0.5
3-E	26.0 (1.02)	0.41 (16.14)	2.5 (0.098)	0.5 (0.020)	10.0	0.5
3-F	26.0 (1.02)	0.41 (16.14)	2.5 (0.098)	0.5 (0.020)	15.0	0.5
3-G	26.0 (1.02)	0.41 (16.14)	2.5 (0.098)	0.5 (0.020)	20.0	0.5
3-H	26.0 (1.02)	0.43 (16.93)	5.0 (0.197)	0.5 (0.020)	0.0	0.5
3-I	26.0 (1.02)	0.41 (16.14)	5.0 (0.197)	0.5 (0.020)	2.5	0.5
3-J	26.0 (1.02)	0.43 (16.93)	5.0 (0.197)	0.5 (0.020)	5.0	0.5
3-K	26.0 (1.02)	0.43 (16.93)	5.0 (0.197)	0.5 (0.020)	7.5	0.5
3-L	26.0 (1.02)	0.46 (18.11)	5.0 (0.197)	0.5 (0.020)	10.0	0.5
3-M	26.0 (1.02)	0.46 (18.11)	5.0 (0.197)	0.5 (0.020)	15.0	0.5
3-N	26.0 (1.02)	0.46 (18.11)	5.0 (0.197)	0.5 (0.020)	20.0	0.5

## A.2 Probes

**Table A.6.** Summary of the Probe Properties Including Wedge Dimensions and Material, and Transducer Dimensions and Frequency Response

Probe ID	L <sub>1</sub> , mm (in.)	L <sub>2</sub> , mm (in.)	L <sub>3</sub> , mm (in.)	L <sub>4</sub> , mm (in.)	Incidence			c <sub>L</sub> (m/s)	c <sub>s</sub> (m/s)	Density (g/cm <sup>3</sup> )	Diameter mm (in.)	Freq. (MHz)	BW %	Mode
					Angle (degrees)	Squint (degrees)	c <sub>L</sub> (m/s)							
1	8.8 (0.346)	9.4 (0.370)	11.4 (0.449)	4.4 (0.173)	36.7	0	2680	1320	1.18	6.35 (0.25)	2.2	64	SV	
2	10.4 (0.409)	10.4 (0.409)	11.4 (0.449)	6.6 (0.260)	47.5	0	2680	1320	1.18	6.35 (0.25)	2.0	69	SV	
3	8.8 (0.346)	9.4 (0.370)	11.4 (0.449)	4.4 (0.173)	36.7	0	2680	1320	1.18	6.35 (0.25)	4.4	73	SV	
4	10.4 (0.409)	10.4 (0.409)	11.4 (0.449)	6.6 (0.260)	47.5	0	2680	1320	1.18	6.35 (0.25)	4.0	76	SV	
5	26.4 (1.039)	18.4 (0.724)	28 (1.102)	26.6 (1.047)	19.24	0	2680	1320	1.18	12.7 (0.50)	2.25	28	L	
6	26.4 (1.039)	18.4 (0.724)	28 (1.102)	26.6 (1.047)	19.24	0	2680	1320	1.18	12.7 (0.50)	4.8	41	L	

**Table A.7.** Summary of the Uncertainties in Measured/Assumed Probe Properties

Probe ID	Incidence										Diameter mm (in.)	Freq. (MHz)	BW %
	L <sub>1</sub> , mm (in.)	L <sub>2</sub> , mm (in.)	L <sub>3</sub> , mm (in.)	L <sub>4</sub> , mm (in.)	Angle (degrees)	Squint (degrees)	c <sub>L</sub> (m/s)	c <sub>s</sub> (m/s)	Density (g/cm <sup>3</sup> )				
1	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	1.0	0.5	50	50	0.1	0.1	0.2 (0.0079)	0.05	2.0
2	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	1.0	0.5	50	50	0.1	0.1	0.2 (0.0079)	0.1	3.5
3	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	1.0	0.5	50	50	0.1	0.1	0.2 (0.0079)	0.1	2.0
4	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	1.0	0.5	50	50	0.1	0.1	0.2 (0.0079)	0.3	13.0
5	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	1.0	0.5	50	50	0.1	0.1	0.2 (0.0079)	0.05	1.0
6	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	0.2 (0.0079)	1.0	0.5	50	50	0.1	0.1	0.2 (0.0079)	0.3	6.0

## **Appendix B**

### **Composite Views of Ultrasonic Results**



## Appendix B

### Composite Views of Ultrasonic Results

In the following, the images generated from the empirical observations and model predictions are shown for all the scenarios described in validation tree from Section 4.0. Those composite views are based on the 2D raster scan.

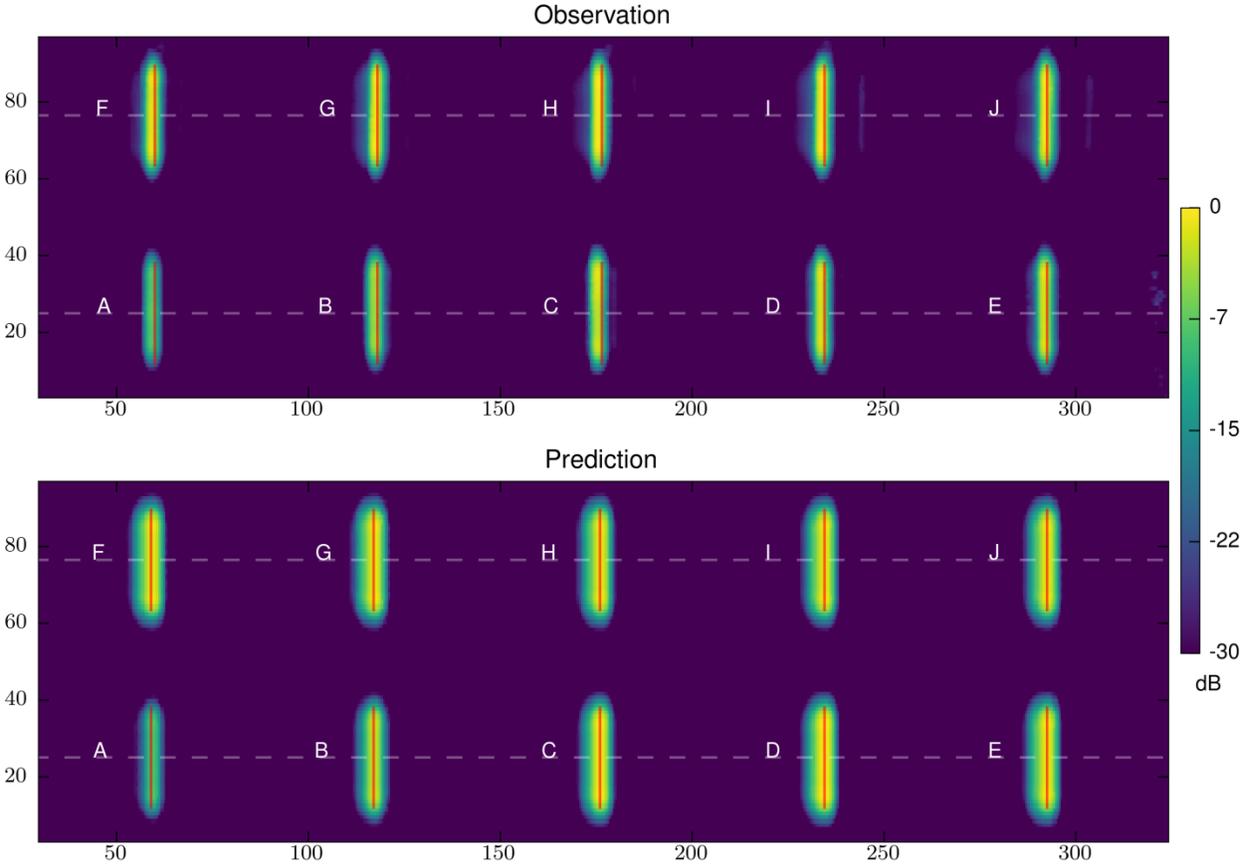
The C-scan views provide a top view of the raster scan. The units for all the coordinates are in millimeters. Note only Probes 1–4 were modeled for raster scan data; therefore, only these probes have C-scan images (see Section 4.3.5 for more details). Dashed horizontal lines represent the mid-point of the flaw and the  $y$ -axis position where the B-scans will be extracted. The red vertical lines represent flaw length as measured by radiography.

The B-scan views provide a cross-sectional view of the raster scan. The cross-section location that was used for plotting the B-scan is indicated in the C-scan images with dashed white lines. On the B-scan, solid white lines indicate the locations of the specimen back-wall and the cross-sectional profile of the flaw. The units for all the coordinates are in millimeters. For each reflector, there are two diagonal dashed lines associated with it. One line is drawn through the corner between the specimen back-wall and the base of the reflector. The other line is drawn through the tip of the reflector. These two lines indicate the locations where the A-scan views have been extracted for each reflector.

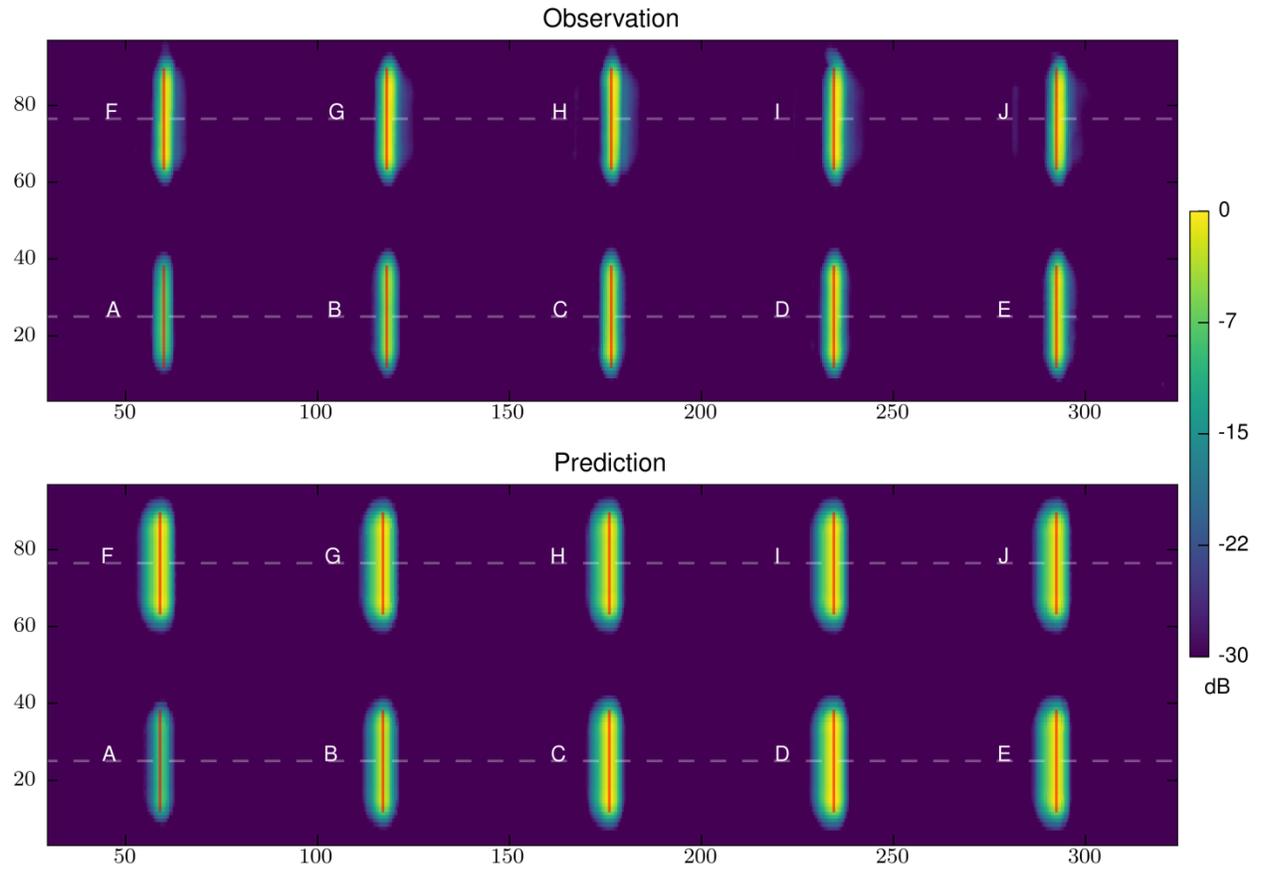
The A-scan views show the RF signal at a single scan position. For each reflector, two A-scan views are shown. One A-scan view represents an ultrasonic beam incident to the corner made by the back-wall and the reflector. The other A-scan view represents the ultrasonic beam incident on the tip of the notch reflector. In the B-scans, diagonal lines are drawn to indicate the positions where the A-scans were extracted. The units of the horizontal axis are in millimeters, representing the true-depth. The vertical axis is unitless, representing the normalized amplitudes, relative to a corner-echo amplitude.

# B.1 C-scan Image Views

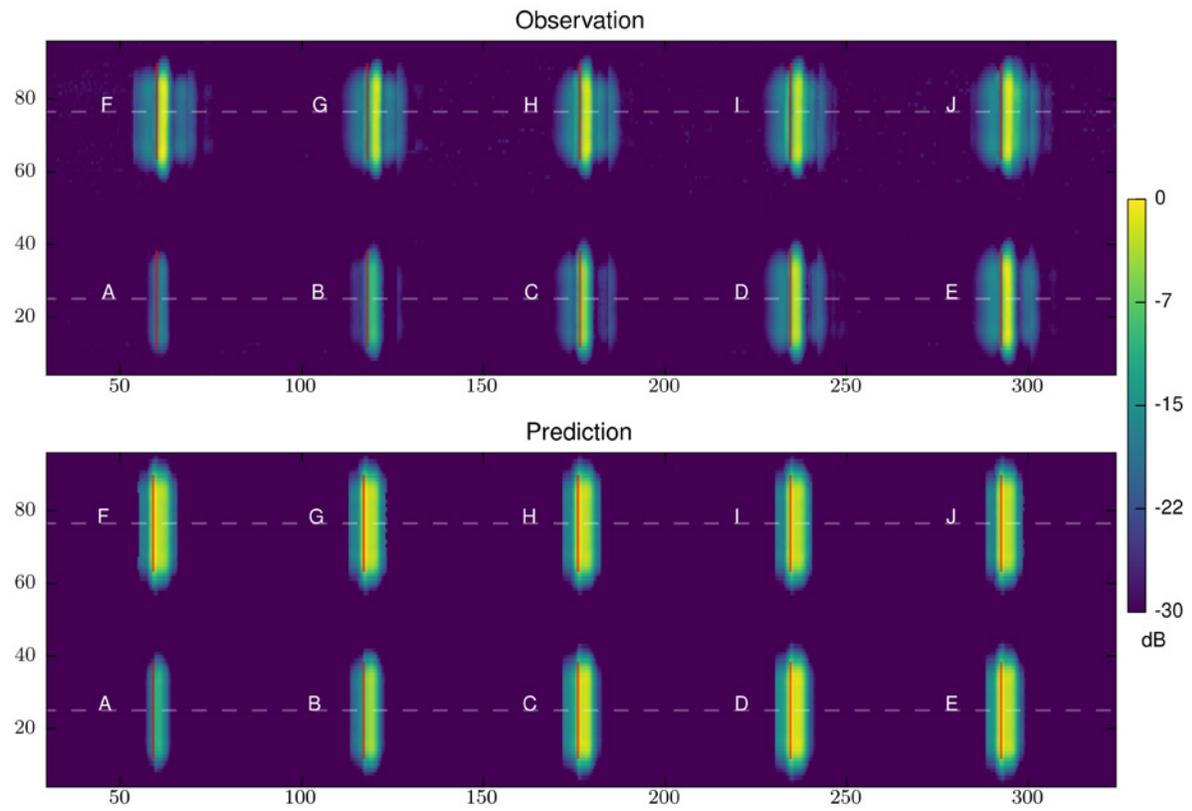
## B.1.1 Specimen 1, Probe 1, Skew 0



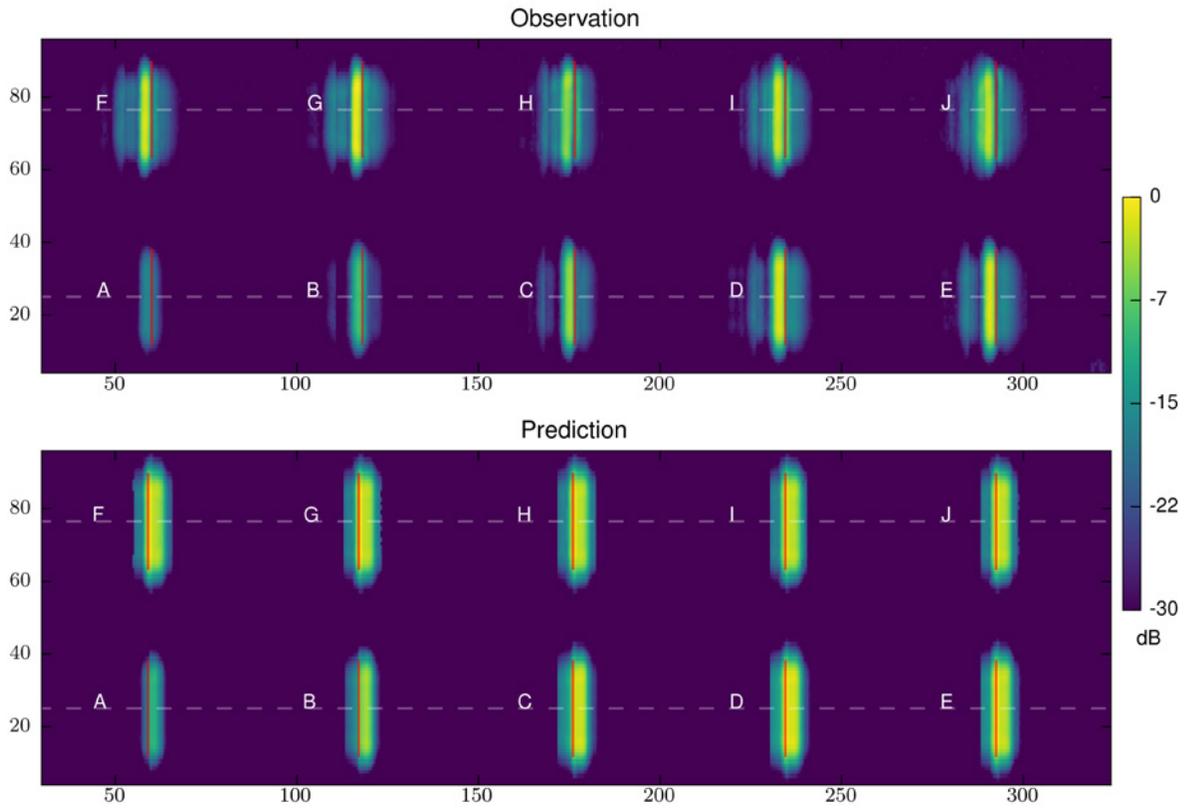
### B.1.2 Specimen 1, Probe 1, Skew 180



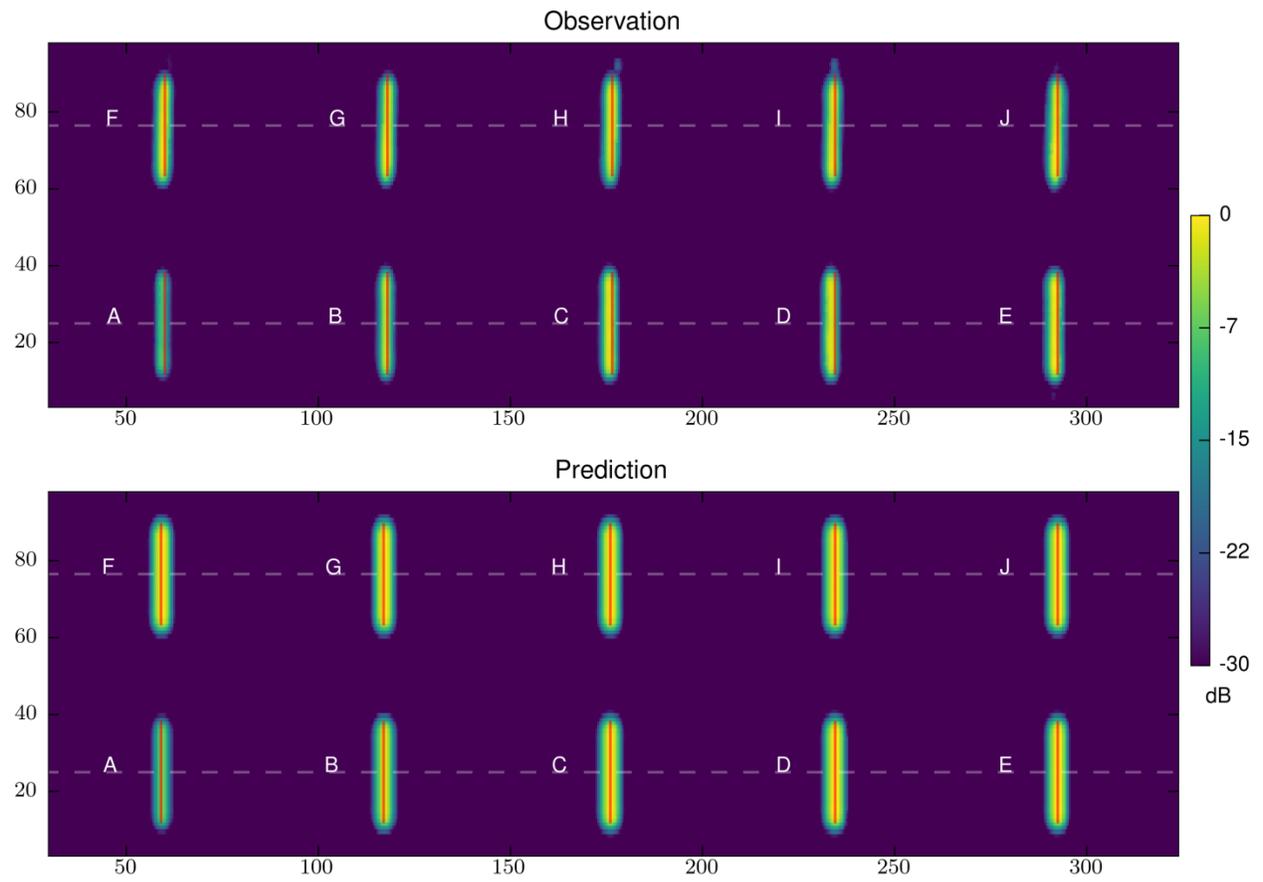
### B.1.3 Specimen 1, Probe 2, Skew 0



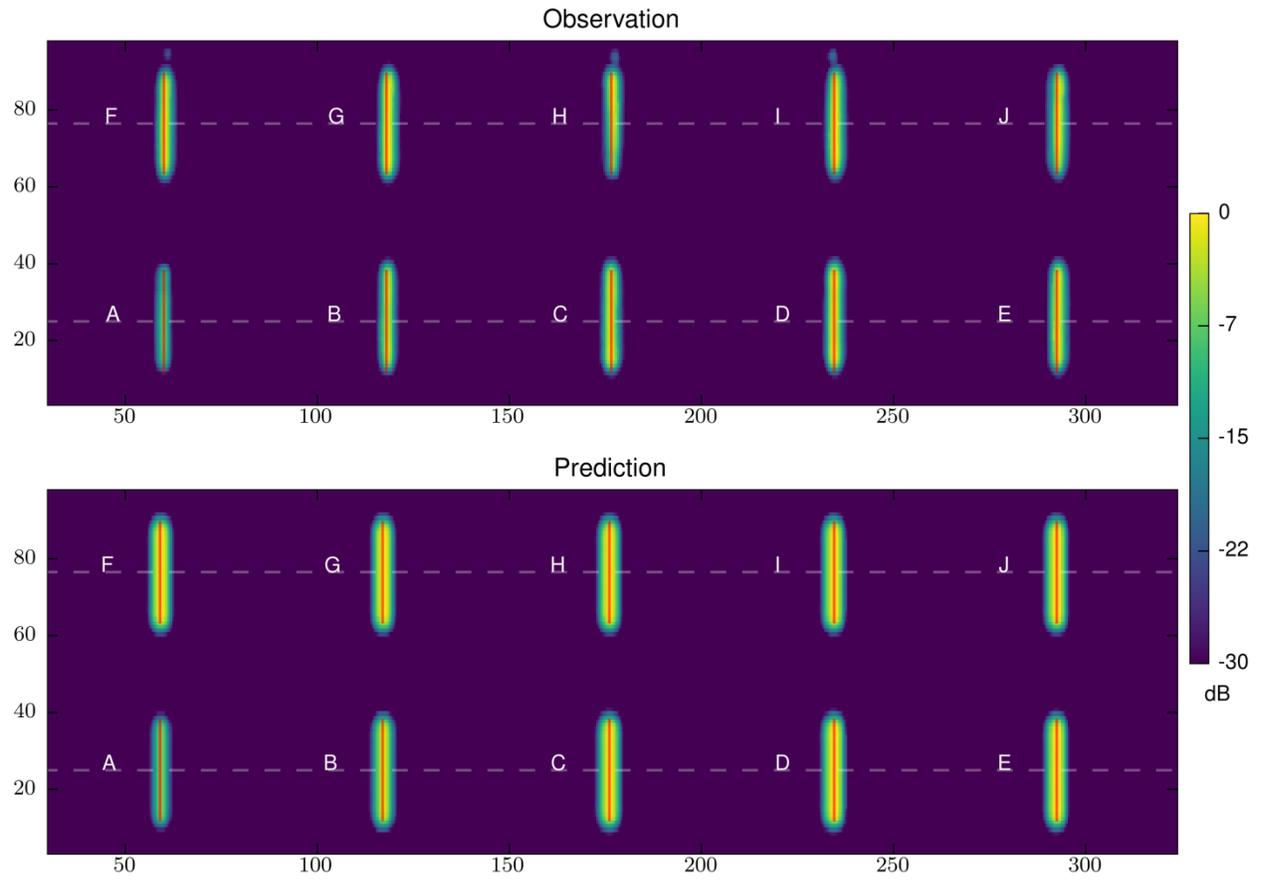
**B.1.4 Specimen 1, Probe 2, Skew 180**



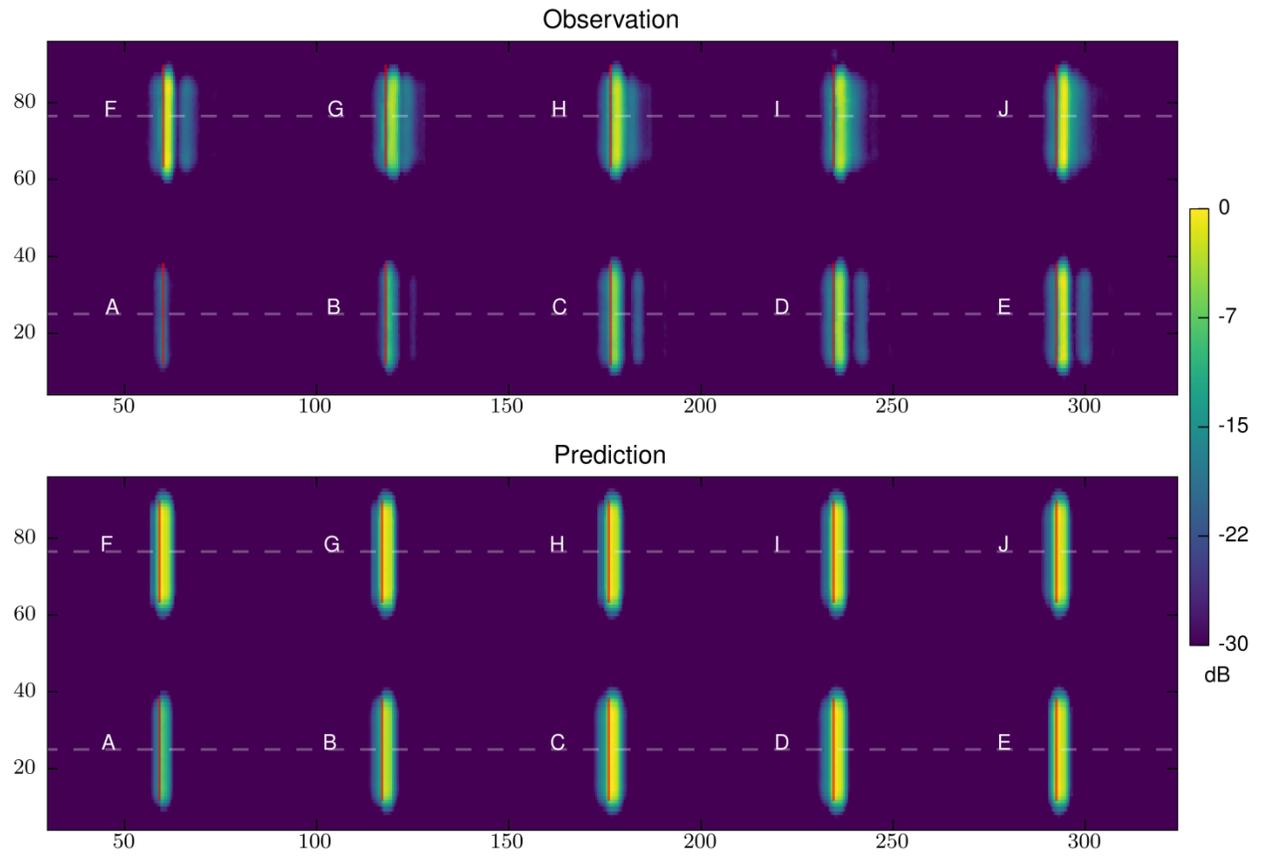
### B.1.5 Specimen 1, Probe 3, Skew 0



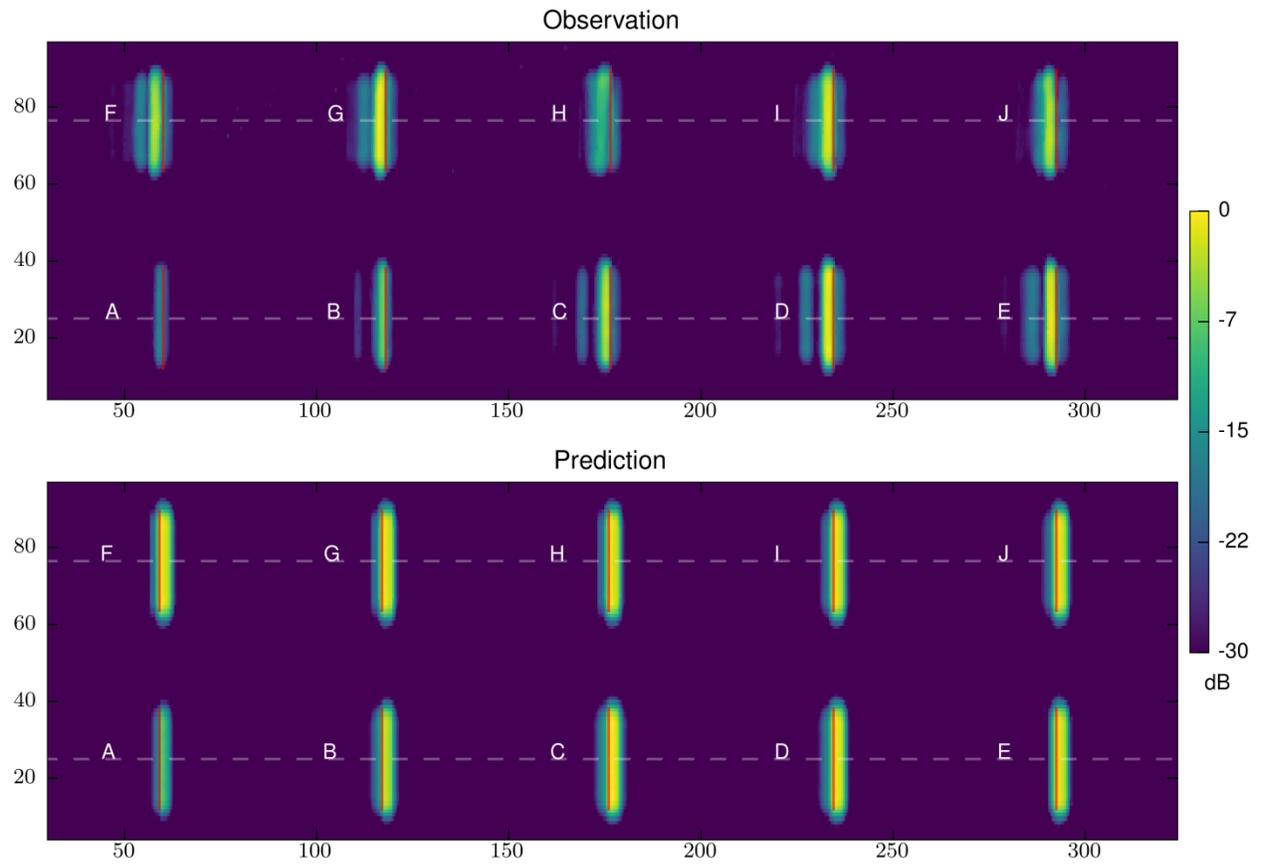
### B.1.6 Specimen 1, Probe 3, Skew 180



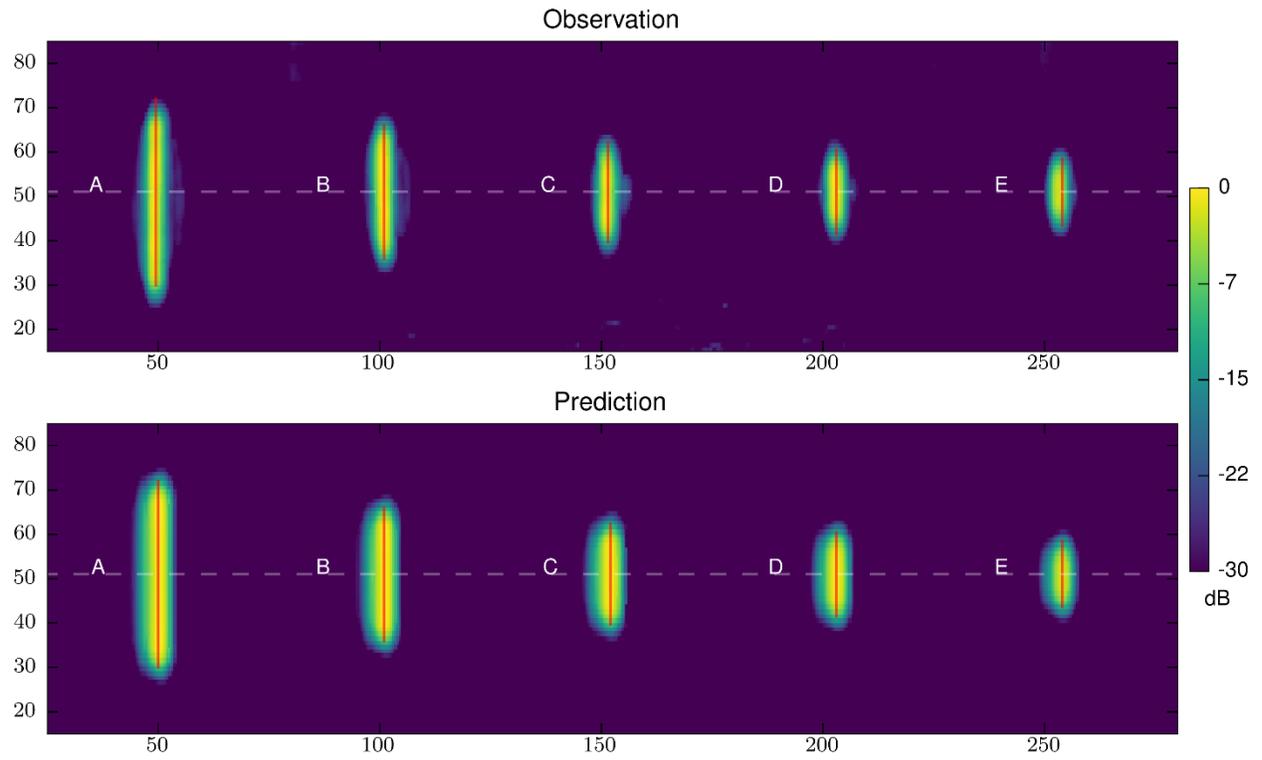
### B.1.7 Specimen 1, Probe 4, Skew 0



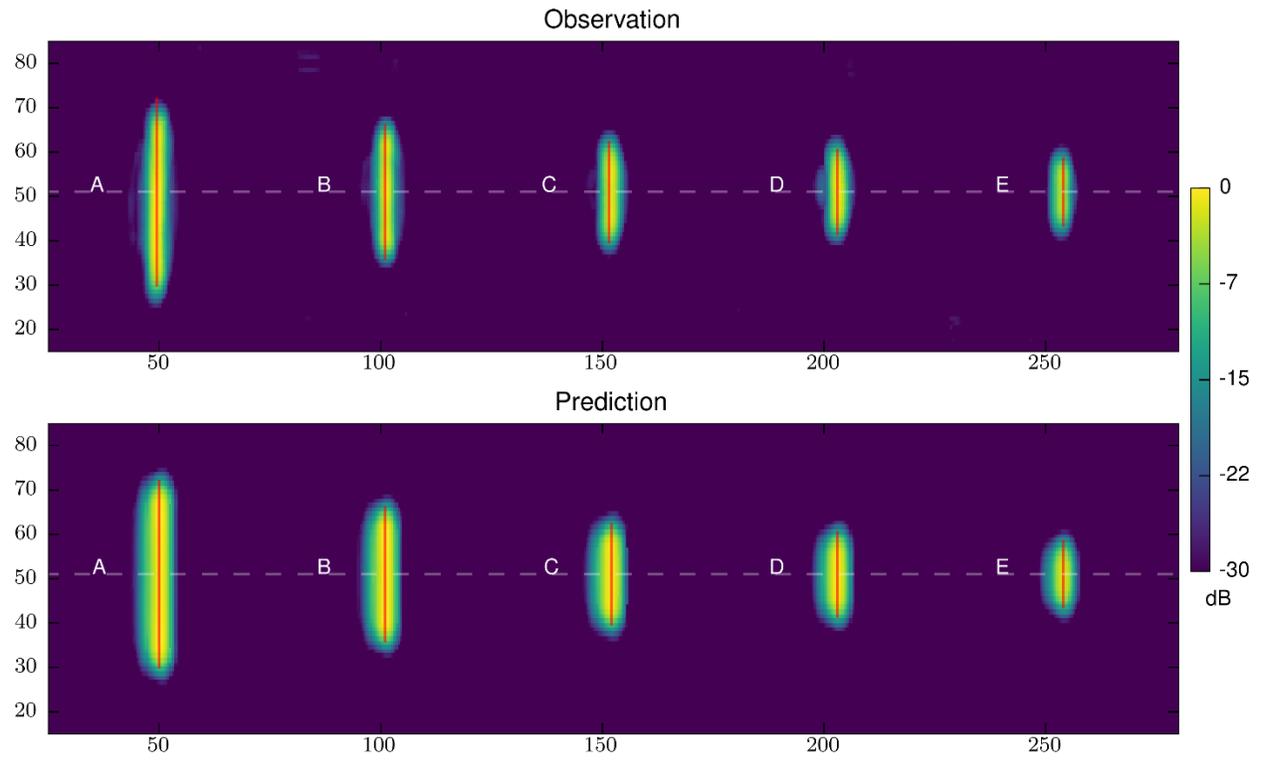
**B.1.8 Specimen 1, Probe 4, Skew 180**



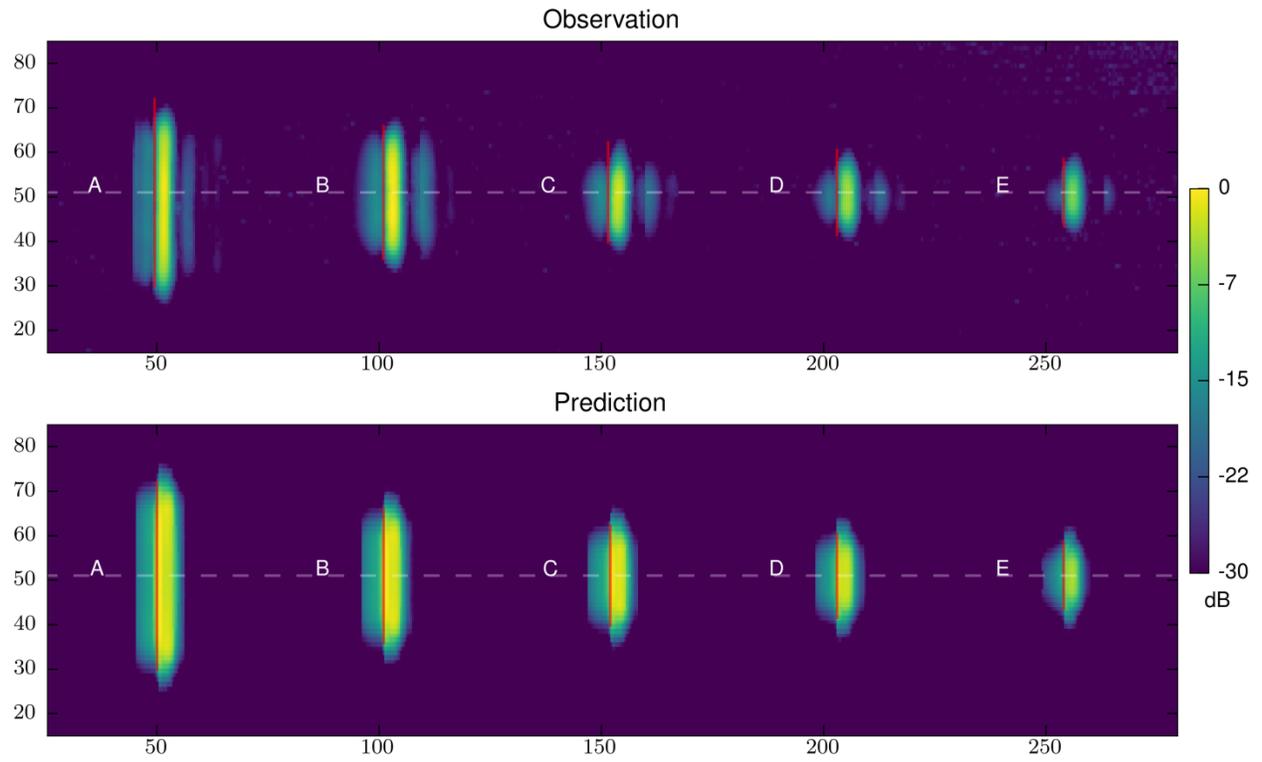
**B.1.9 Specimen 2, Probe 1, Skew 0**



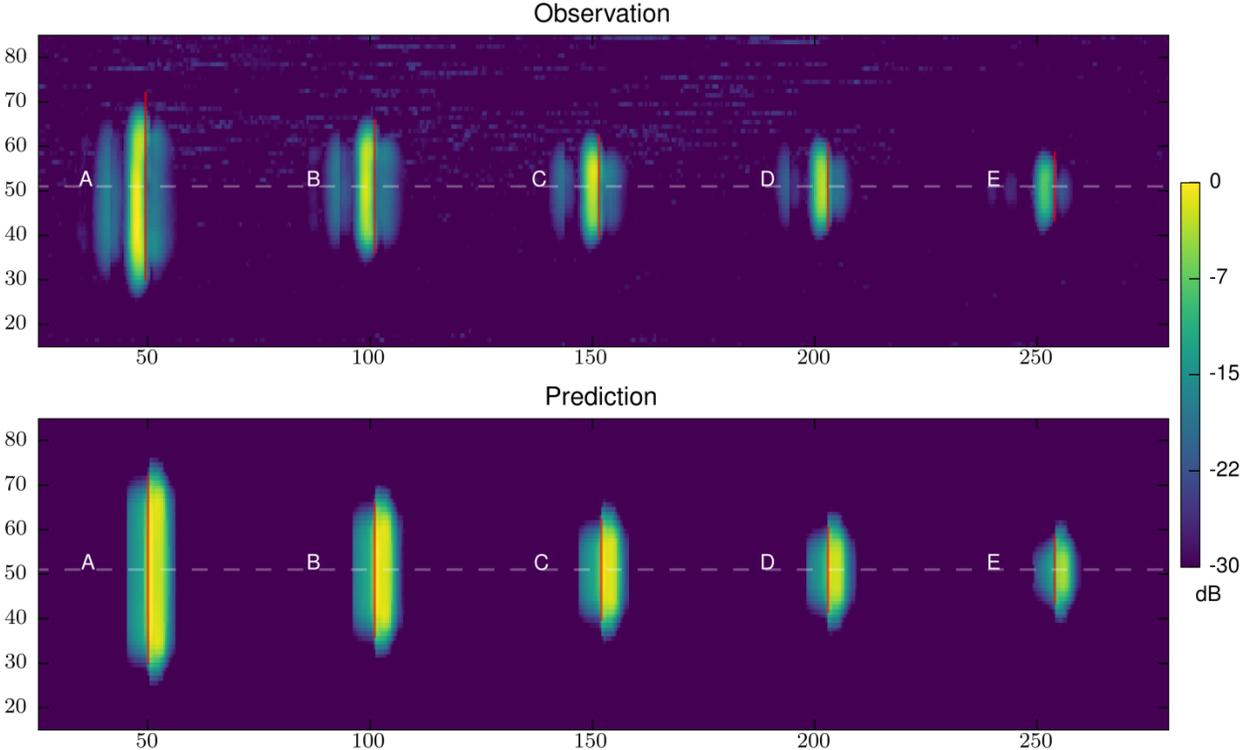
### B.1.10 Specimen 2, Probe 1, Skew 180



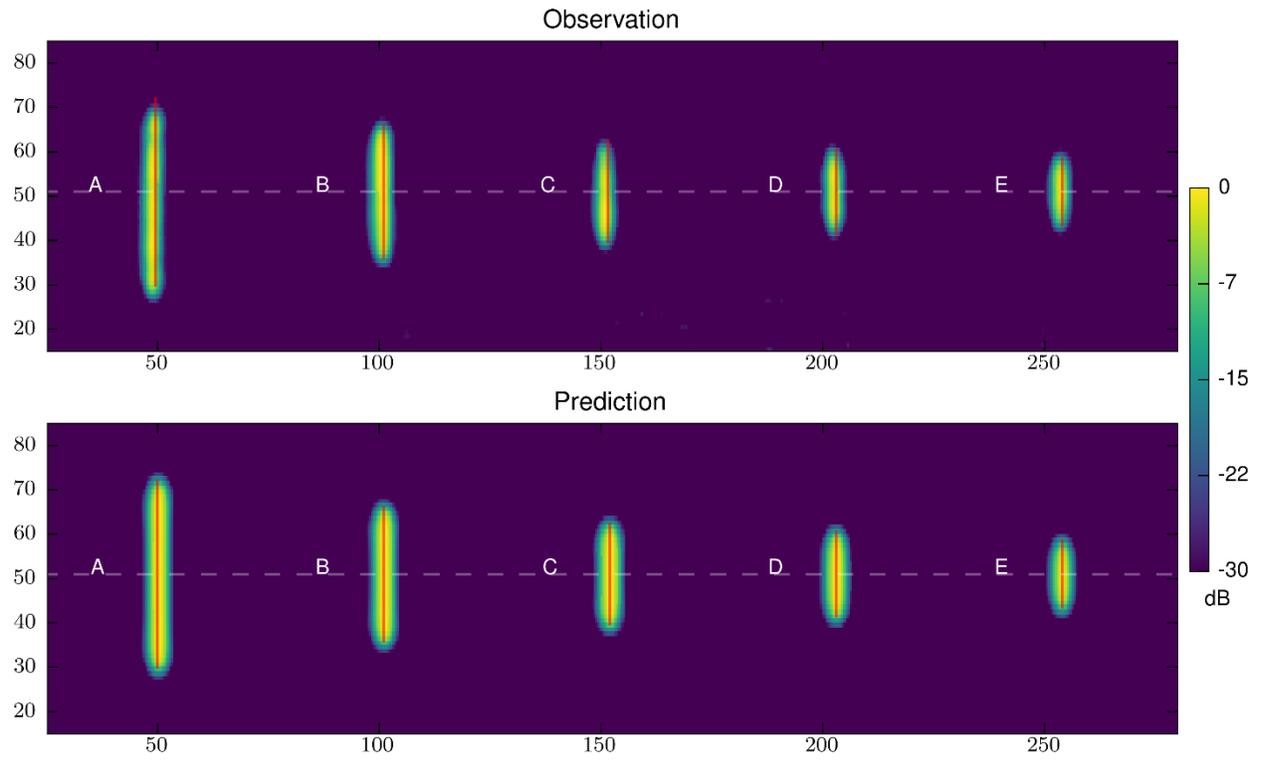
### B.1.11 Specimen 2, Probe 2, Skew 0



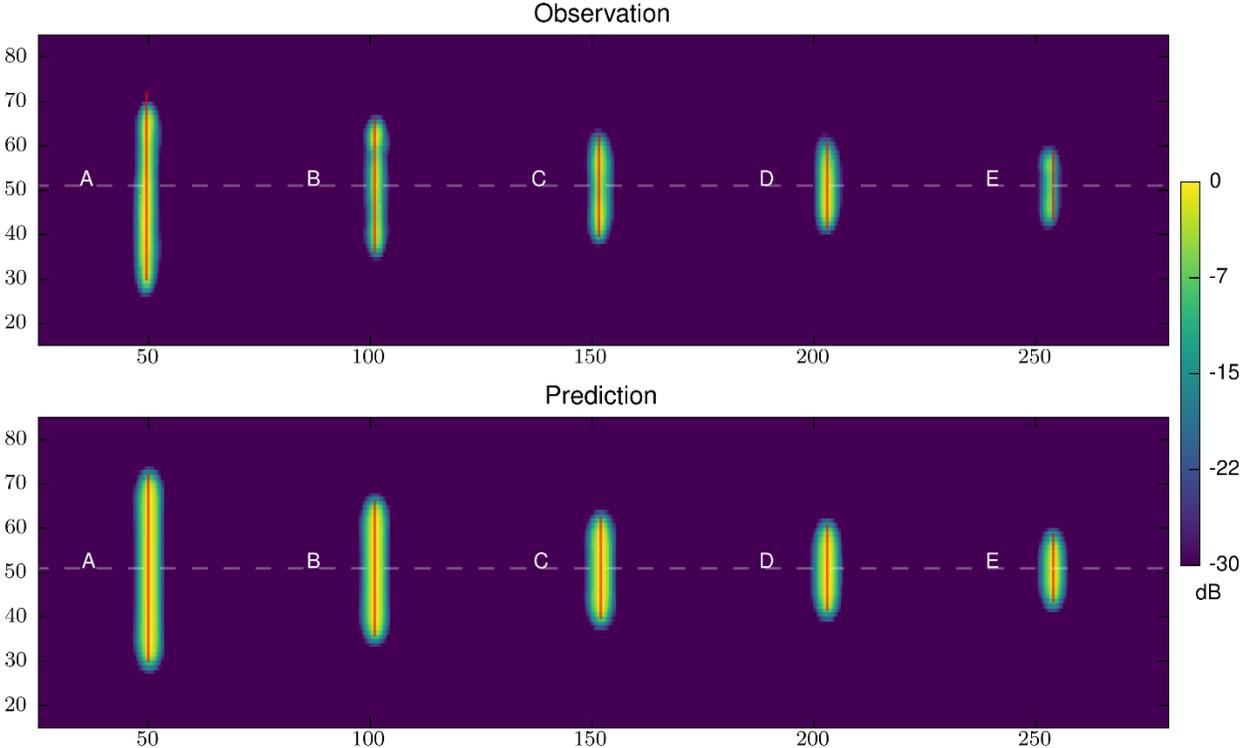
**B.1.12 Specimen 2, Probe 2, Skew 180**



