

Physically Consistent Free-Hand Calibration for Polynomial Camera Models

P. Sperotto^{1,*}, B. Watz¹, D. Hess¹

¹: Dantec Dynamics A/S, Denmark

*Corresponding author: pietro.sperotto@dantecdynamics.com

Keywords: Calibration, Volumetric measurements, Multi-camera calibration.

ABSTRACT

Polynomial camera models, such as the Soloff formulation, are widely used in volumetric imaging and particle-based velocimetry due to their flexibility. However, they lack clear separation between intrinsic and extrinsic parameters, which makes classical calibration challenging due to the necessity of a traversing system. In this work, we introduce a free-hand multi-camera calibration pipeline that reproduces the physically consistent behavior of pinhole-based methods. While, preserving the expressive power of purely mathematical camera models. The approach is demonstrated on the Soloff model but is applicable to any projection function that can be expressed as a nonlinear basis $f(\mathbf{X})$ multiplied by a coefficient matrix $\mathbf{A}$. Both synthetic and experimental results confirm that the proposed formulation preserves geometric consistency.

1. Introduction

The third-order polynomial calibration, also known as the Soloff