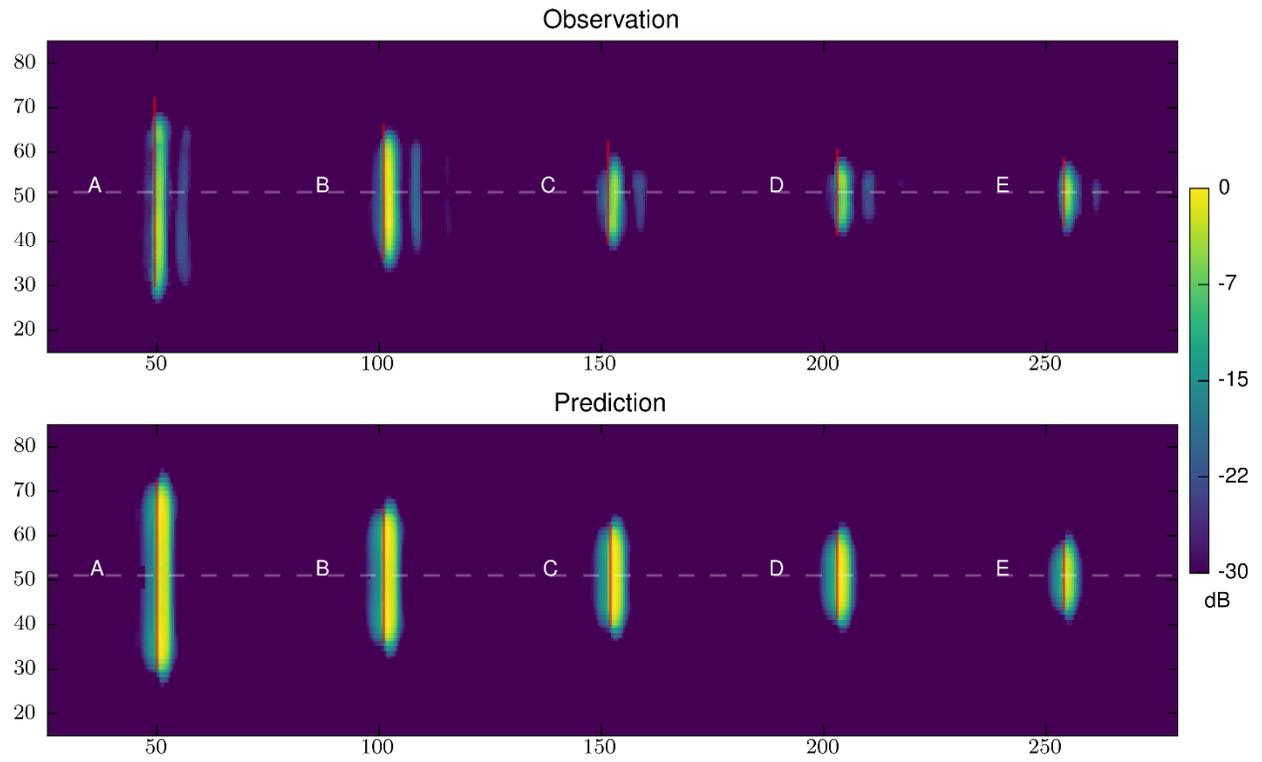
### B.1.13 Specimen 2, Probe 3, Skew 0



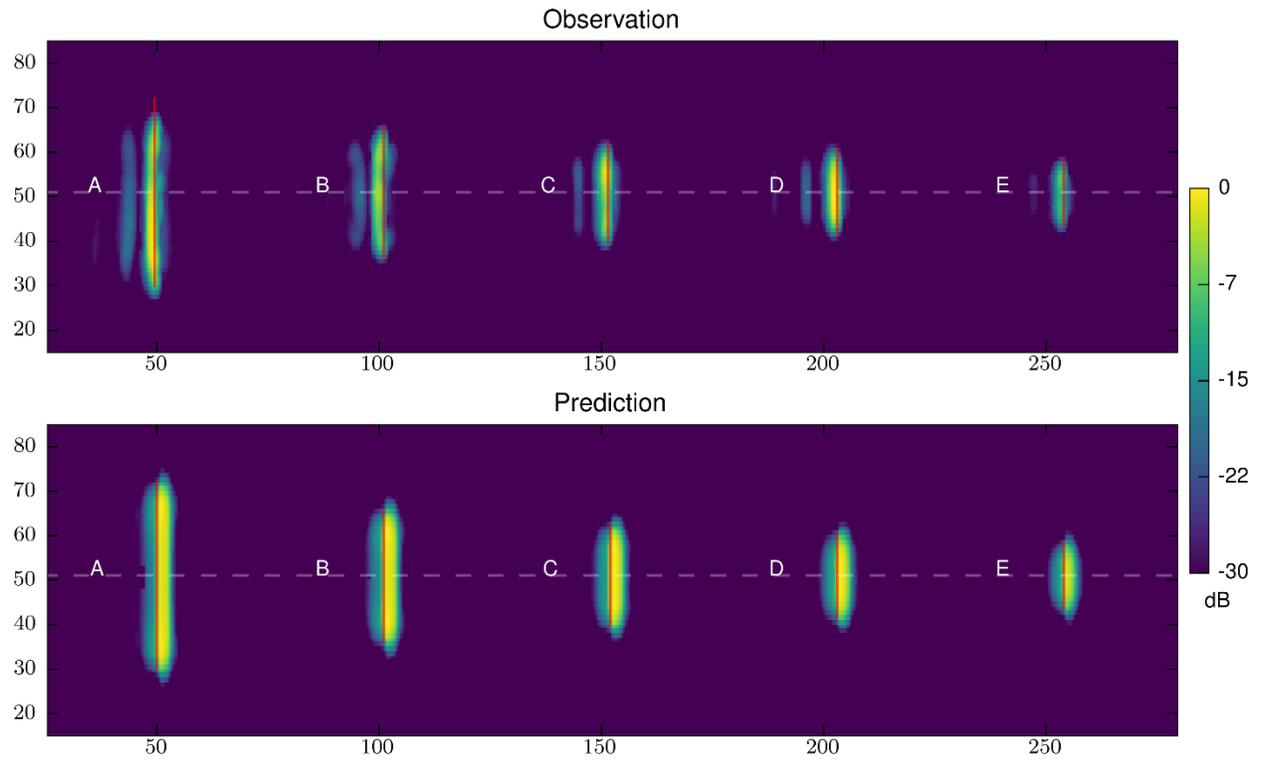
**B.1.14 Specimen 2, Probe 3, Skew 180**



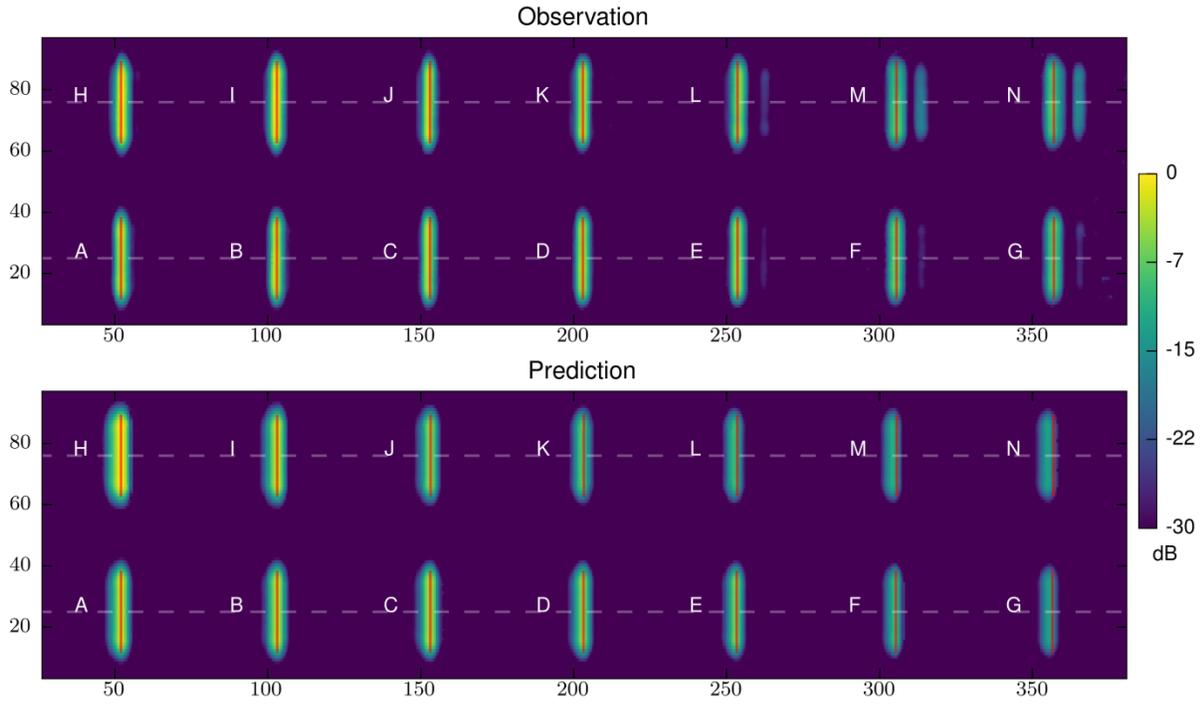
### B.1.15 Specimen 2, Probe 4, Skew 0



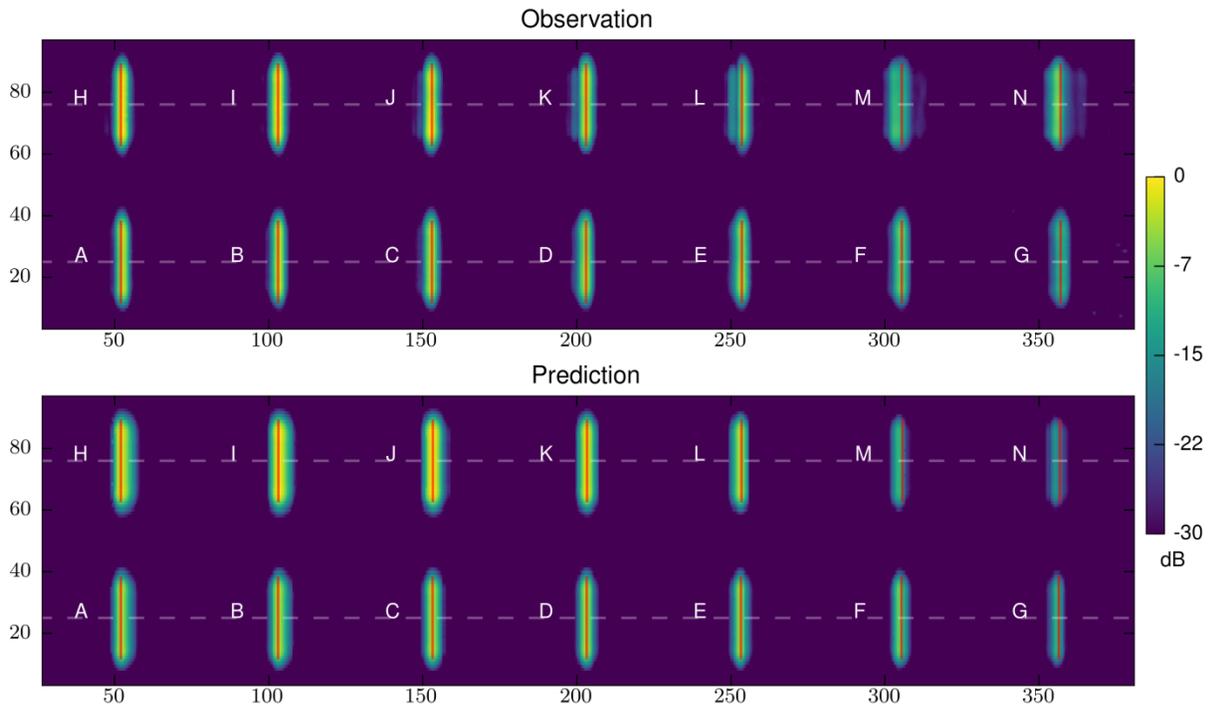
### B.1.16 Specimen 2, Probe 4, Skew 180



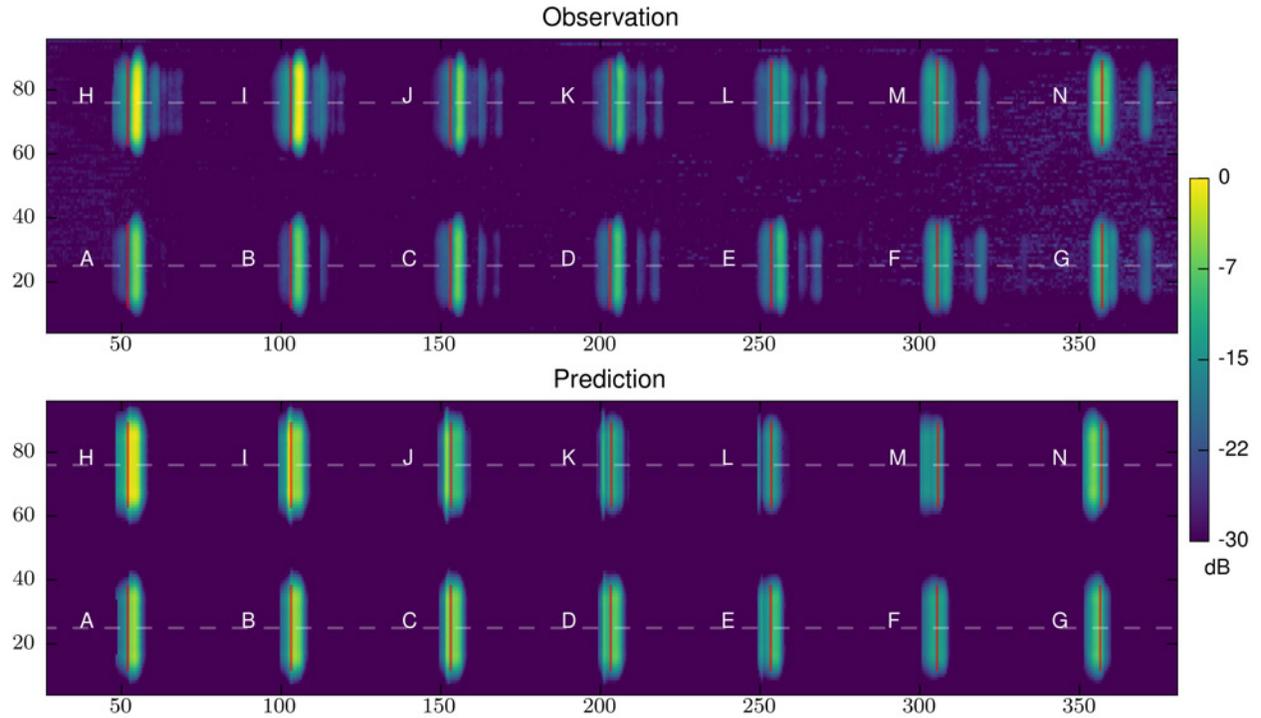
### B.1.17 Specimen 3, Probe 1, Skew 0



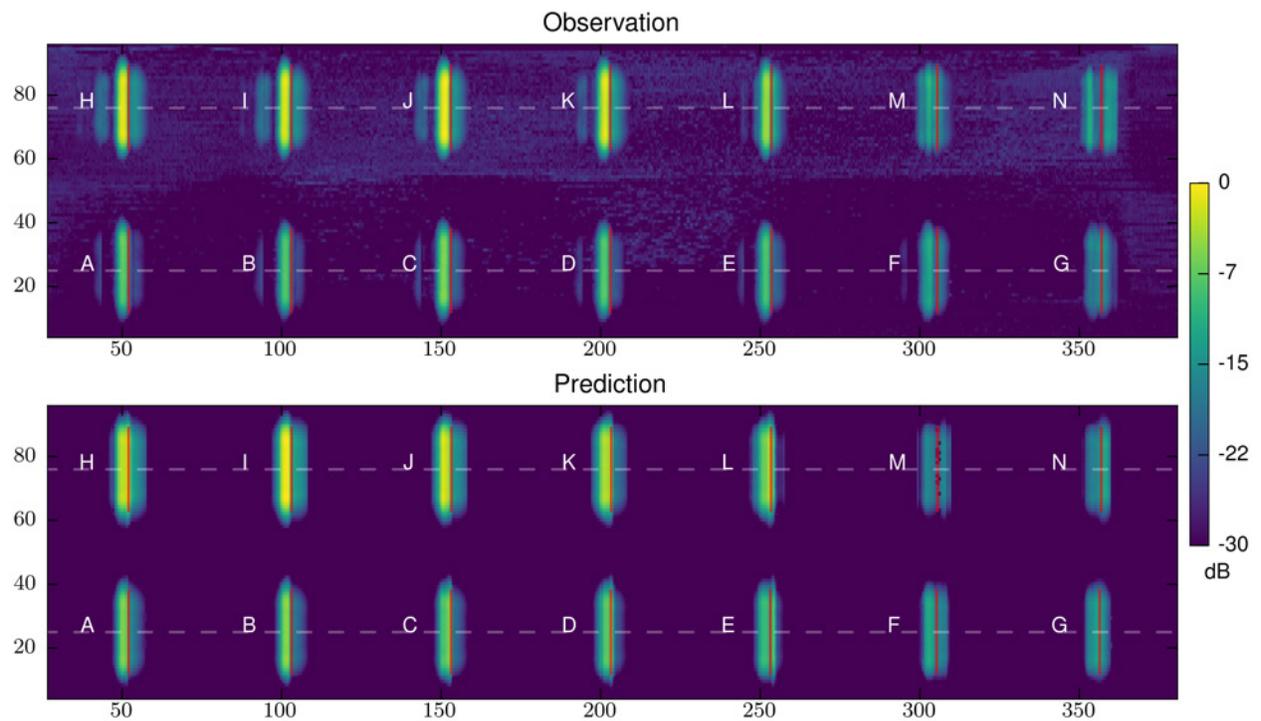
### B.1.18 Specimen 3, Probe 1, Skew 180



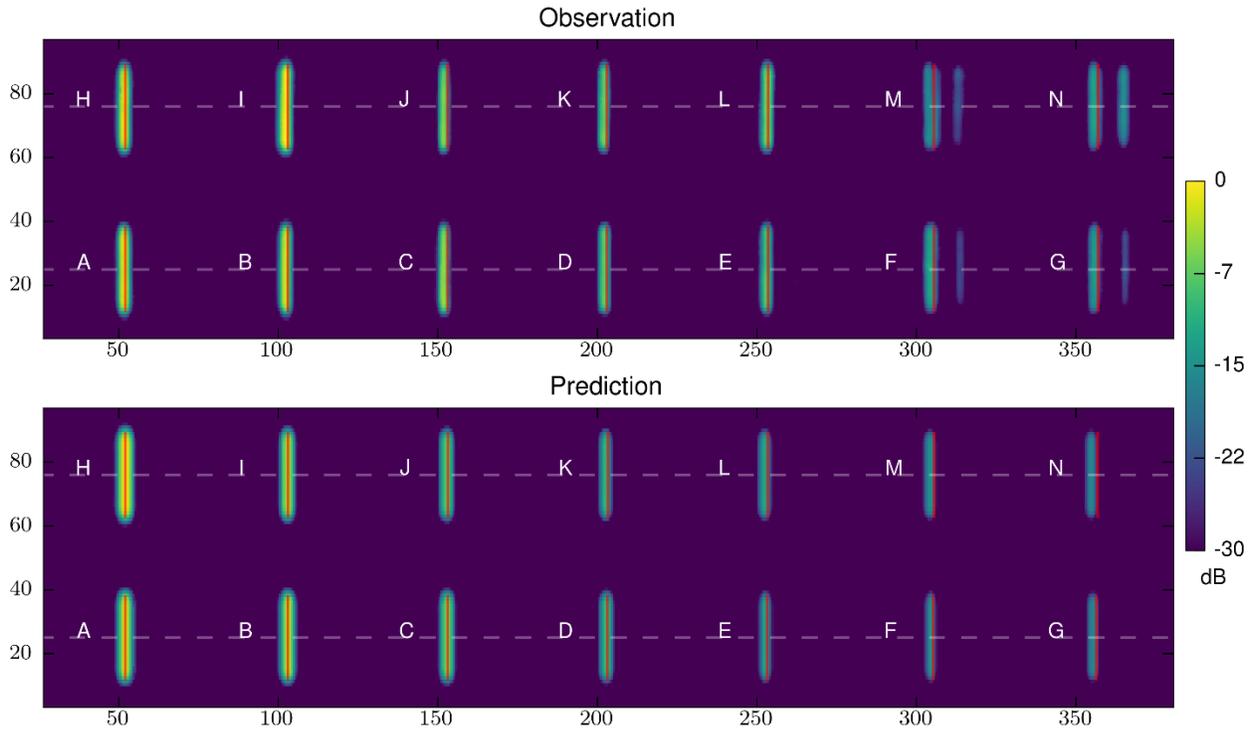
### B.1.19 Specimen 3, Probe 2, Skew 0



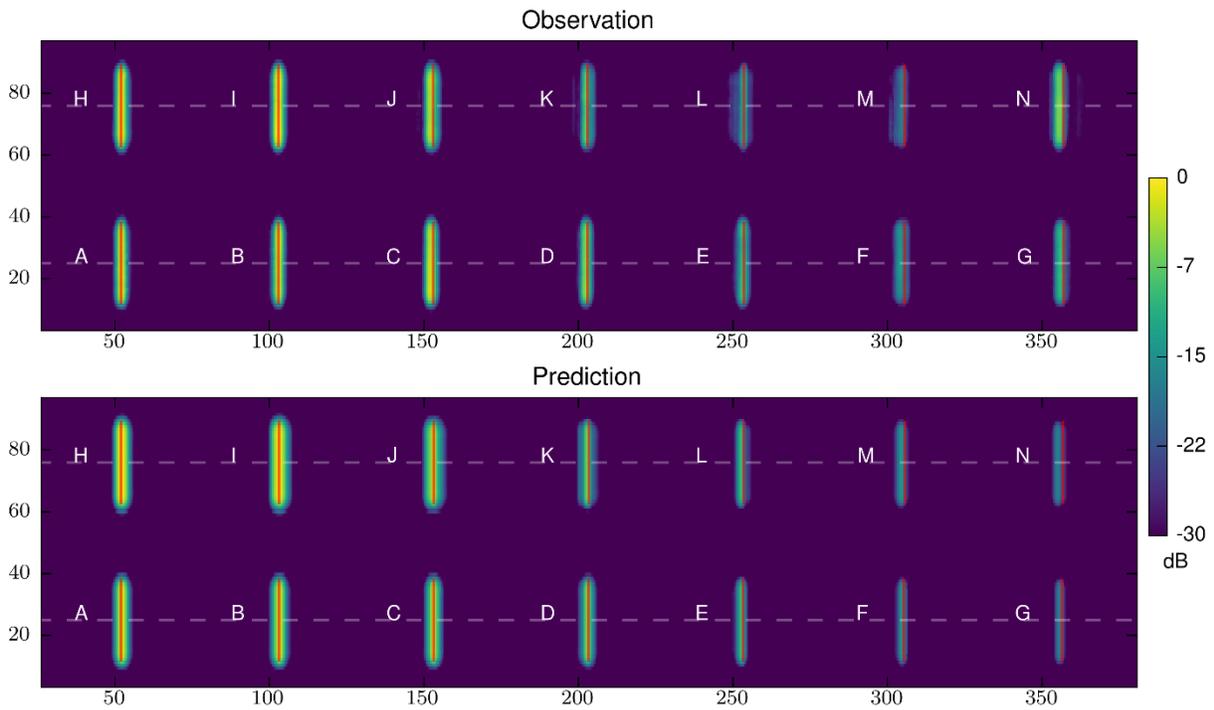
### B.1.20 Specimen 3, Probe 2, Skew 180



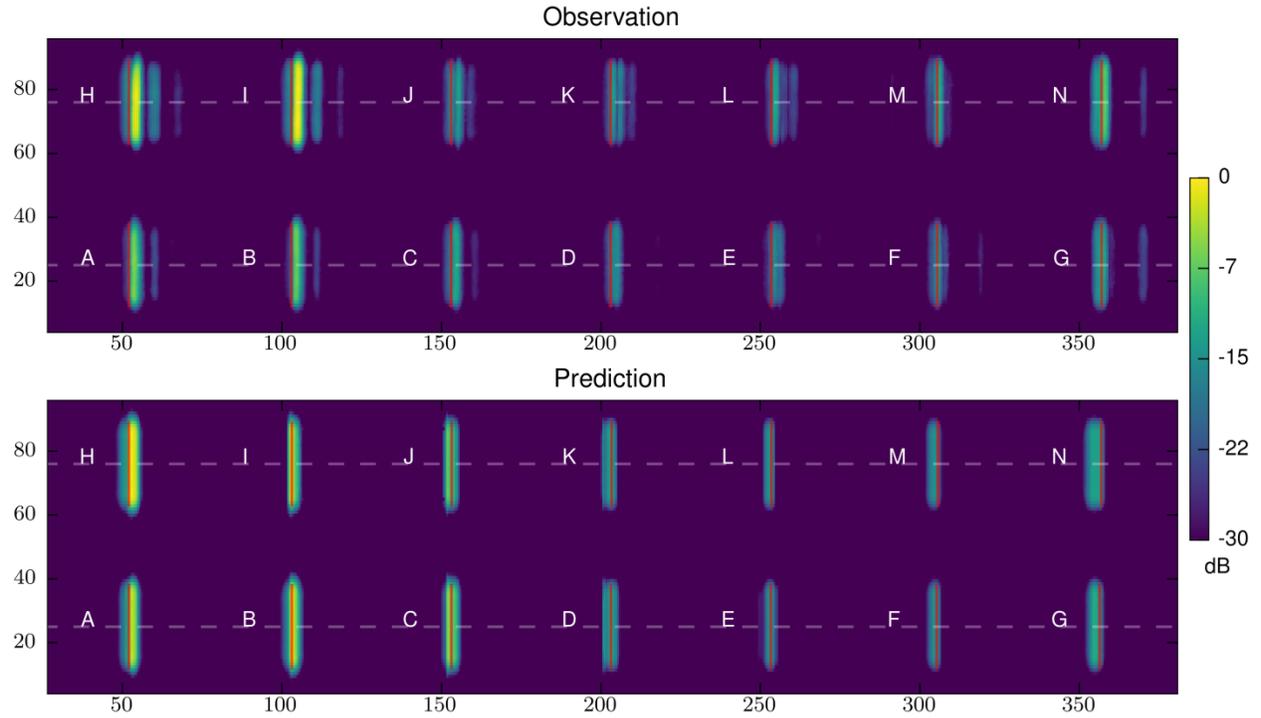
### B.1.21 Specimen 3, Probe 3, Skew 0



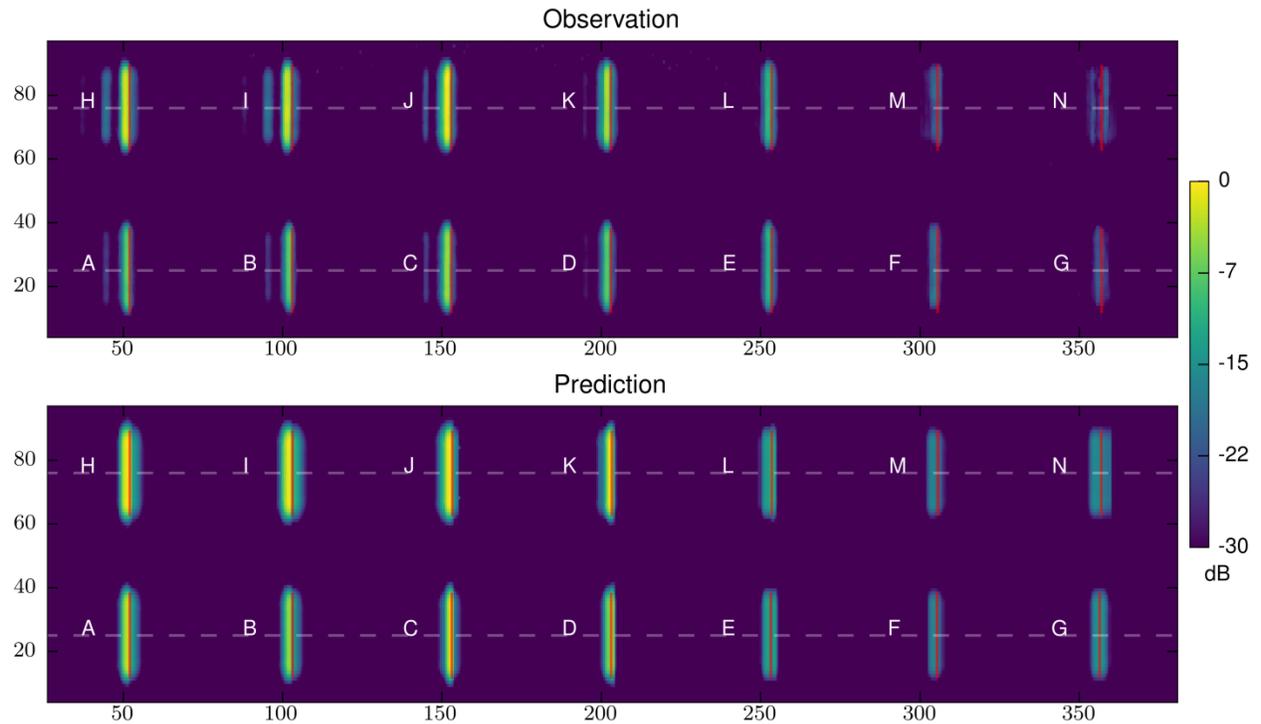
### B.1.22 Specimen 3, Probe 3, Skew 180



### B.1.23 Specimen 3, Probe 4, Skew 0

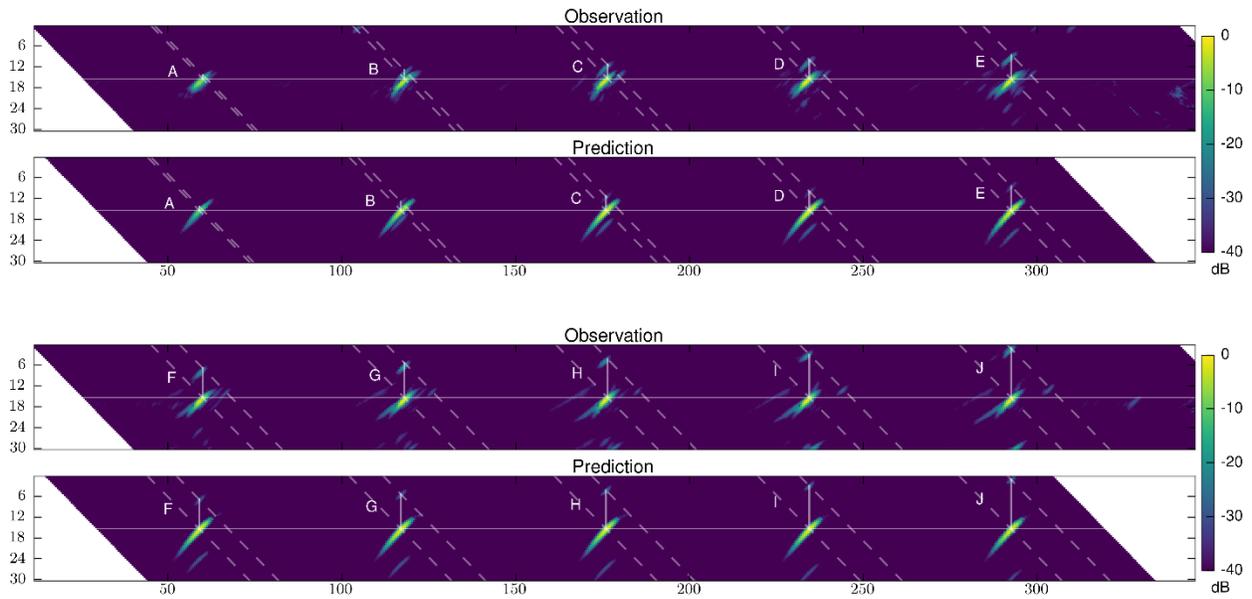


### B.1.24 Specimen 3, Probe 4, Skew 180

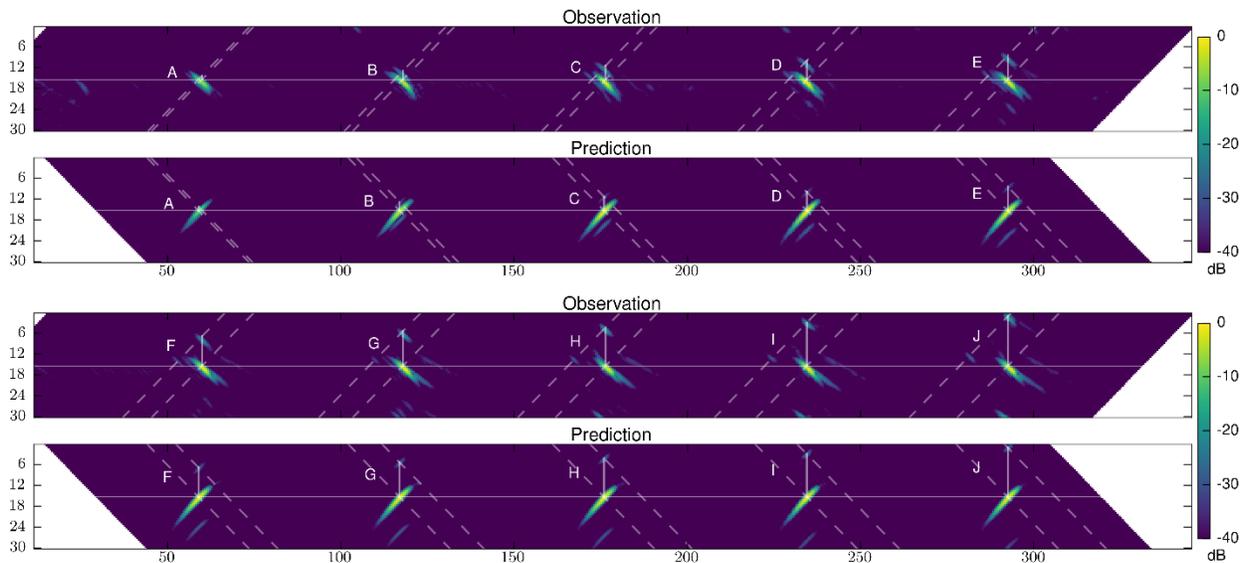


## B.2 B-scan Image Views

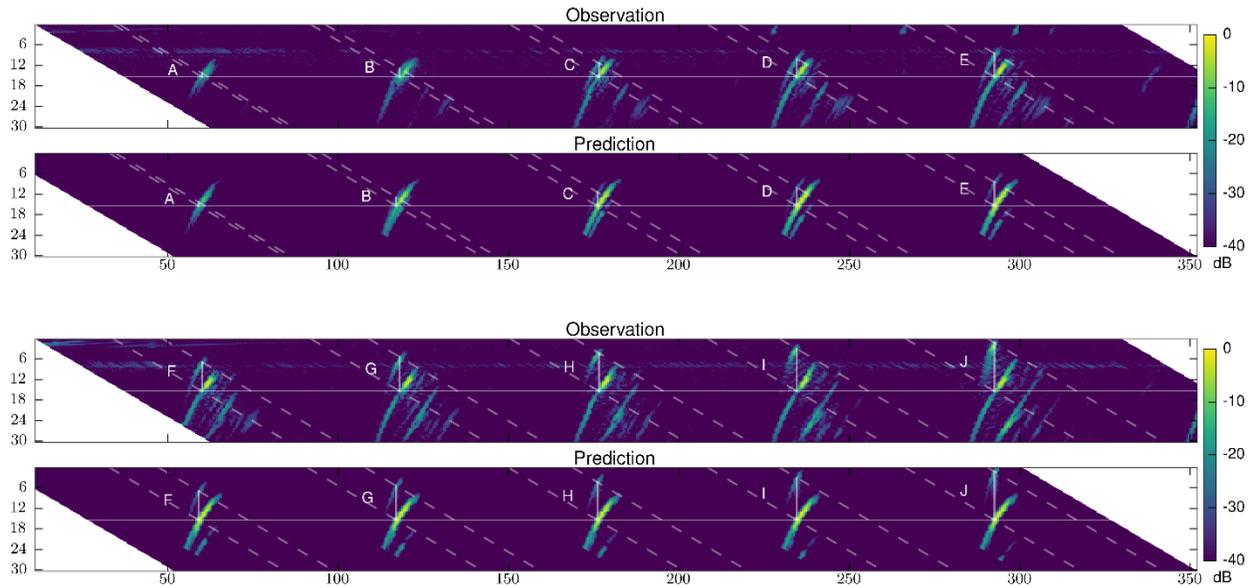
### B.2.1 Specimen 1, Probe 1, Skew 0



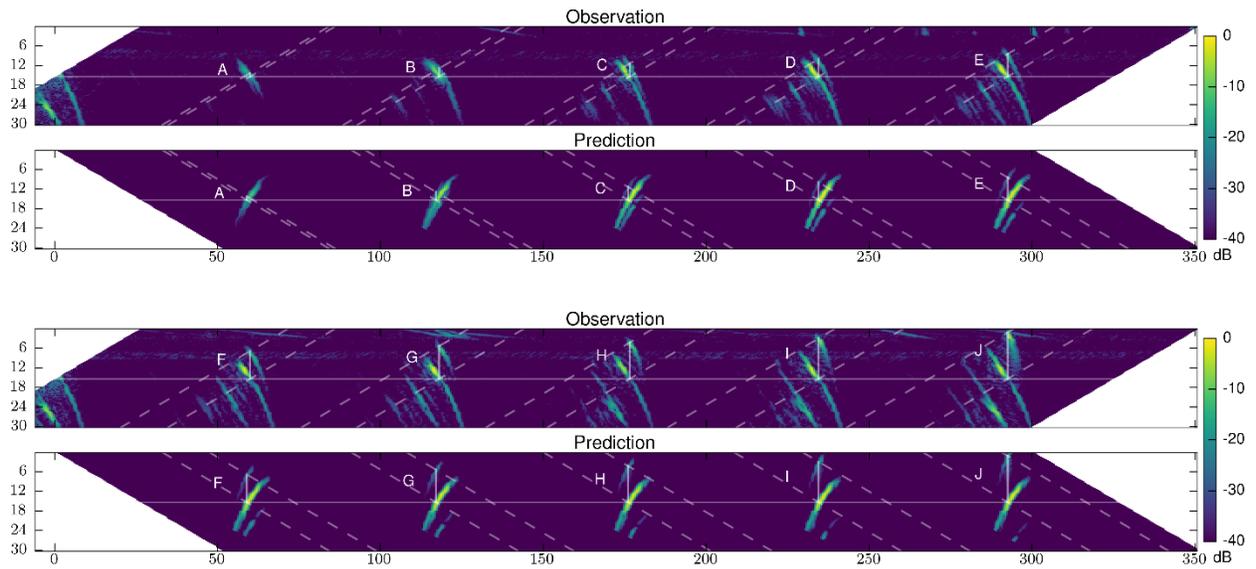
### B.2.2 Specimen 1, Probe 1, Skew 180



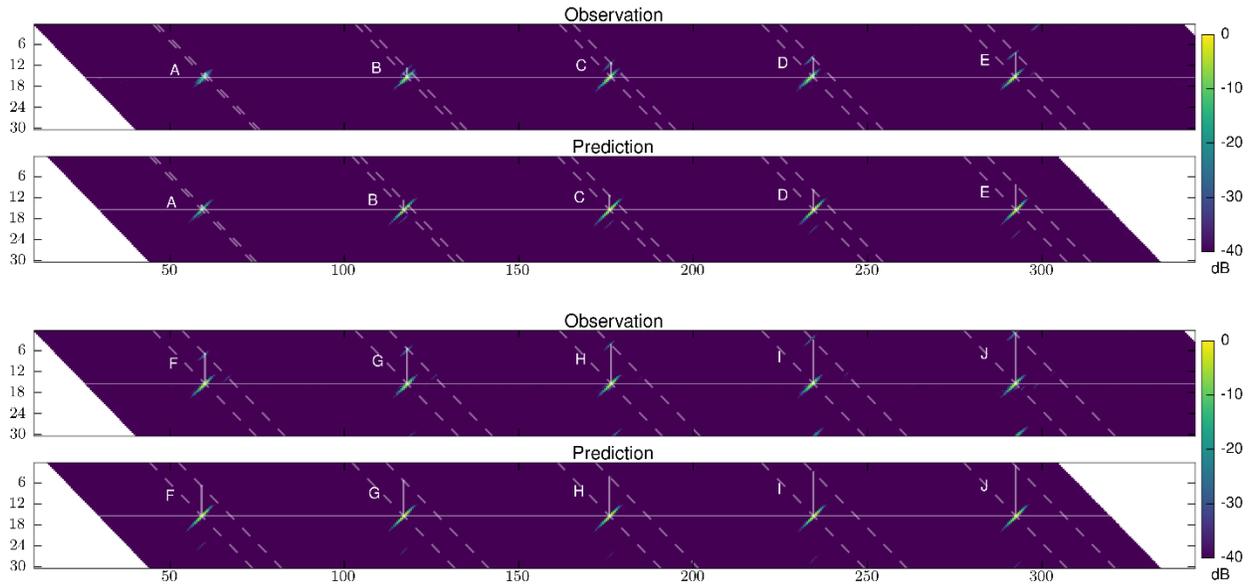
### B.2.3 Specimen 1, Probe 2, Skew 0



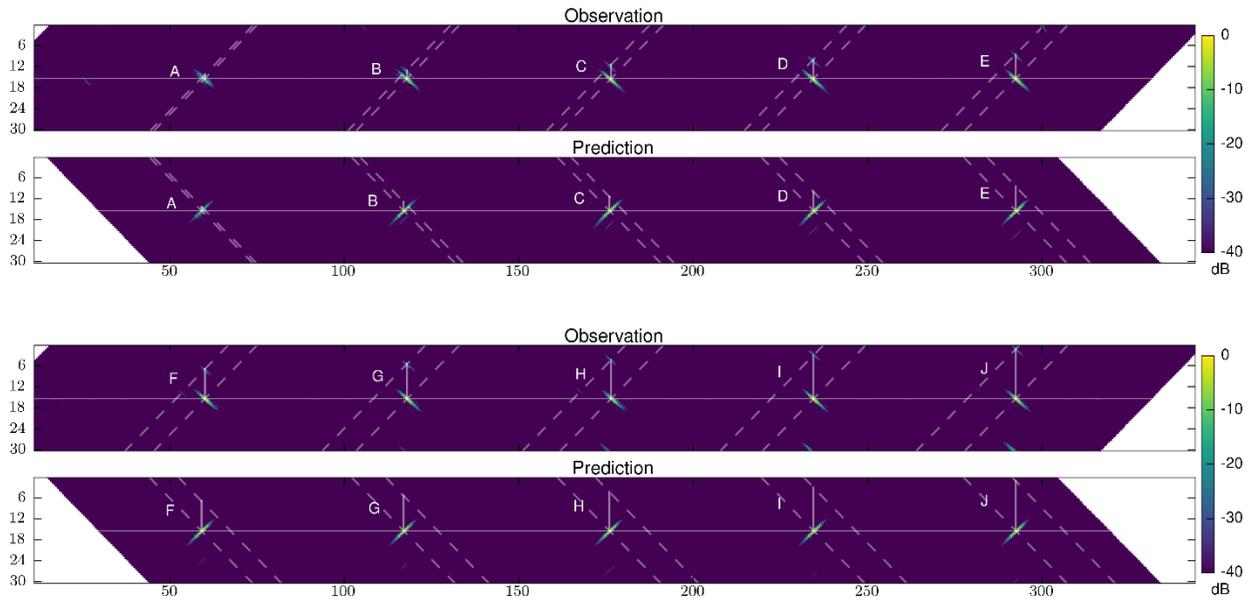
### B.2.4 Specimen 1, Probe 2, Skew 180



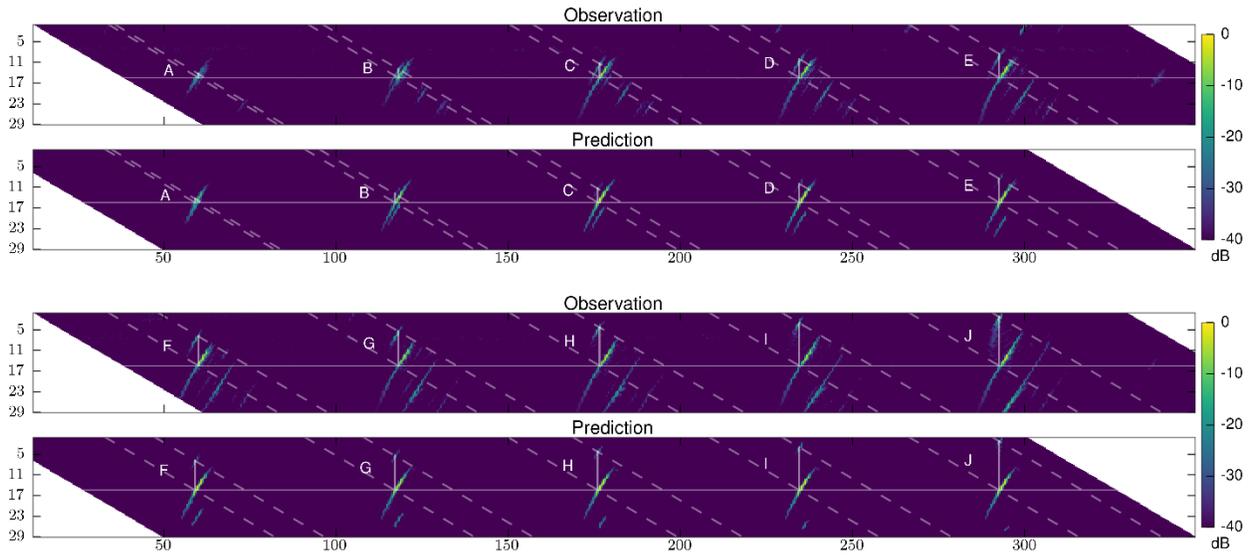
### B.2.5 Specimen 1, Probe 3, Skew 0



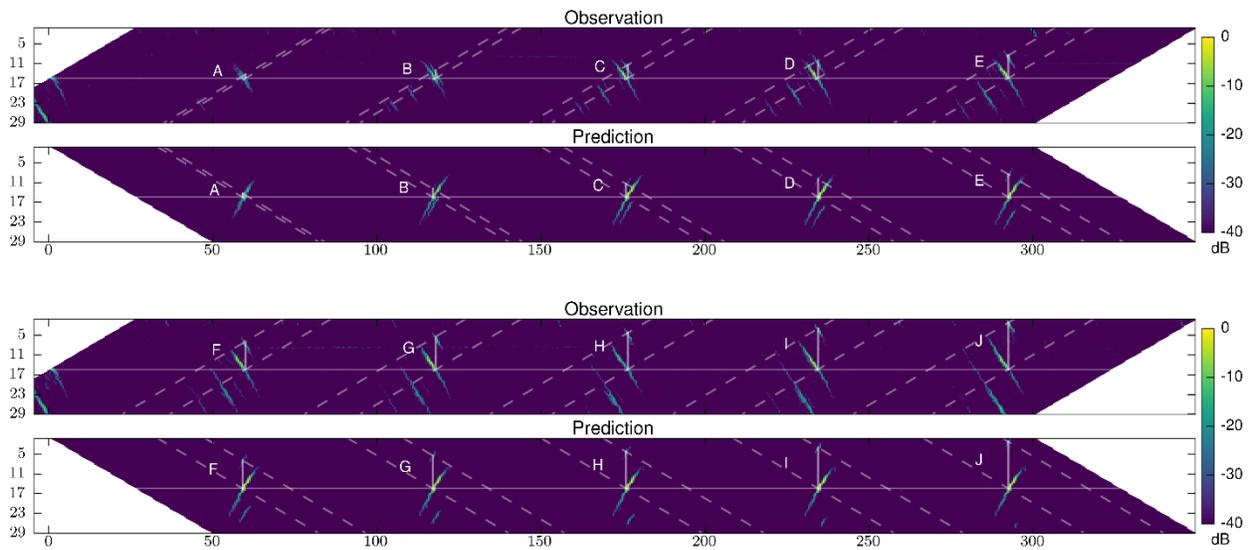
### B.2.6 Specimen 1, Probe 3, Skew 180



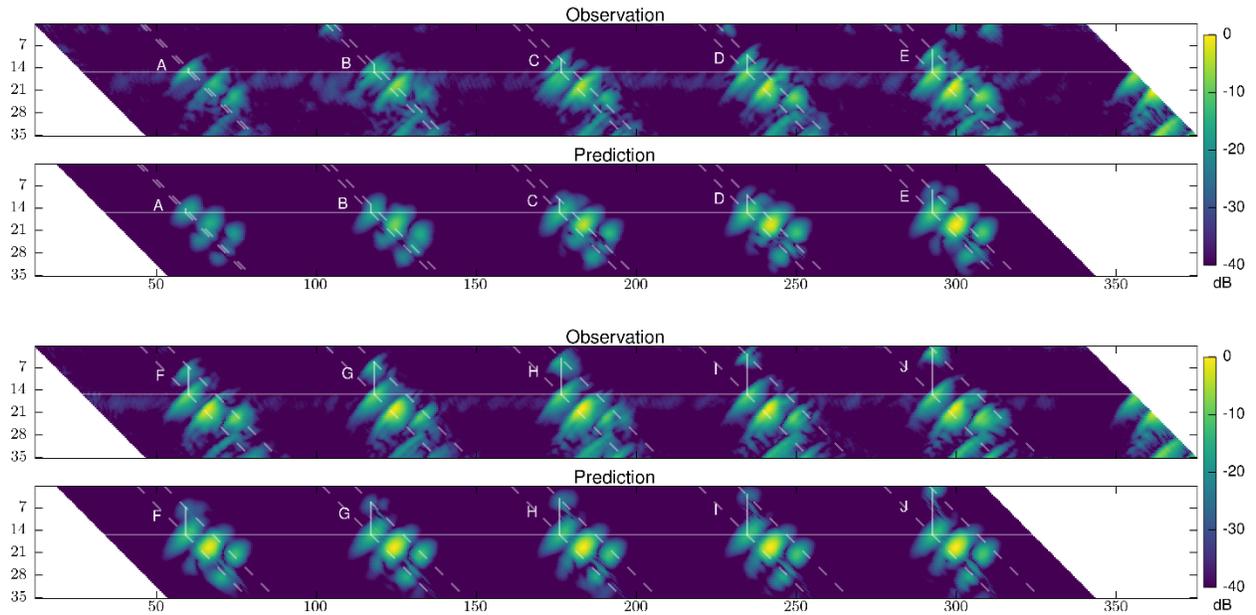
### B.2.7 Specimen 1, Probe 4, Skew 0



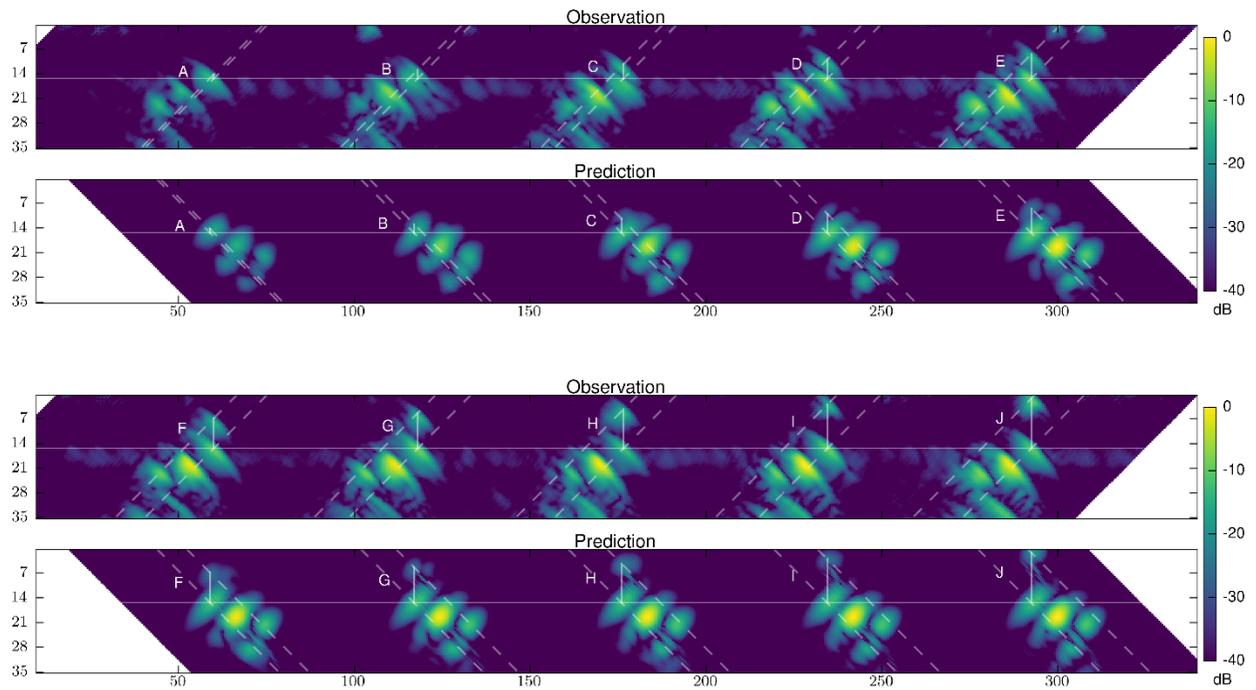
### B.2.8 Specimen 1, Probe 4, Skew 180



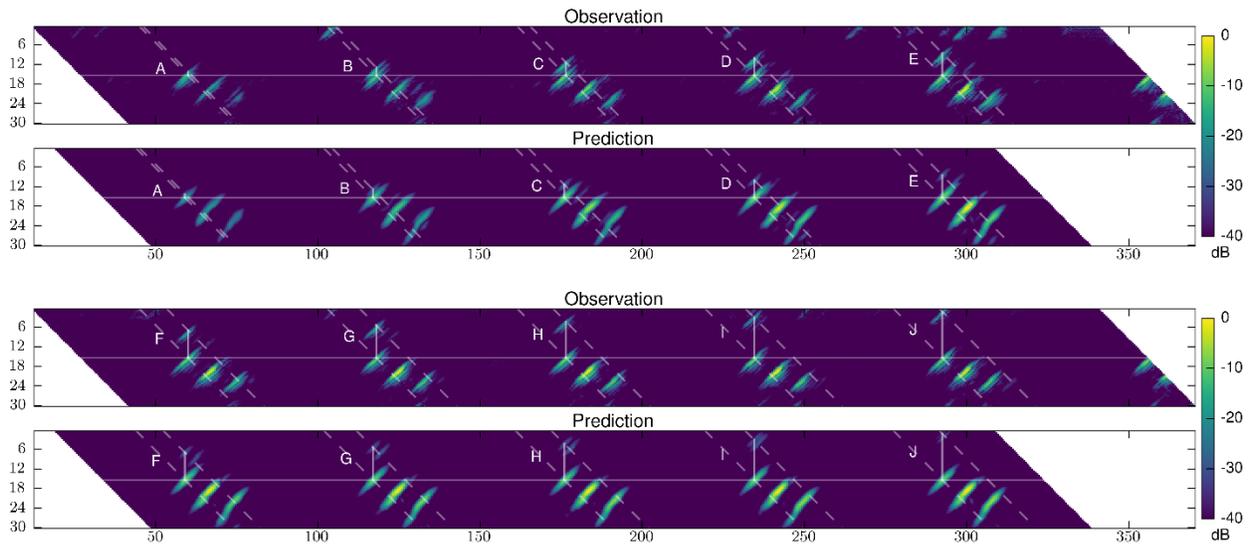
### B.2.9 Specimen 1, Probe 5, Skew 0



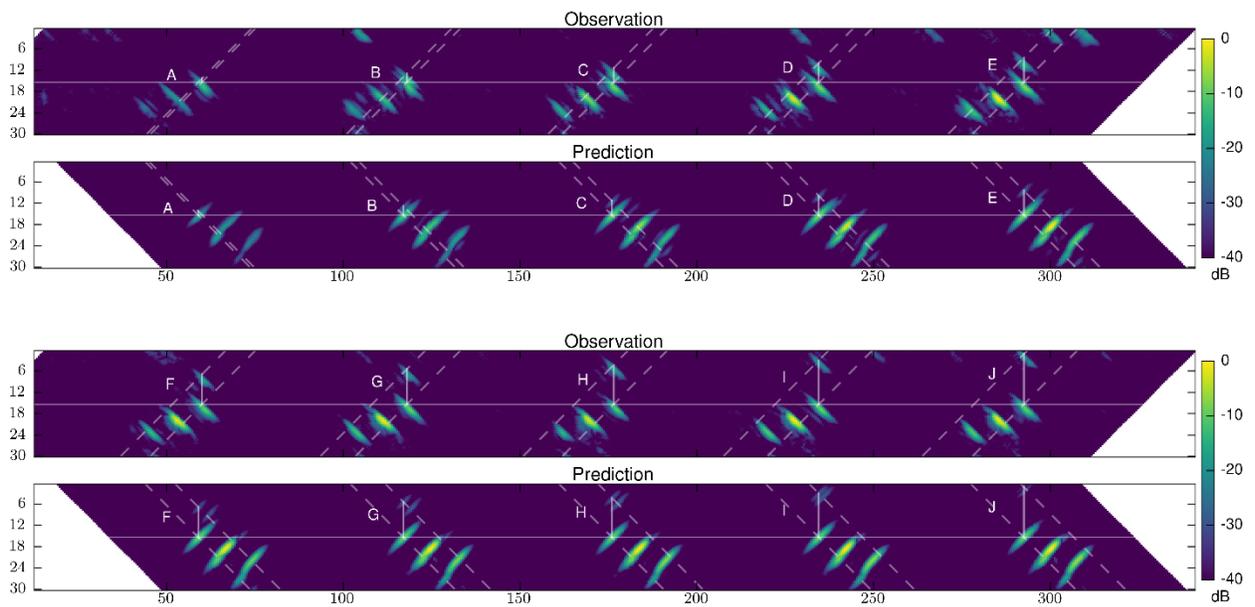
### B.2.10 Specimen 1, Probe 5, Skew 180



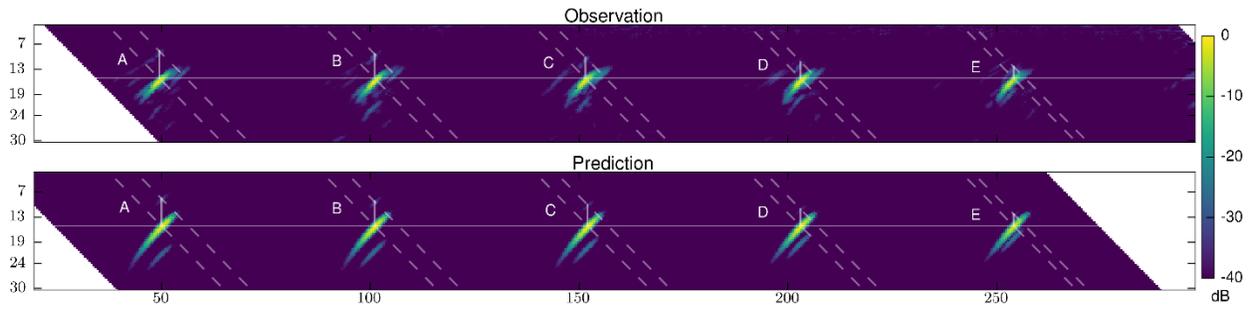
### B.2.11 Specimen 1, Probe 6, Skew 0



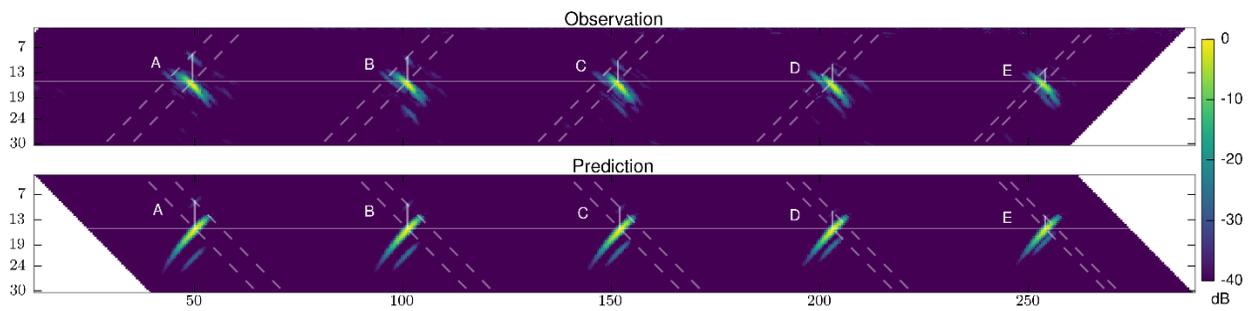
### B.2.12 Specimen 1, Probe 6, Skew 180



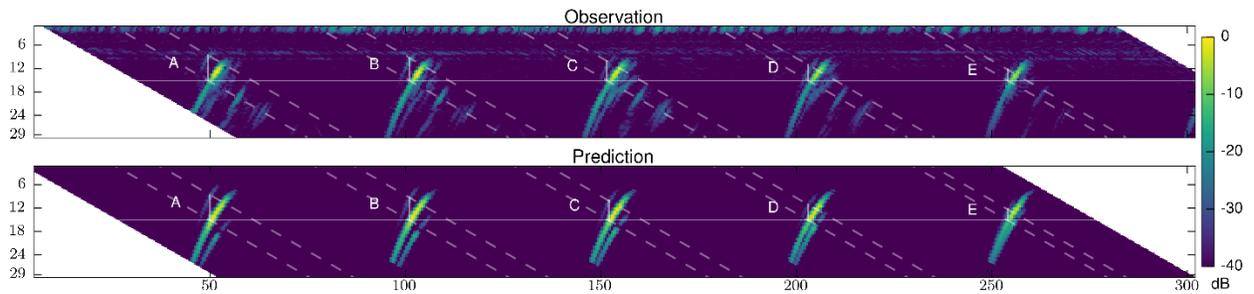
### B.2.13 Specimen 2, Probe 1, Skew 0



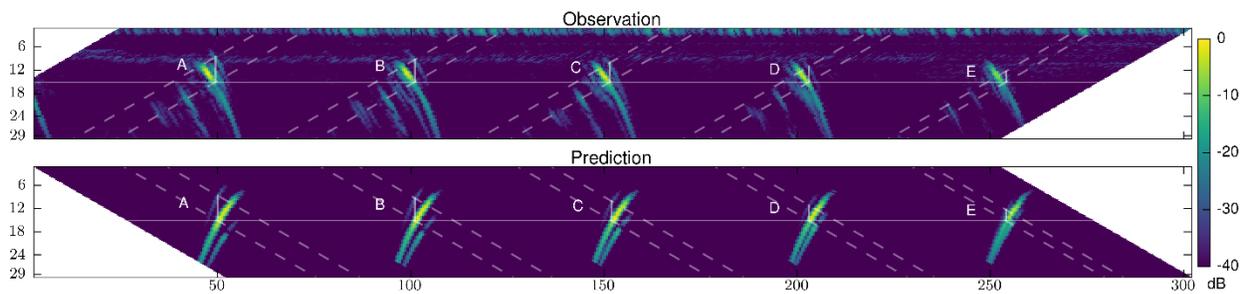
### B.2.14 Specimen 2, Probe 1, Skew 180



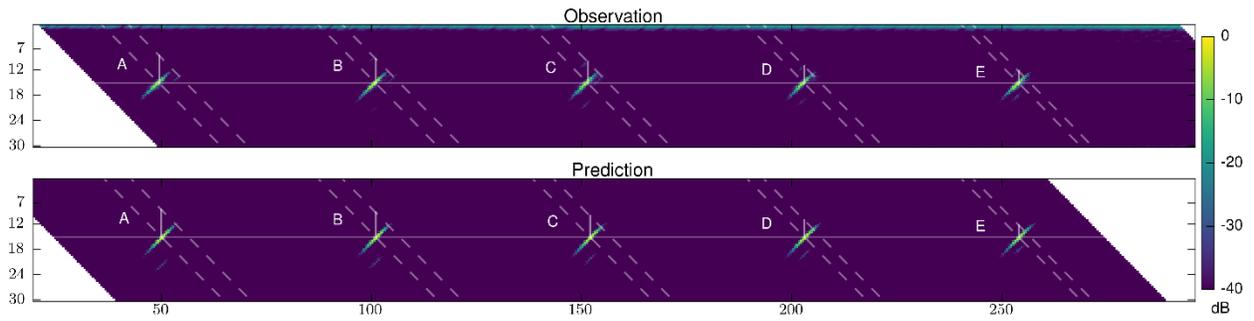
### B.2.15 Specimen 2, Probe 2, Skew 0



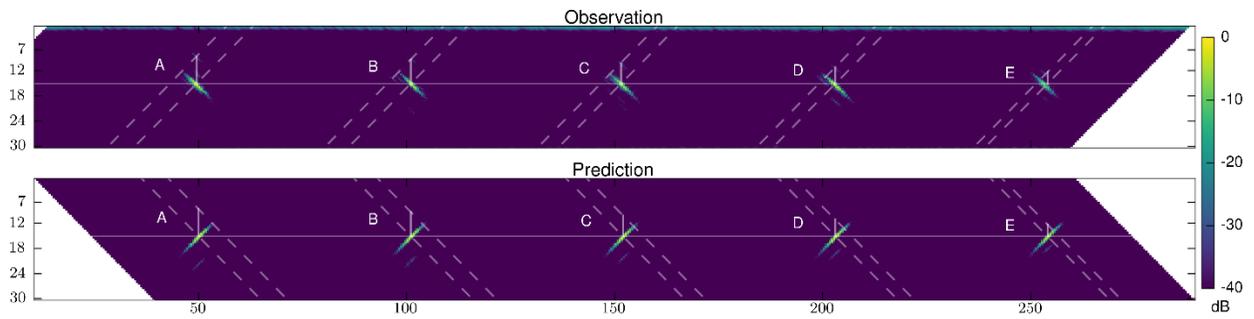
### B.2.16 Specimen 2, Probe 2, Skew 180



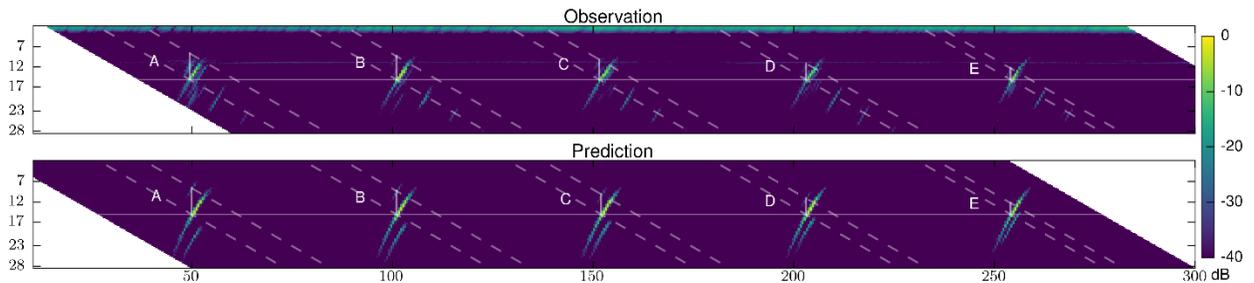
### B.2.17 Specimen 2, Probe 3, Skew 0



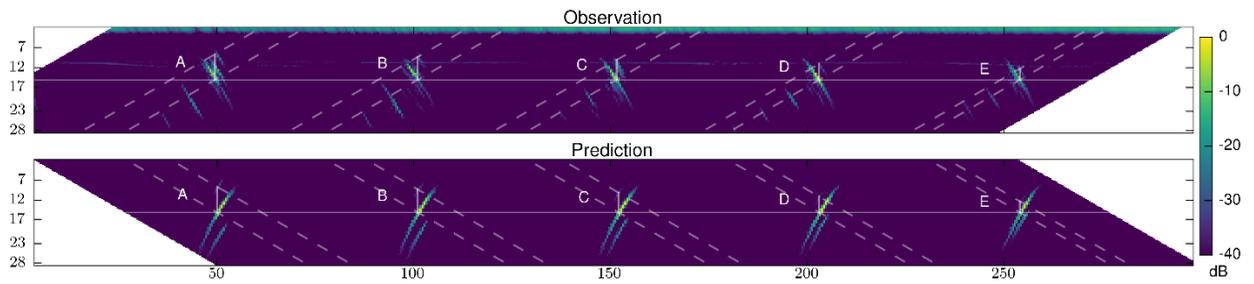
### B.2.18 Specimen 2, Probe 3, Skew 180



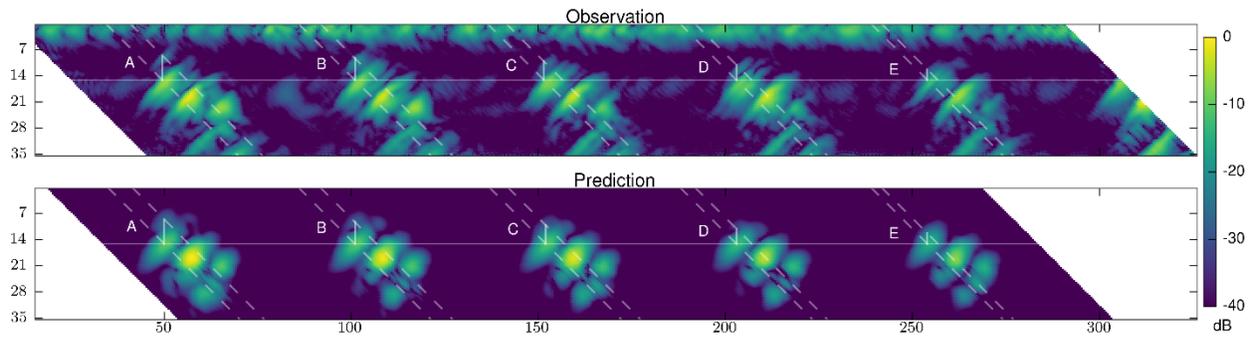
### B.2.19 Specimen 2, Probe 4, Skew 0



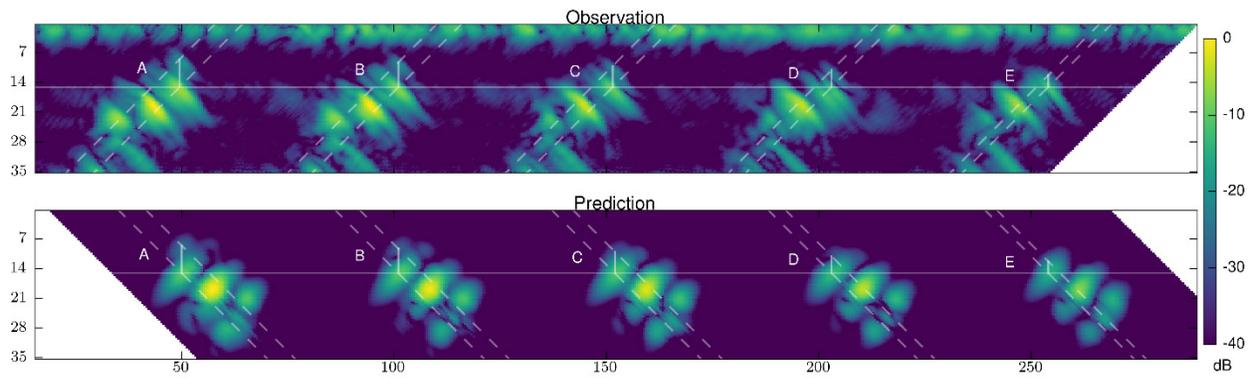
### B.2.20 Specimen 2, Probe 4, Skew 180



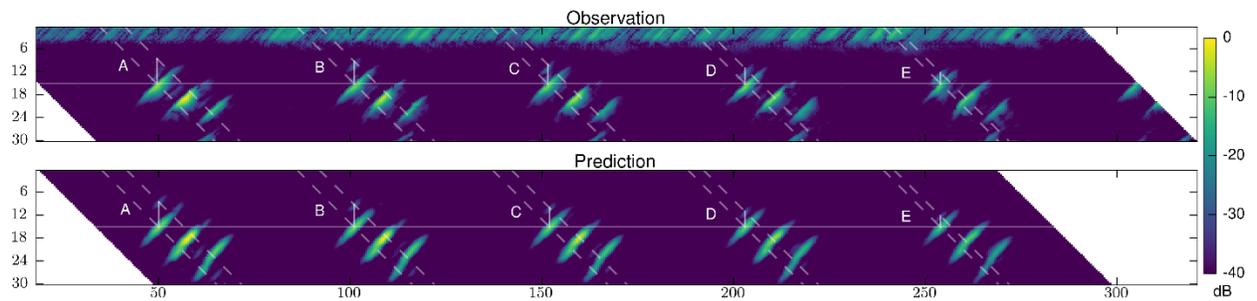
### B.2.21 Specimen 2, Probe 5, Skew 0



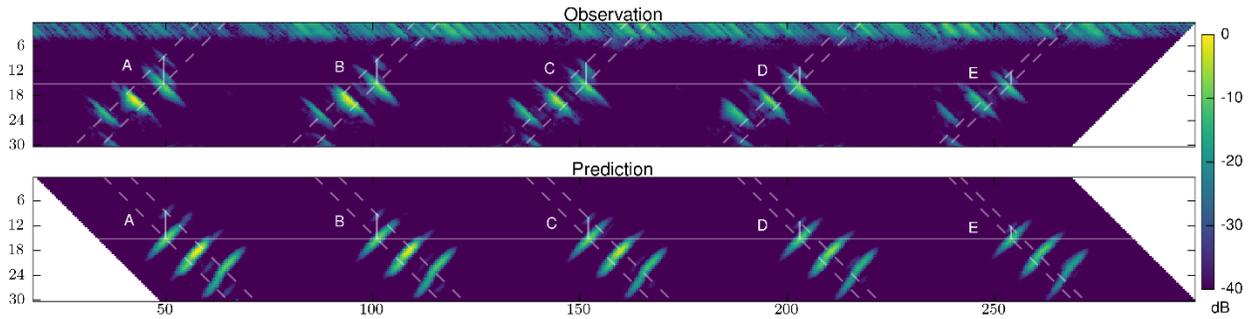
### B.2.22 Specimen 2, Probe 5, Skew 180



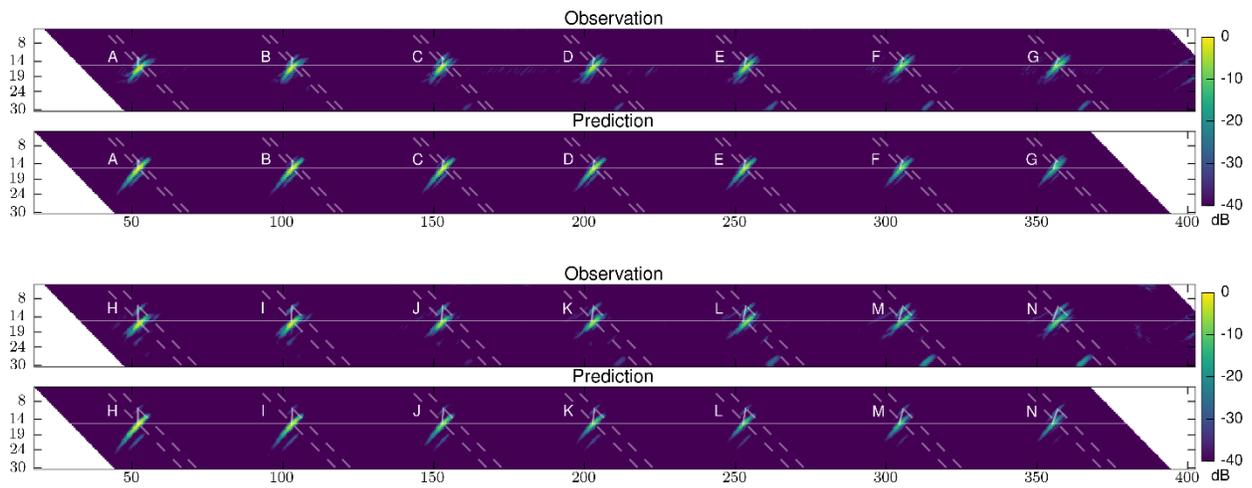
### B.2.23 Specimen 2, Probe 6, Skew 0



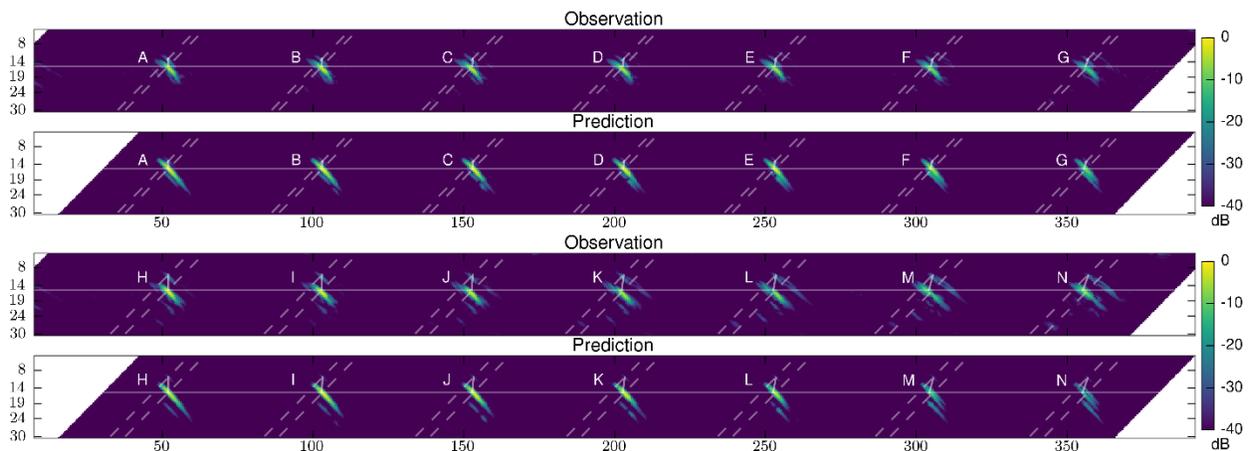
### B.2.24 Specimen 2, Probe 6, Skew 180



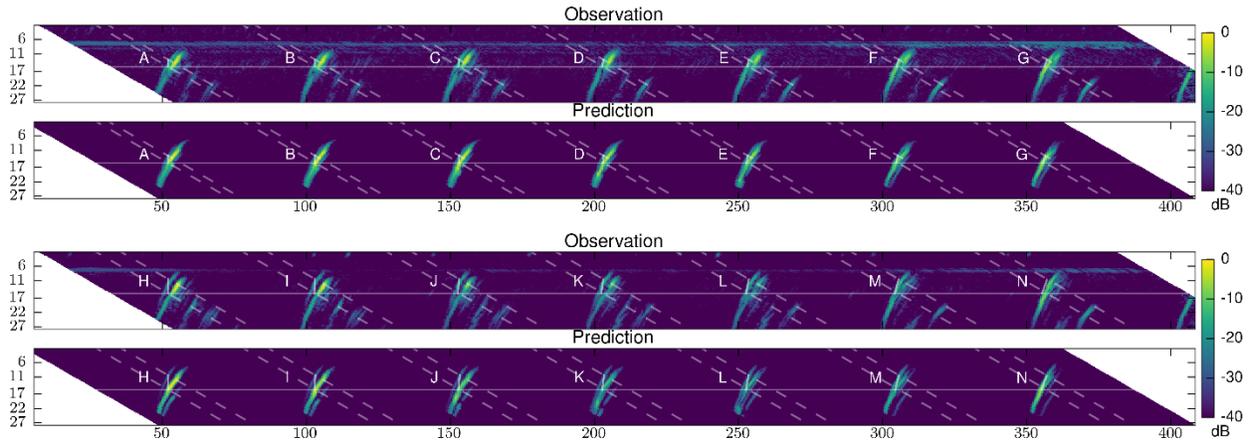
### B.2.25 Specimen 3, Probe 1, Skew 0



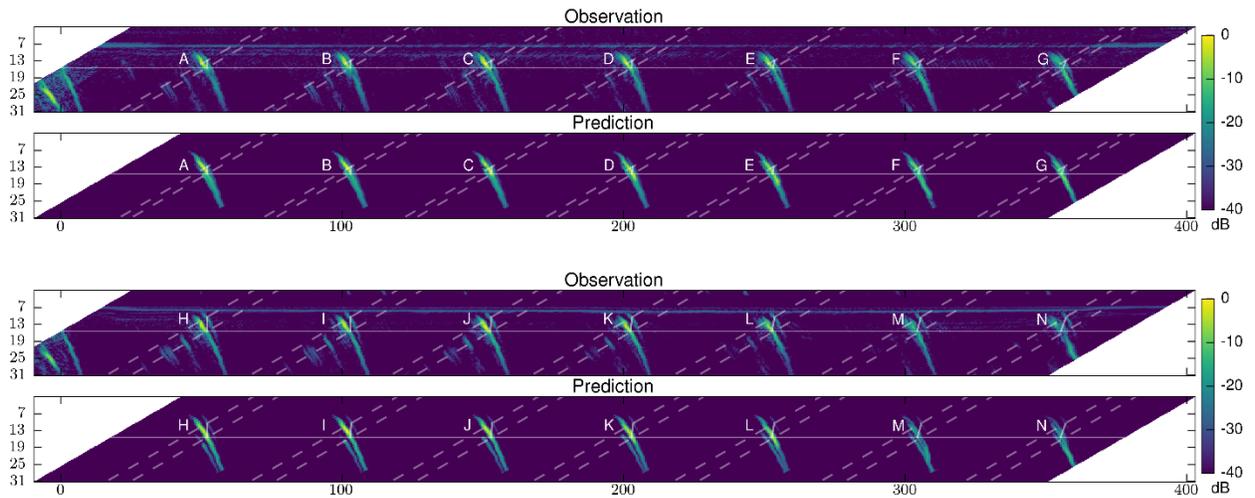
### B.2.26 Specimen 3, Probe 1, Skew 180



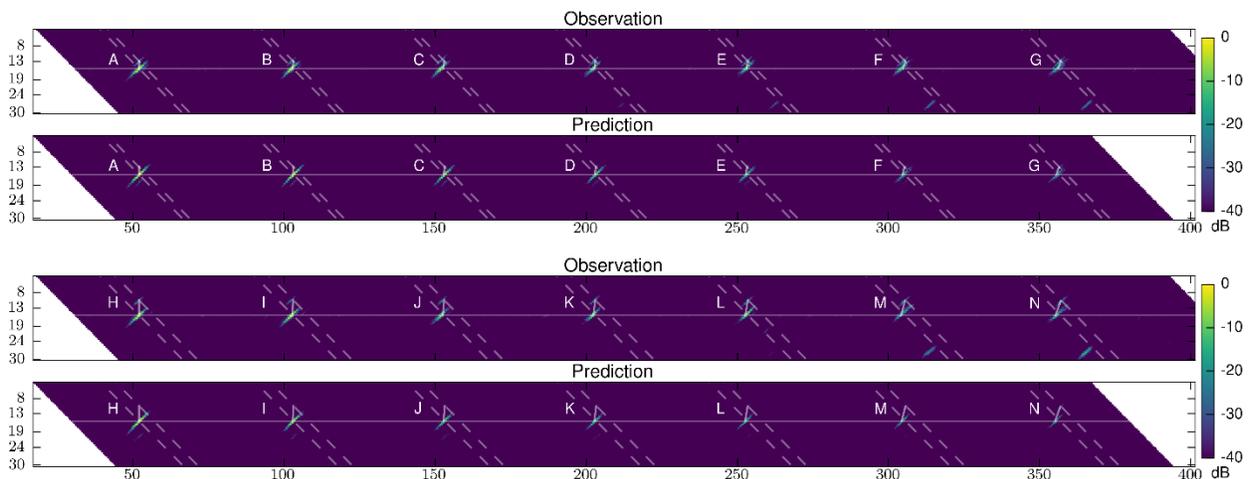
### B.2.27 Specimen 3, Probe 2, Skew 0



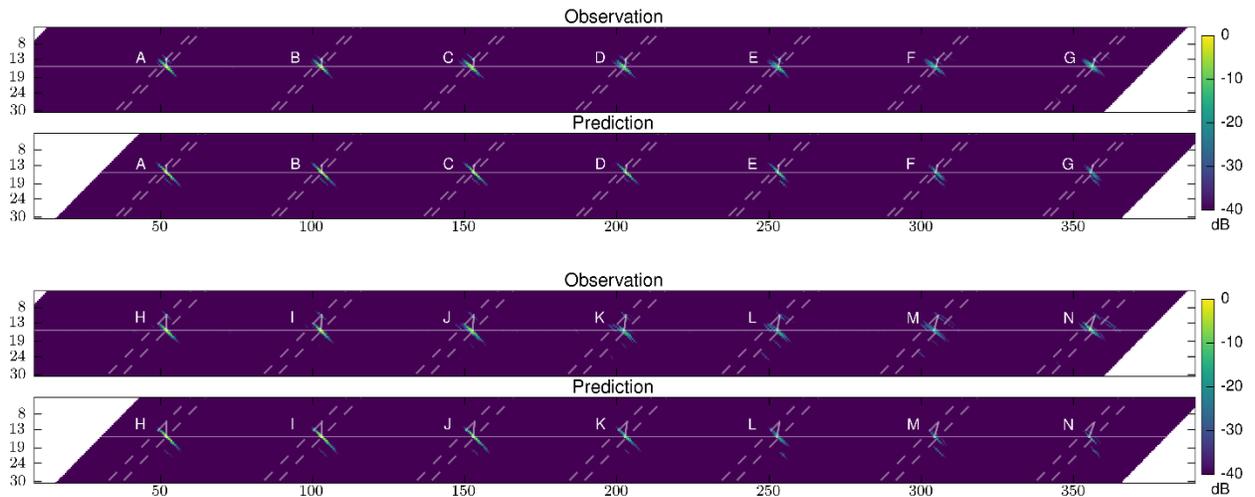
### B.2.28 Specimen 3, Probe 2, Skew 180



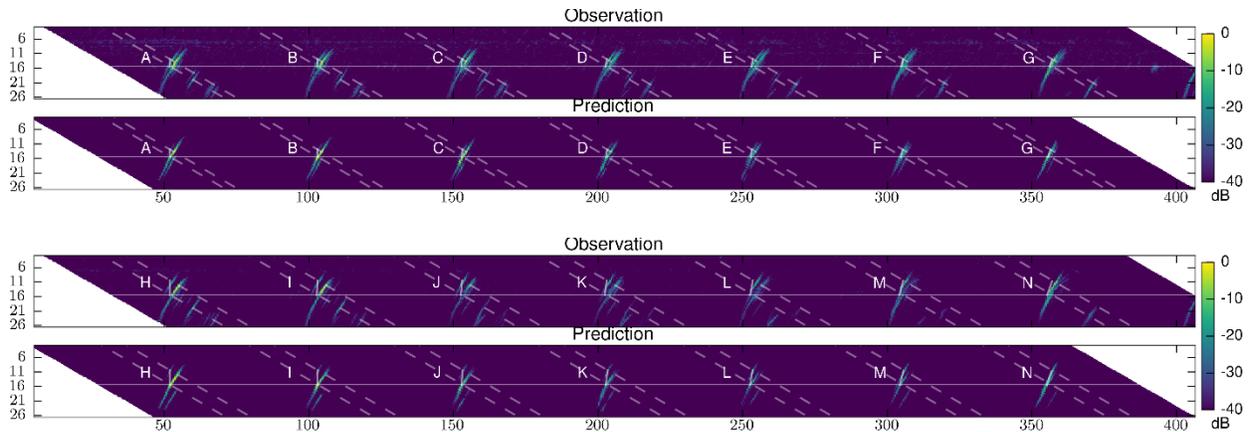
### B.2.29 Specimen 3, Probe 3, Skew 0



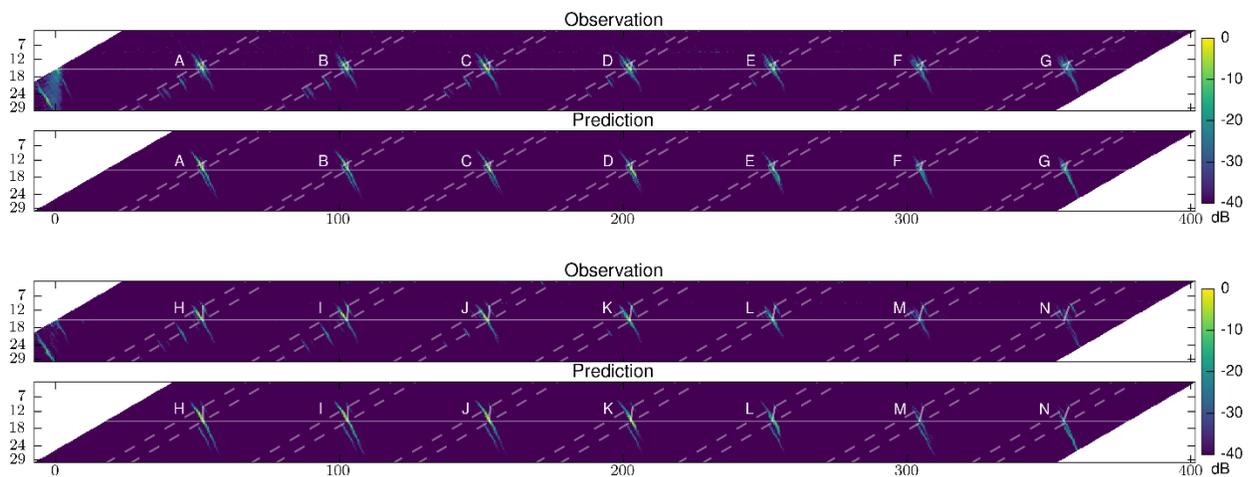
### B.2.30 Specimen 3, Probe 3, Skew 180



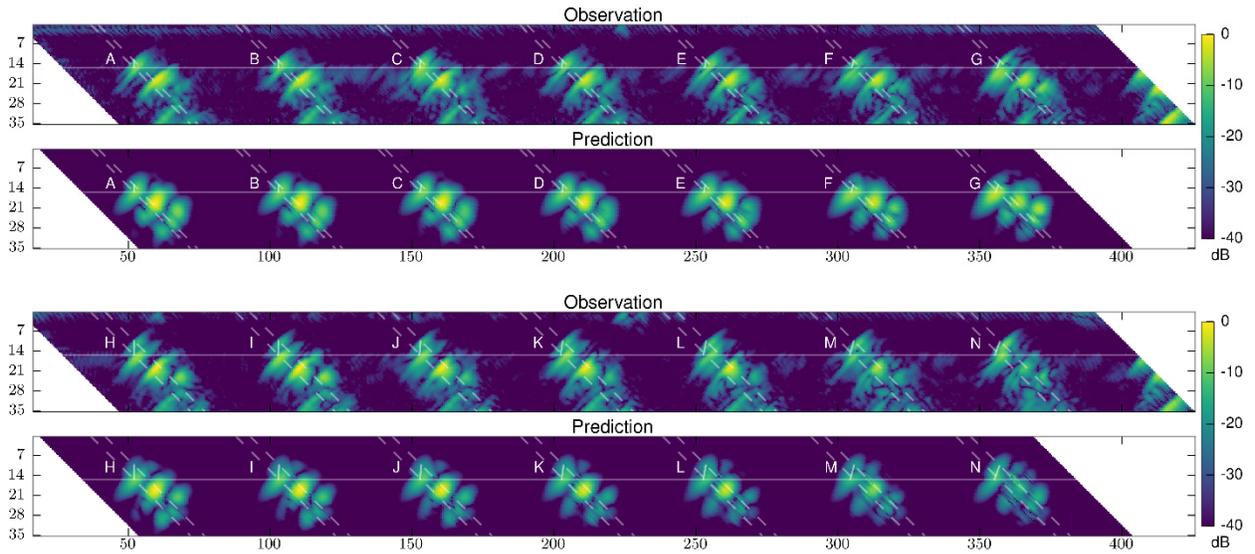
### B.2.31 Specimen 3, Probe 4, Skew 0



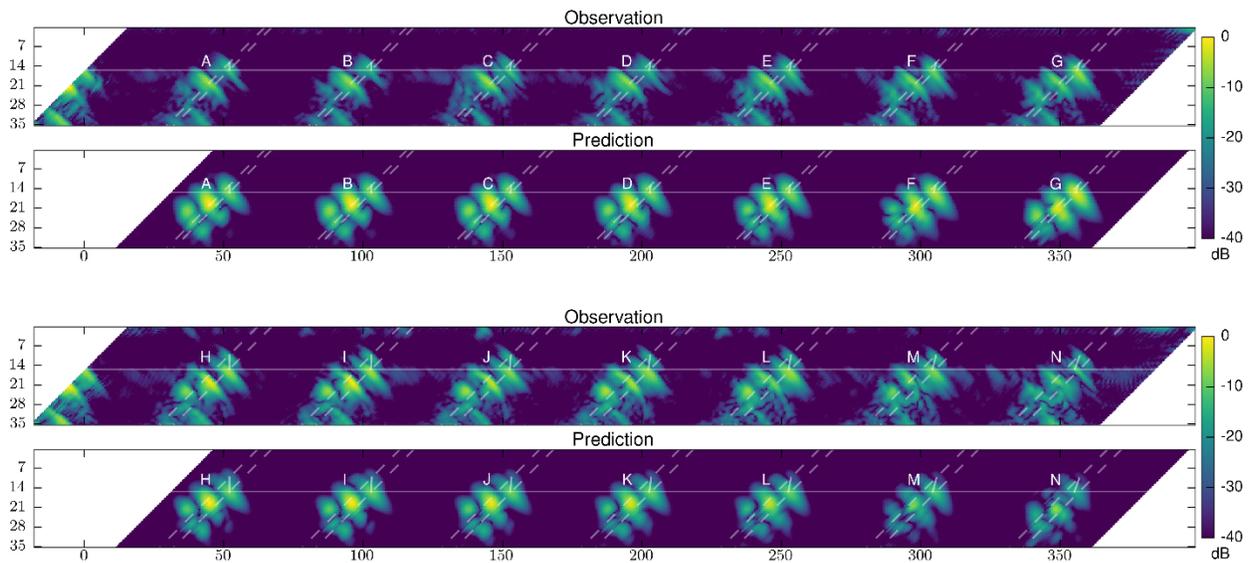
### B.2.32 Specimen 3, Probe 4, Skew 180



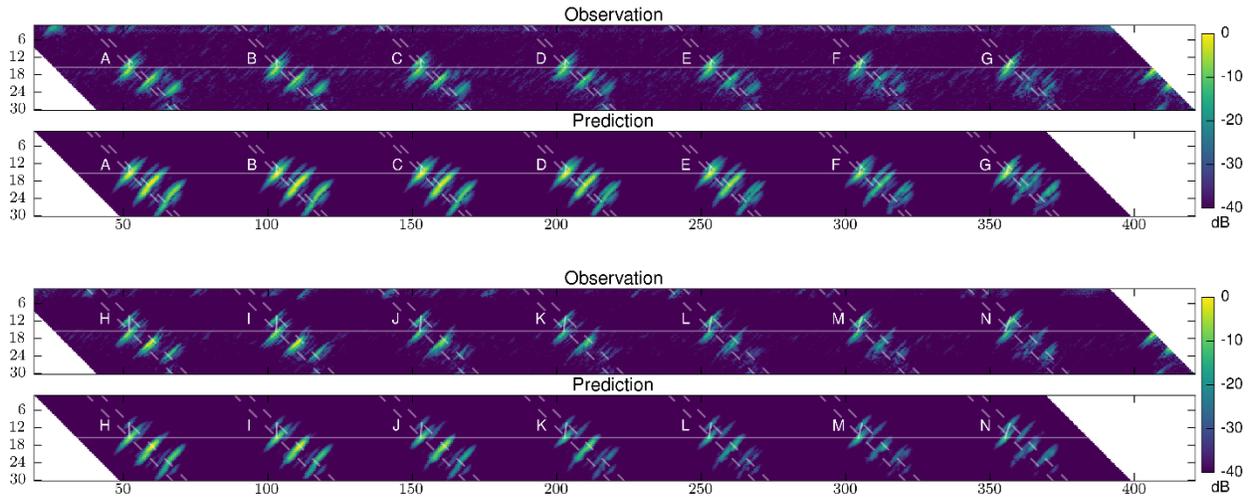
### B.2.33 Specimen 3, Probe 5, Skew 0



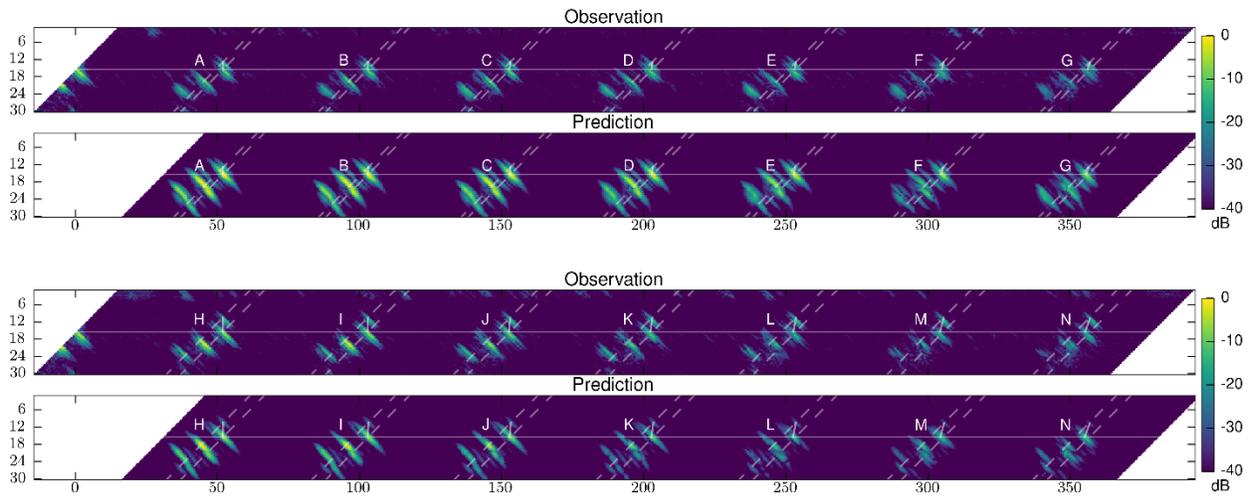
### B.2.34 Specimen 3, Probe 5, Skew 180



### B.2.35 Specimen 3, Probe 6, Skew 0

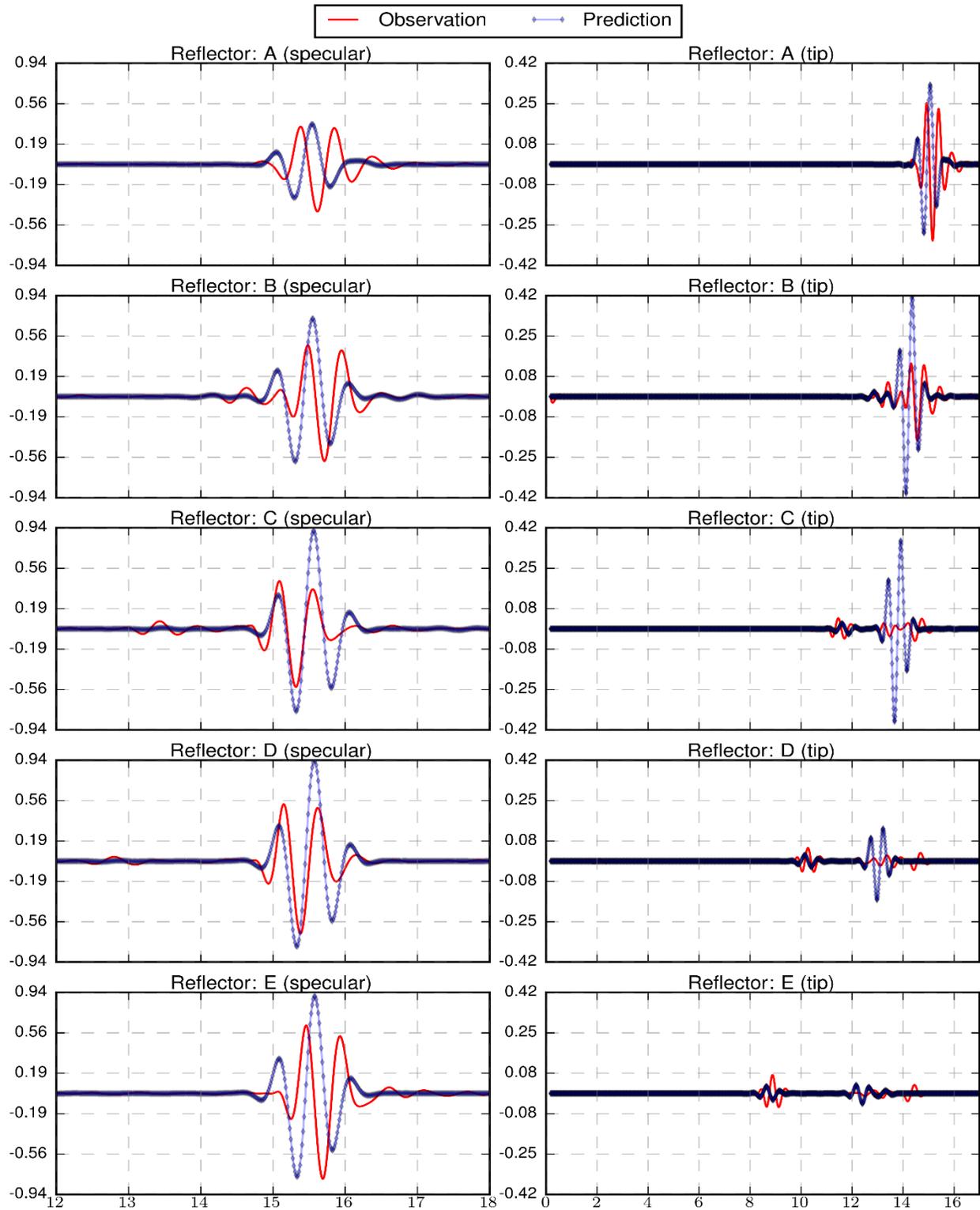


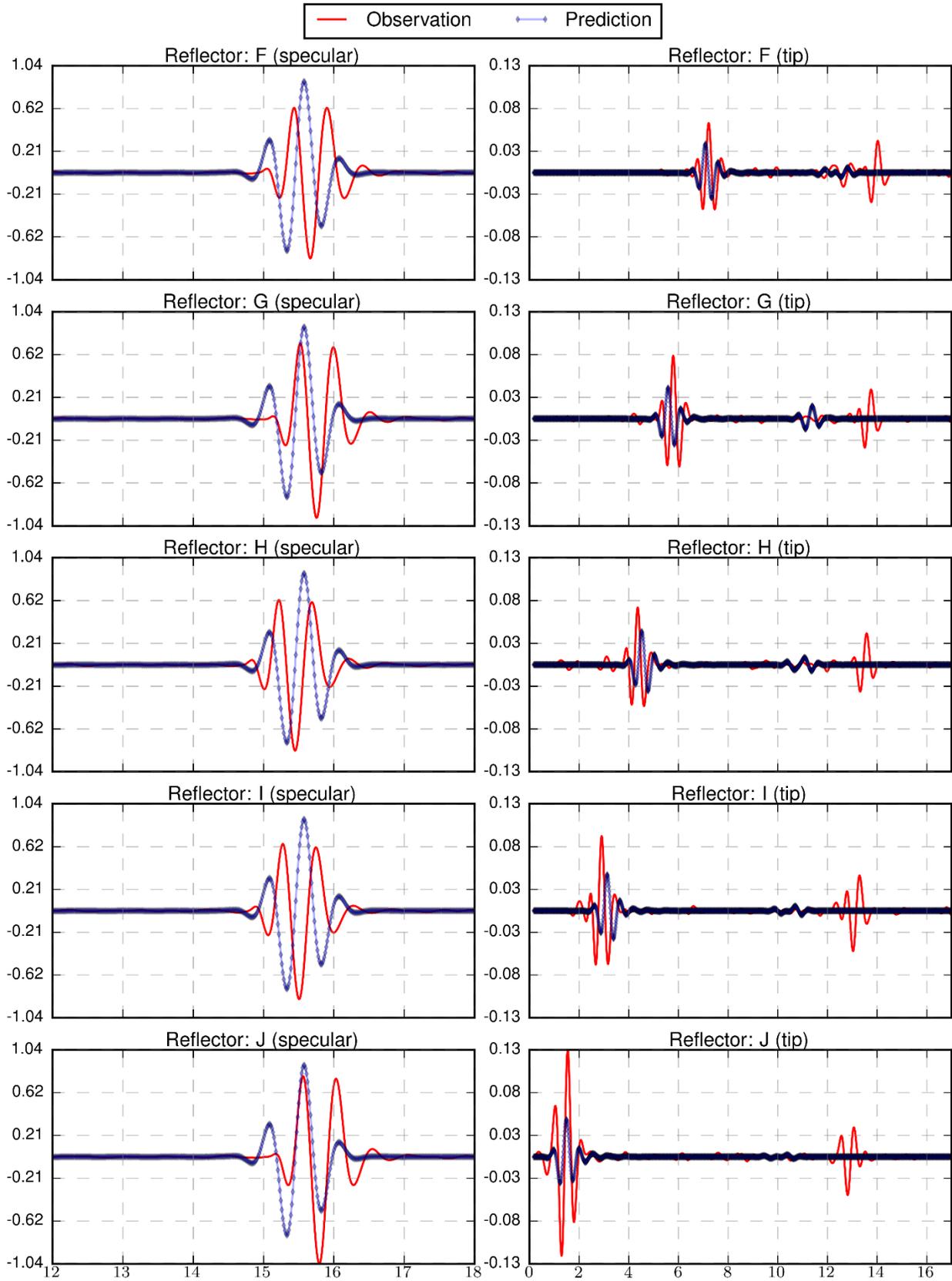
### B.2.36 Specimen 3, Probe 6, Skew 180



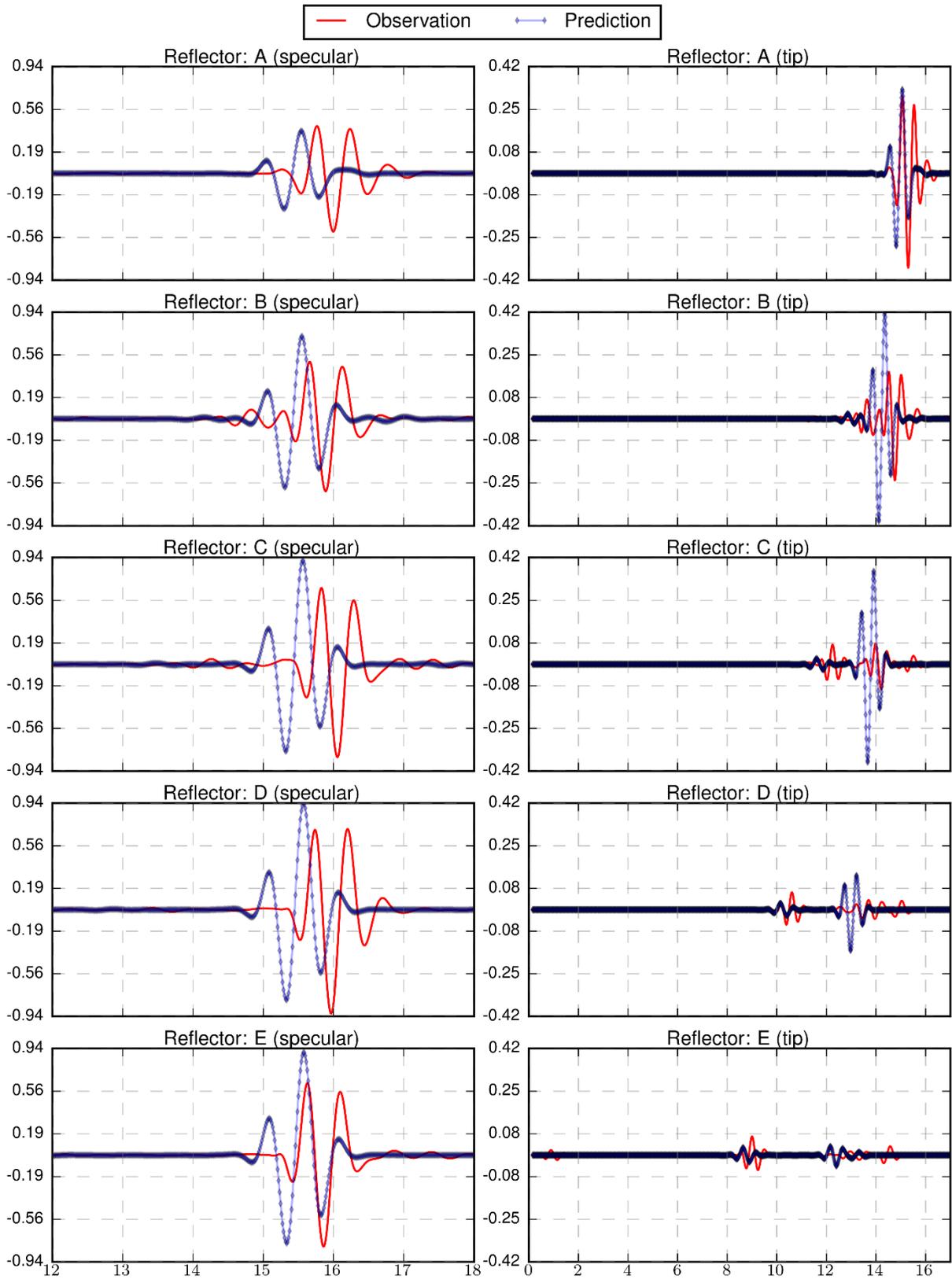
## B.3 A-scan Views

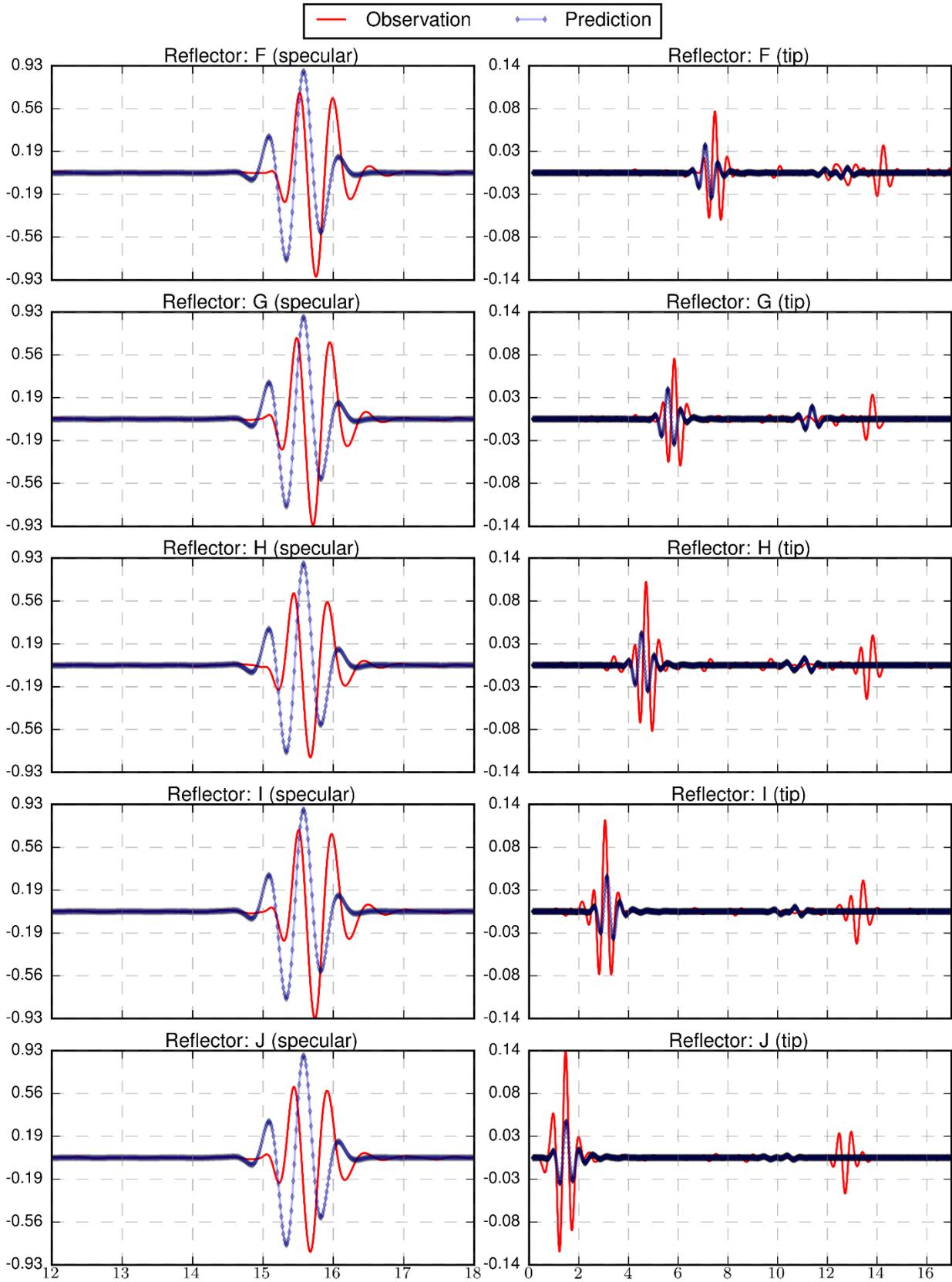
### B.3.1 Specimen 1, Probe 1, Skew 0



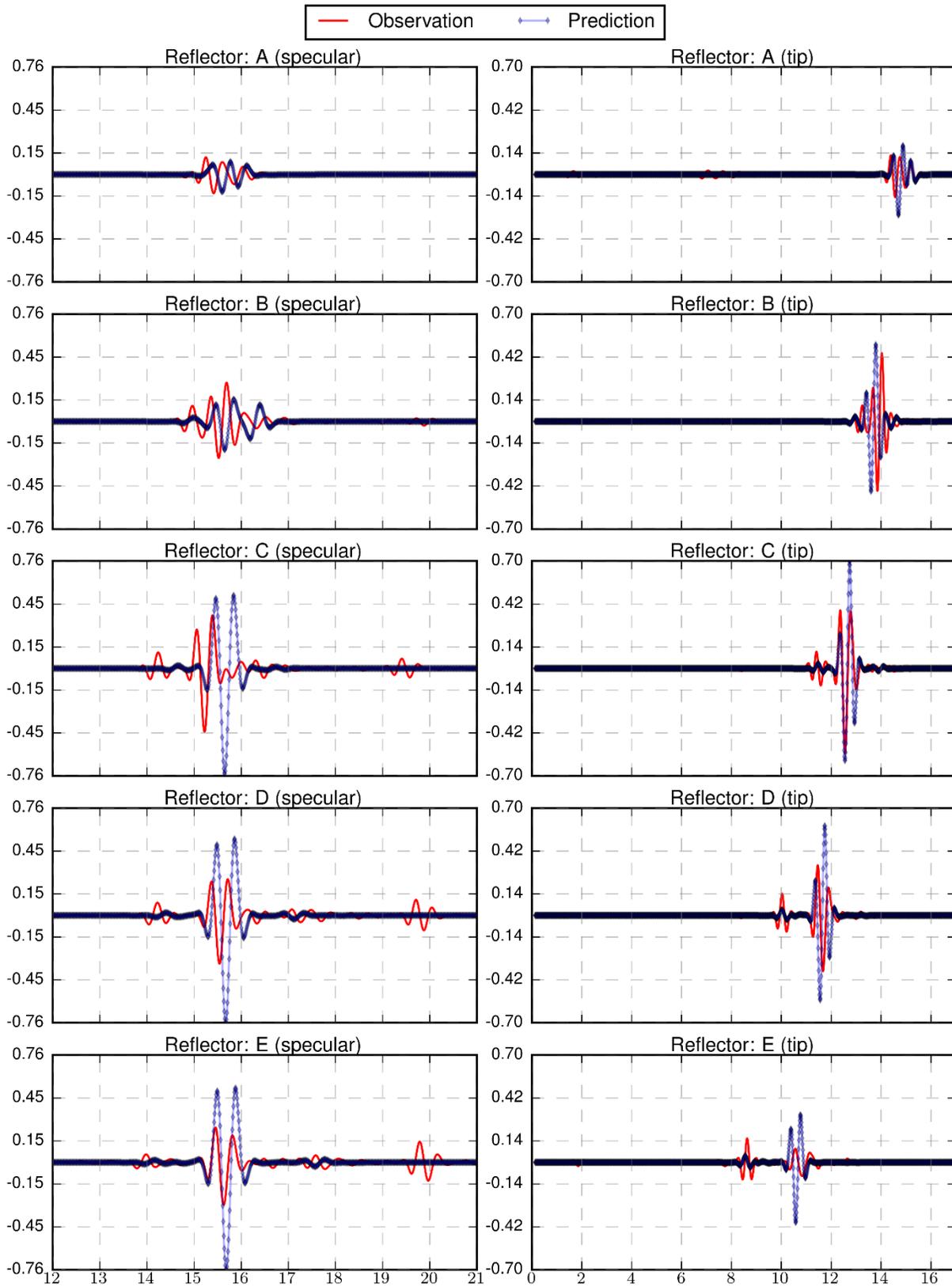


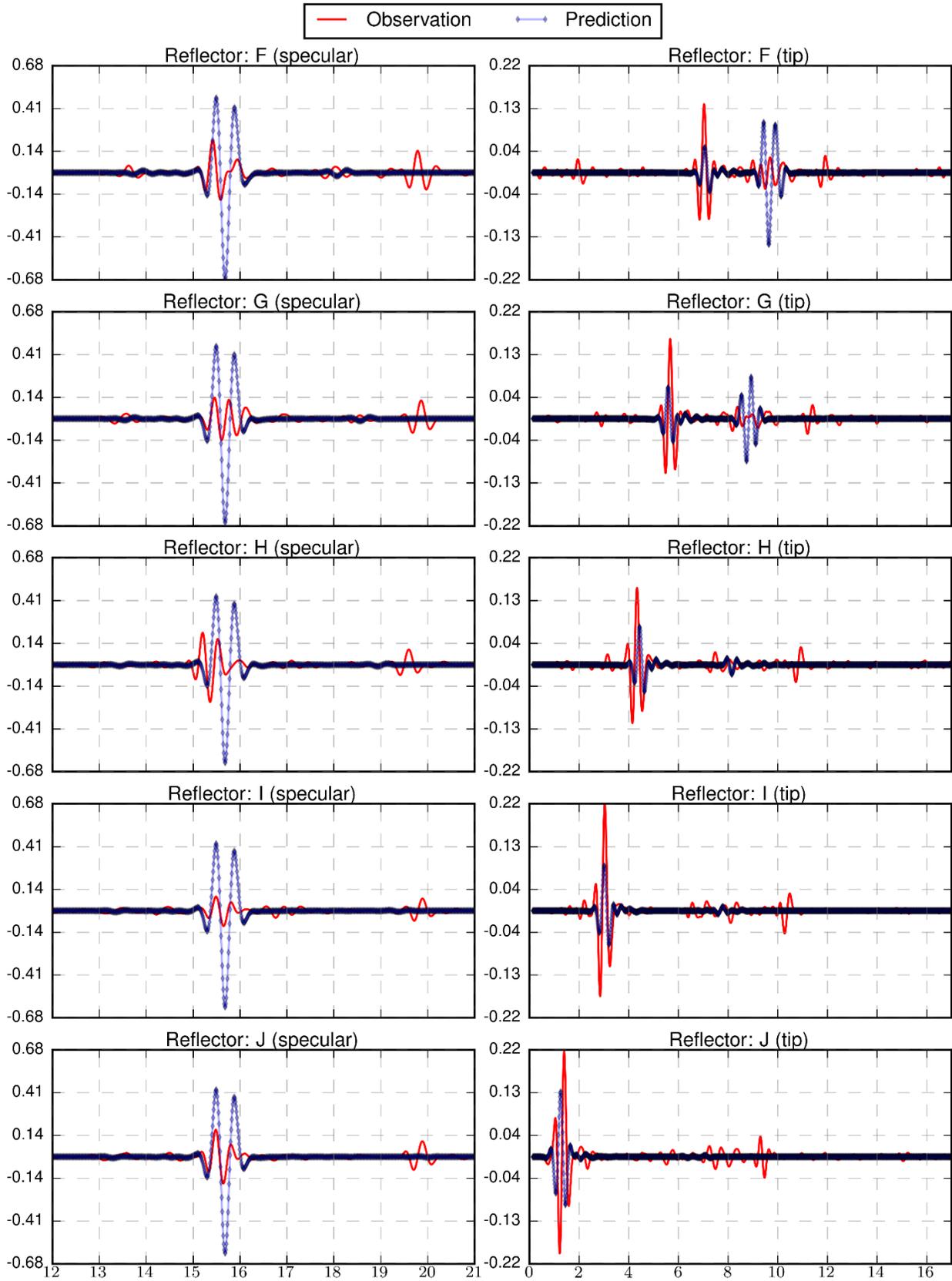
### B.3.2 Specimen 1, Probe 1, Skew 180



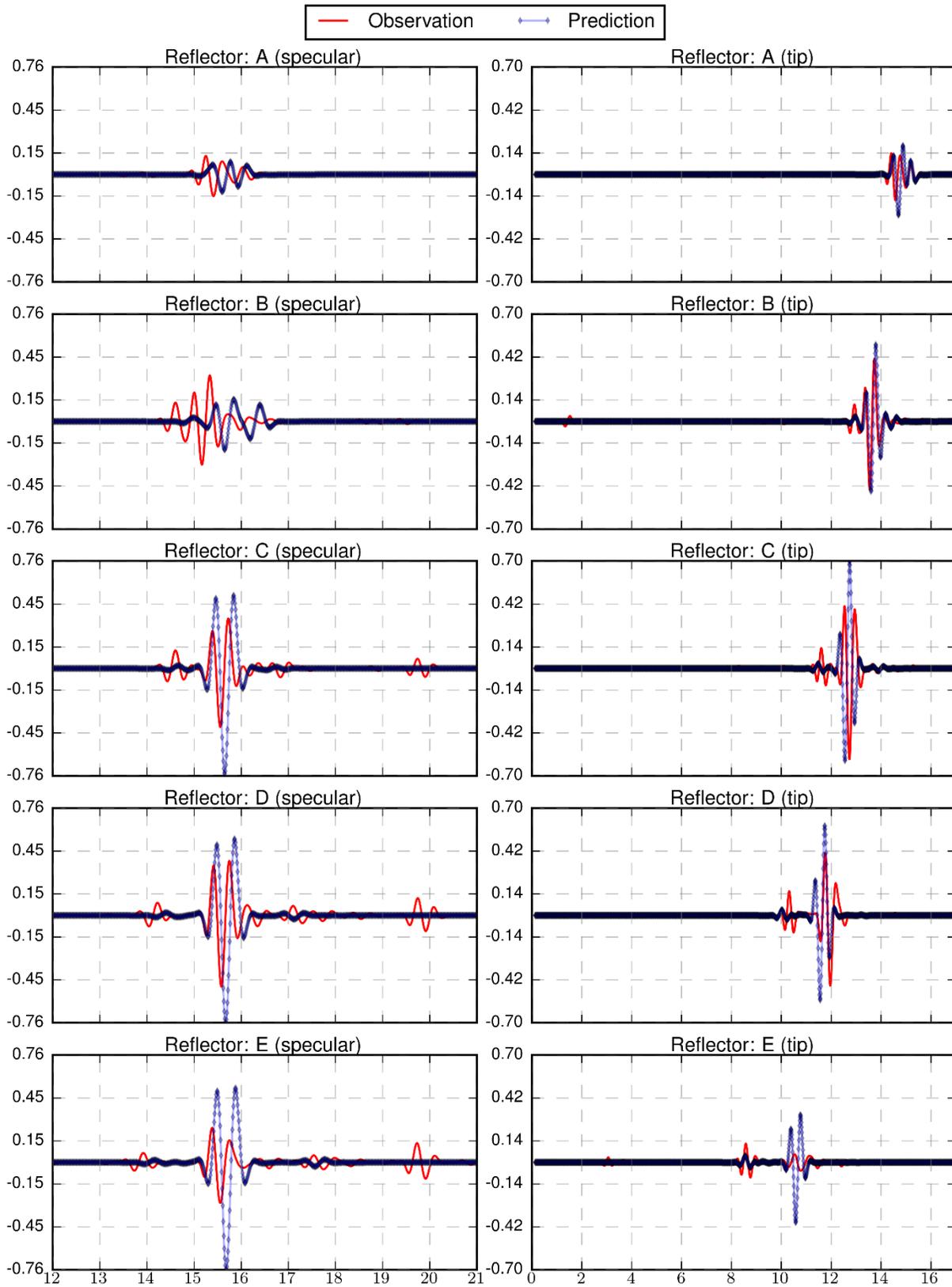


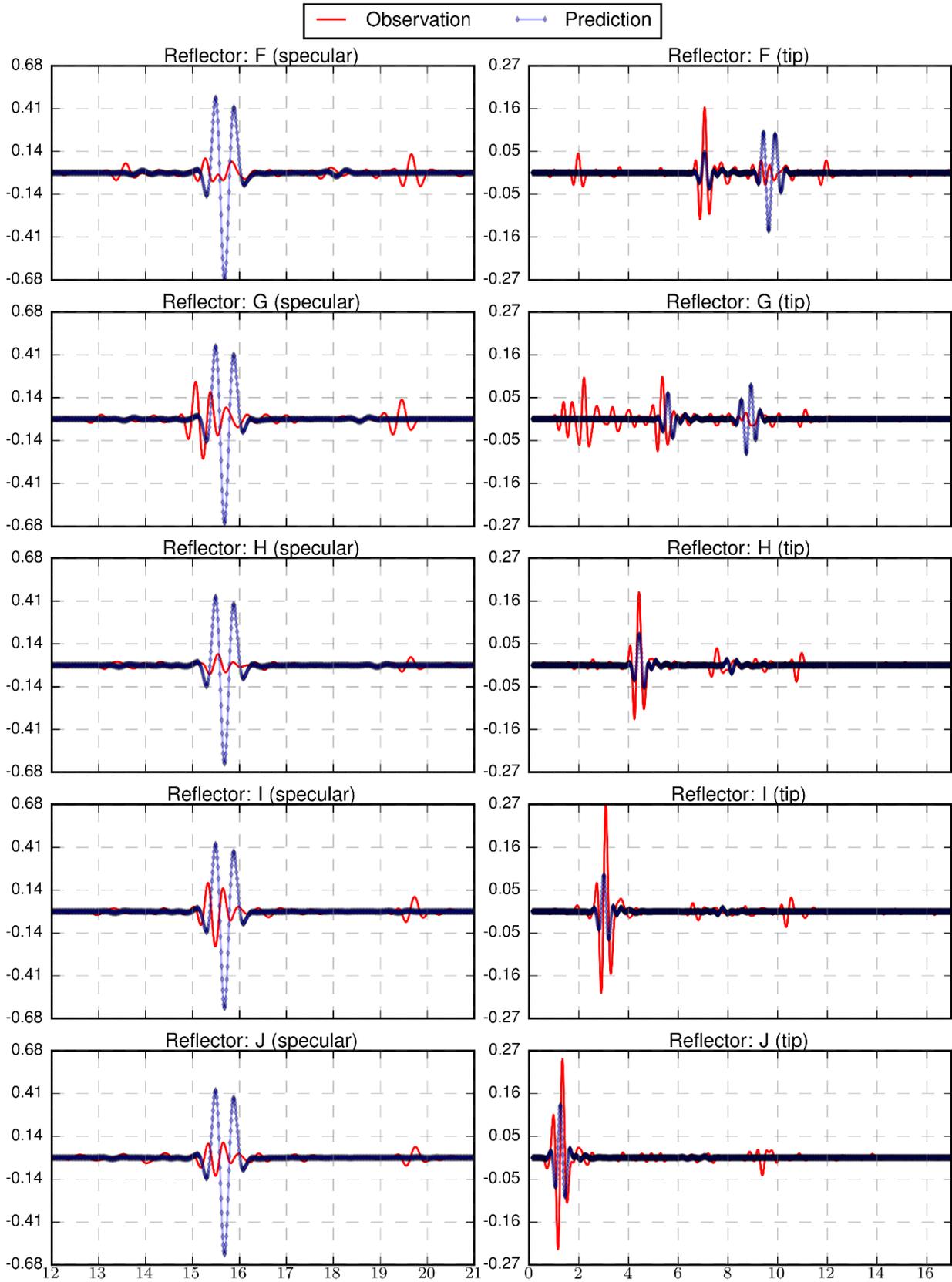
### B.3.3 Specimen 1, Probe 2, Skew 0



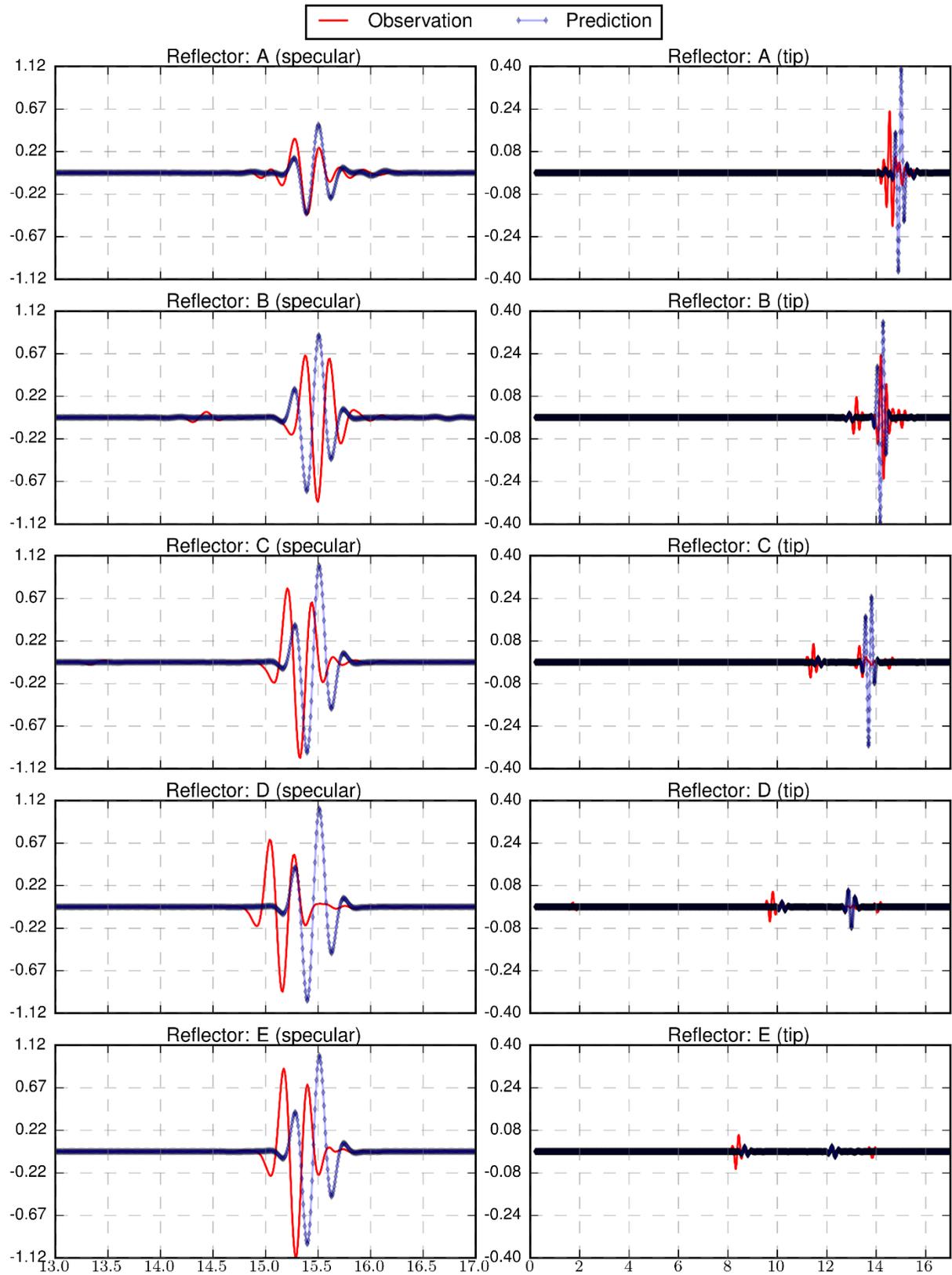


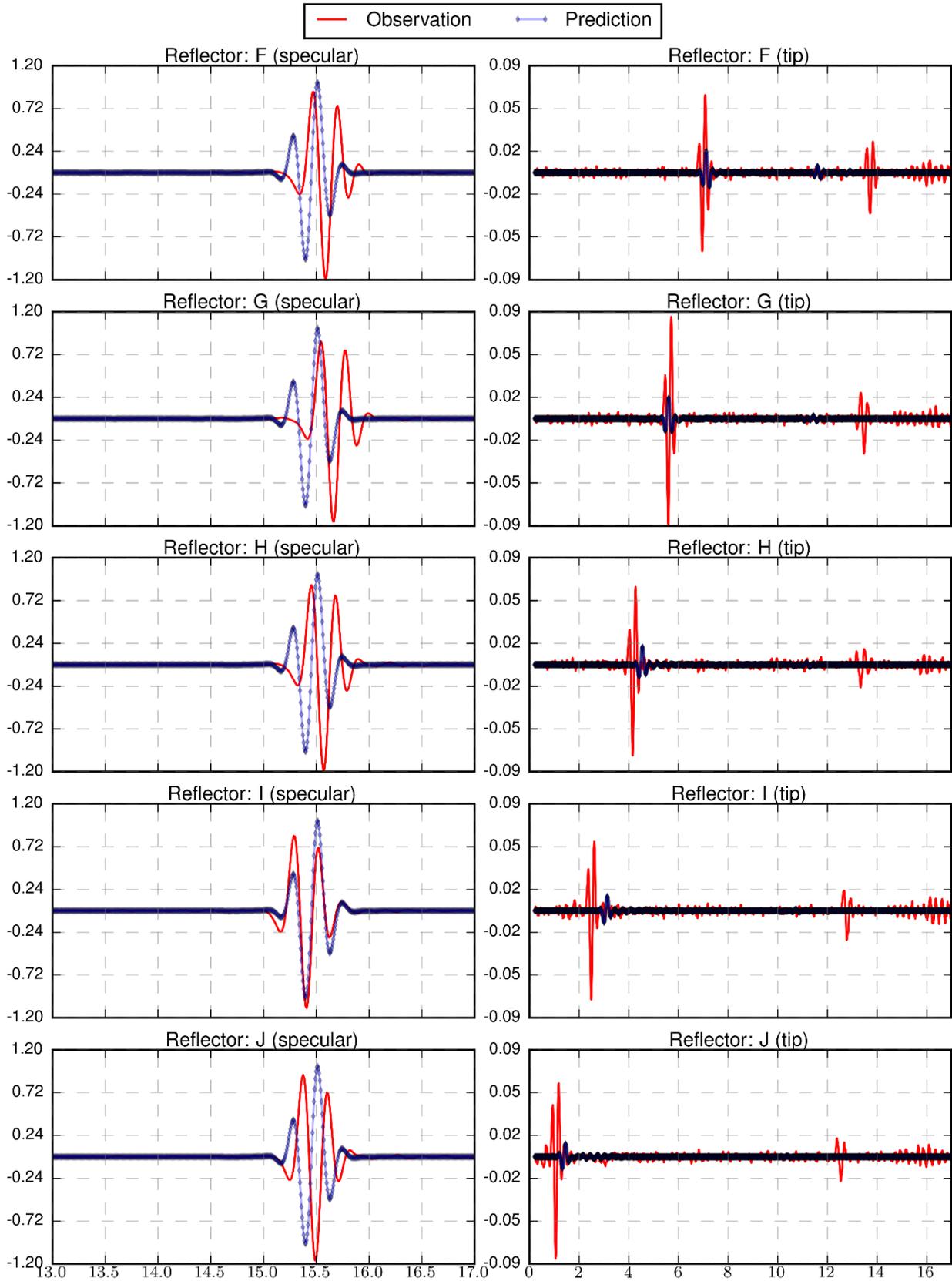
### B.3.4 Specimen 1, Probe 2, Skew 180



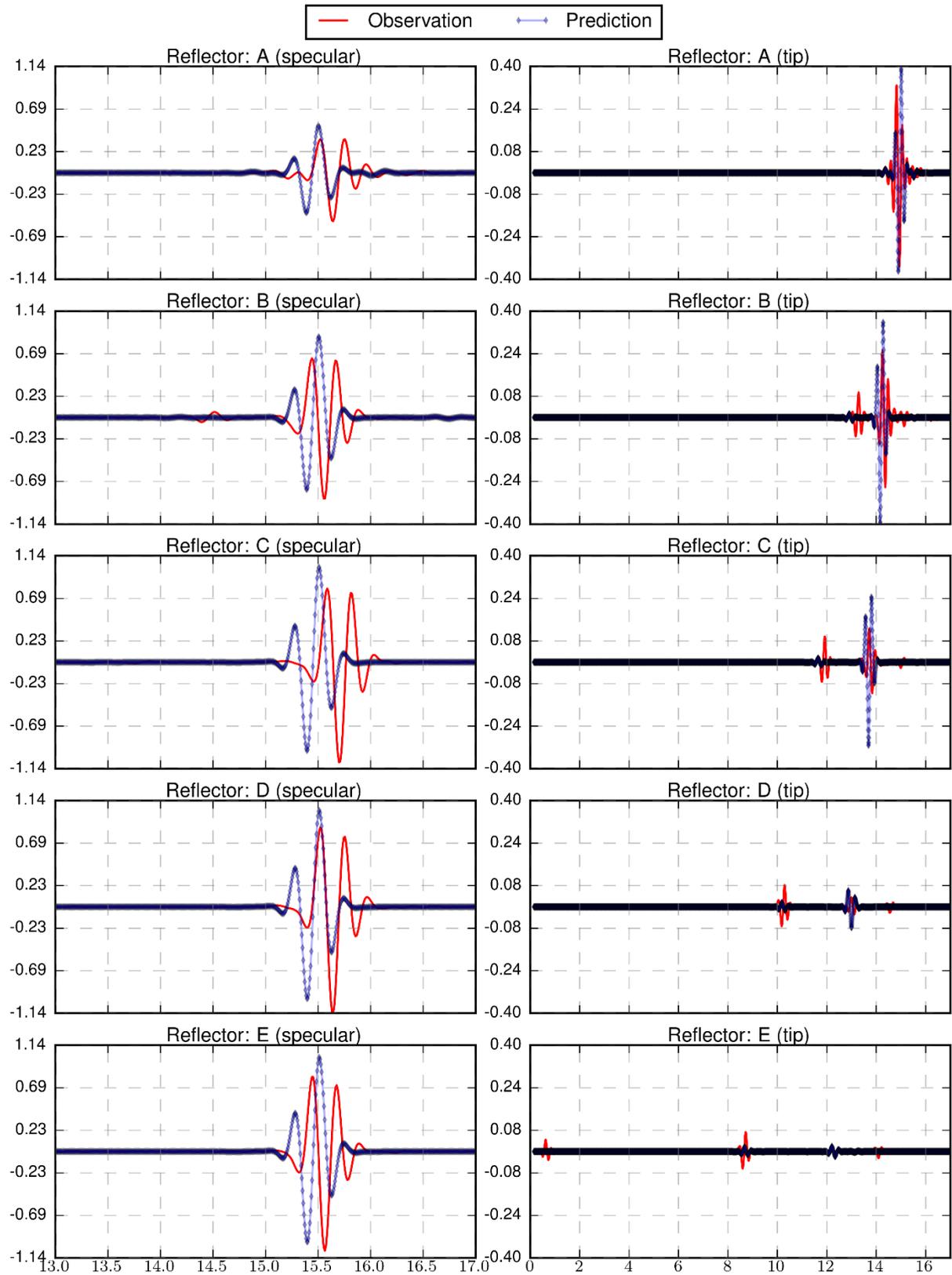


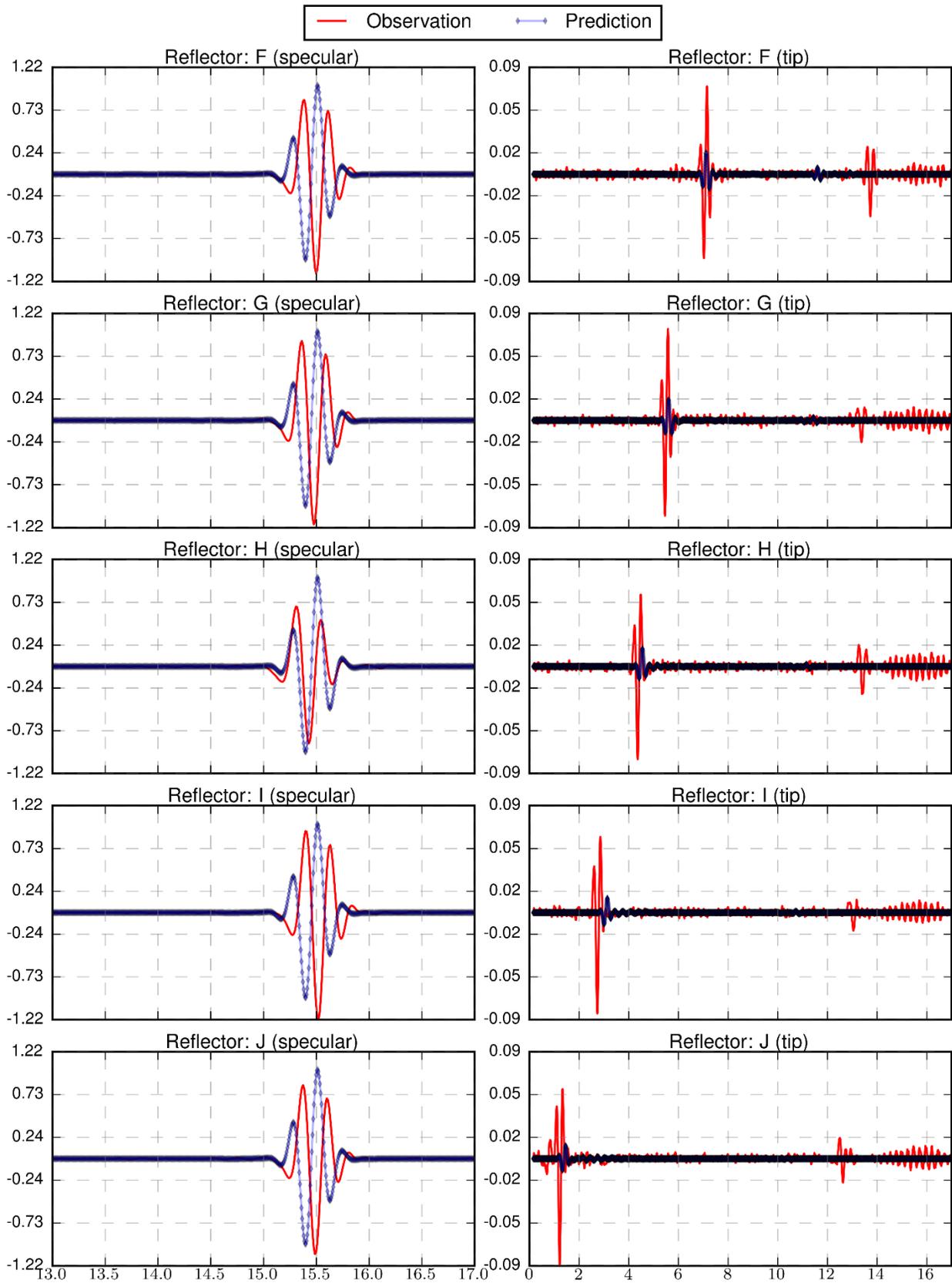
### B.3.5 Specimen 1, Probe 3, Skew 0



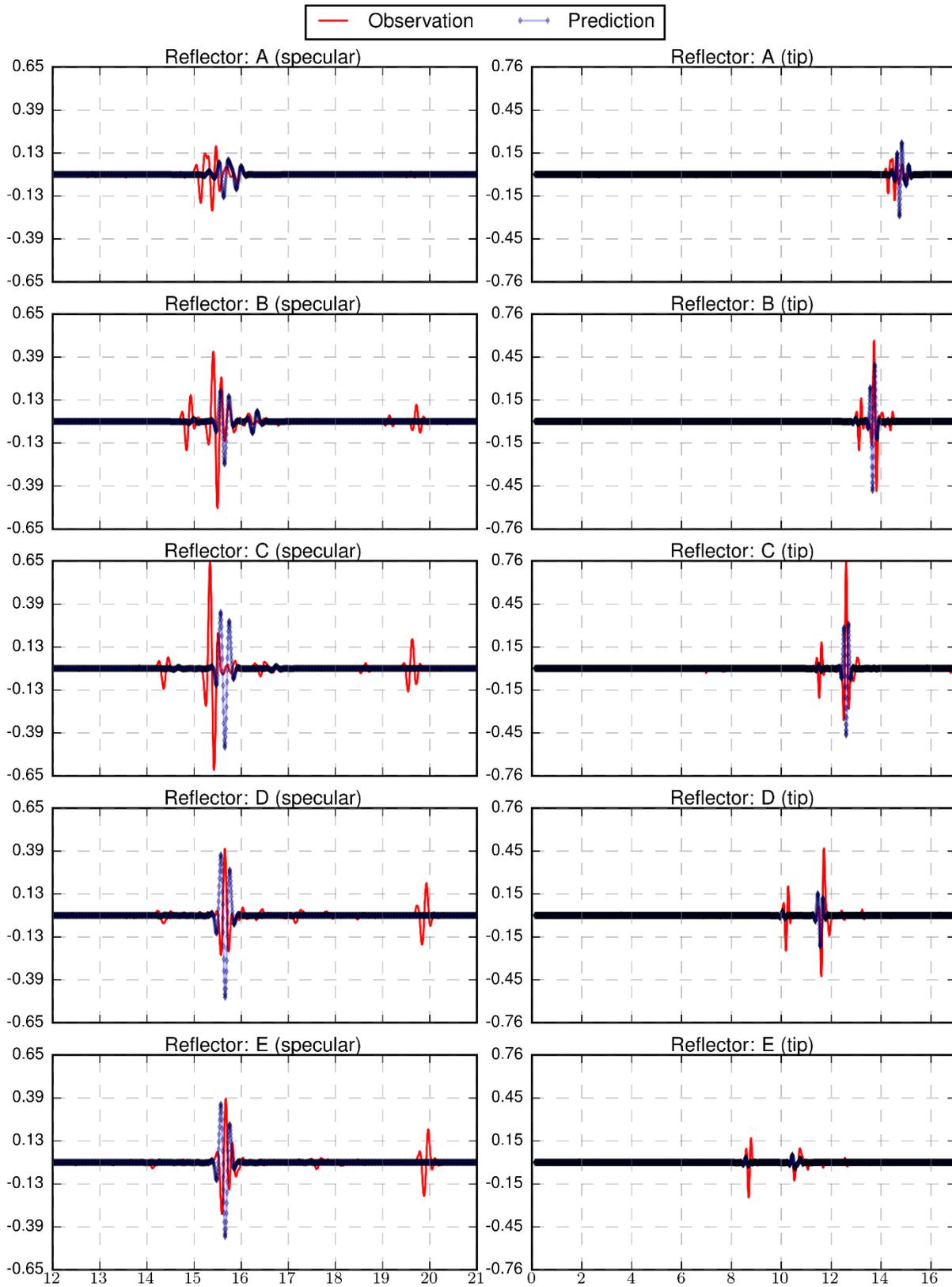


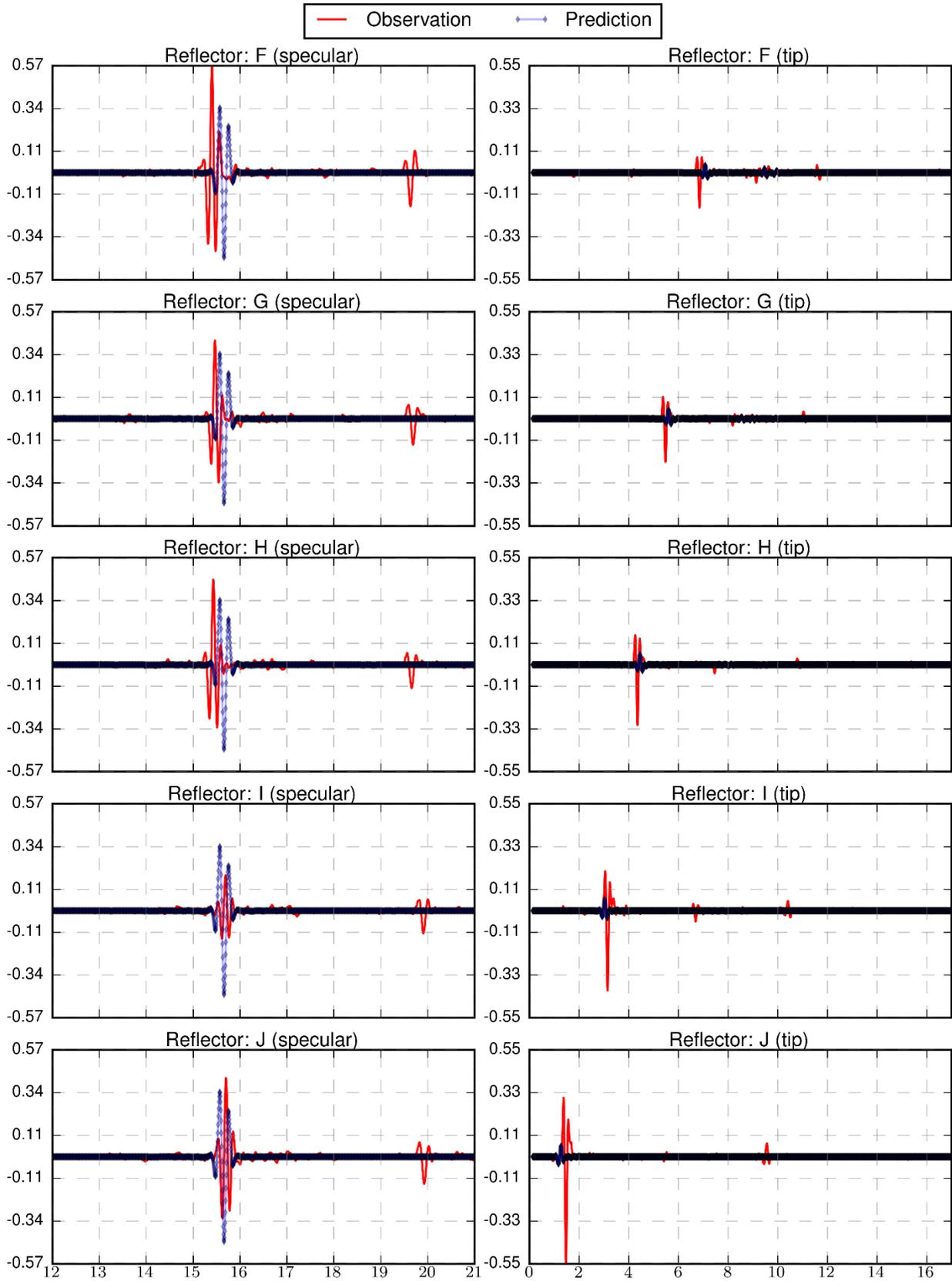
### B.3.6 Specimen 1, Probe 3, Skew 180



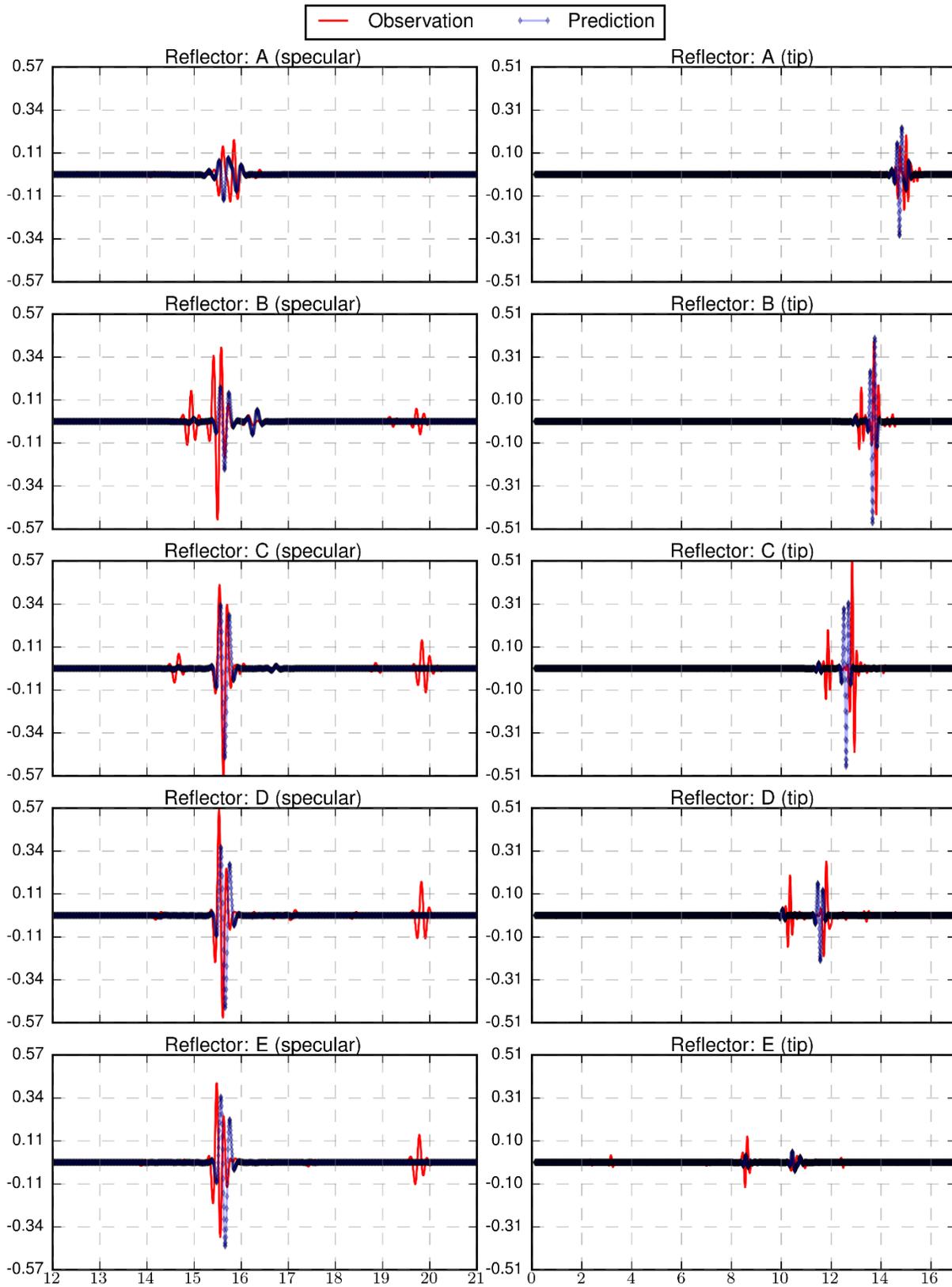


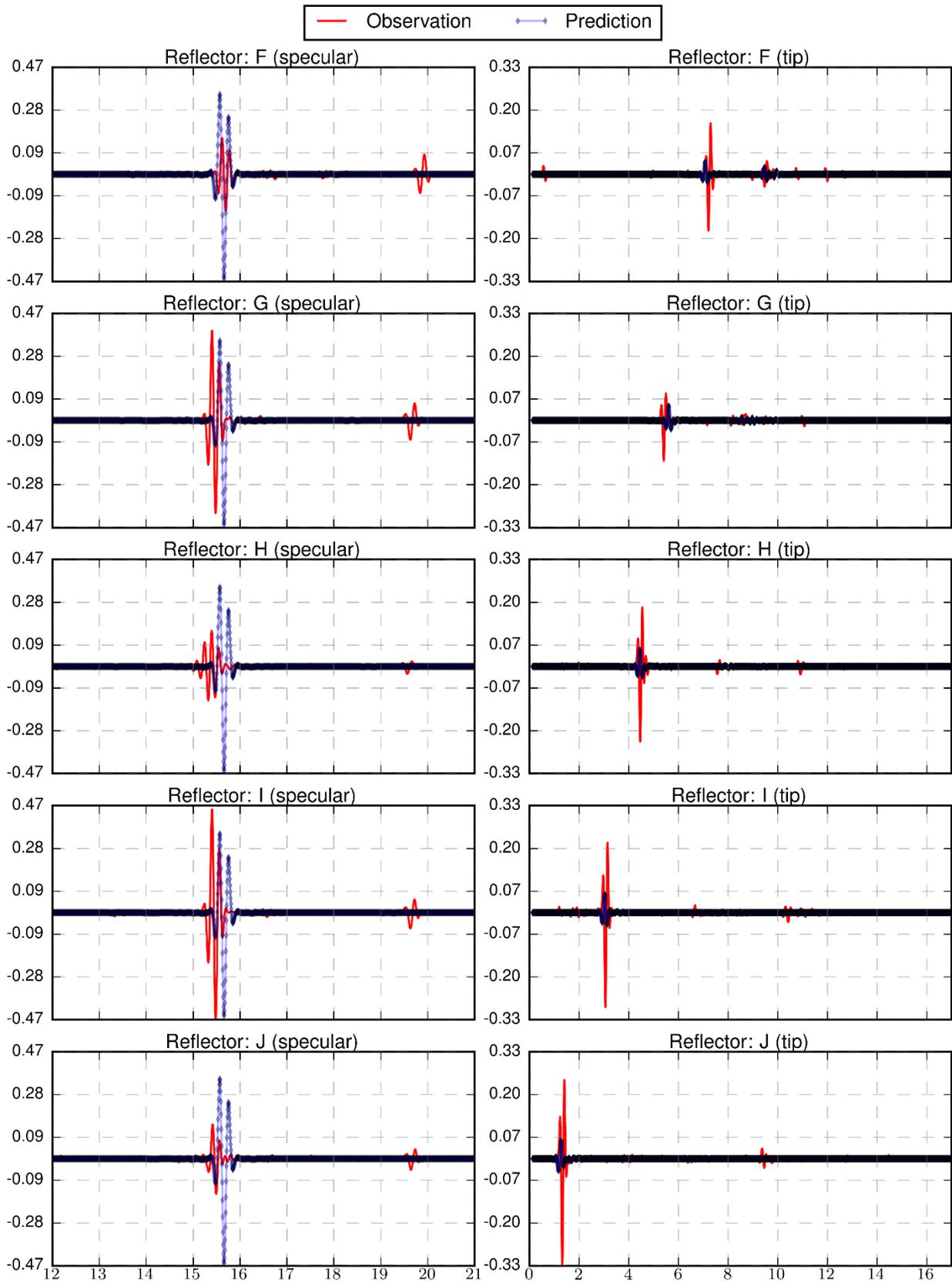
### B.3.7 Specimen 1, Probe 4, Skew 0



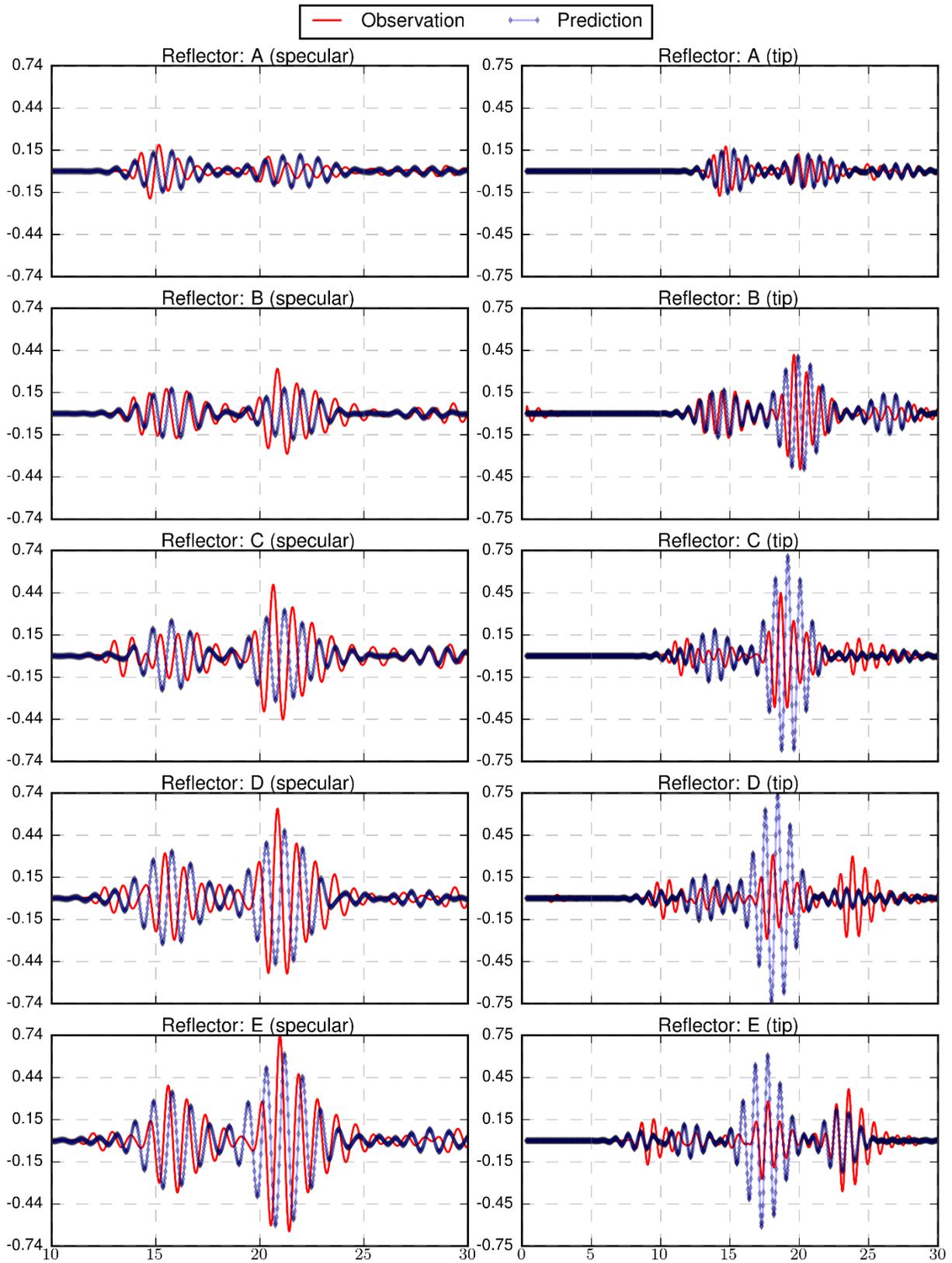


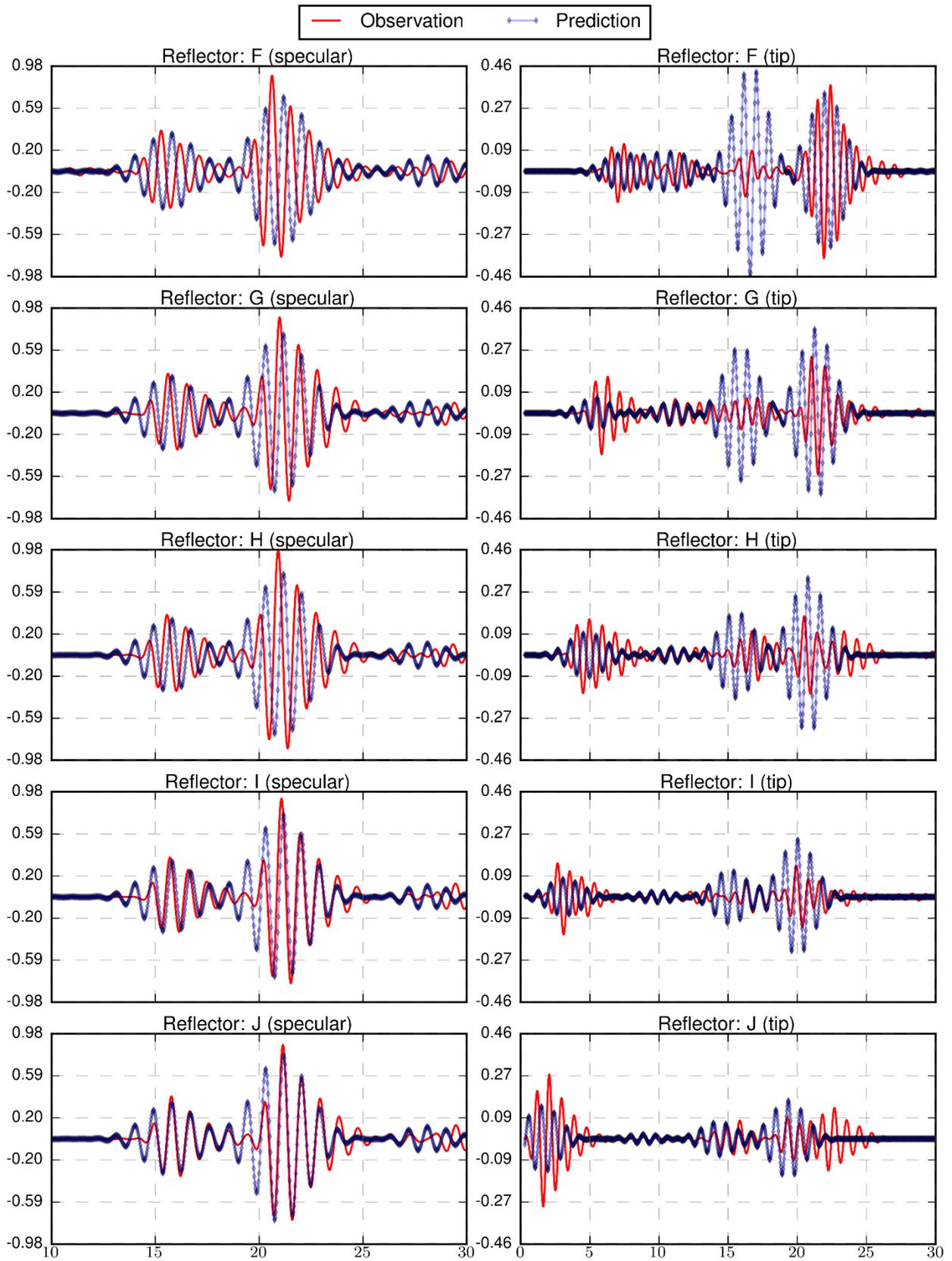
### B.3.8 Specimen 1, Probe 4, Skew 180



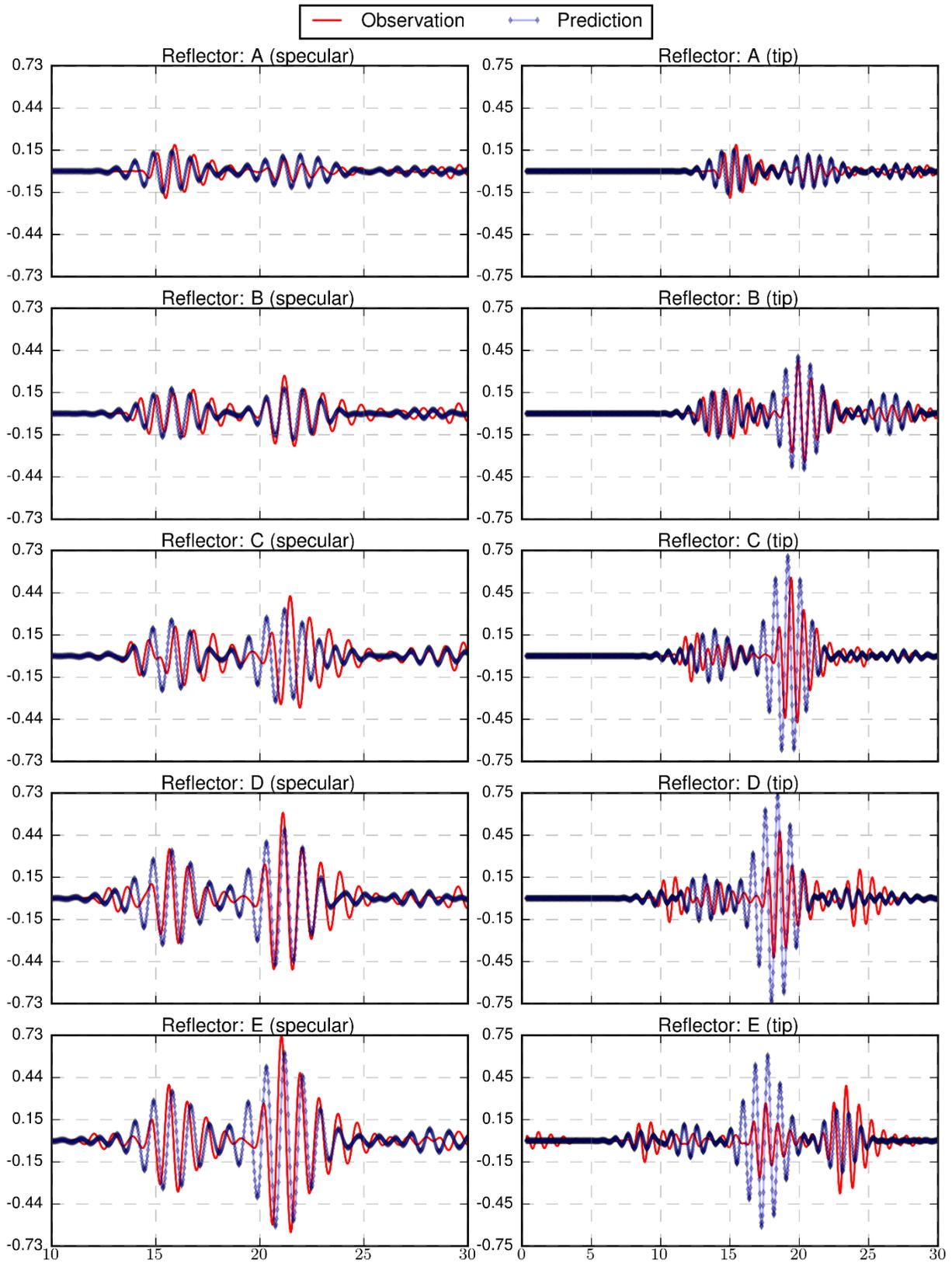


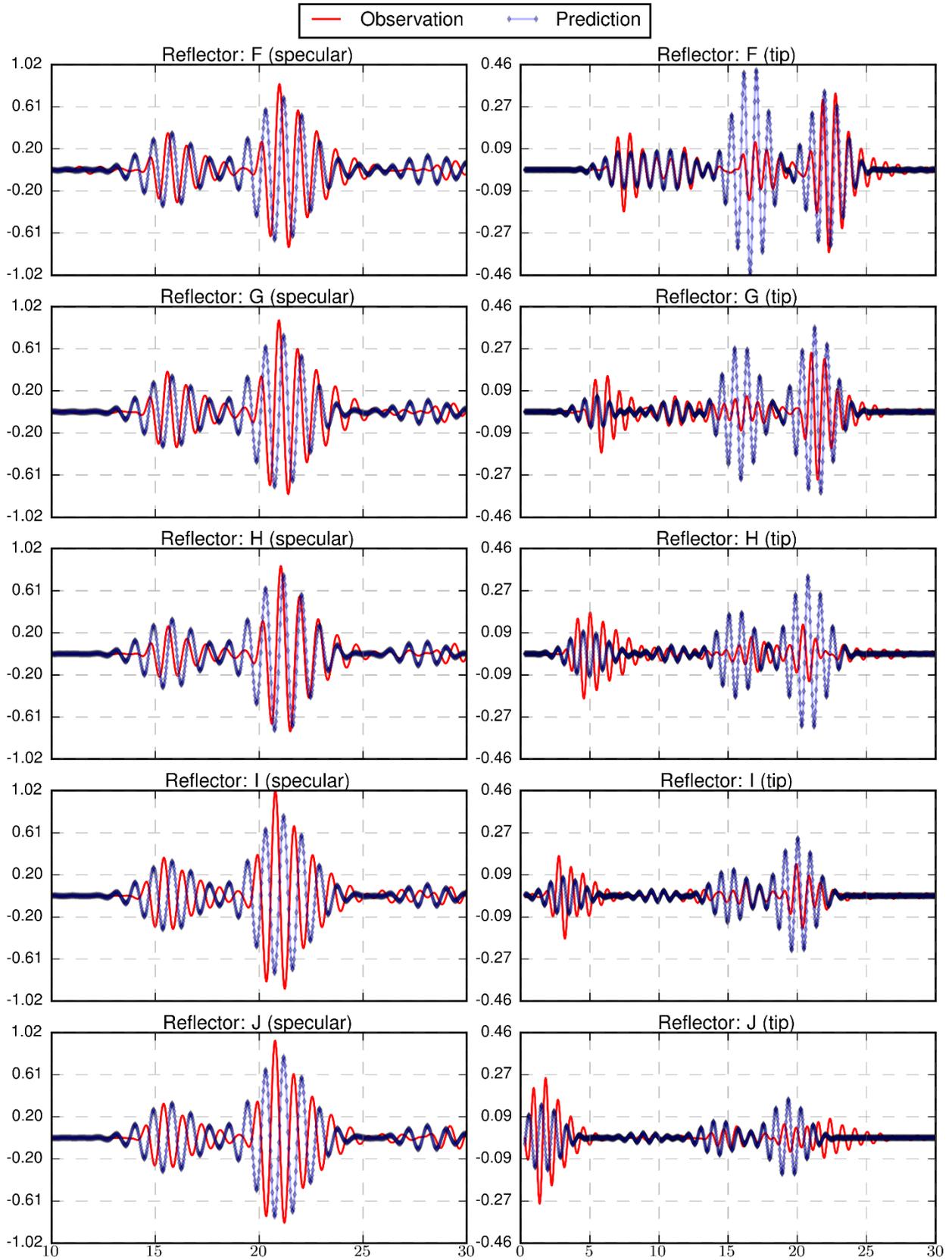
### B.3.9 Specimen 1, Probe 5, Skew 0



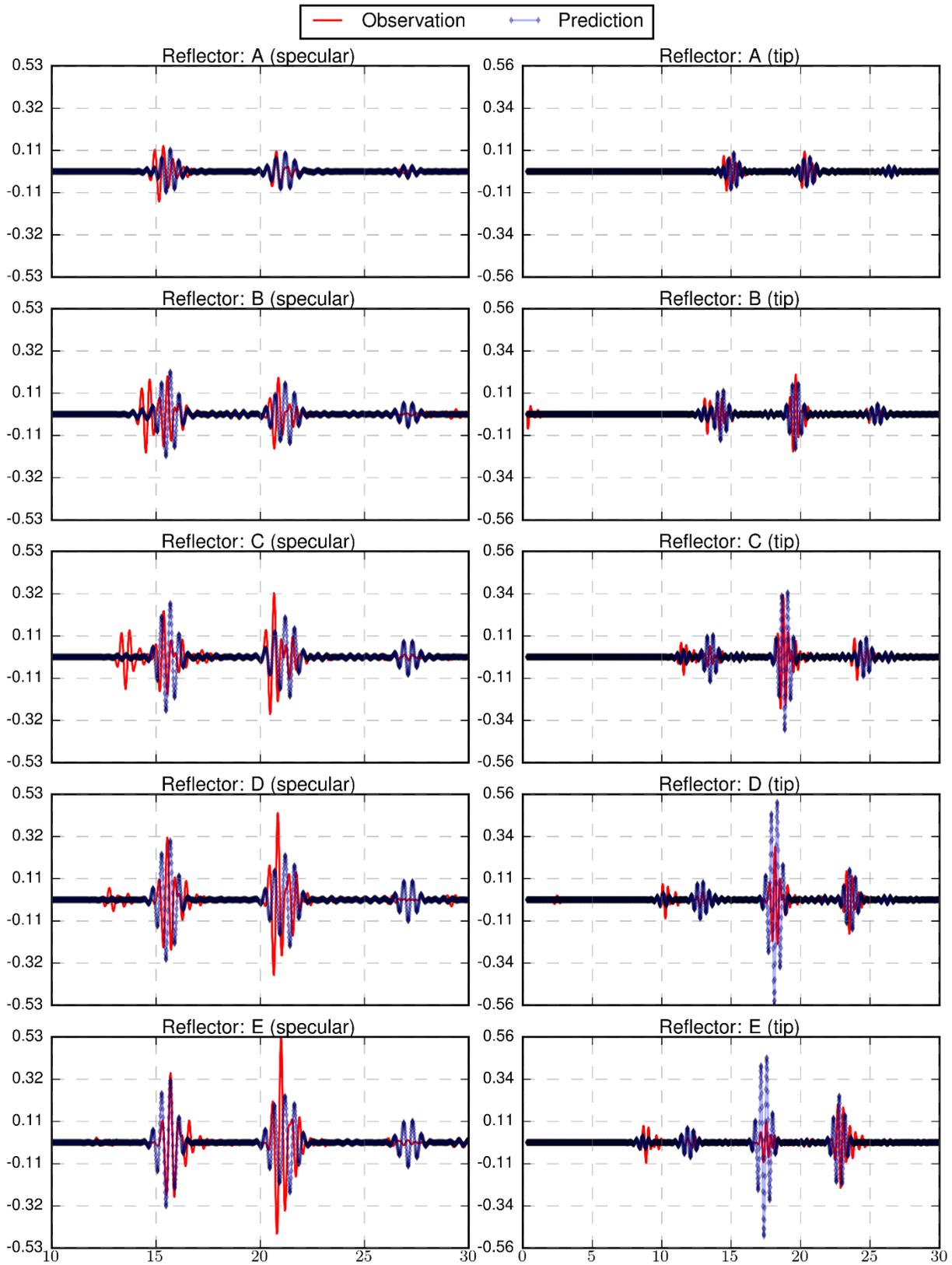


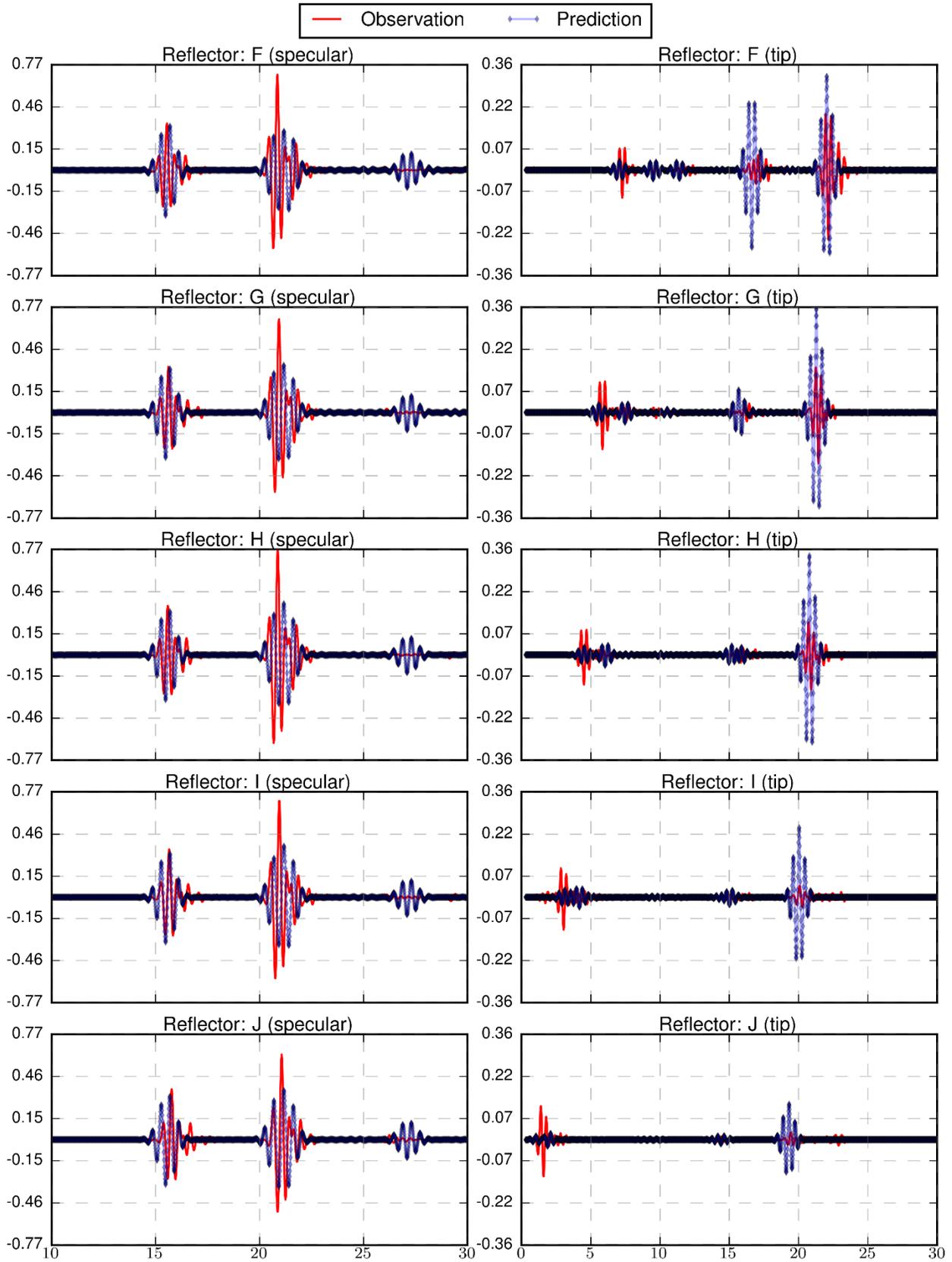
### B.3.10 Specimen 1, Probe 5, Skew 180



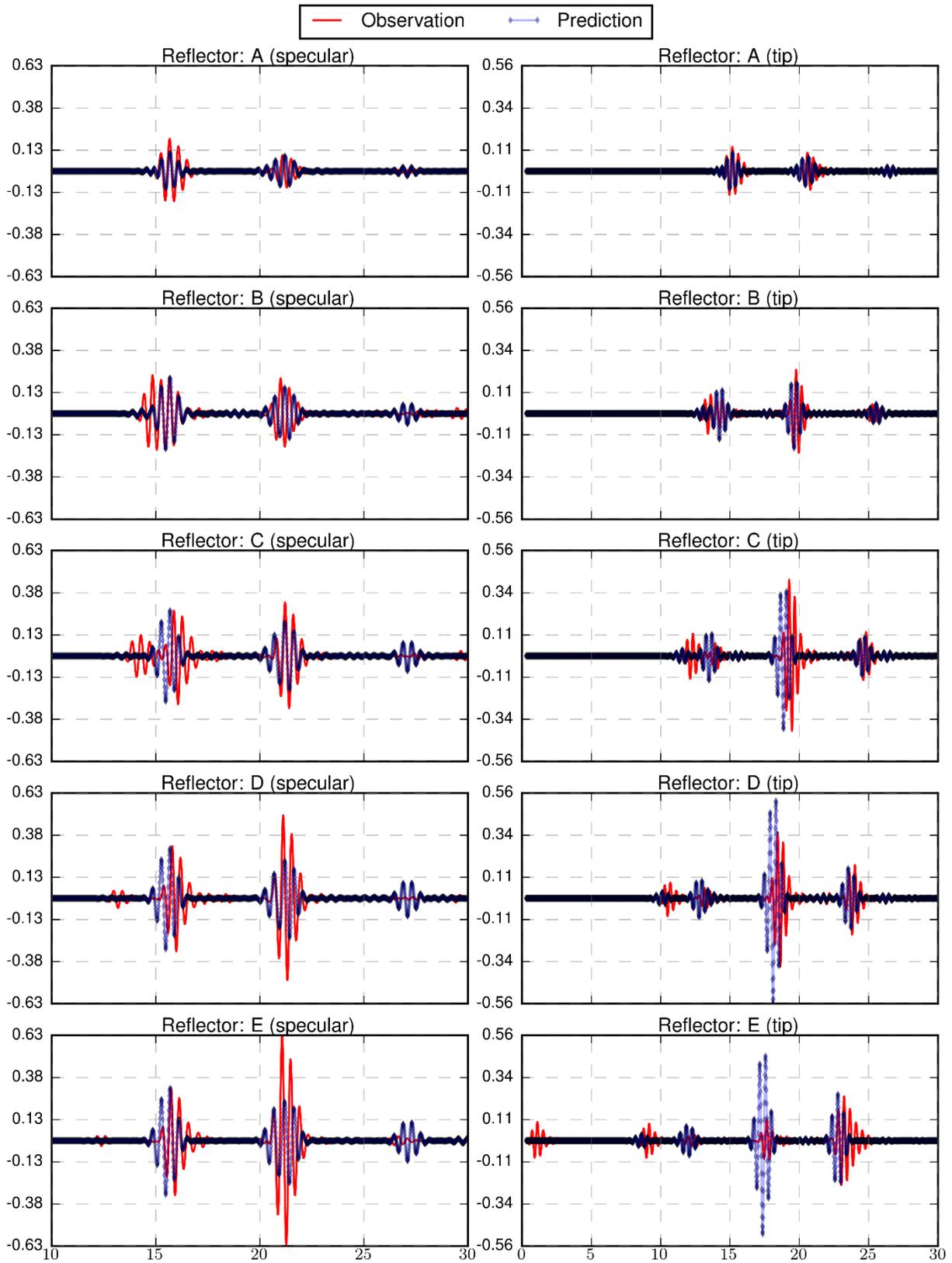


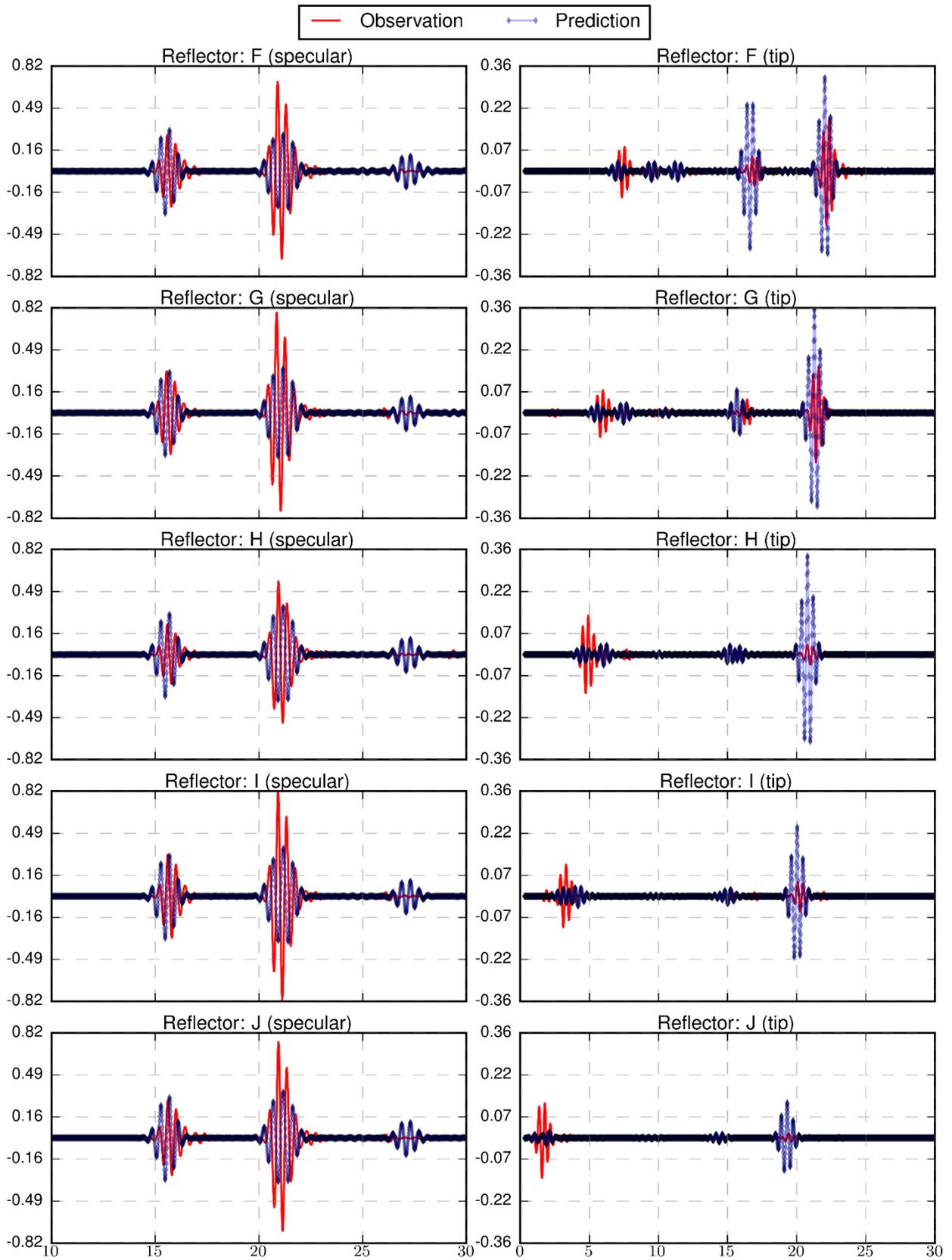
### B.3.11 Specimen 1, Probe 6, Skew 0



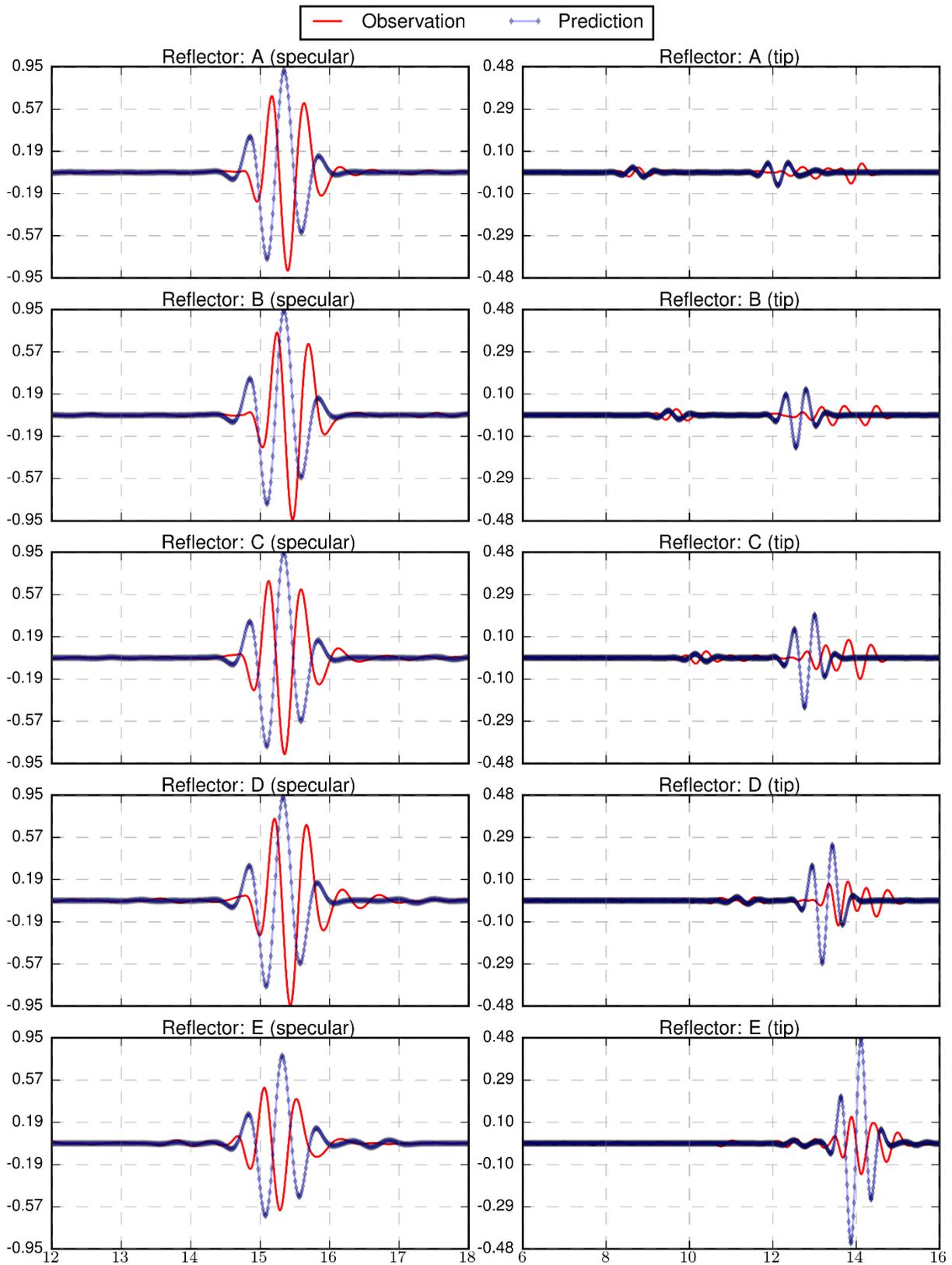


### B.3.12 Specimen 1, Probe 6, Skew 180

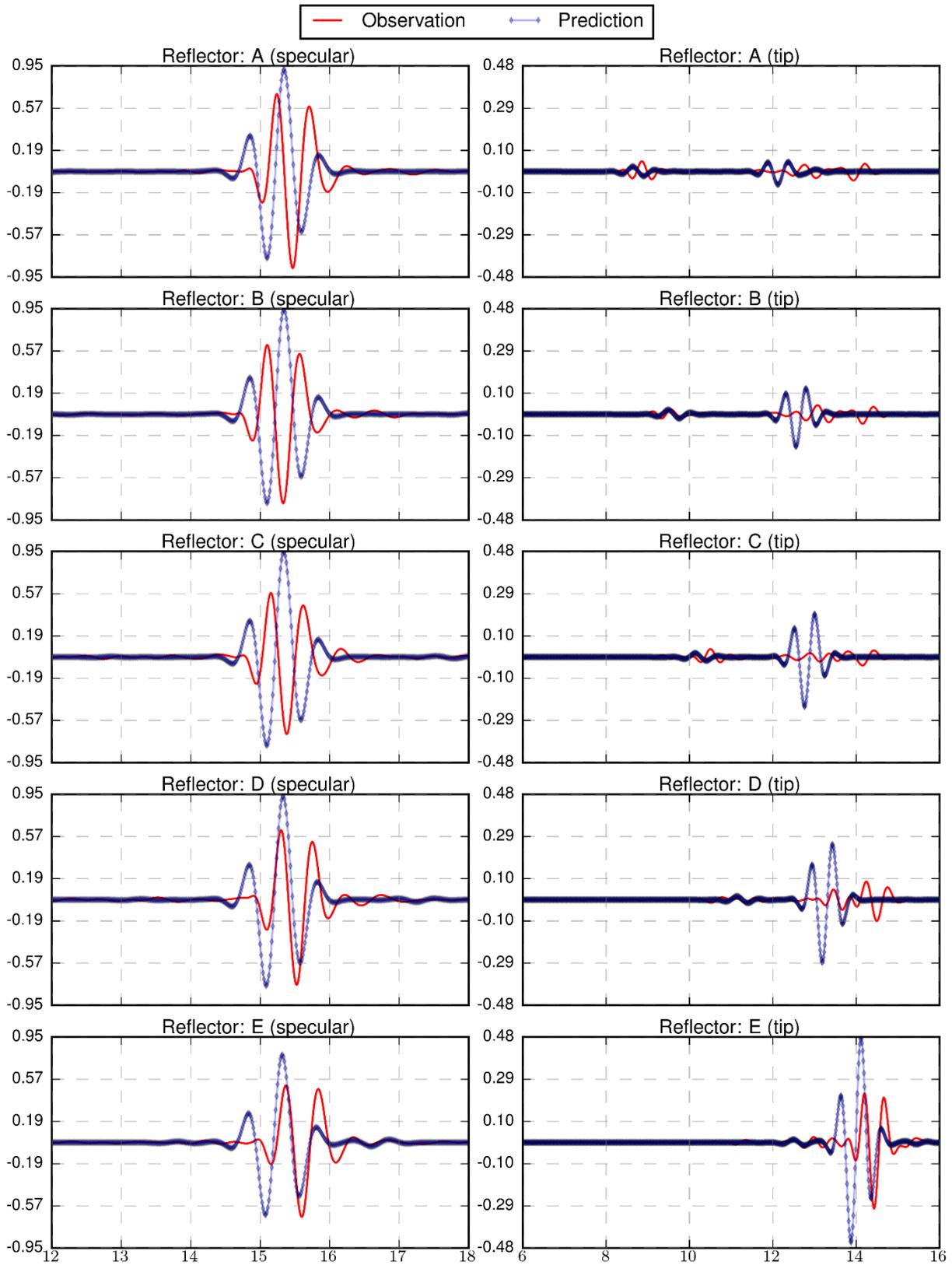




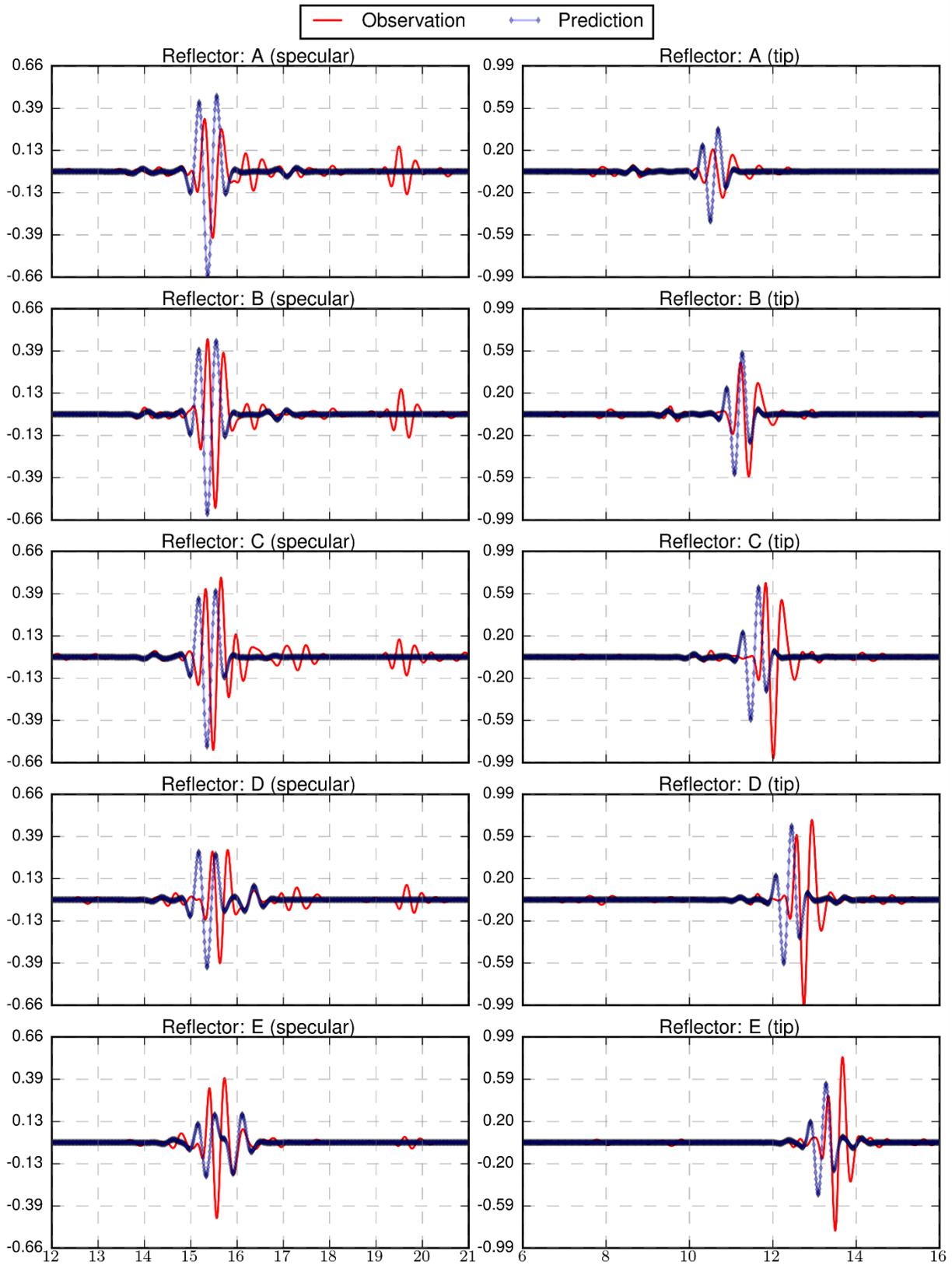
### B.3.13 Specimen 2, Probe 1, Skew 0



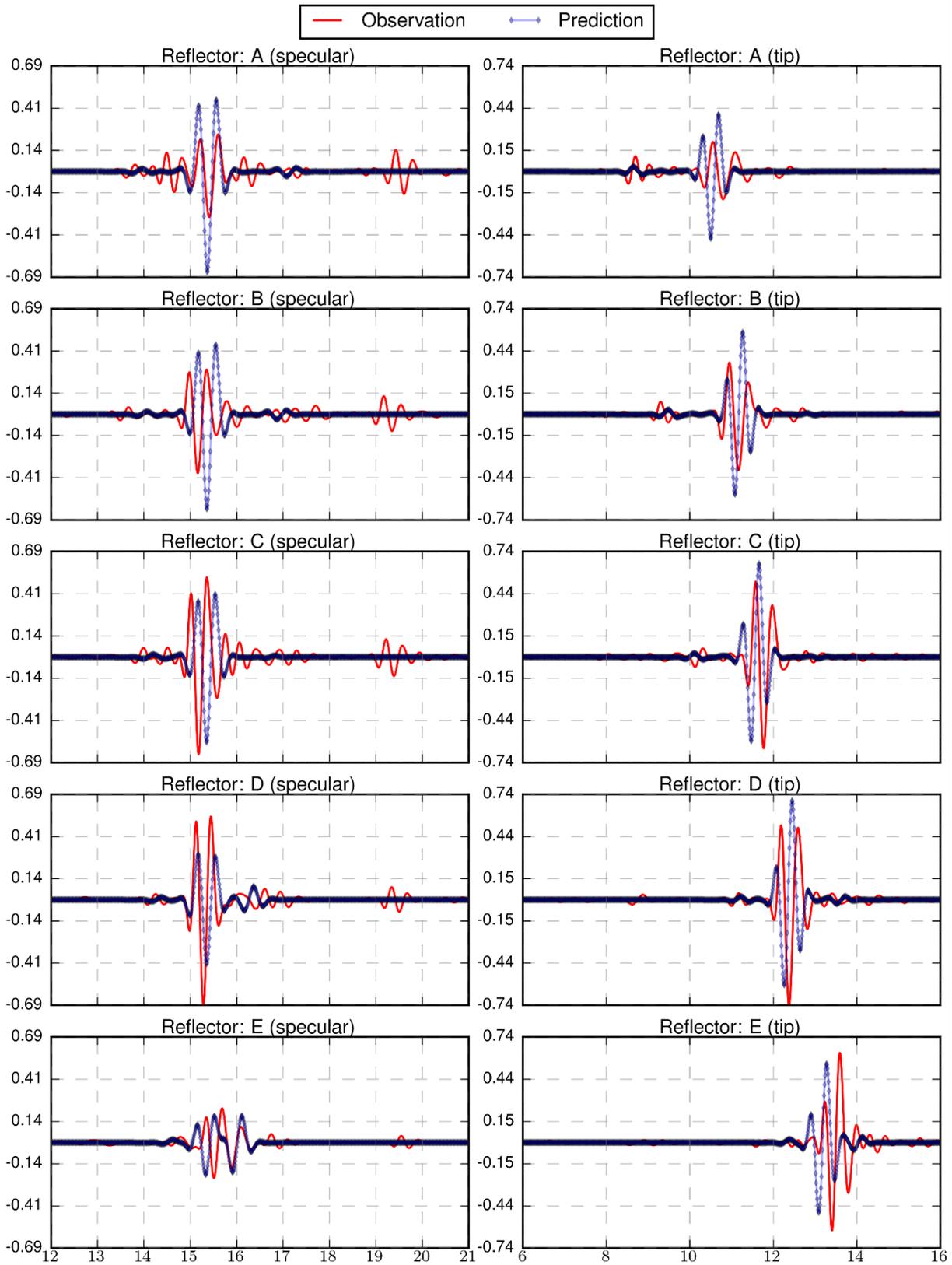
### B.3.14 Specimen 2, Probe 1, Skew 180



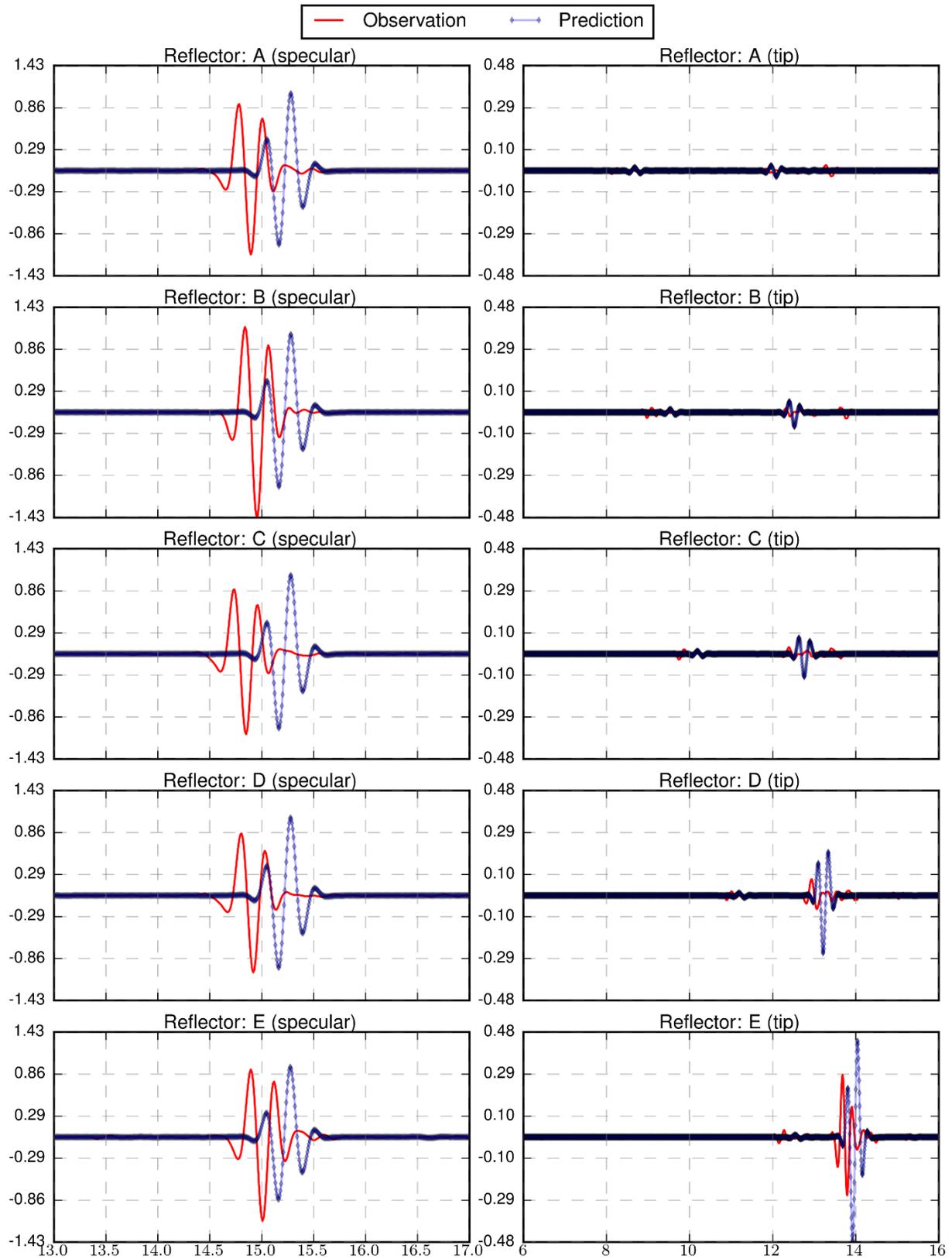
### B.3.15 Specimen 2, Probe 2, Skew 0



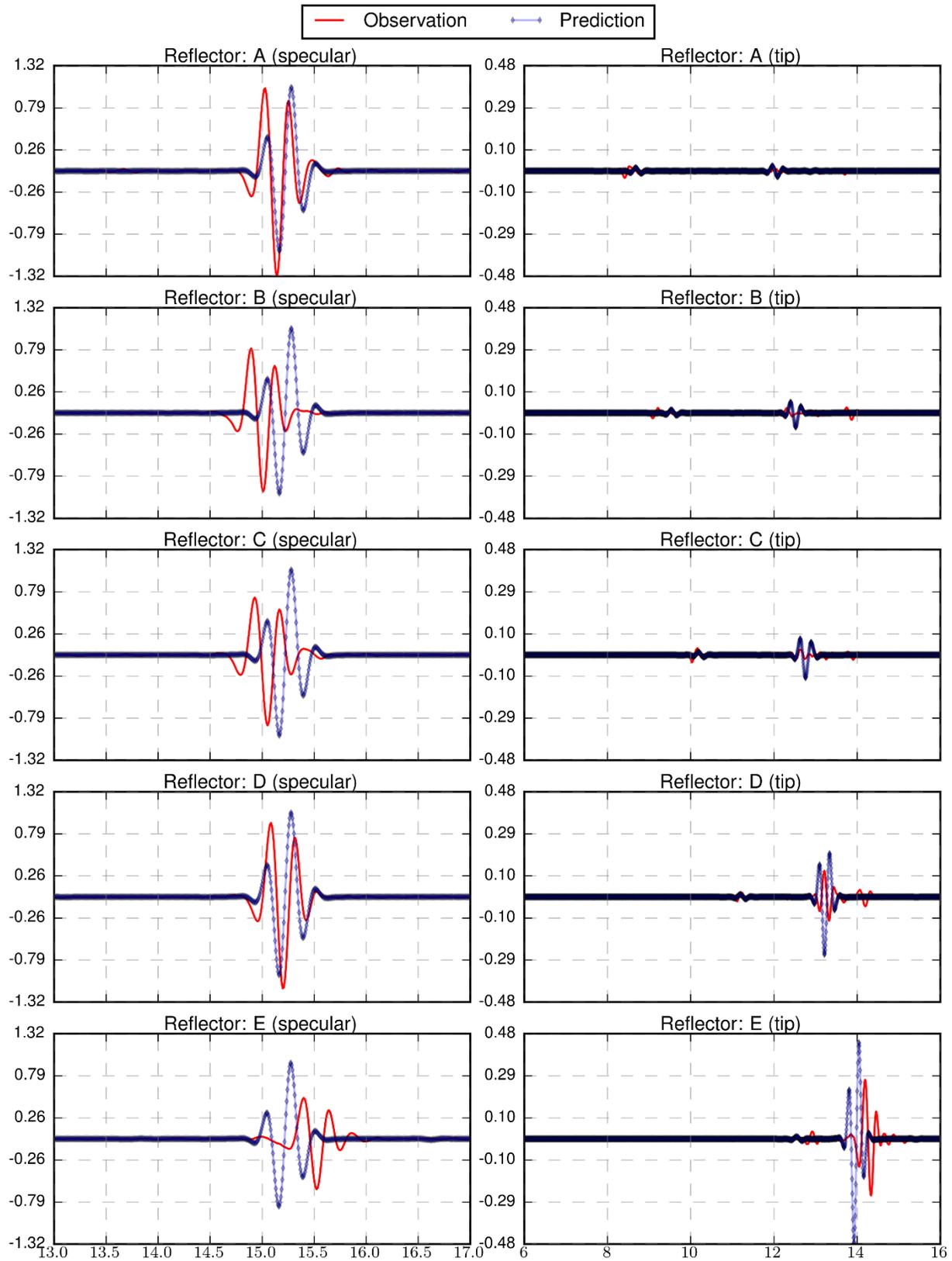
### B.3.16 Specimen 2, Probe 2, Skew 180



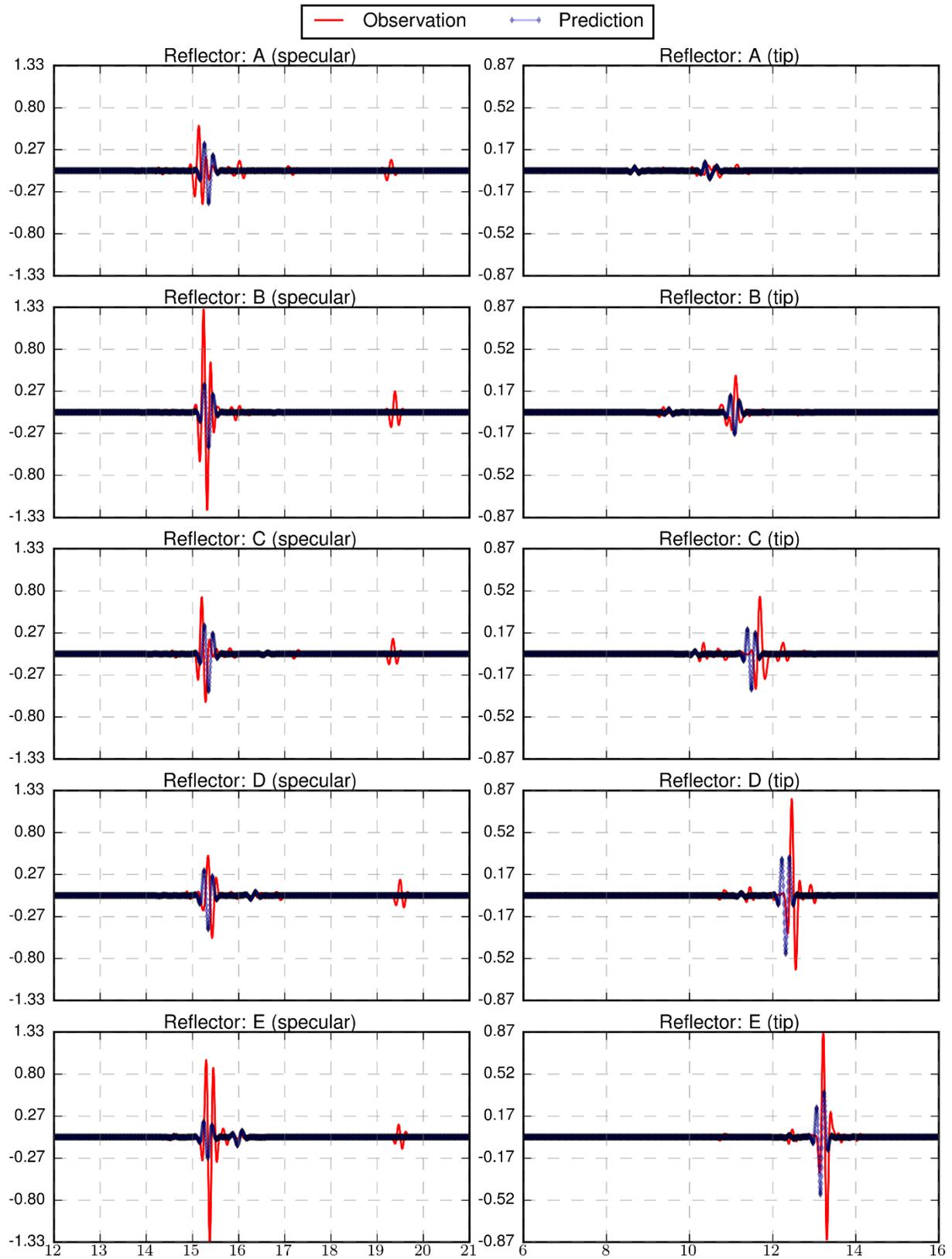
### B.3.17 Specimen 2, Probe 3, Skew 0



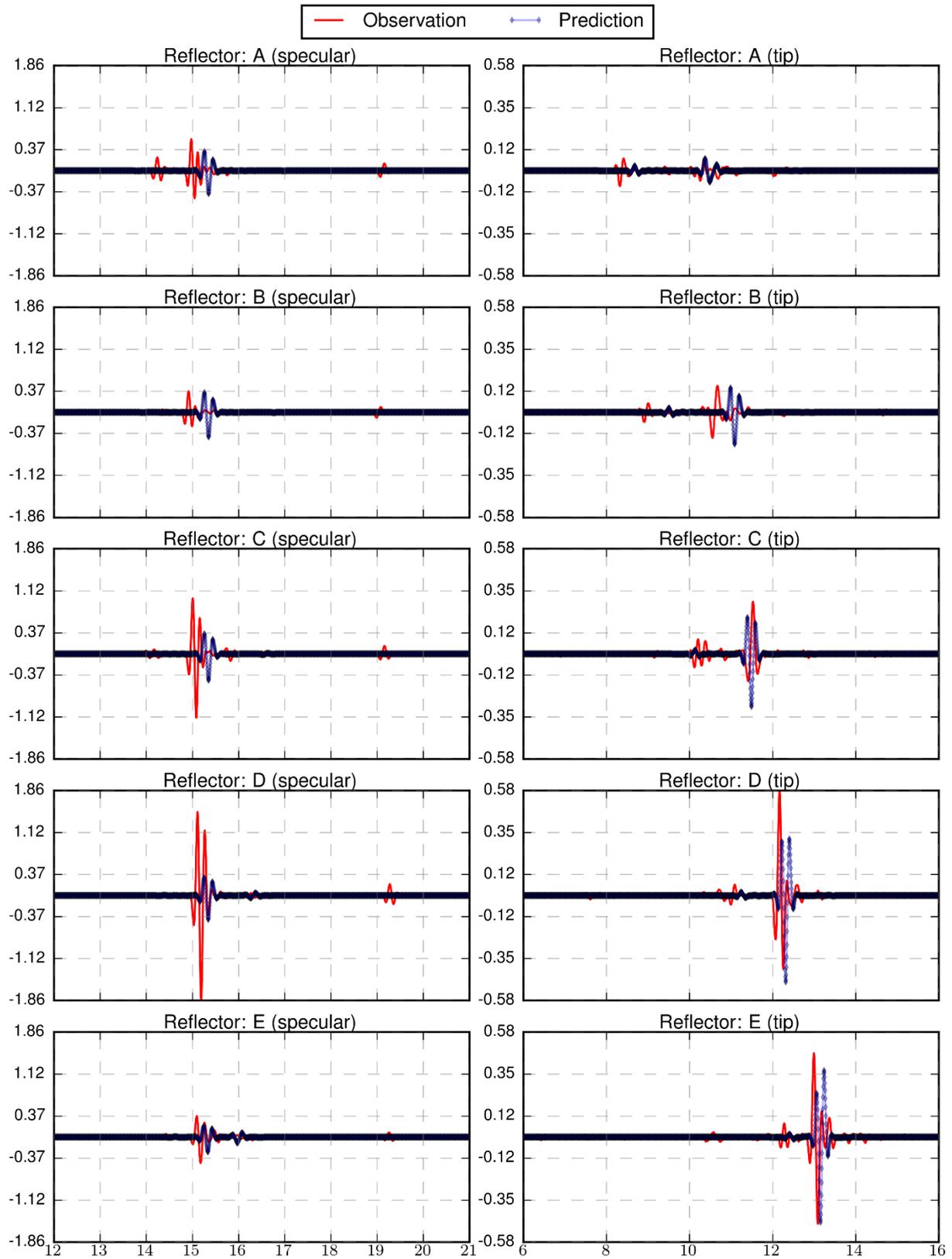
### B.3.18 Specimen 2, Probe 3, Skew 180



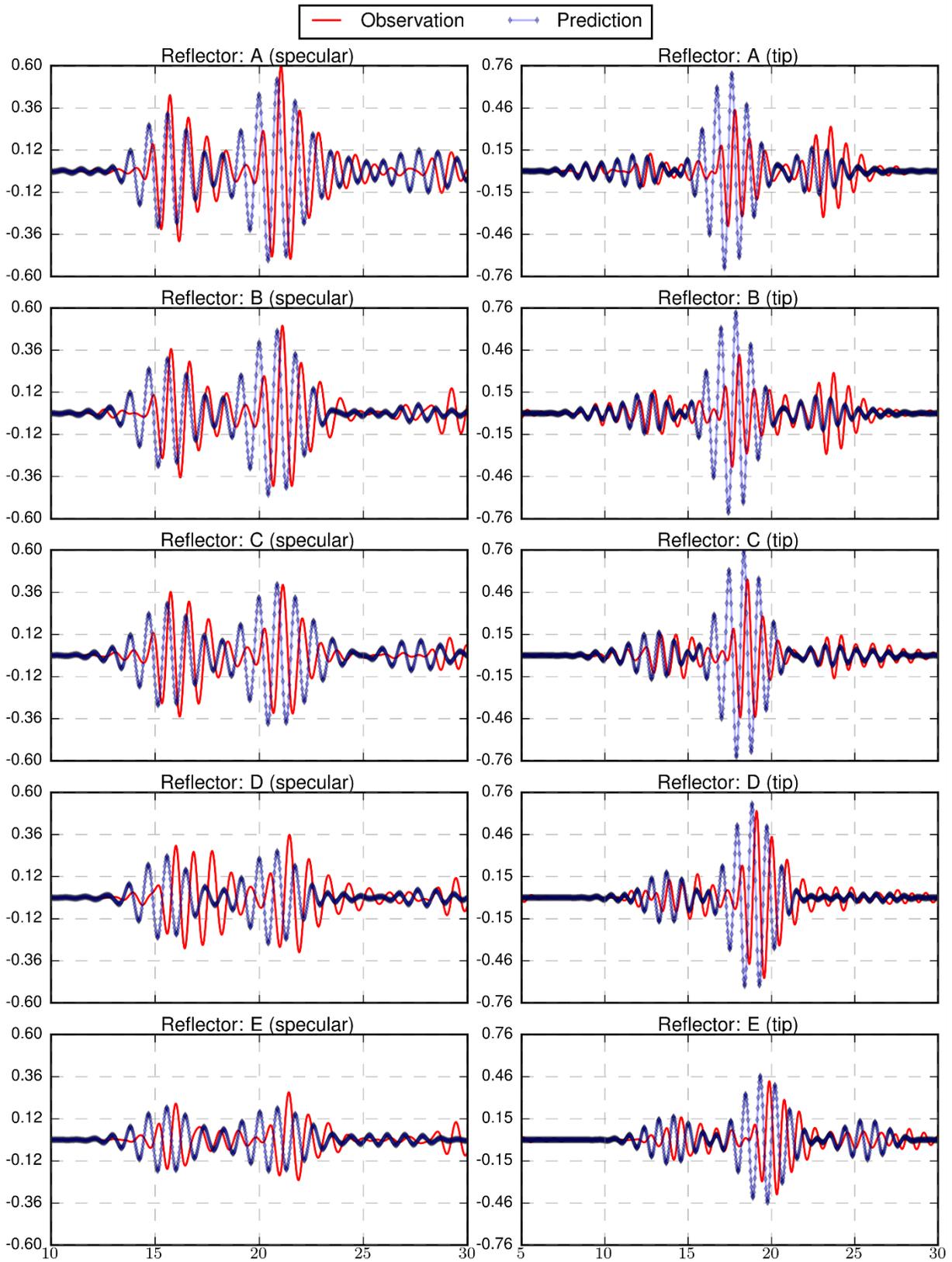
### B.3.19 Specimen 2, Probe 4, Skew 0



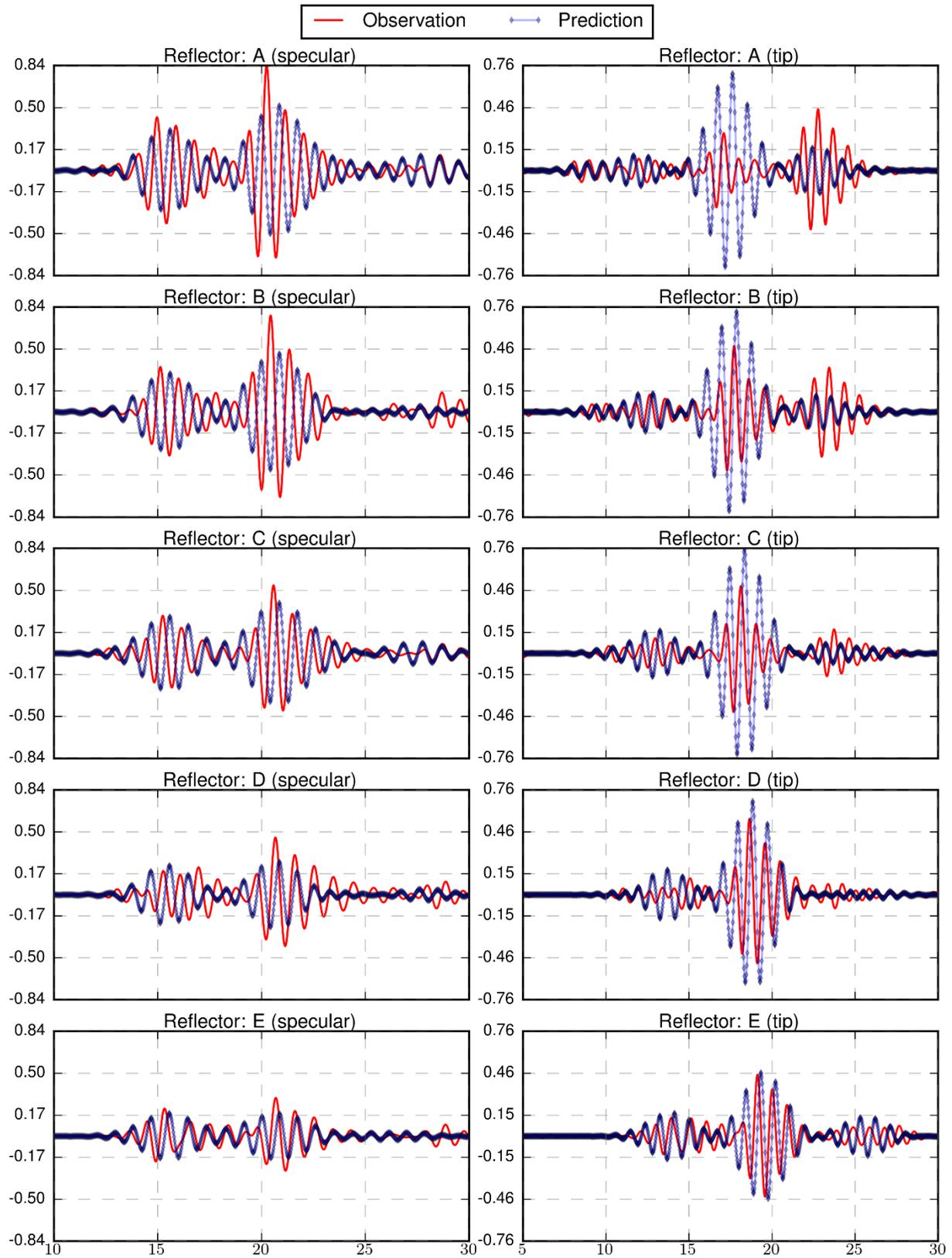
### B.3.20 Specimen 2, Probe 4, Skew 180



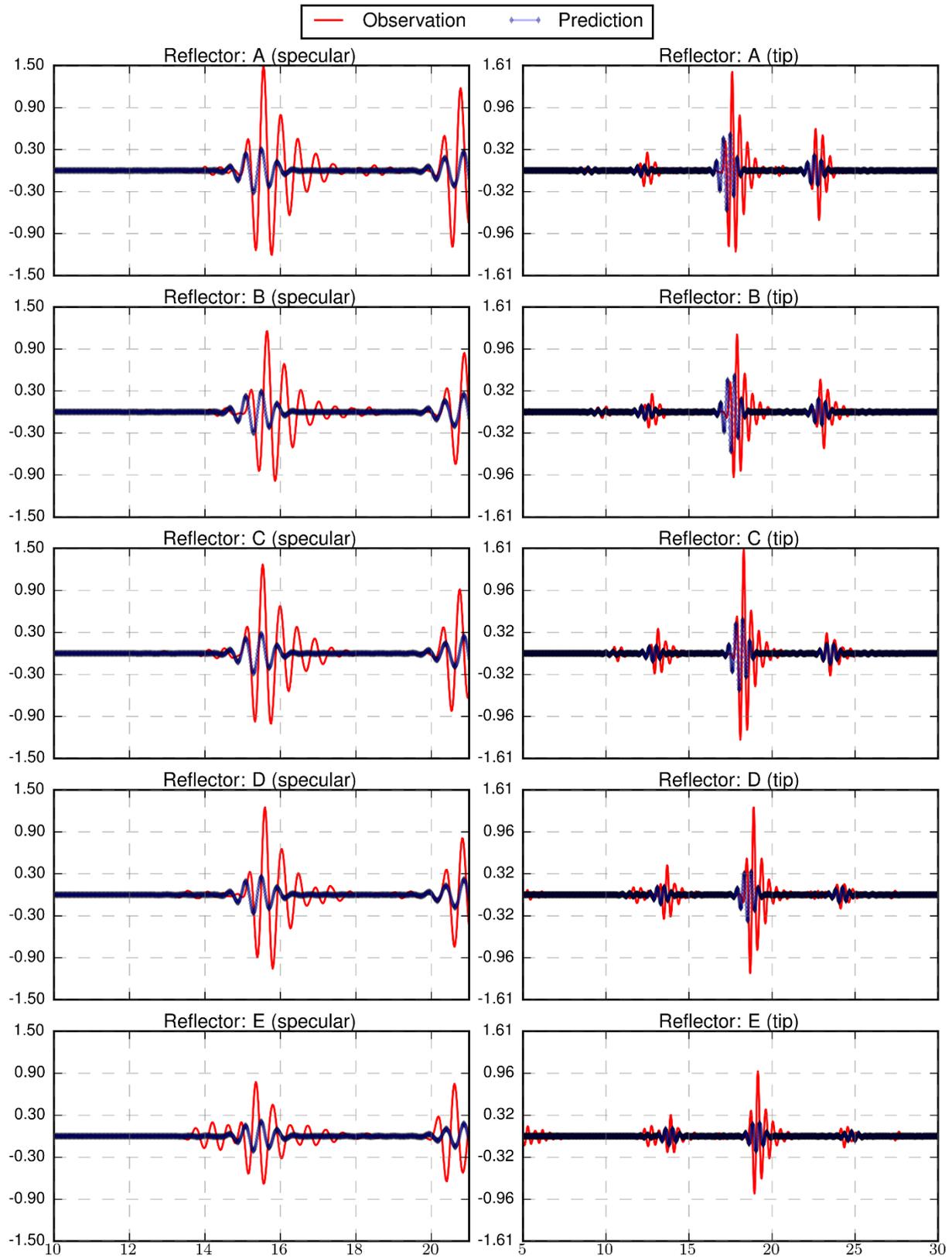
### B.3.21 Specimen 2, Probe 5, Skew 0



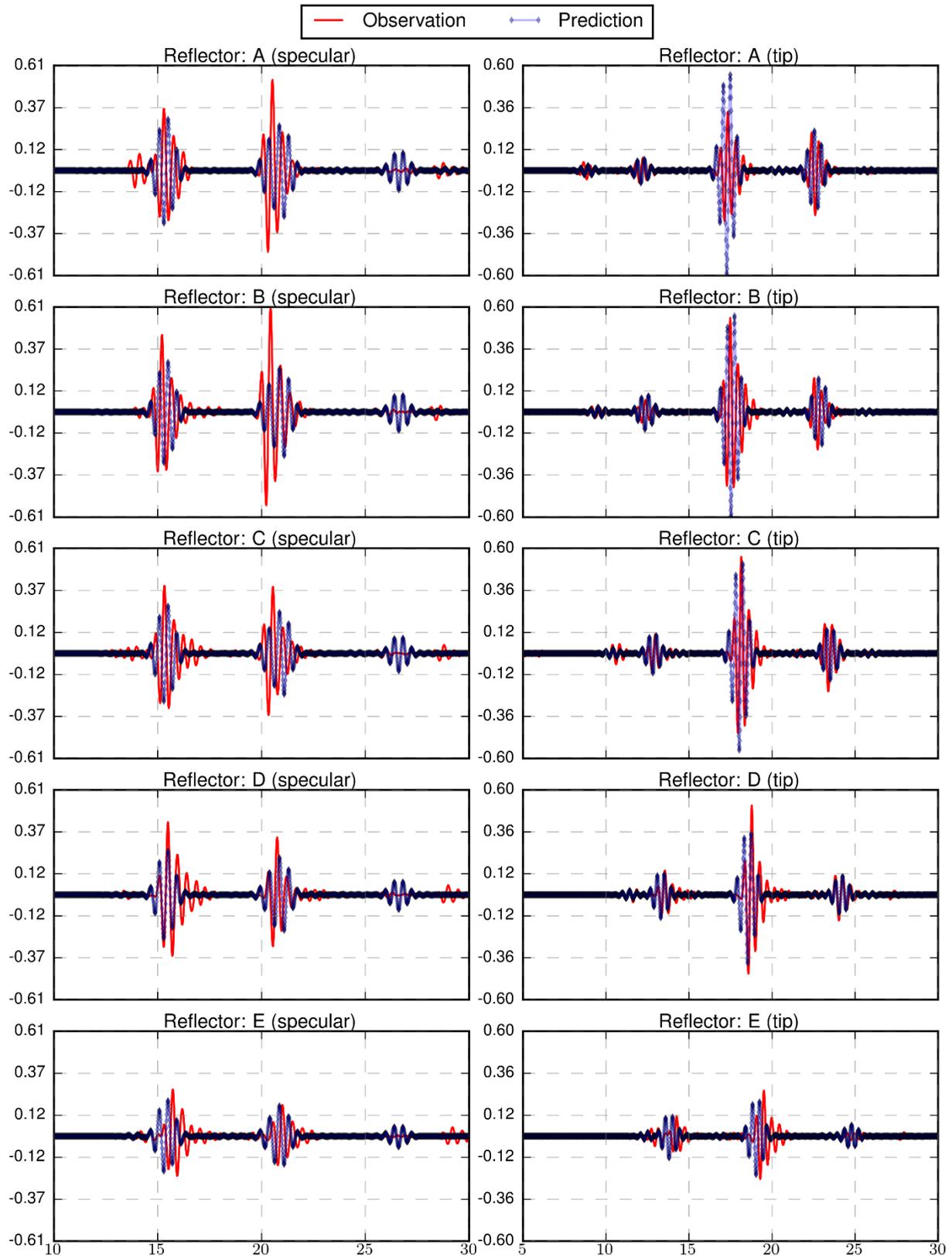
### B.3.22 Specimen 2, Probe 5, Skew 180



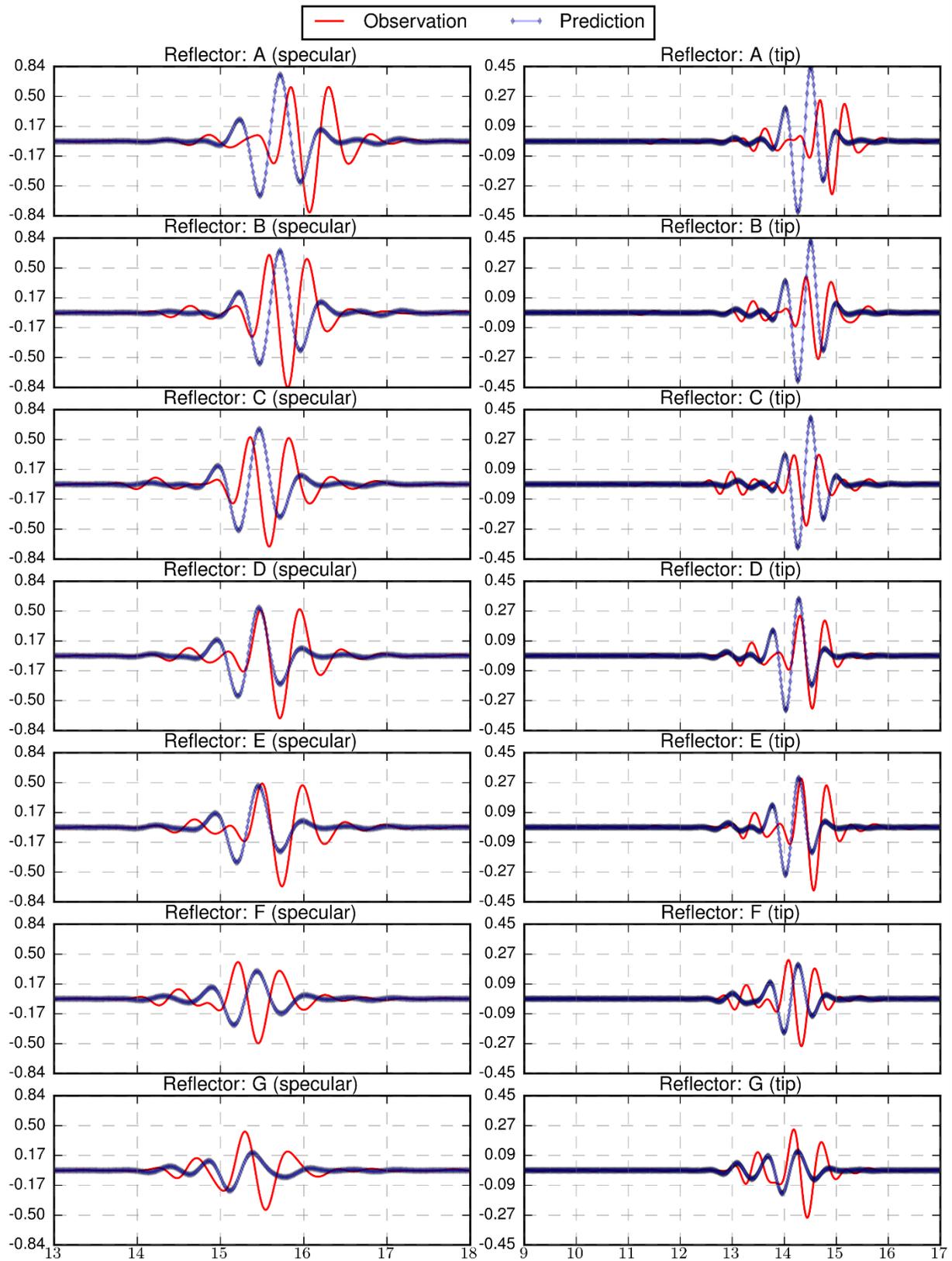
### B.3.23 Specimen 2, Probe 6, Skew 0

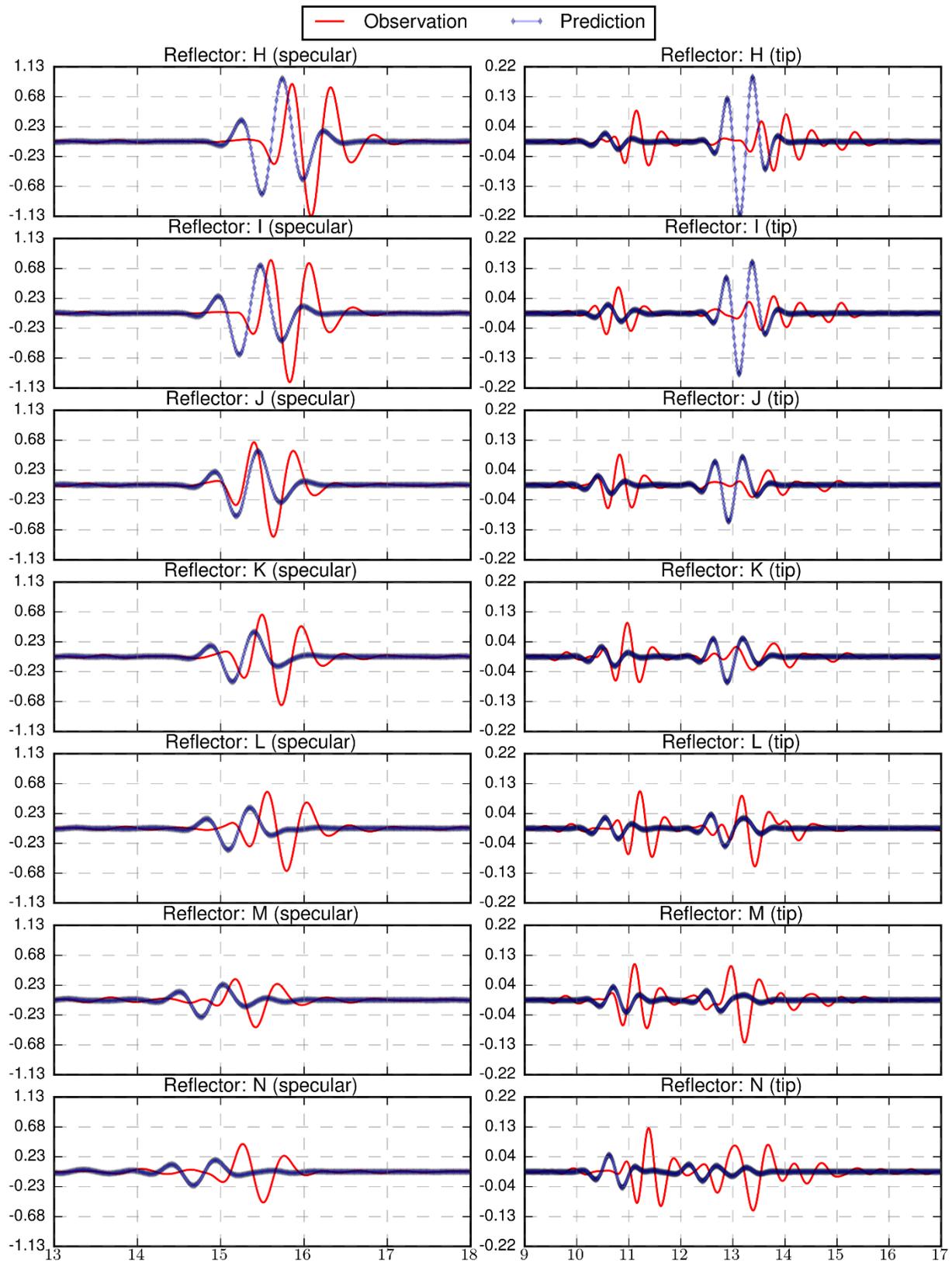


### B.3.24 Specimen 2, Probe 6, Skew 180

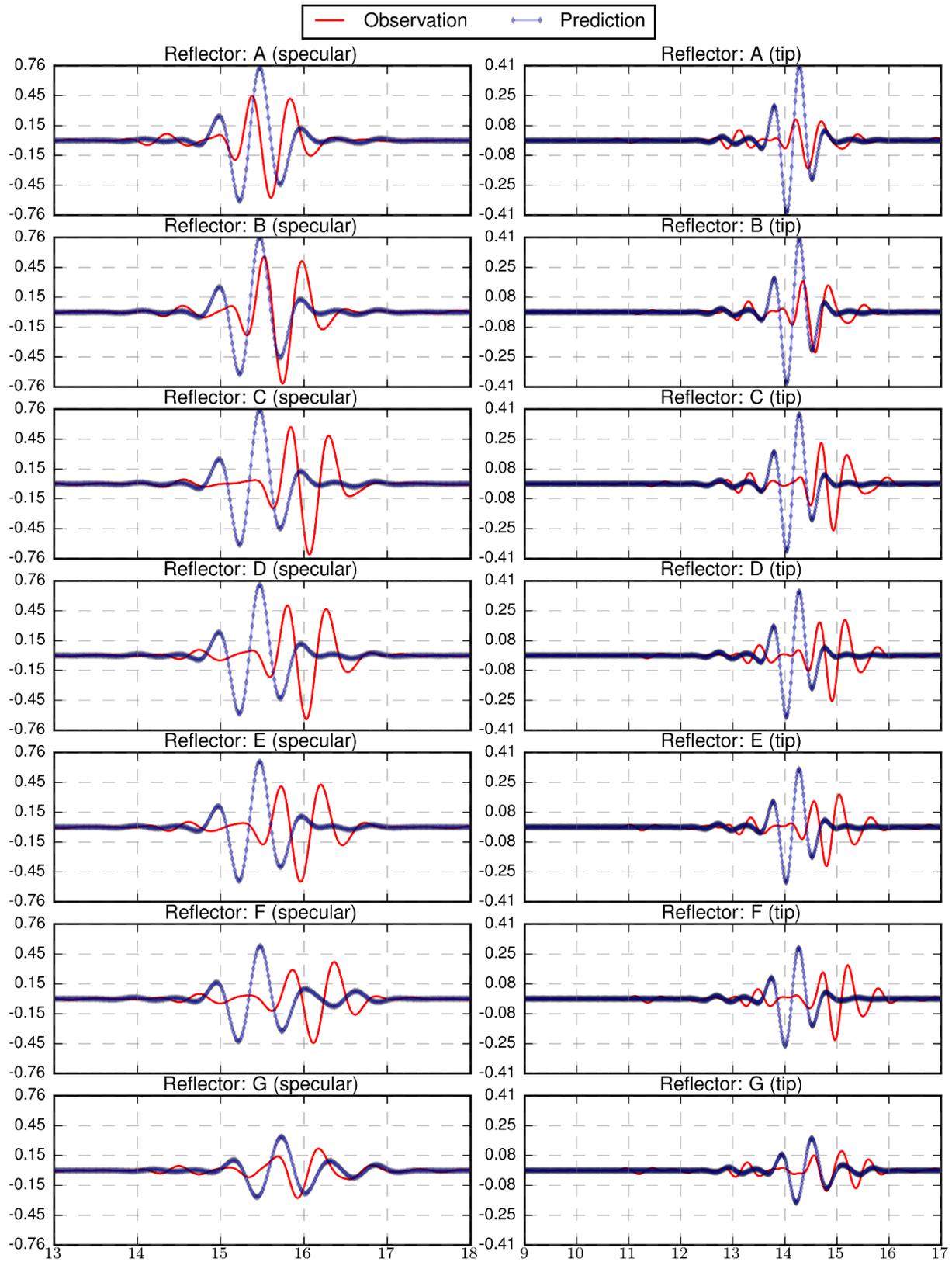


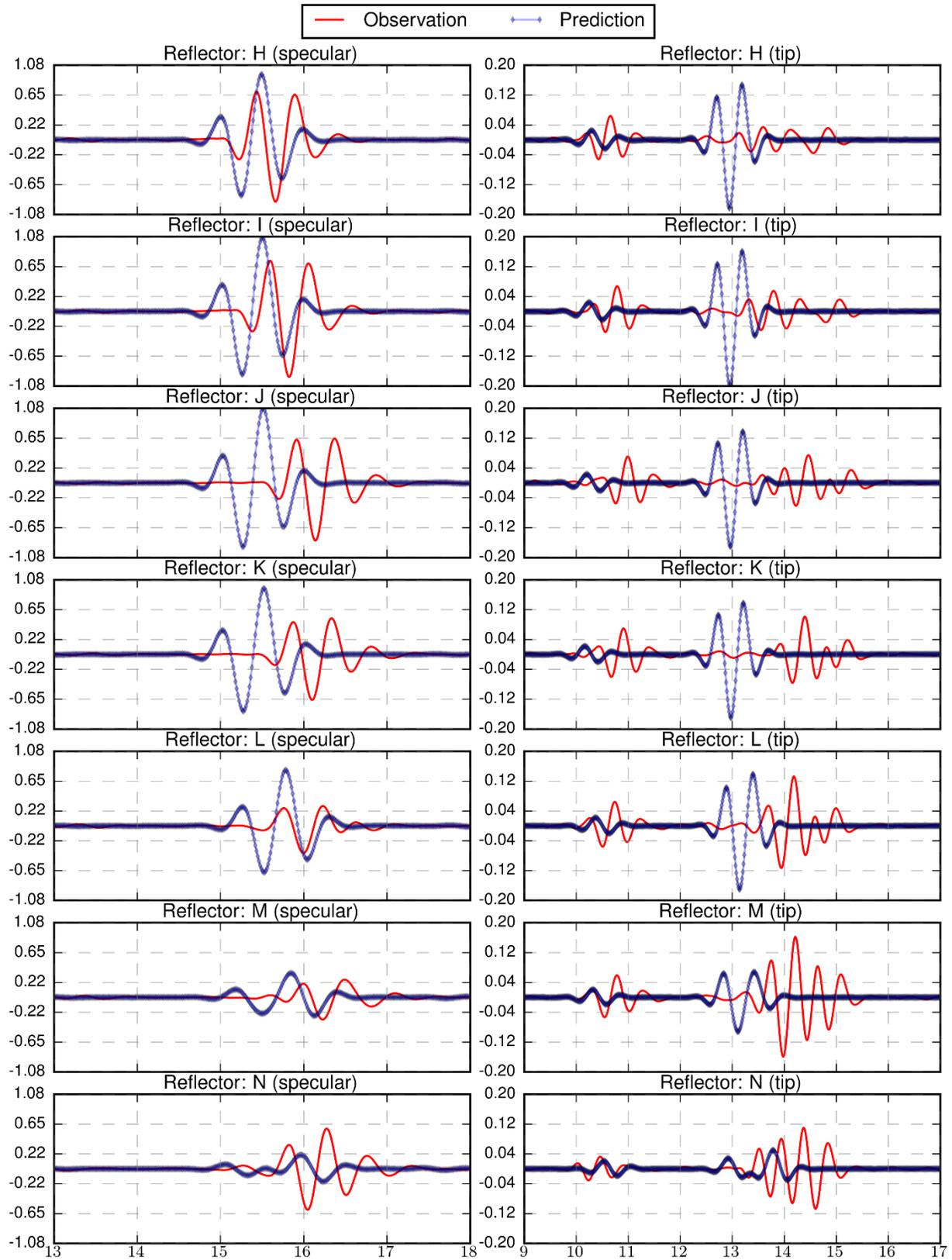
### B.3.25 Specimen 3, Probe 1, Skew 0



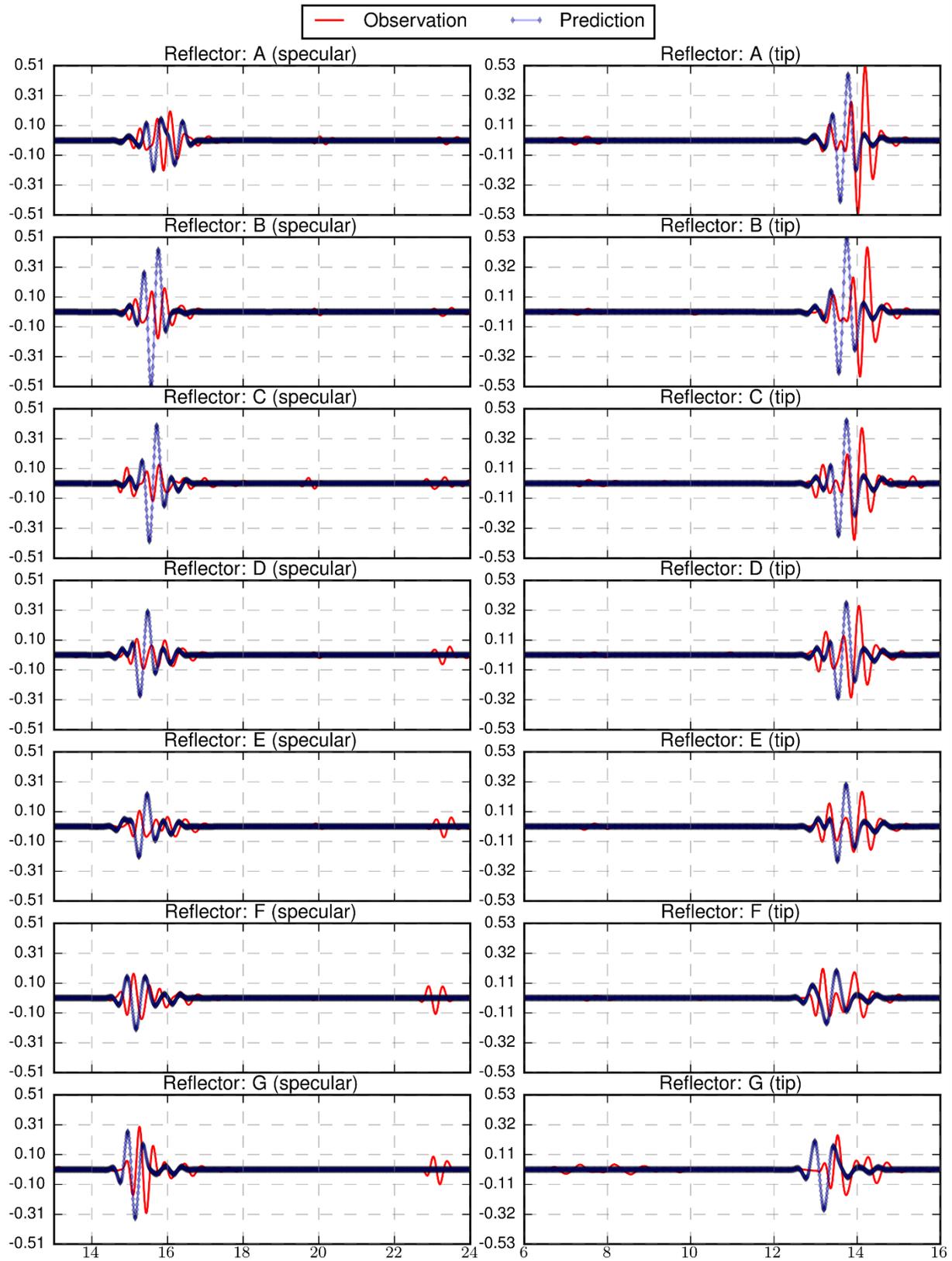


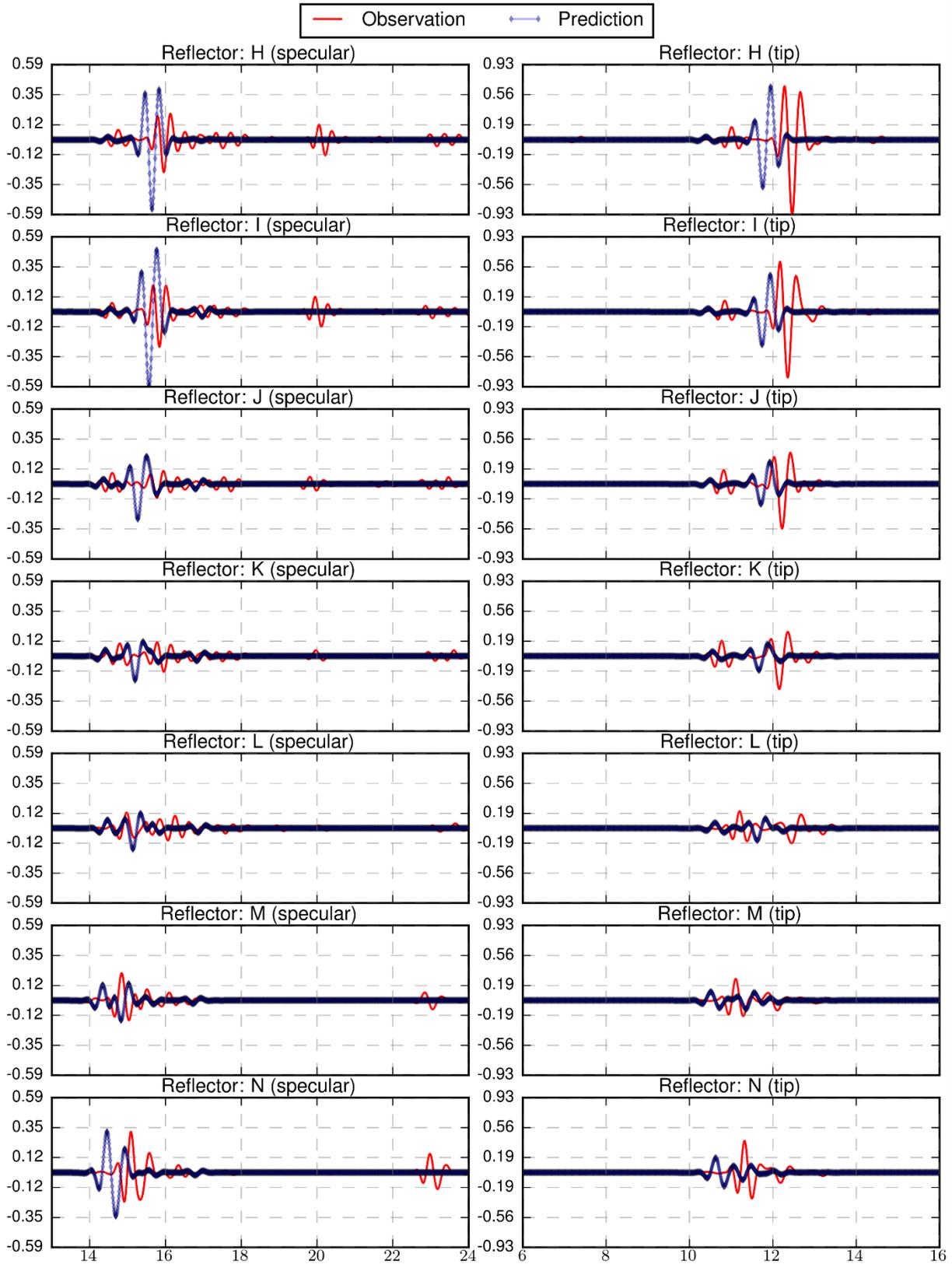
### B.3.26 Specimen 3, Probe 1, Skew 180



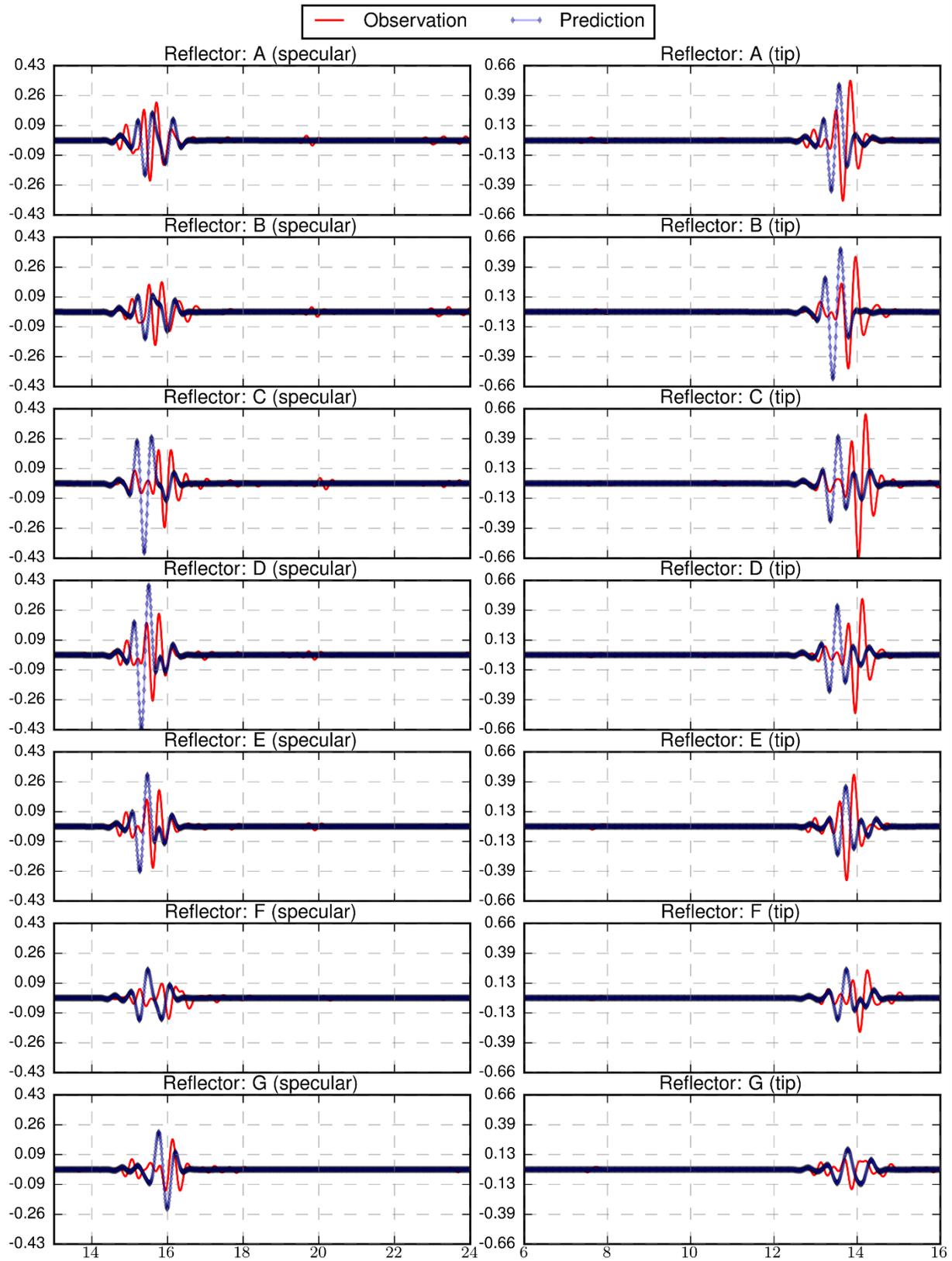


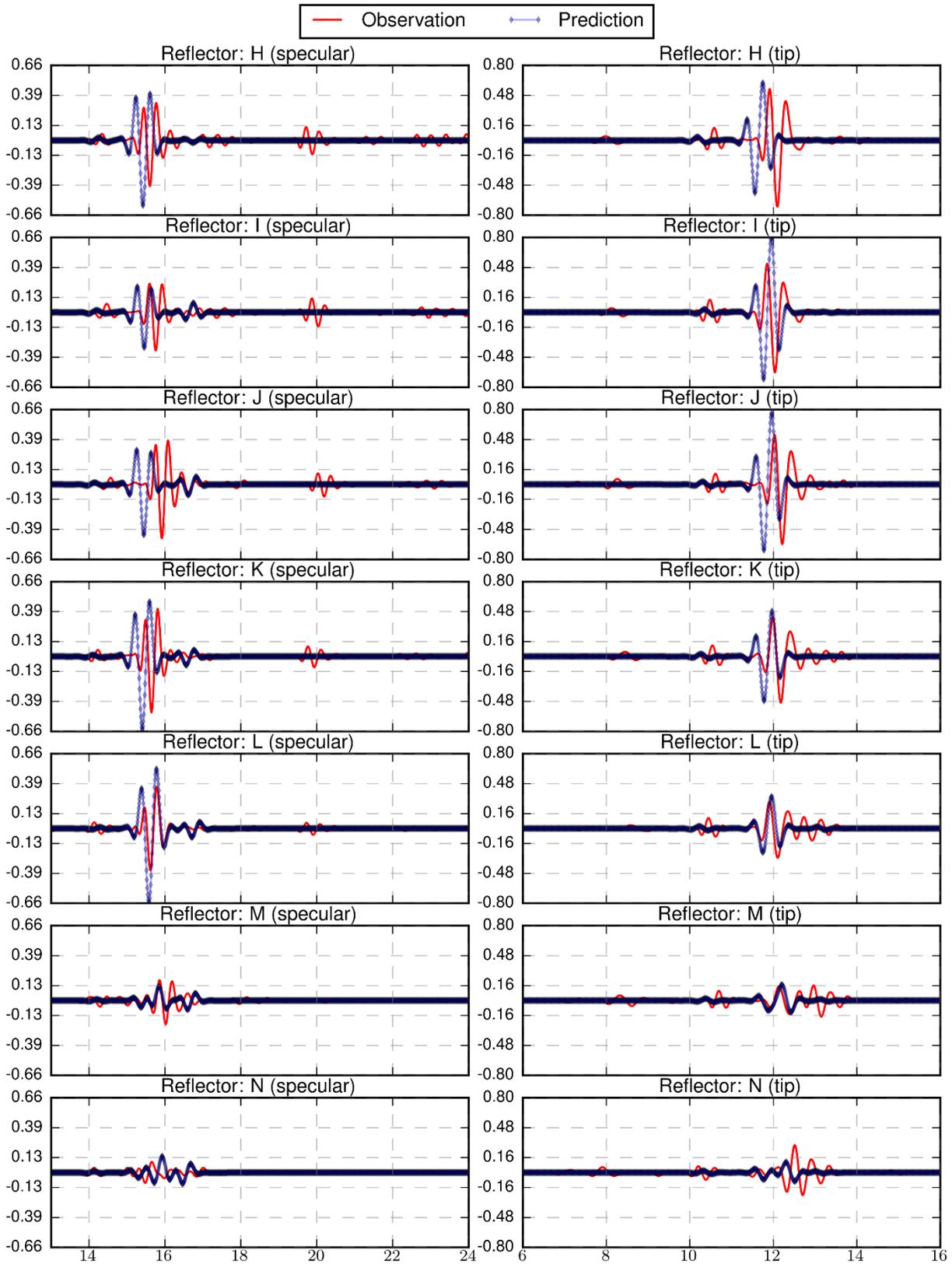
### B.3.27 Specimen 3, Probe 2, Skew 0



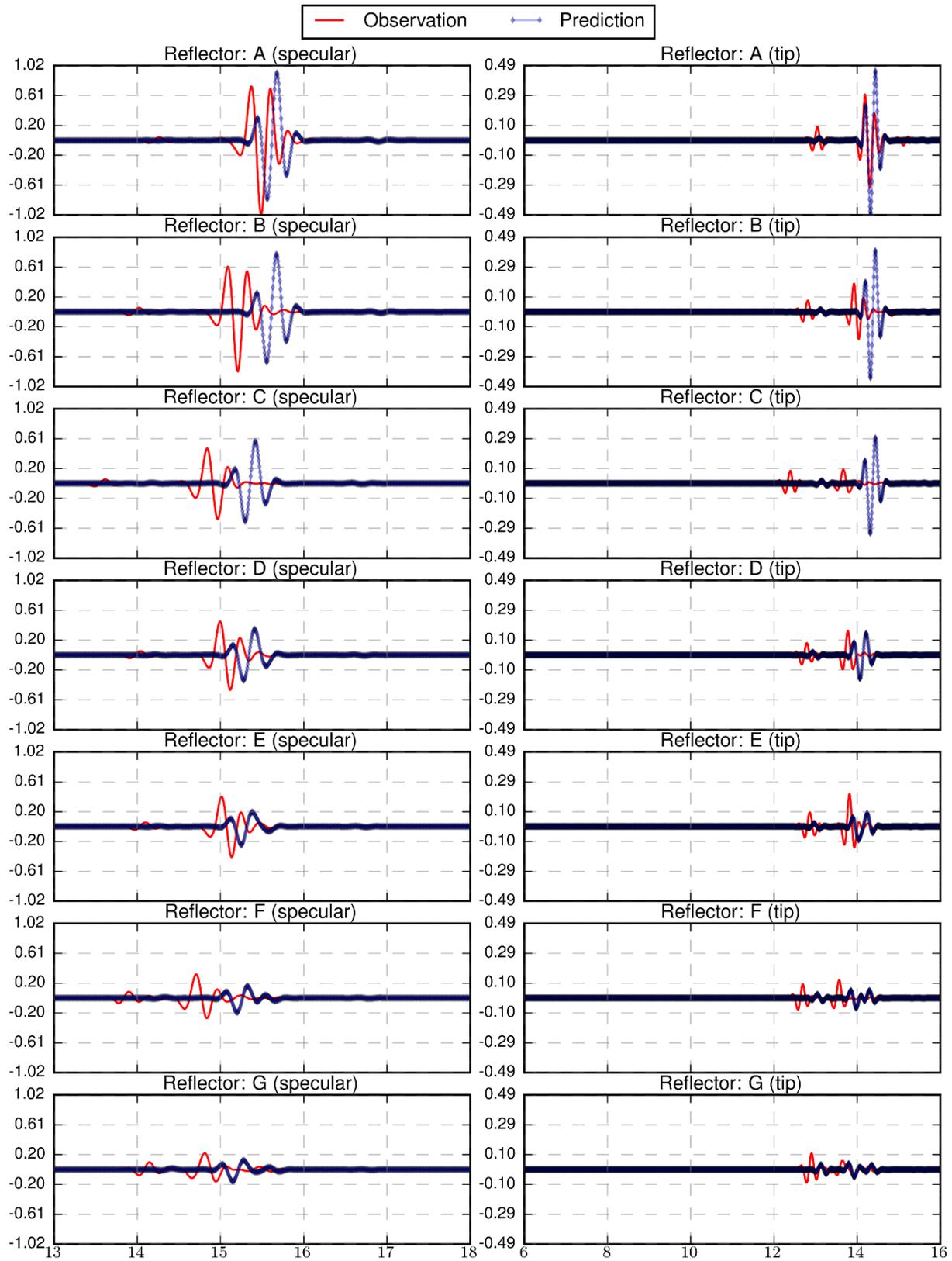


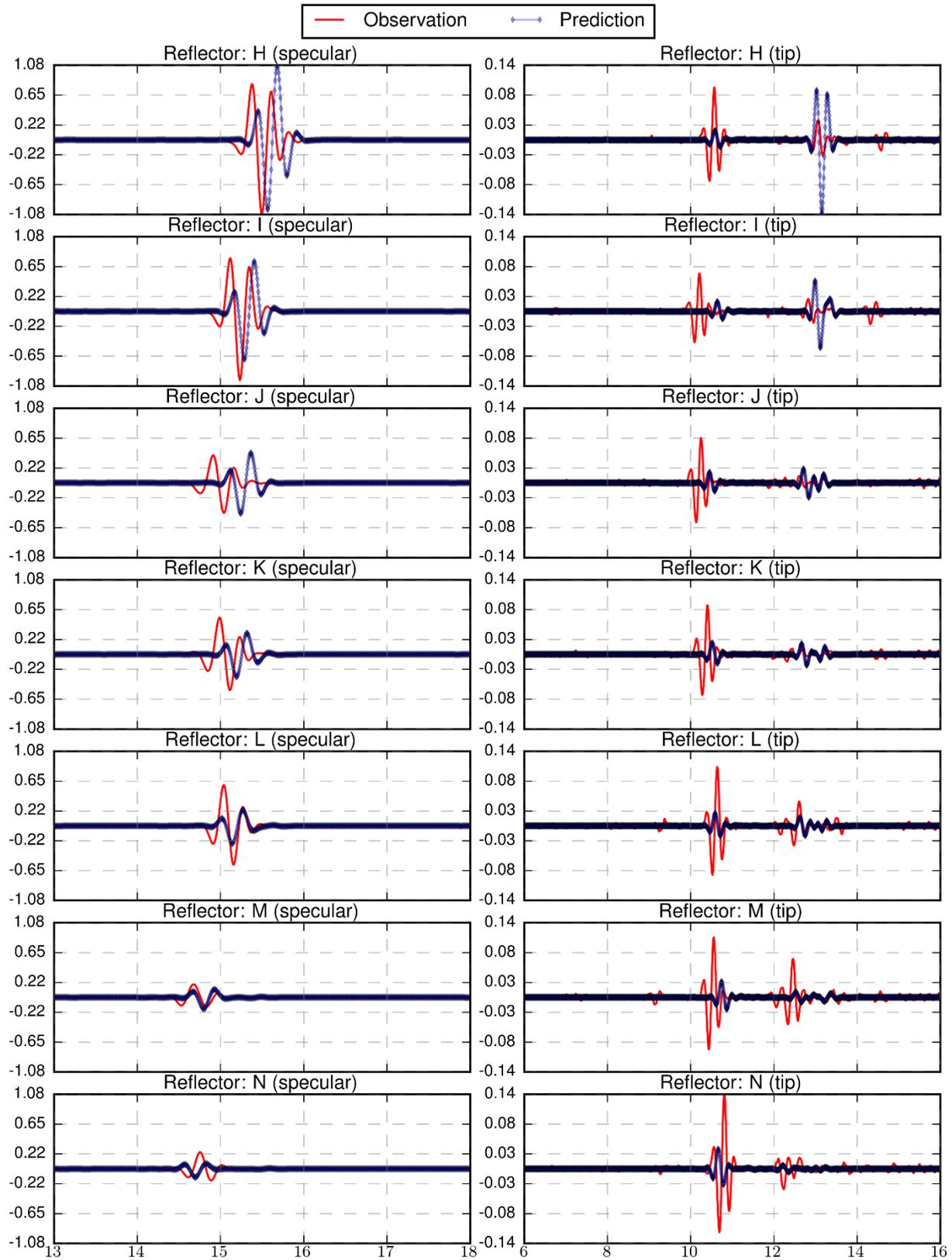
### B.3.28 Specimen 3, Probe 2, Skew 180



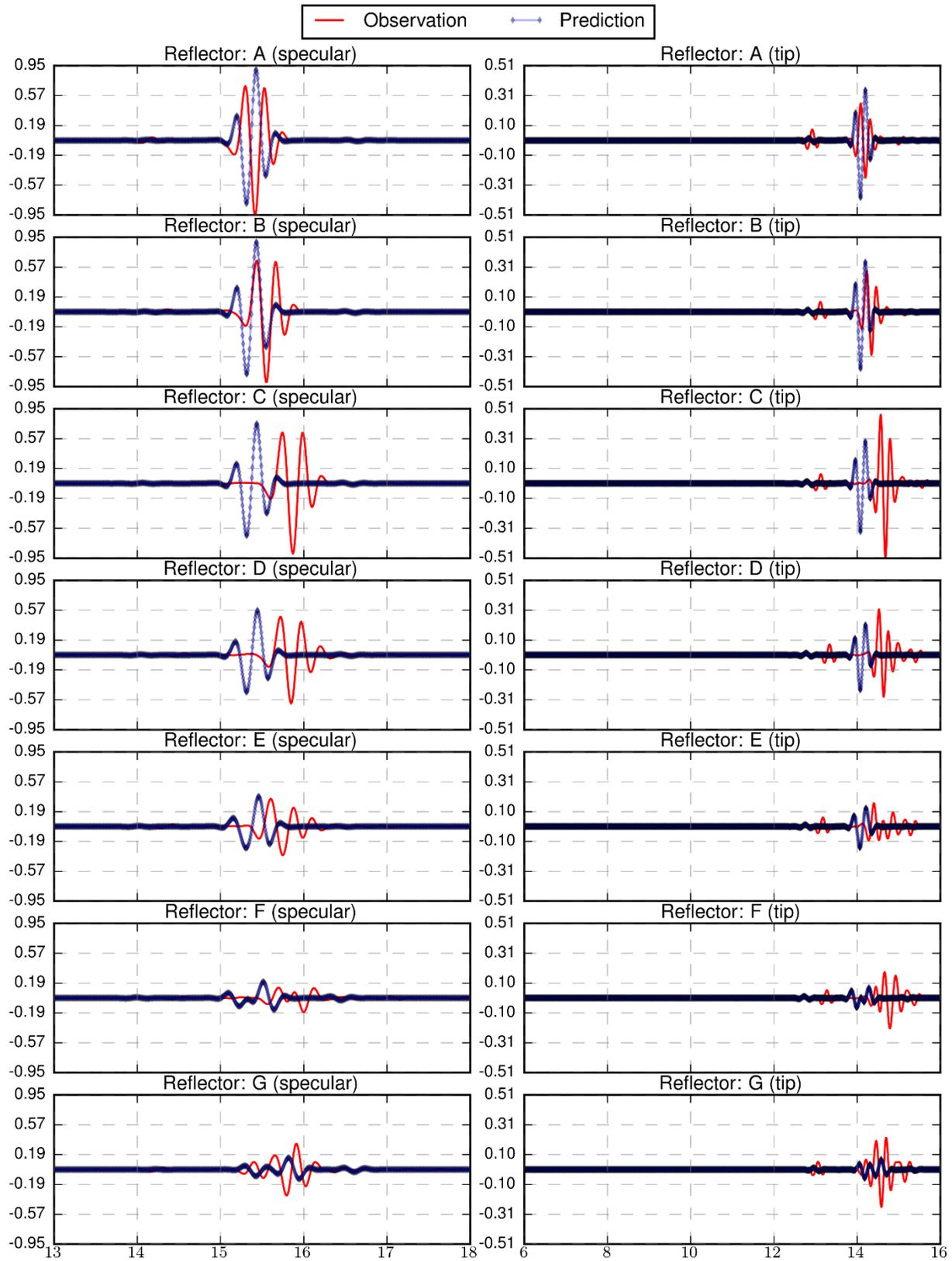


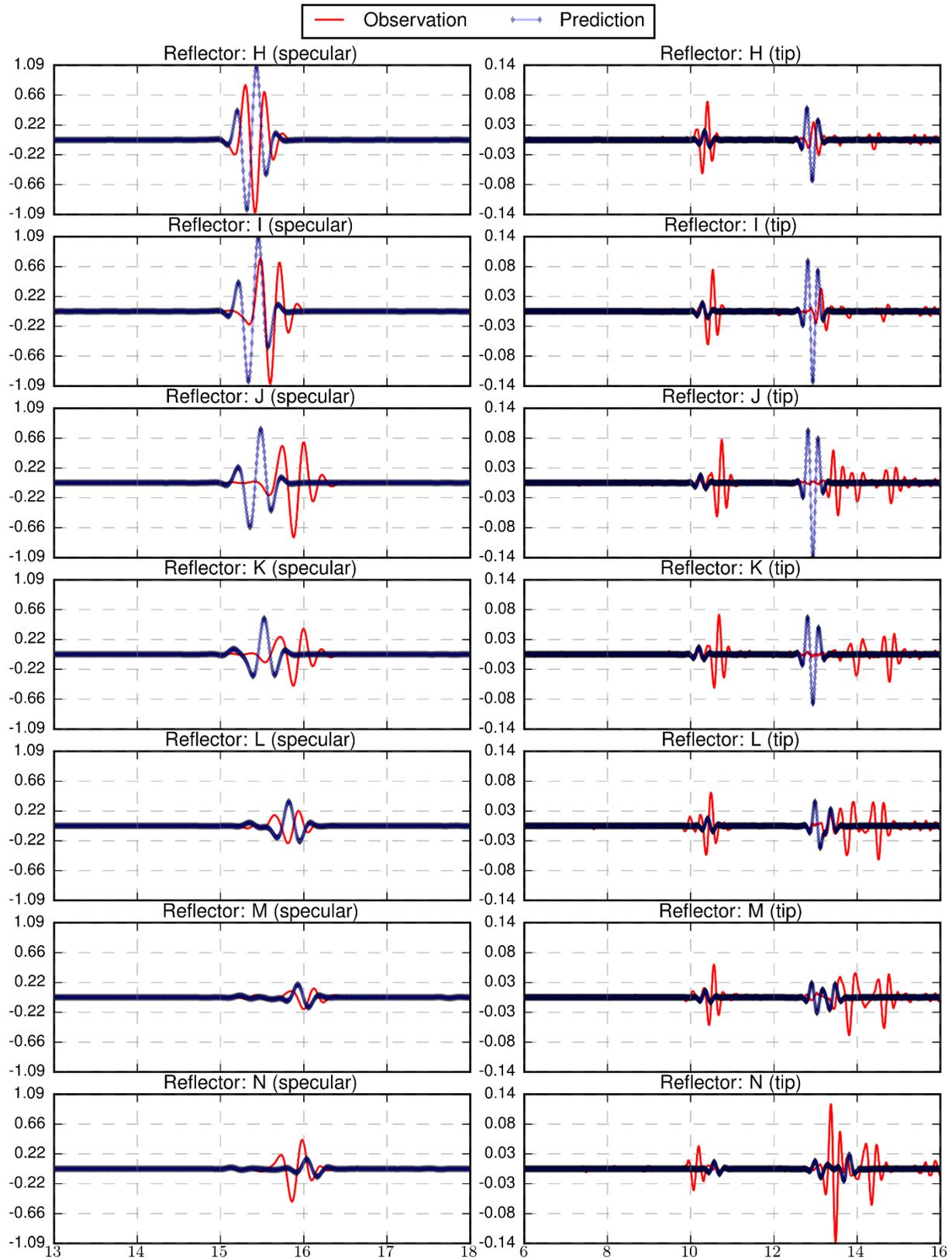
### B.3.29 Specimen 3, Probe 3, Skew 0



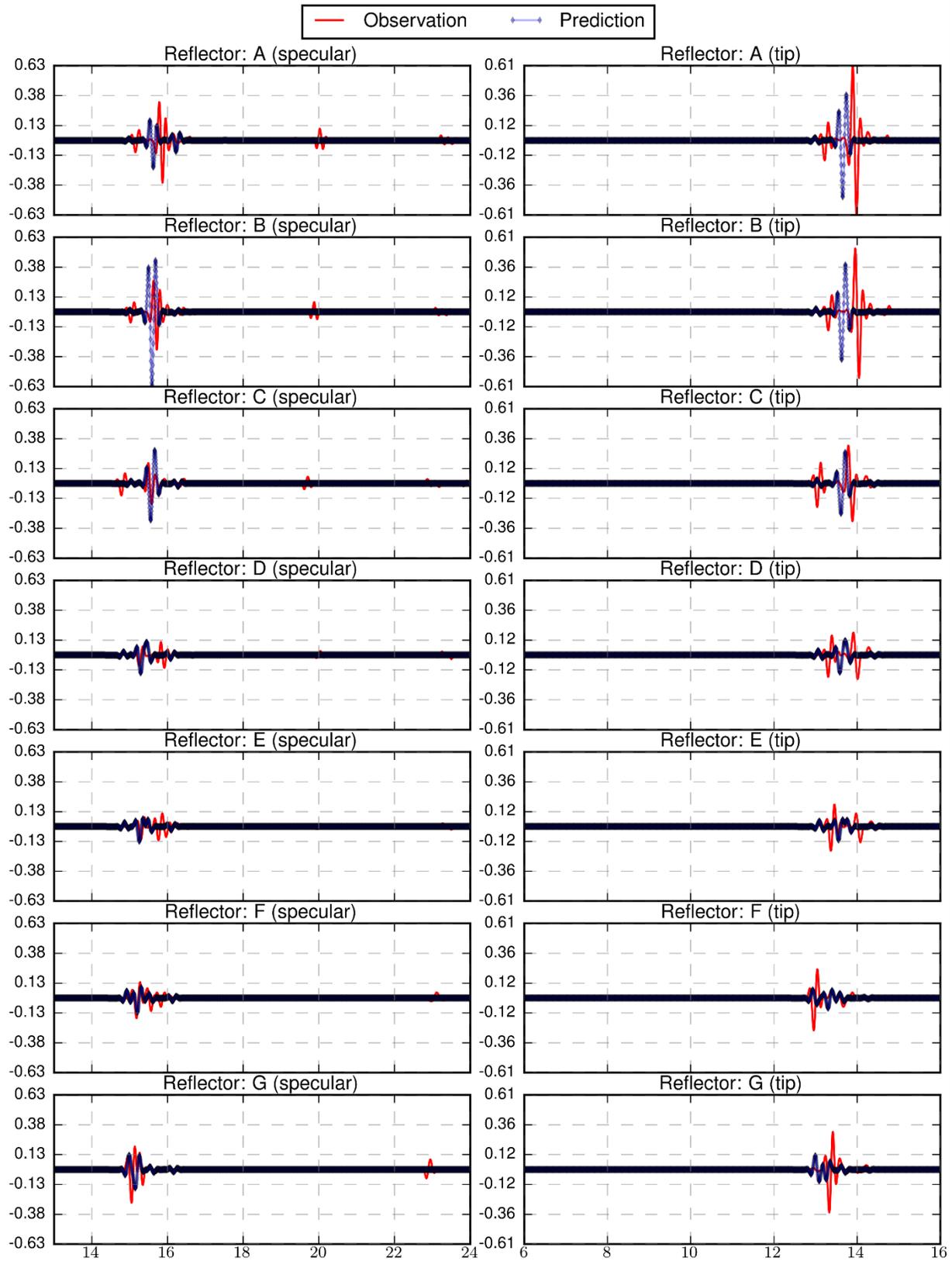


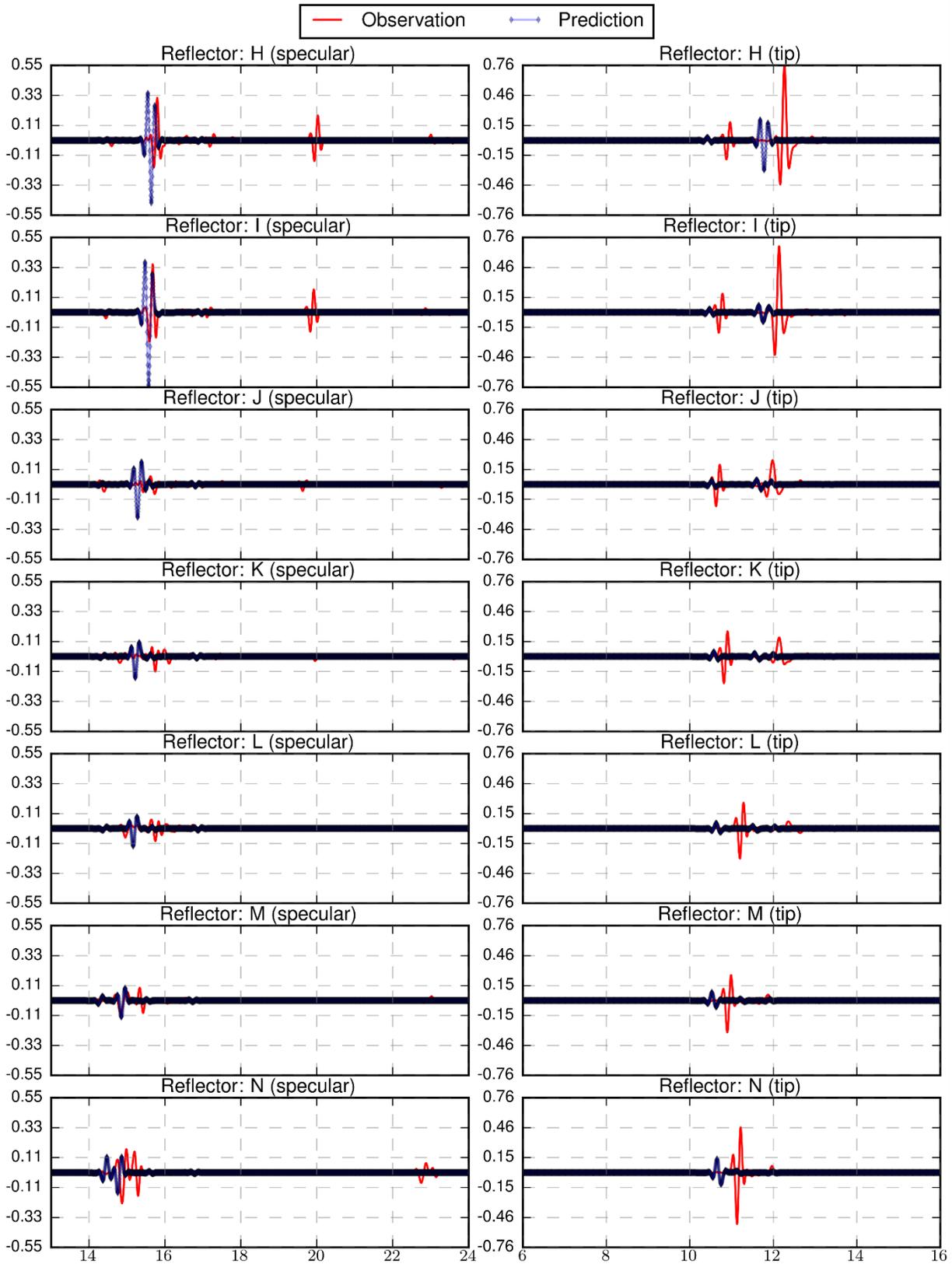
### B.3.30 Specimen 3, Probe 3, Skew 180



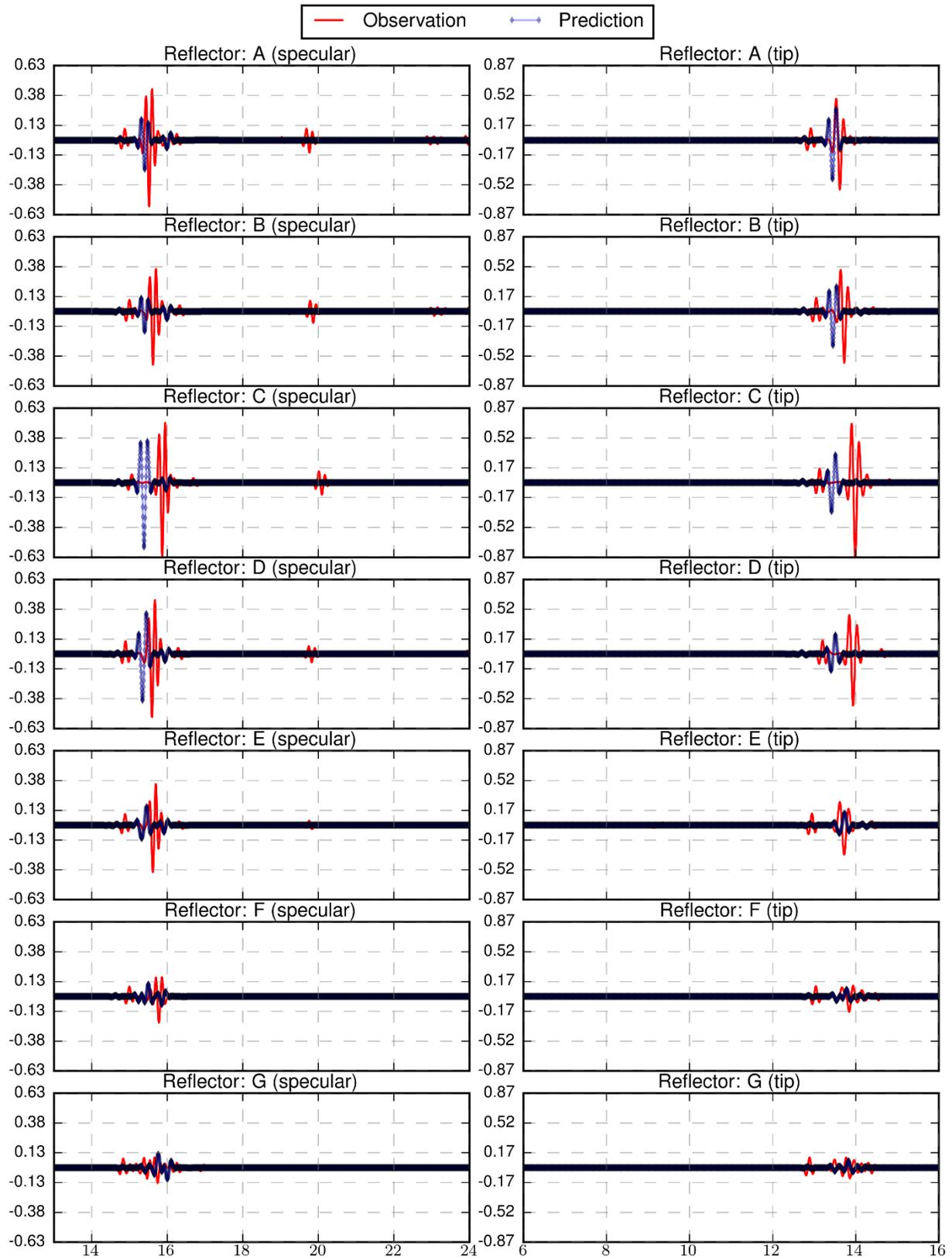


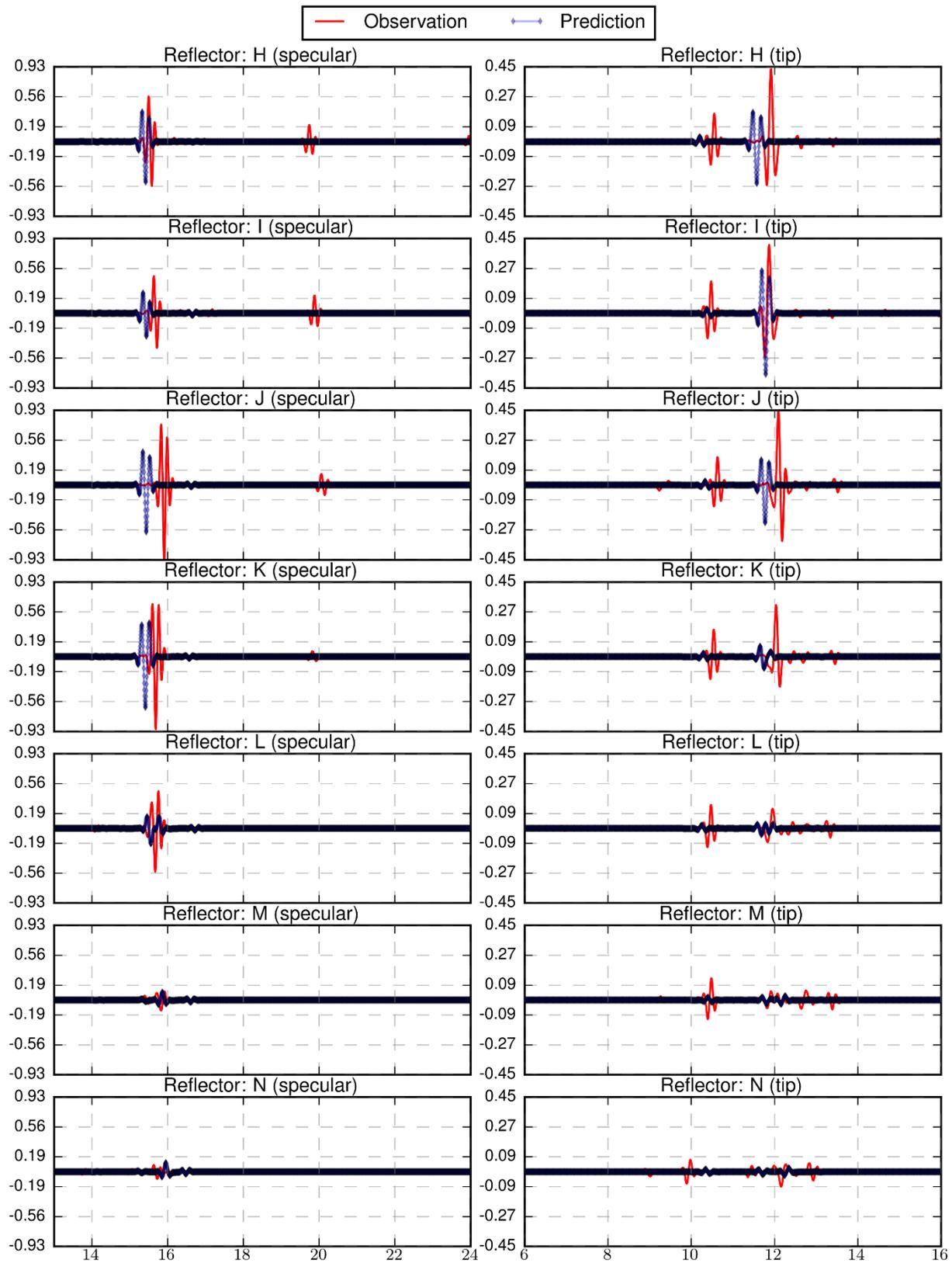
### B.3.31 Specimen 3, Probe 4, Skew 0



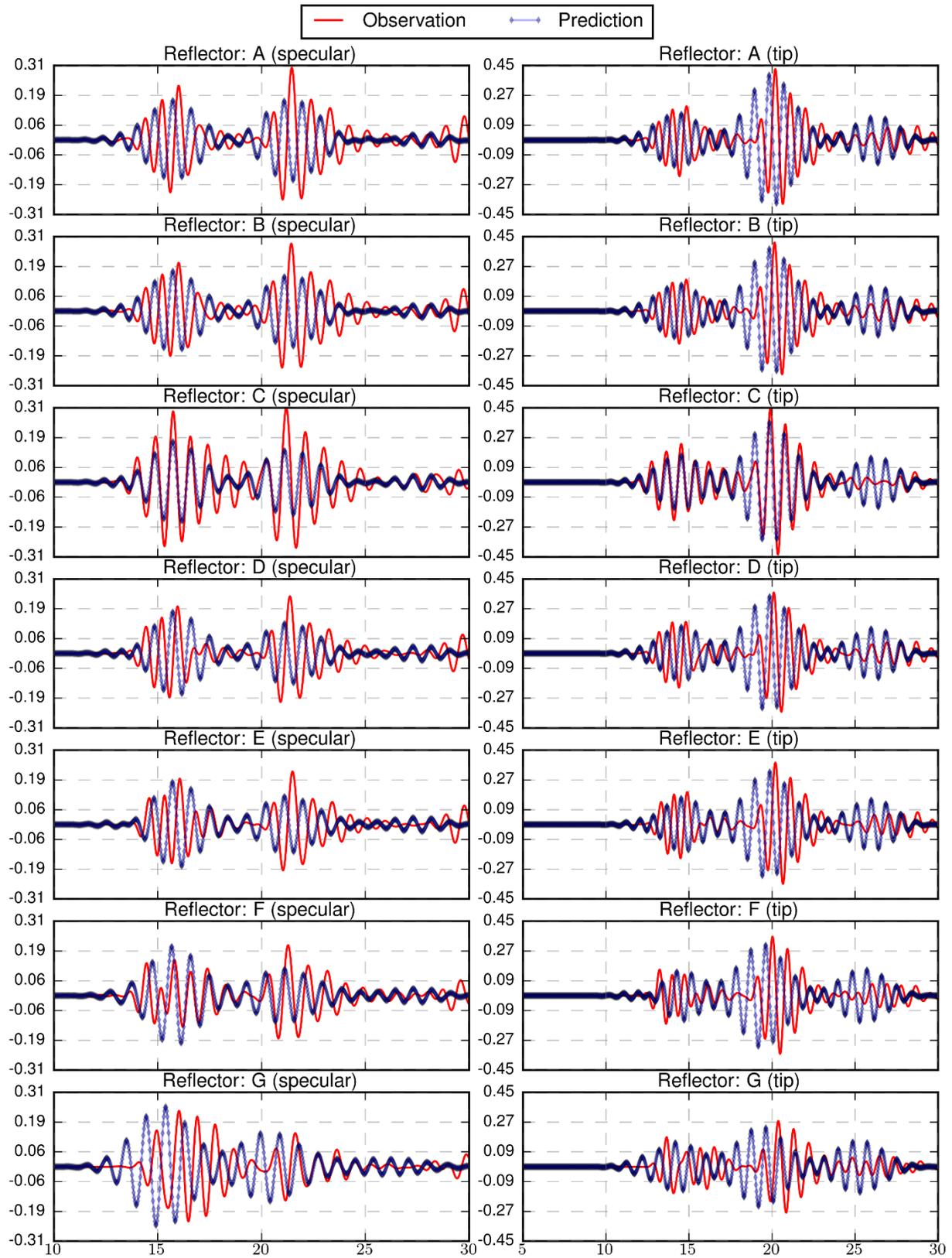


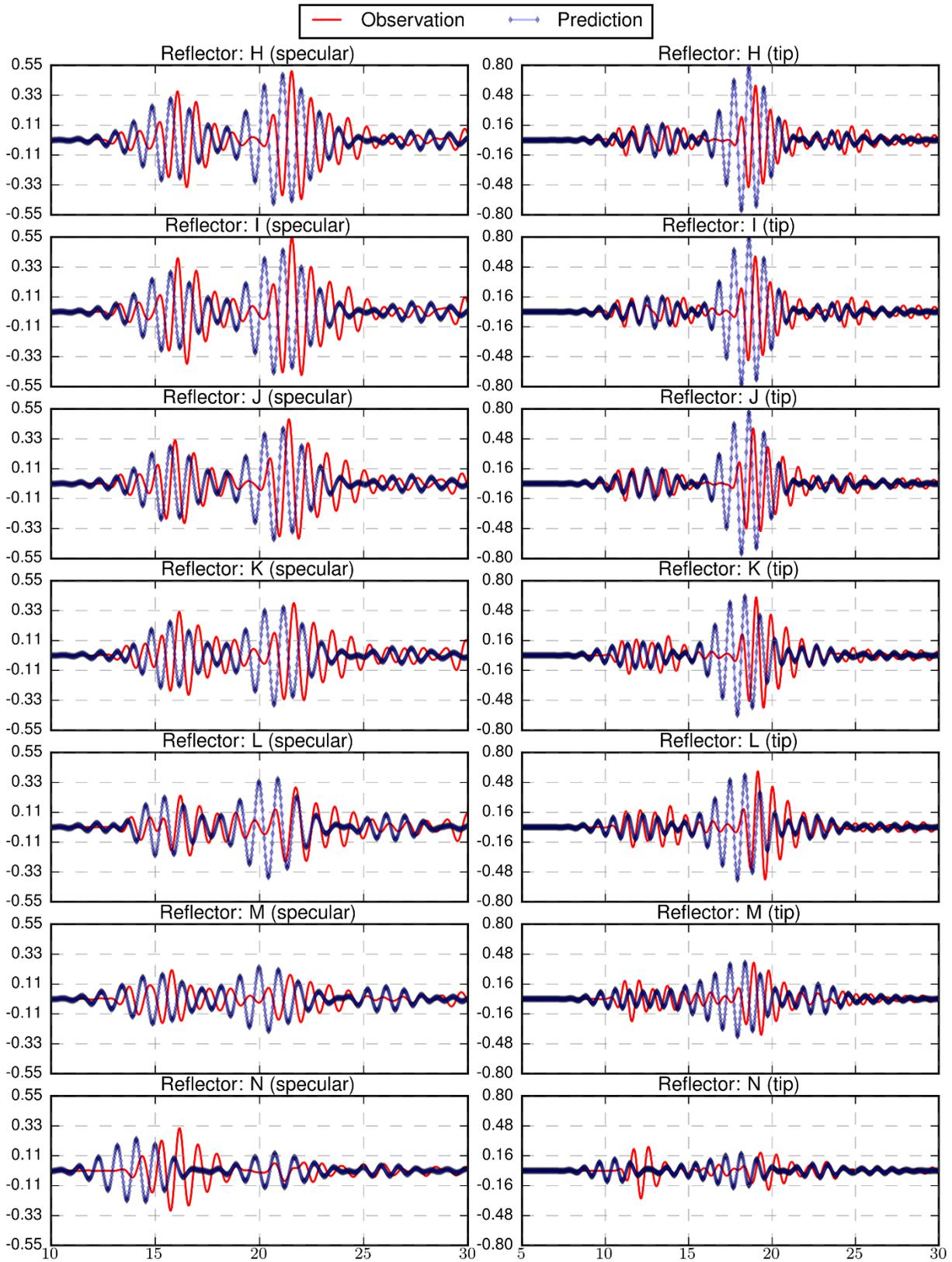
### B.3.32 Specimen 3, Probe 4, Skew 180



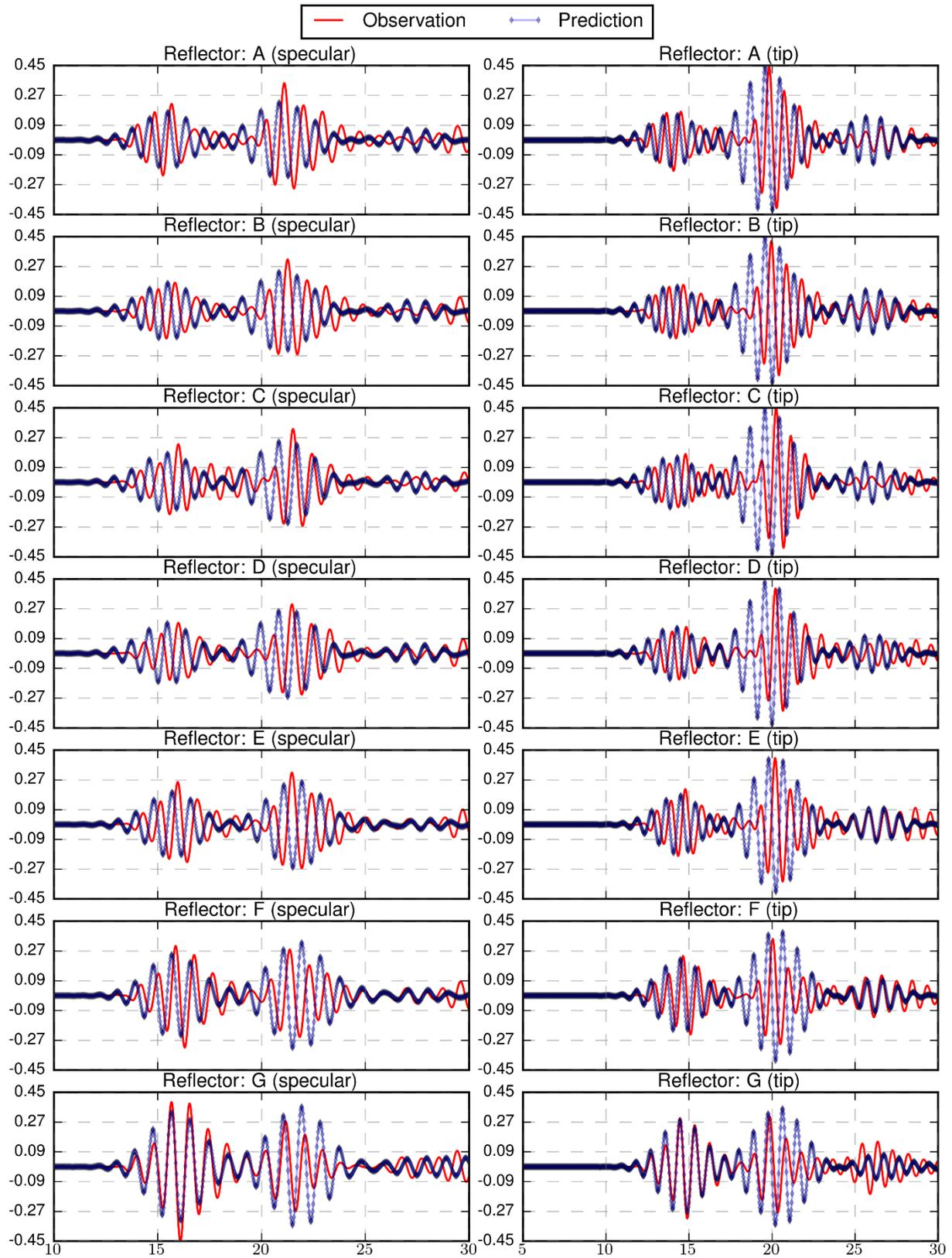


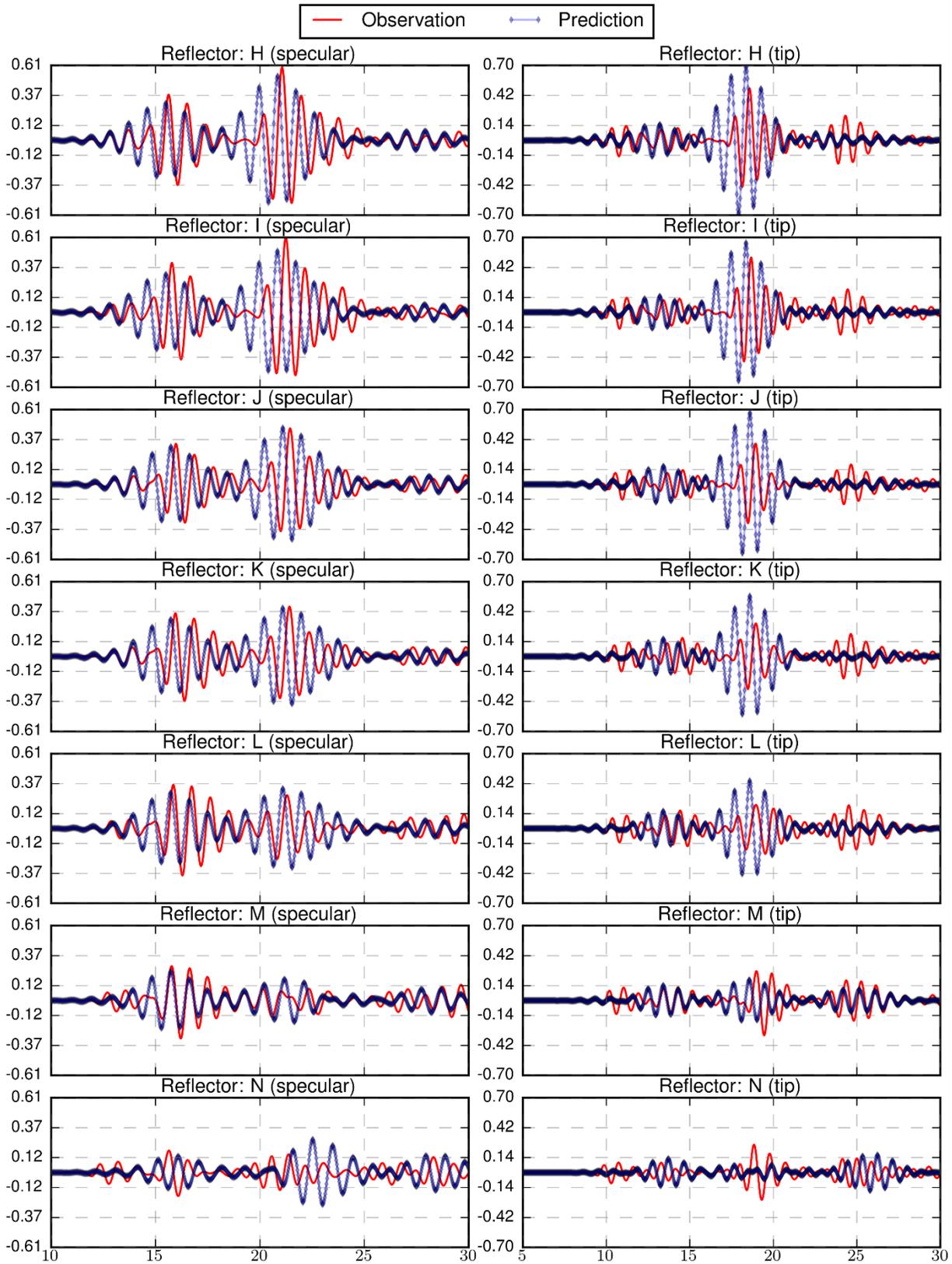
### B.3.33 Specimen 3, Probe 5, Skew 0



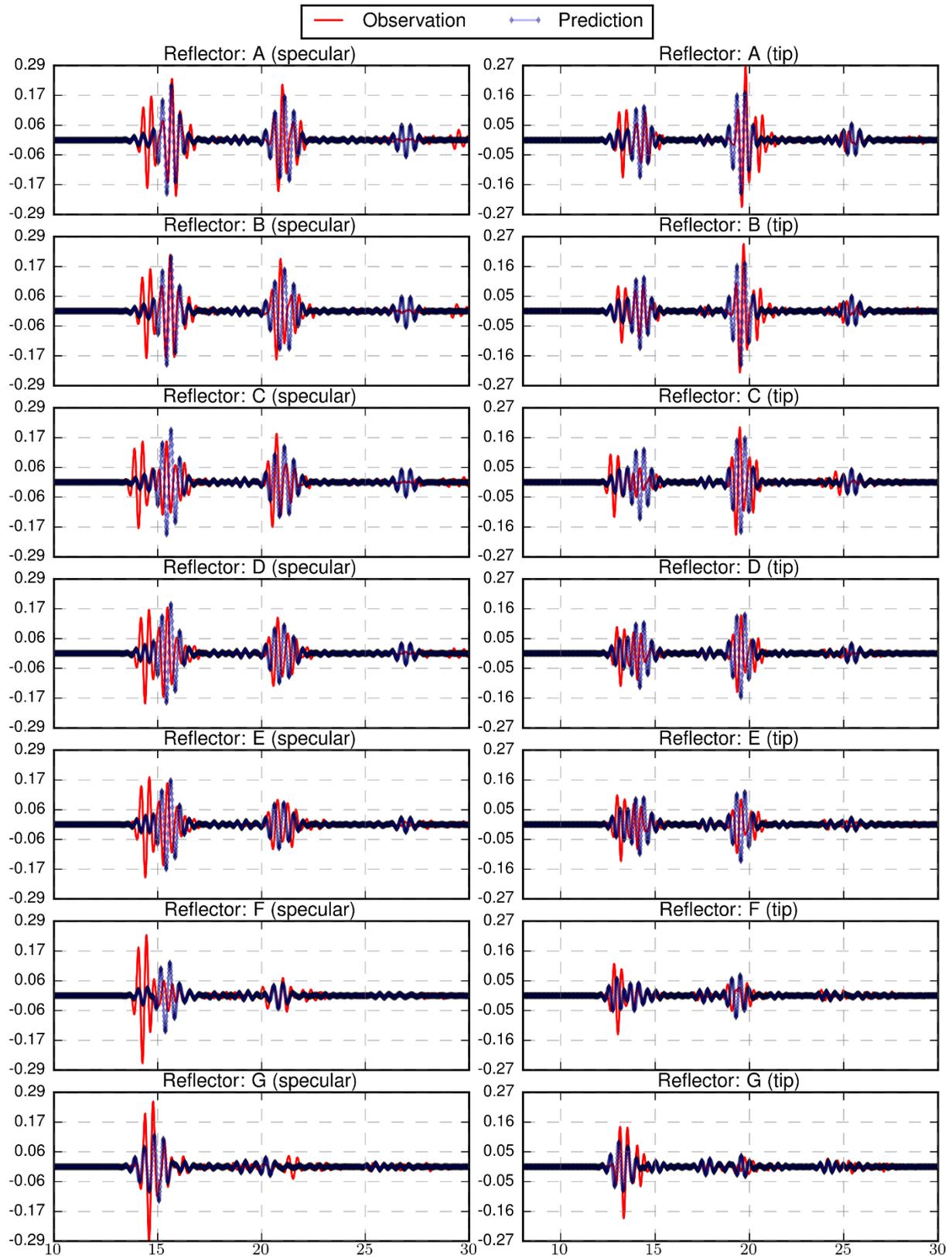


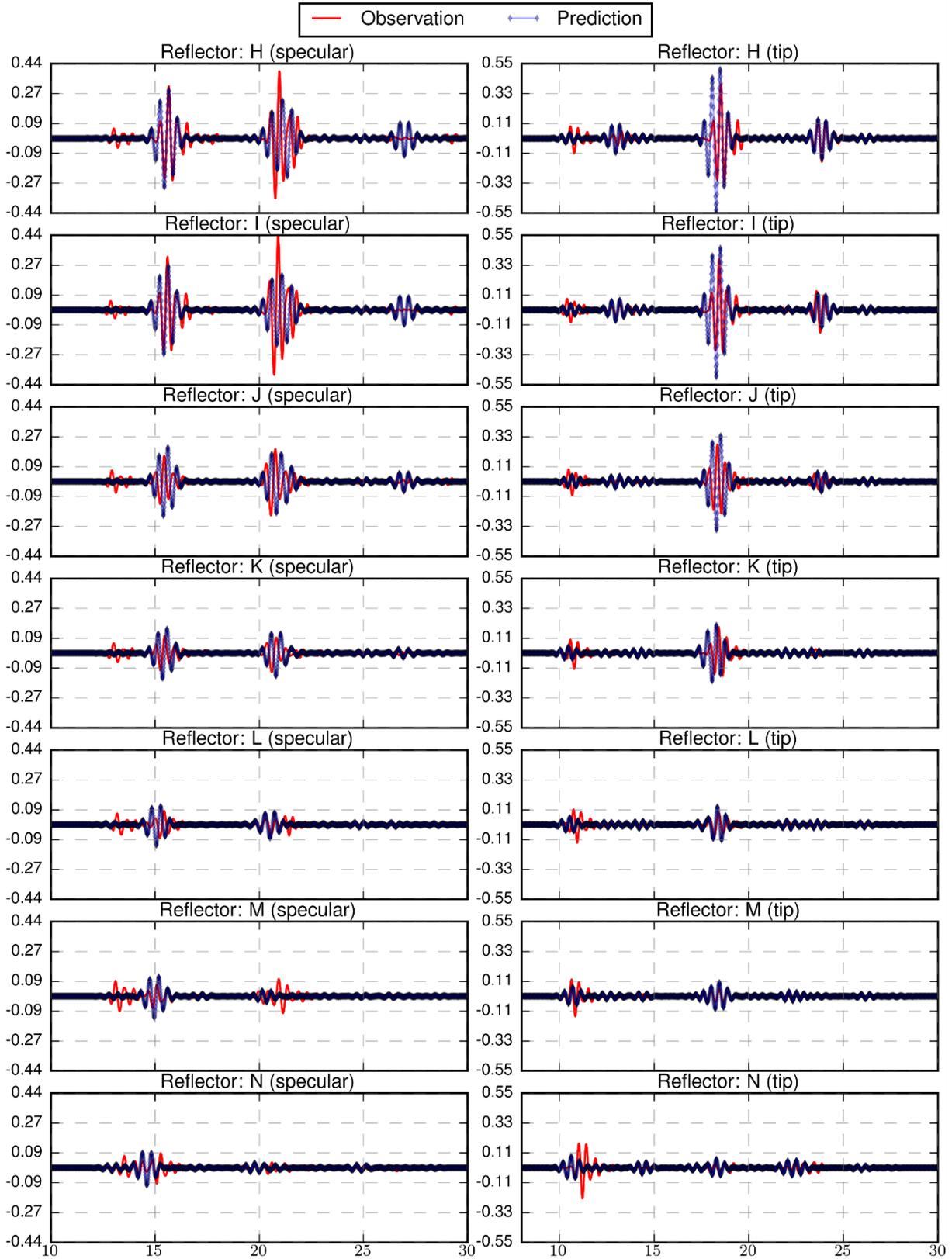
### B.3.34 Specimen 3, Probe 5, Skew 180



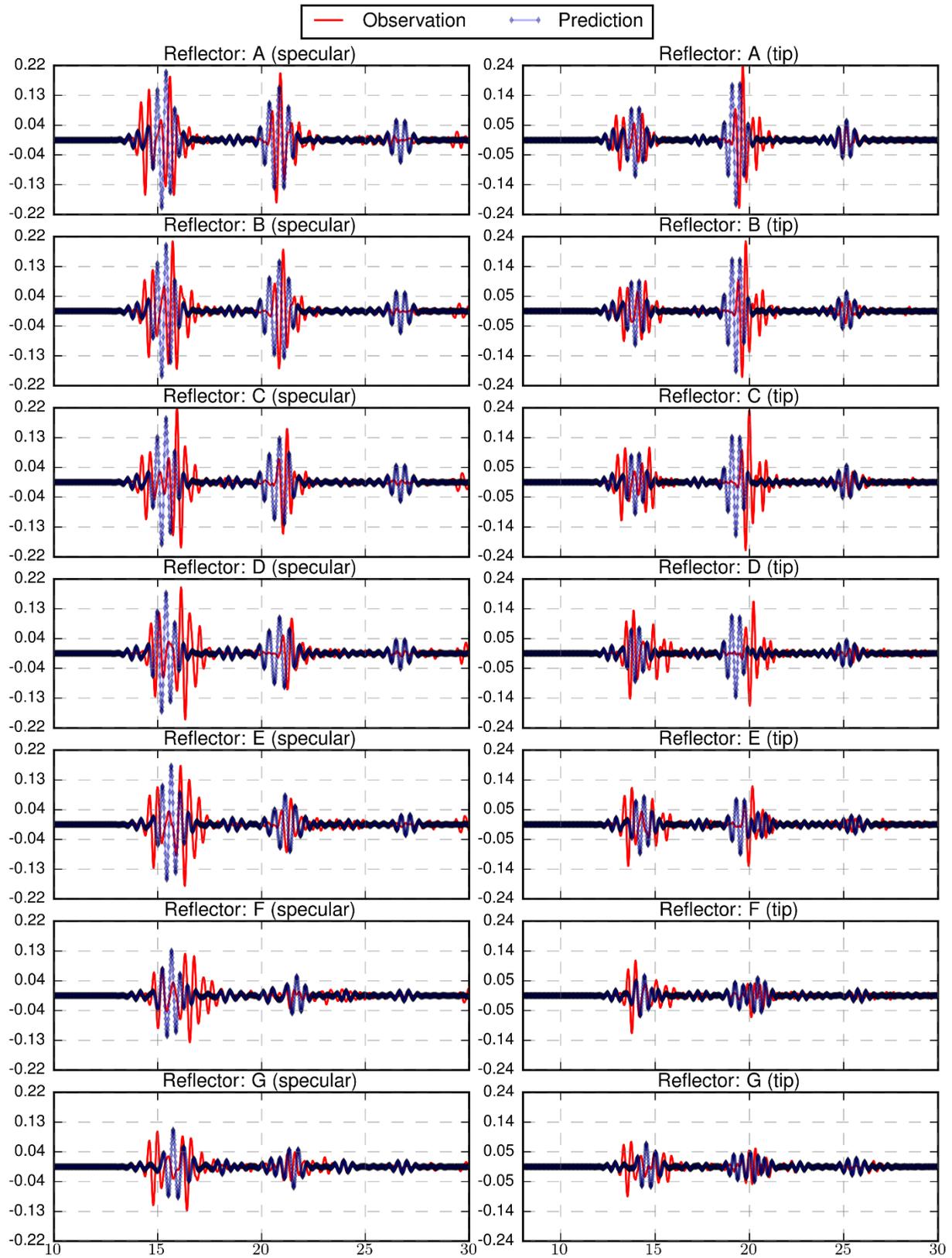


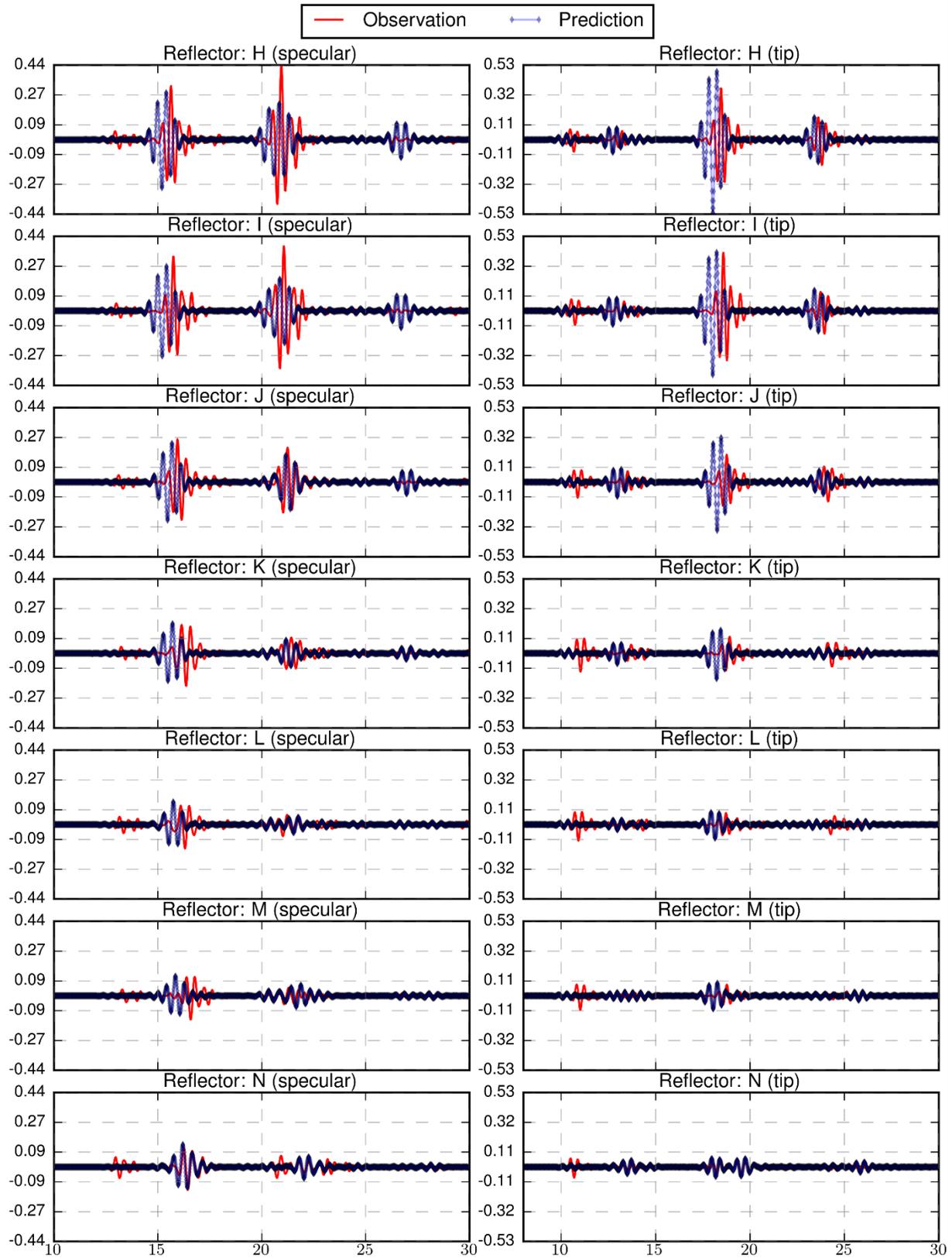
### B.3.35 Specimen 3, Probe 6, Skew 0





### B.3.36 Specimen 3, Probe 6, Skew 180







## **Appendix C**

### **Validation Metric Results**

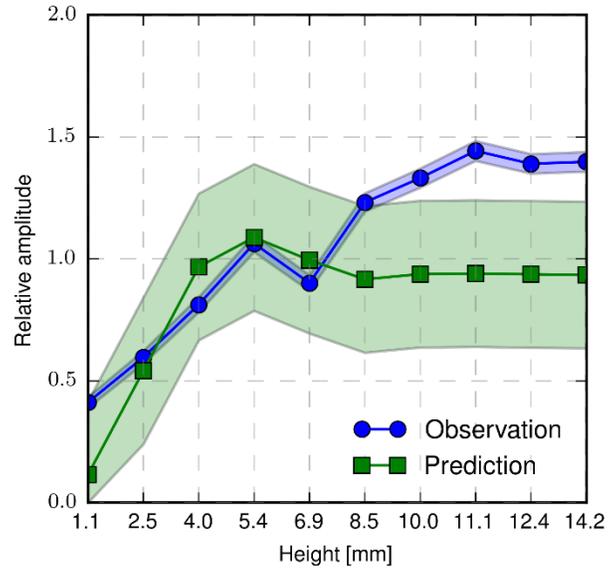


# Appendix C

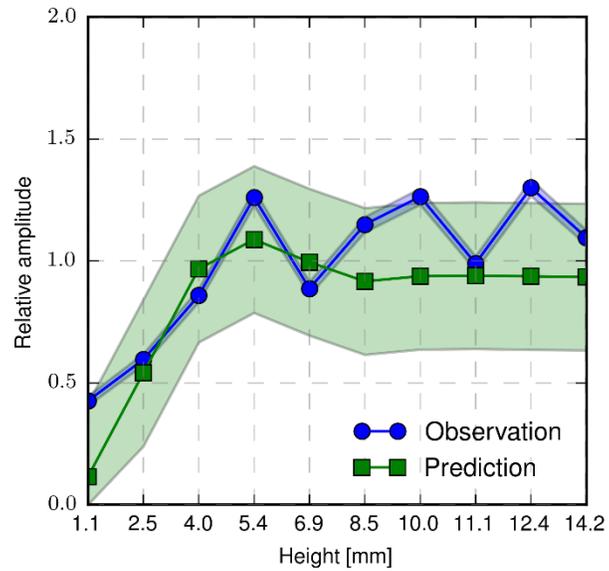
## Validation Metric Results

### C.1 Specimen 1

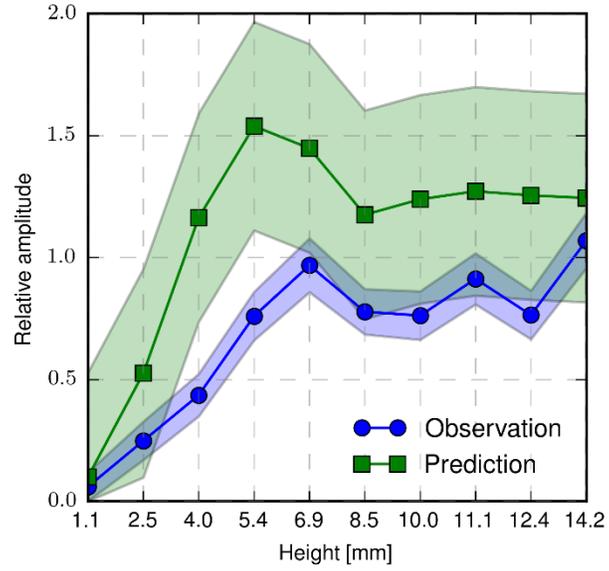
#### C.1.1 Probe 1, Skew 0



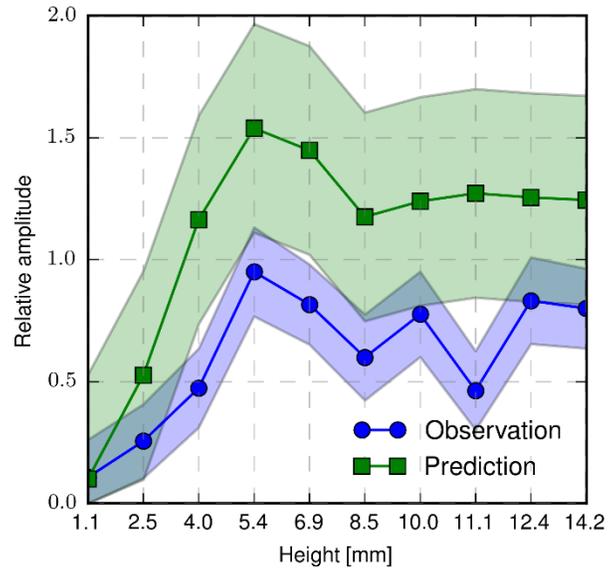
#### C.1.2 Probe 1, Skew 180



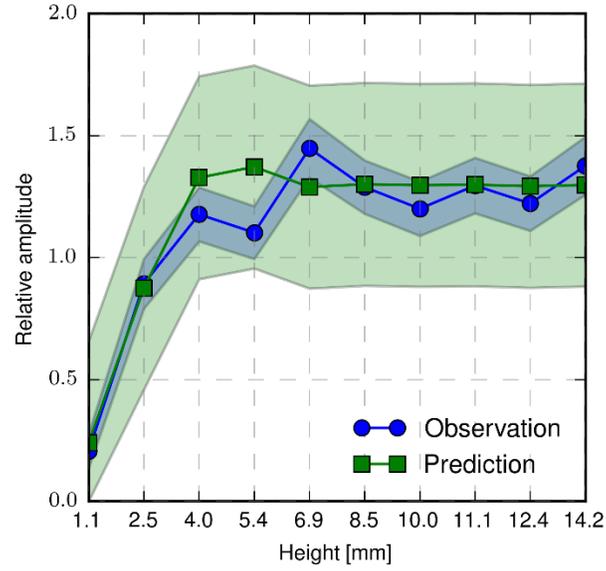
### C.1.3 Probe 2, Skew 0



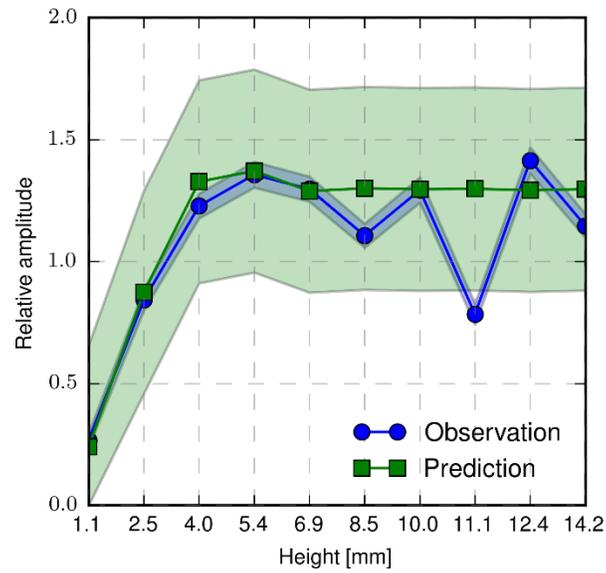
### C.1.4 Probe 2, Skew 180



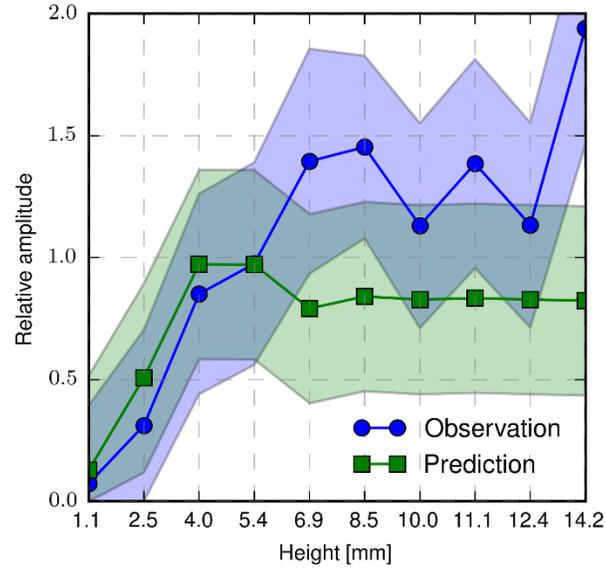
### C.1.5 Probe 3, Skew 0



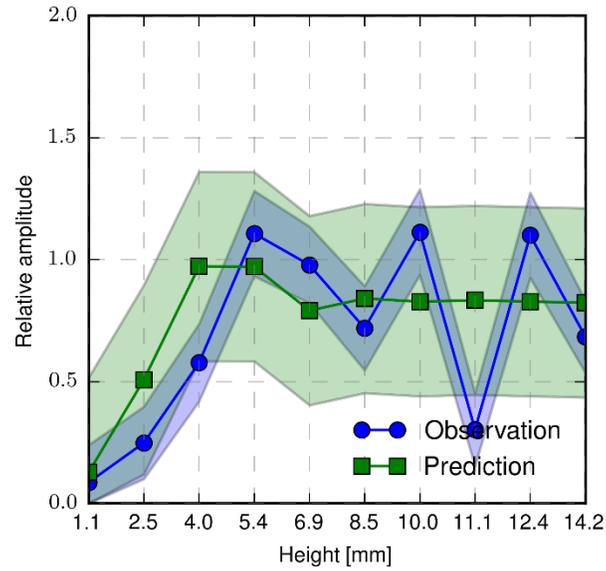
### C.1.6 Probe 3, Skew 180



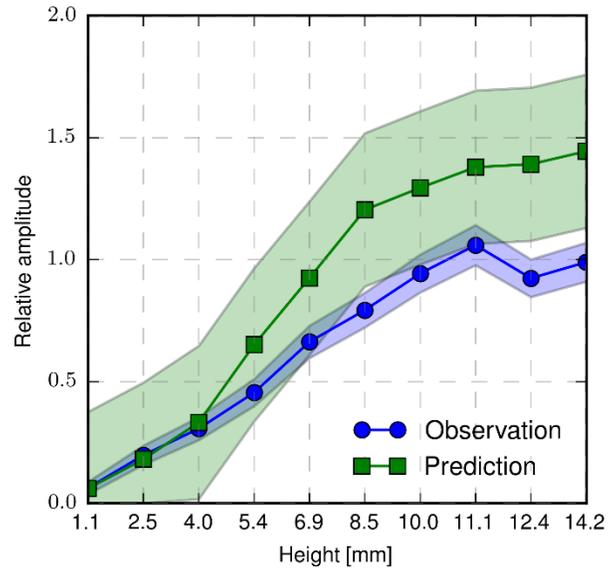
### C.1.7 Probe 4, Skew 0



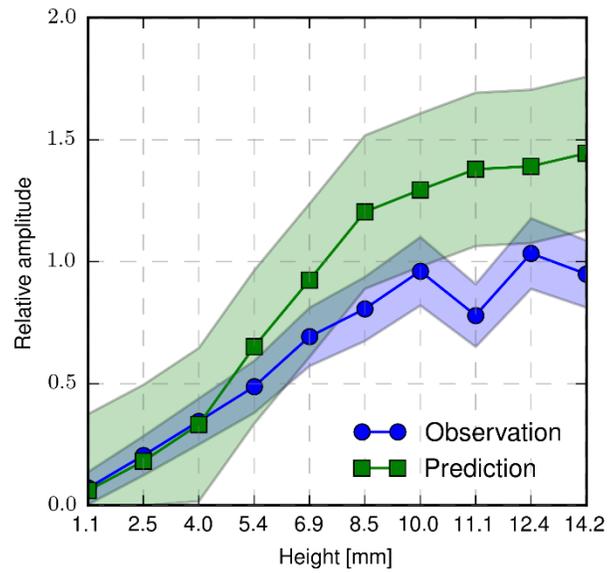
### C.1.8 Probe 4, Skew 180



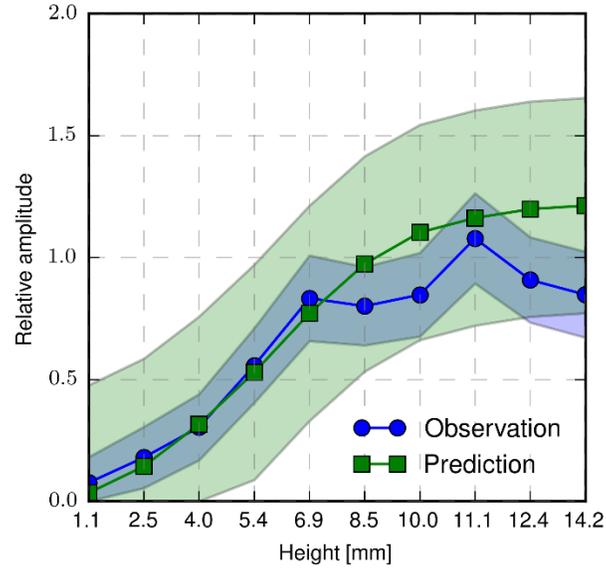
### C.1.9 Probe 5, Skew 0



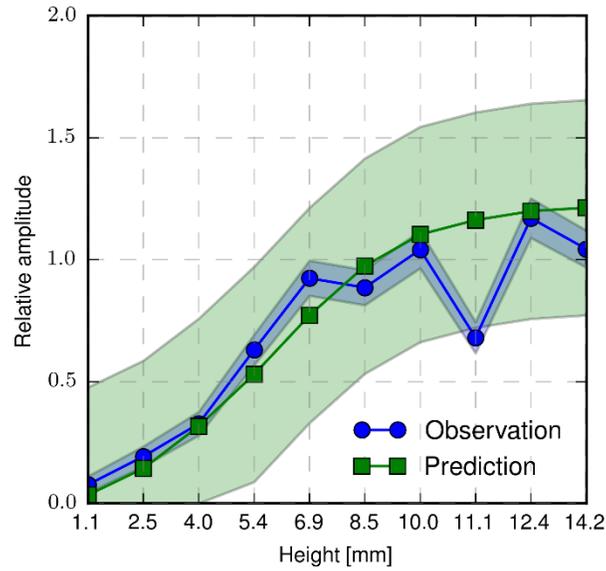
### C.1.10 Probe 5, Skew 180



### C.1.11 Probe 6, Skew 0

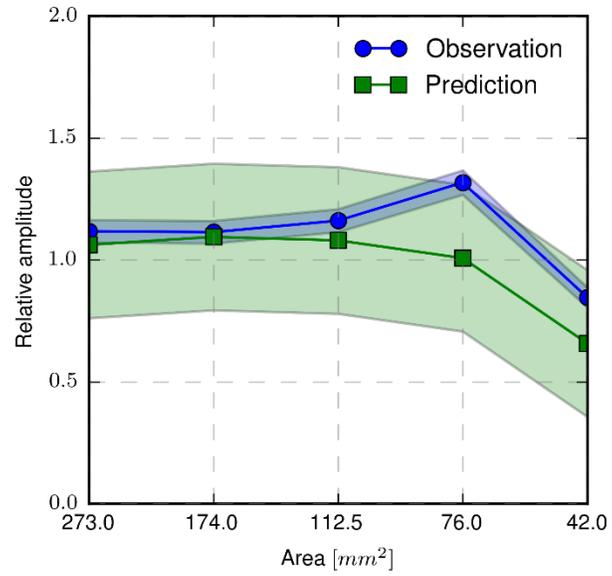


### C.1.12 Probe 6, Skew 180

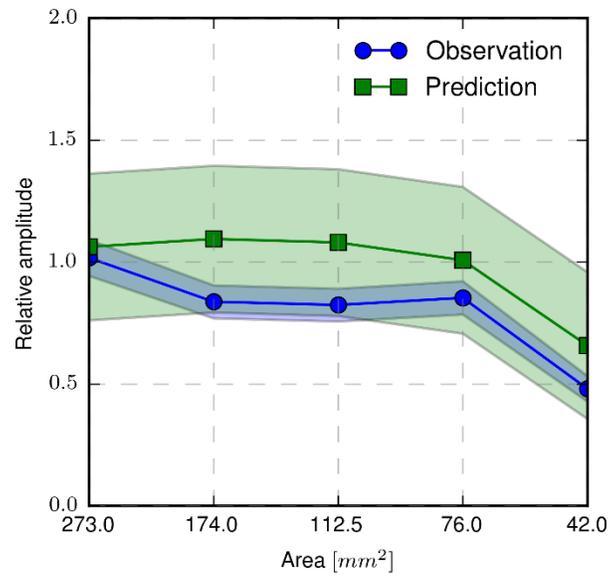


## C.2 Specimen 2

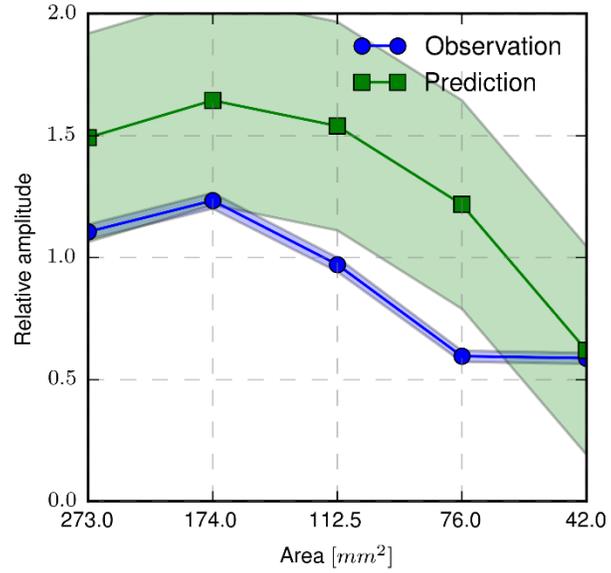
### C.2.1 Probe 1, Skew 0



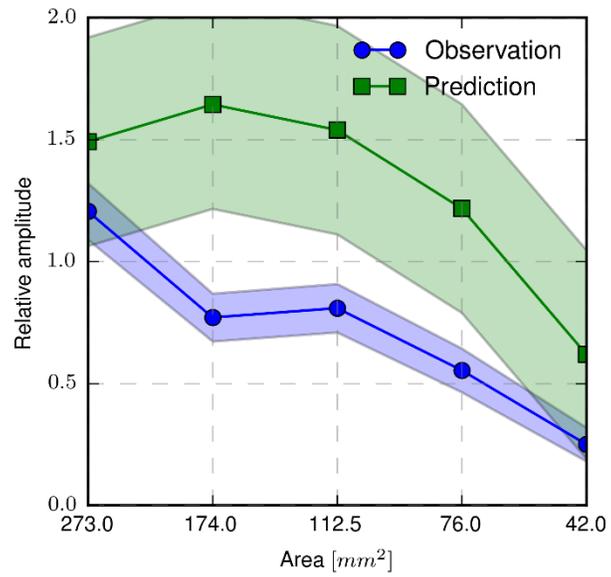
### C.2.2 Probe 1, Skew 180



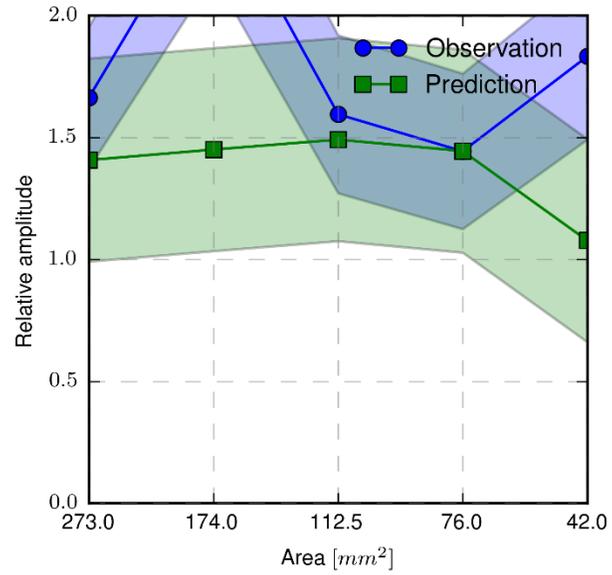
### C.2.3 Probe 2, Skew 0



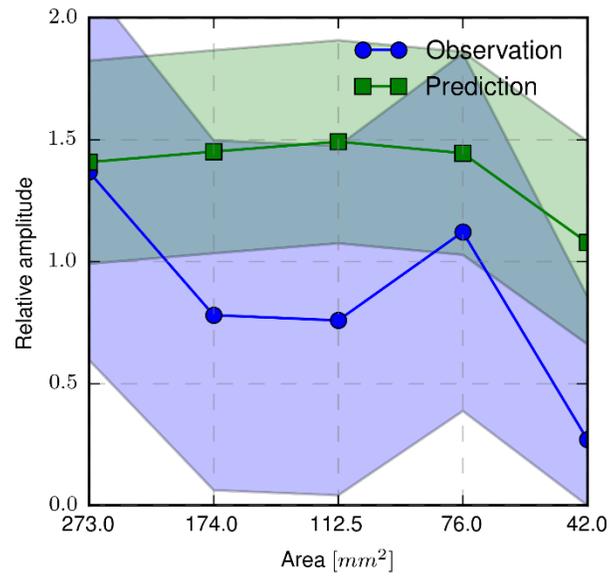
### C.2.4 Probe 2, Skew 180



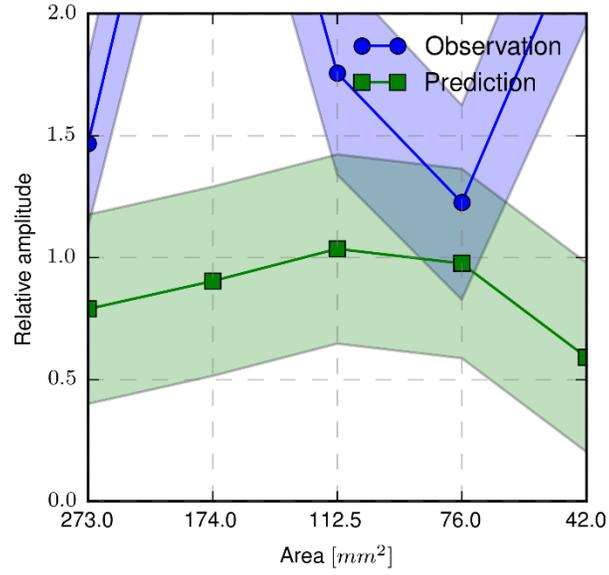
### C.2.5 Probe 3, Skew 0



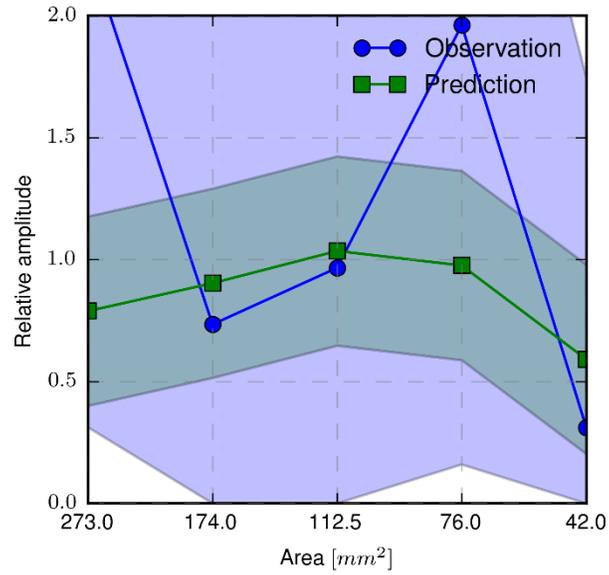
### C.2.6 Probe 3, Skew 180



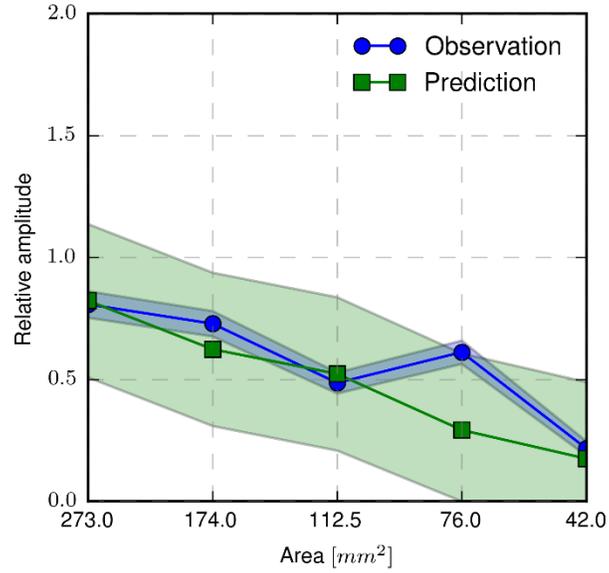
### C.2.7 Probe 4, Skew 0



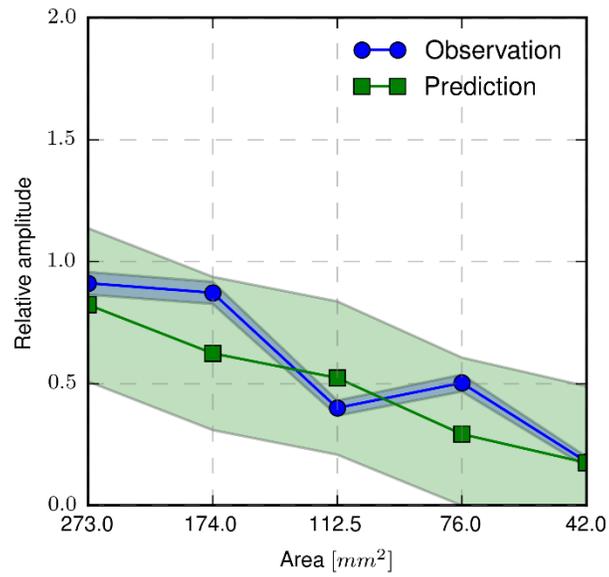
### C.2.8 Probe 4, Skew 180



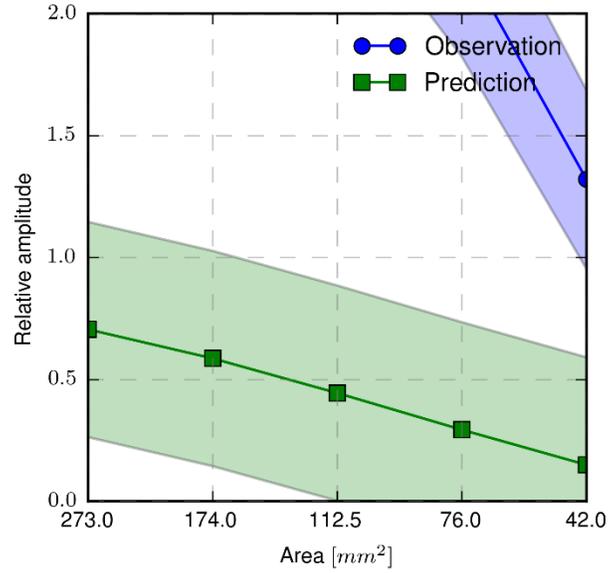
### C.2.9 Probe 5, Skew 0



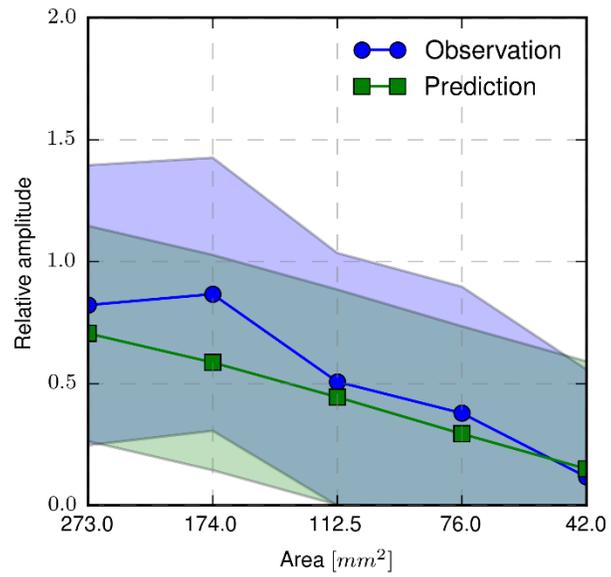
### C.2.10 Probe 5, Skew 180



### C.2.11 Probe 6, Skew 0

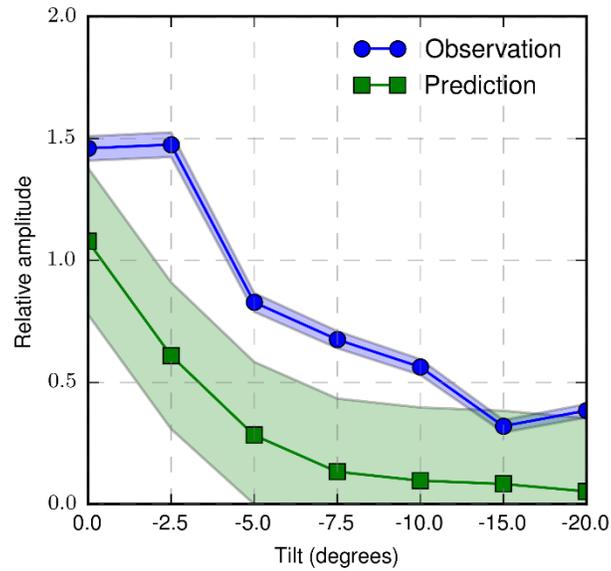
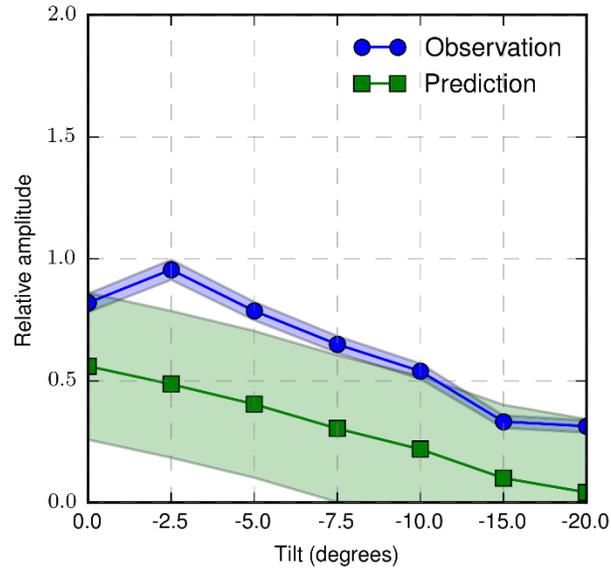


### C.2.12 Probe 6, Skew 180

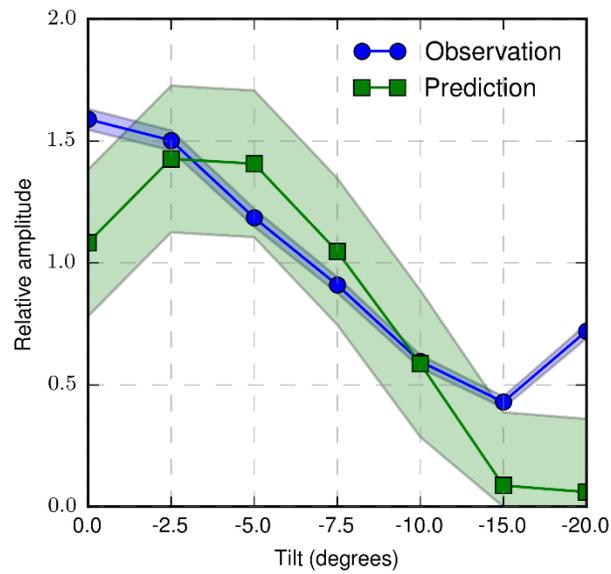
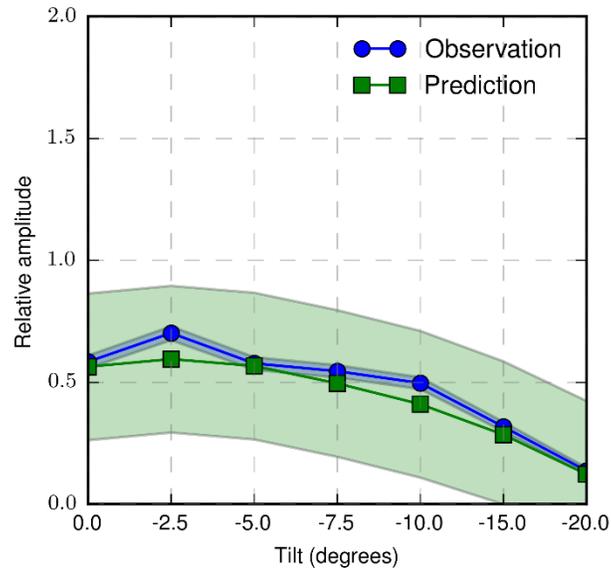


### C.3 Specimen 3

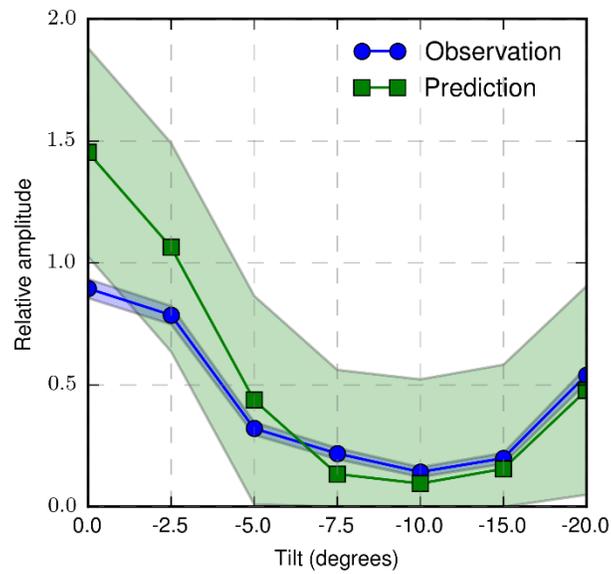
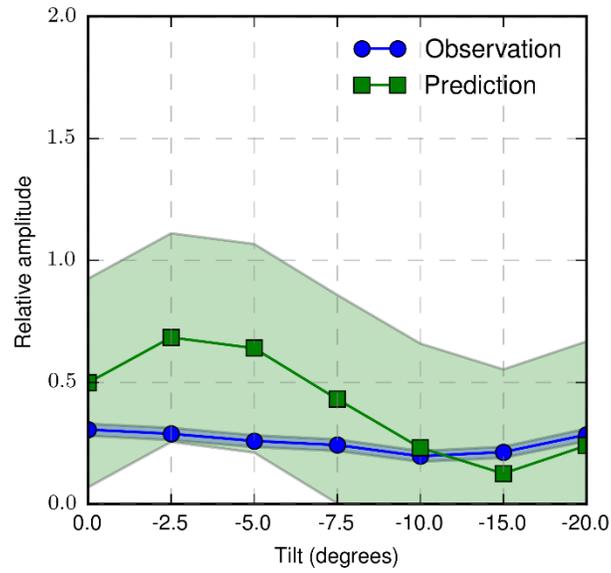
#### C.3.1 Probe 1, Skew 0 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))



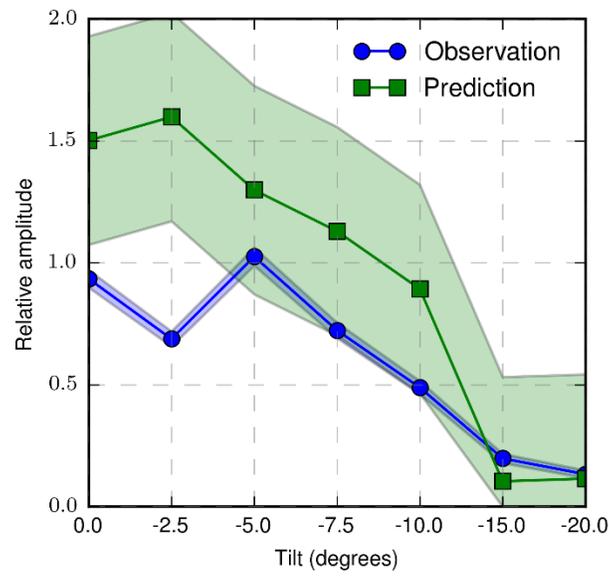
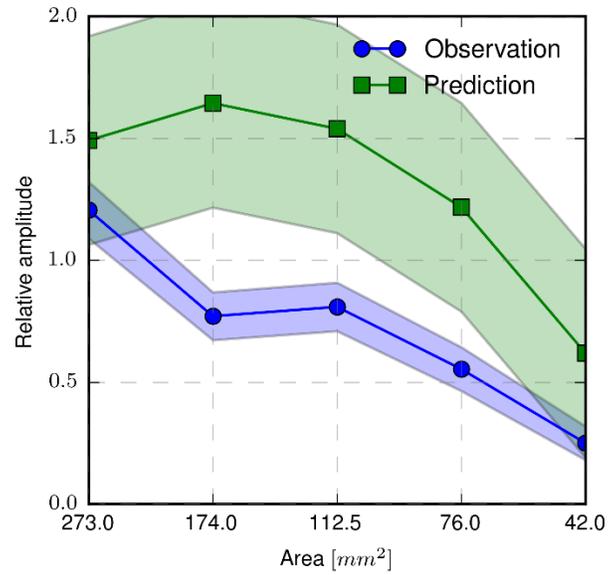
**C.3.2 Probe 1, Skew 180 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))**



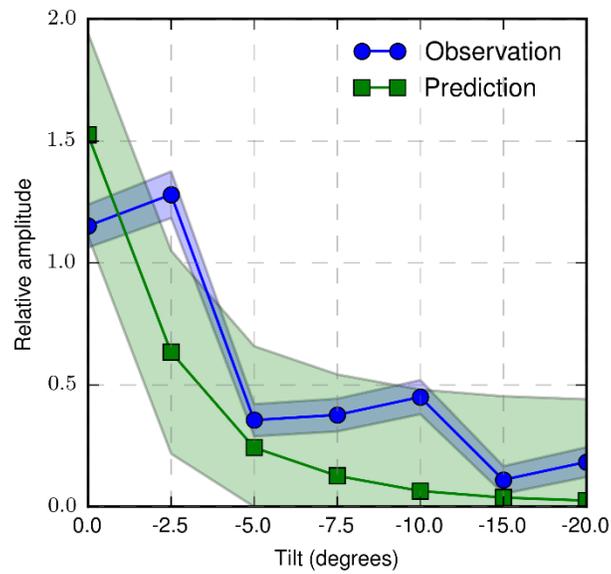
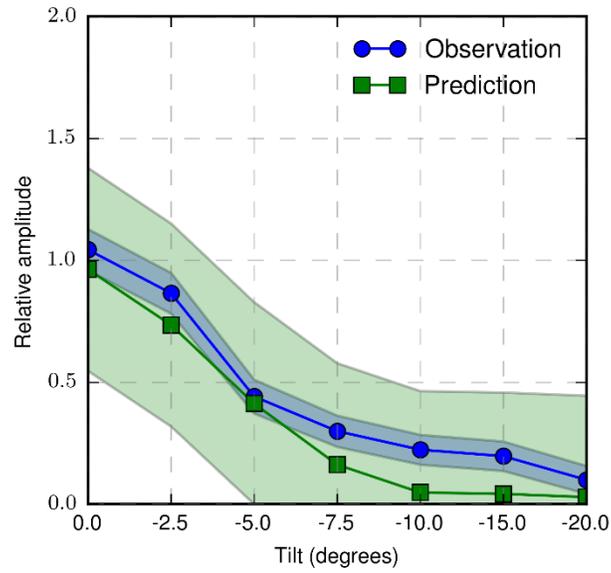
**C.3.3 Probe 2, Skew 0 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))**



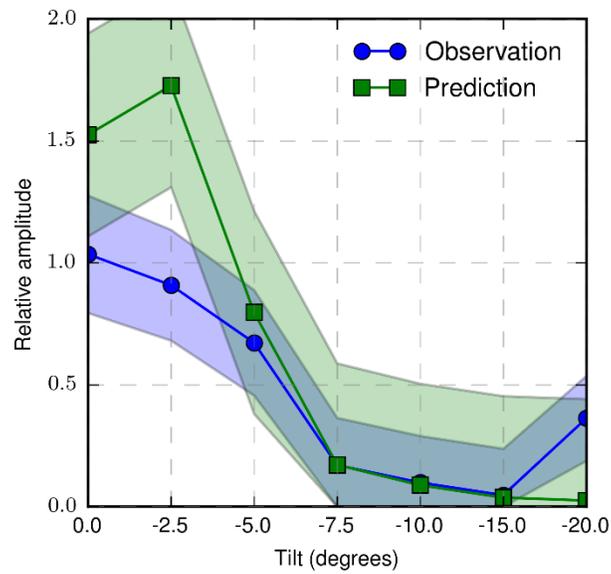
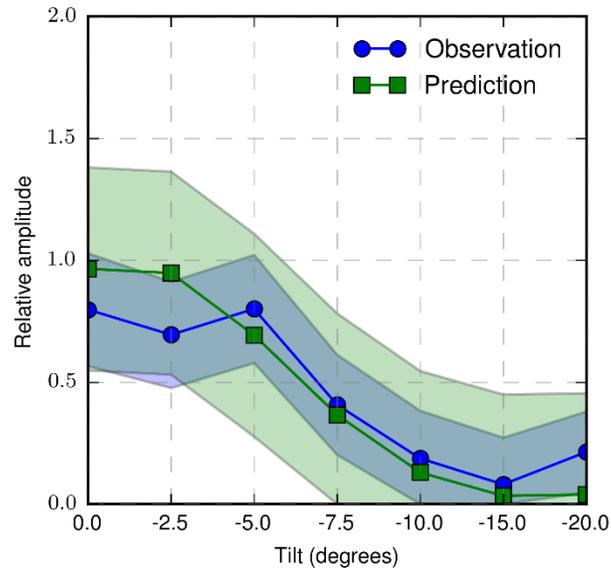
**C.3.4 Probe 2, Skew 180 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))**



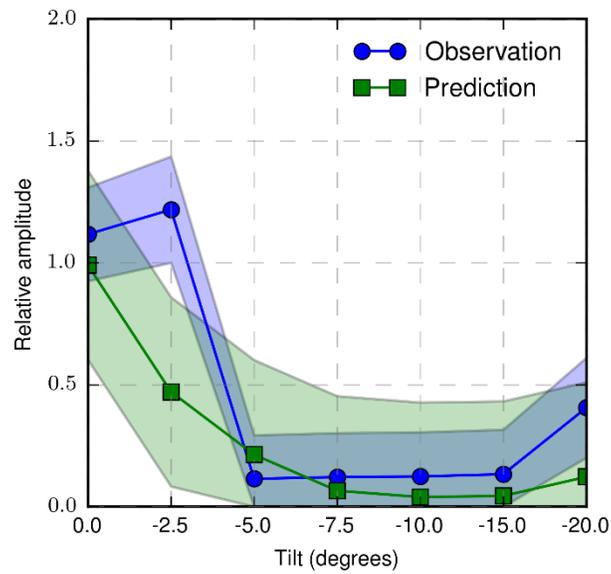
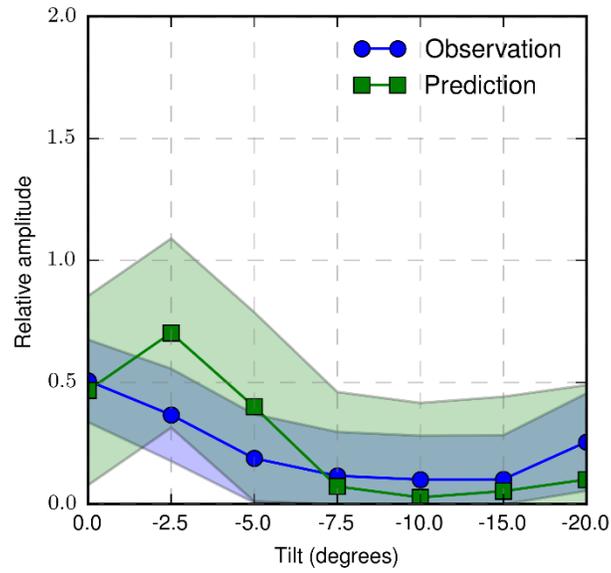
**C.3.5 Probe 3, Skew 0 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))**



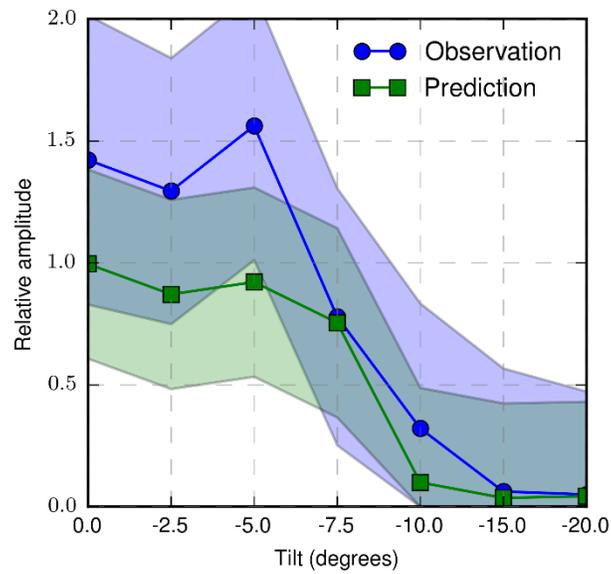
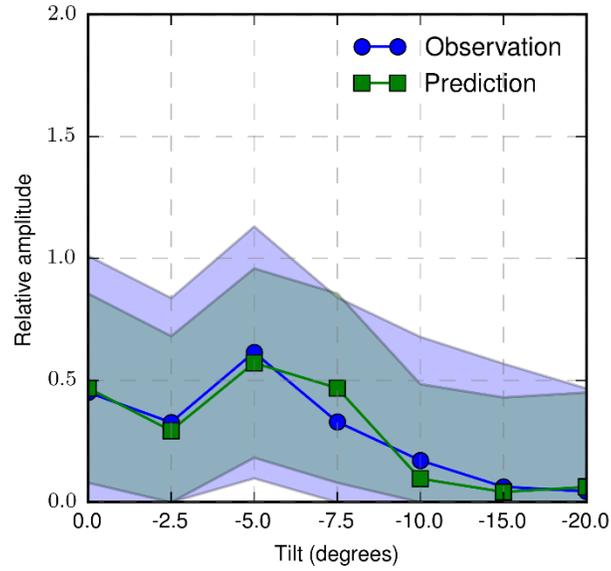
**C.3.6 Probe 3, Skew 180 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))**



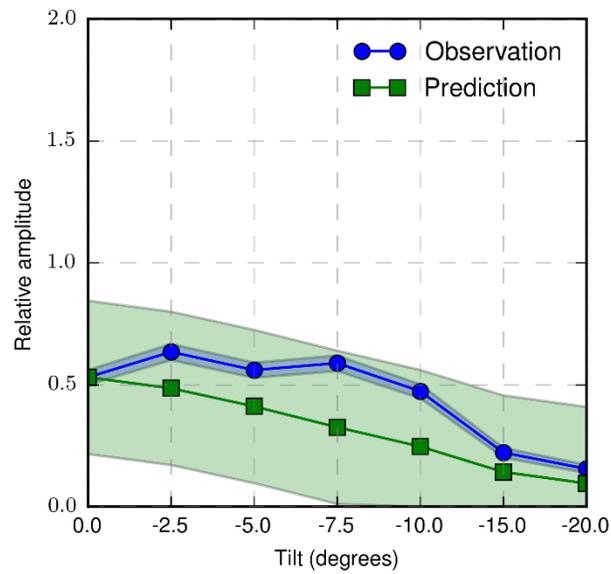
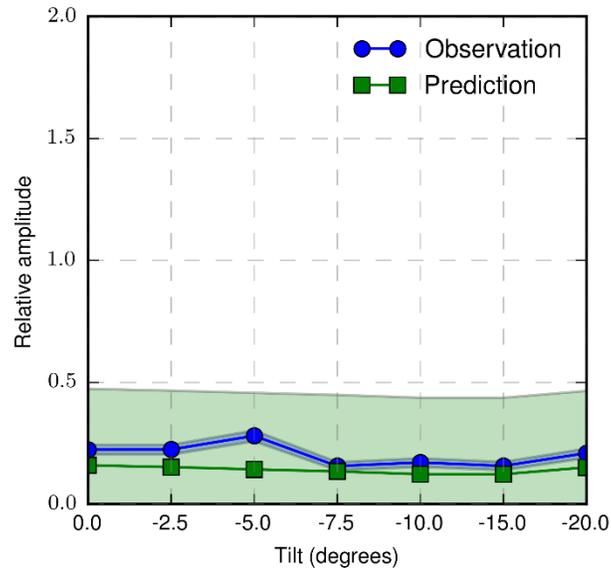
**C.3.7 Probe 4, Skew 0 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))**



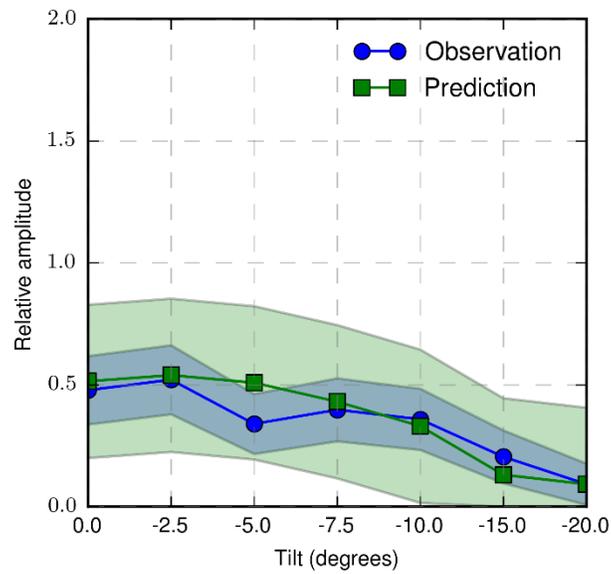
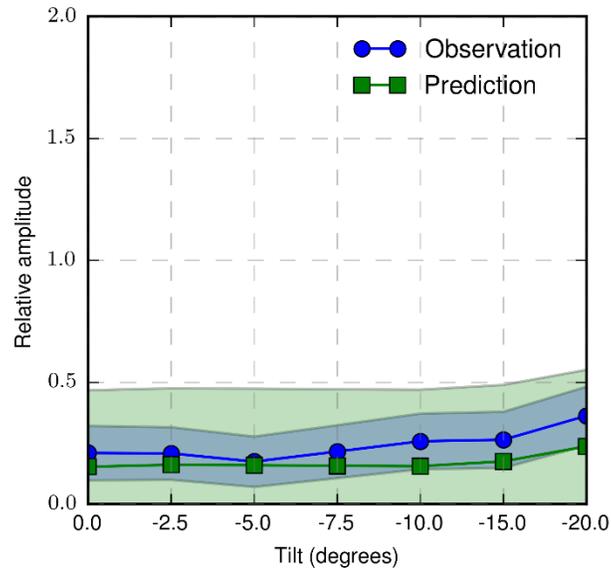
**C.3.8 Probe 4, Skew 180 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))**



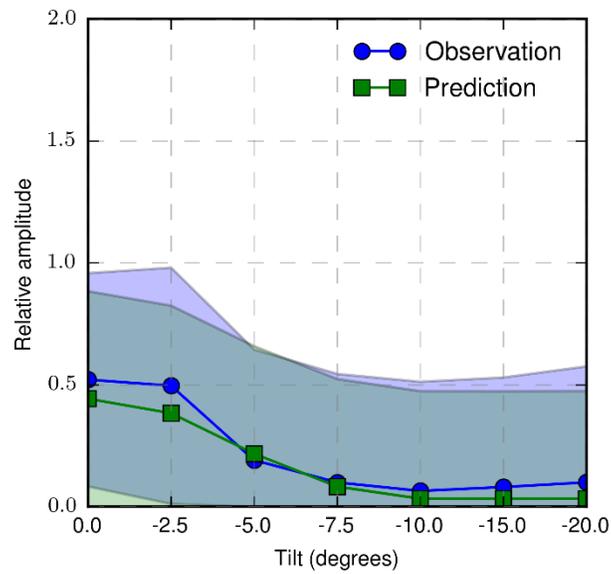
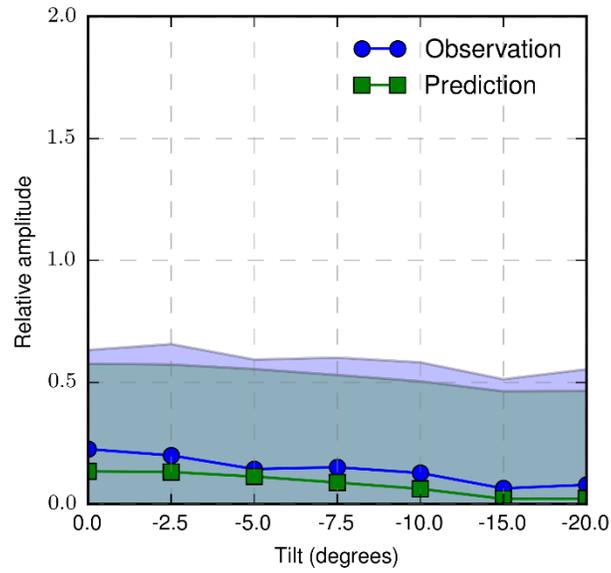
**C.3.9 Probe 5, Skew 0 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))**



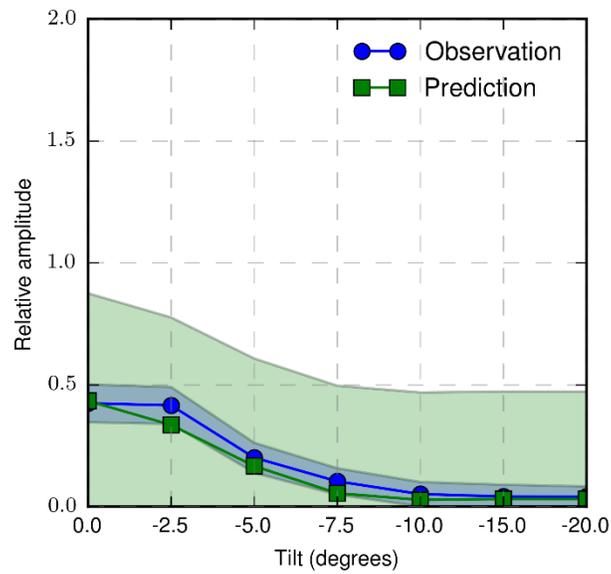
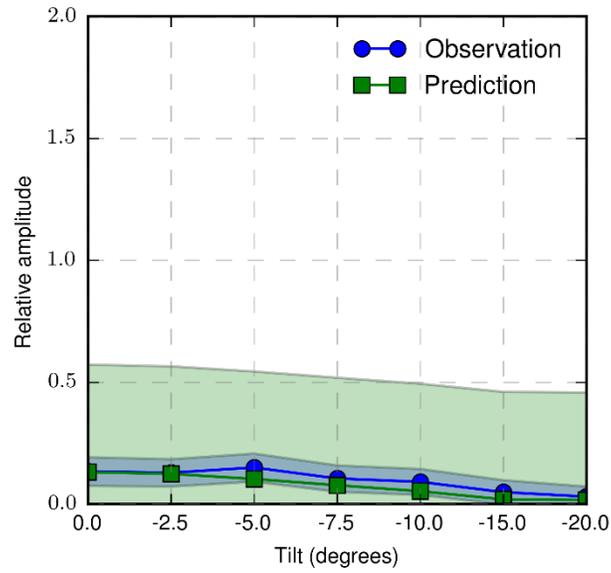
**C.3.10 Probe 5, Skew 180 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))**



**C.3.11 Probe 6, Skew 0 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))**



**C.3.12 Probe 6, Skew 180 (top: flaw depth 2.5 mm (0.10 in.), bottom: flaw depth 5.0 mm (0.20 in.))**







**Pacific Northwest**  
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

902 Battelle Boulevard  
P.O. Box 999  
Richland, WA 99352  
1-888-375-PNNL (7665)

[www.pnnl.gov](http://www.pnnl.gov)



Prepared for the U.S. Nuclear Regulatory Commission  
under a Related Services Agreement with the U.S. Department of Energy  
CONTRACT DE-AC05-76RL01830

U.S. DEPARTMENT OF  
**ENERGY**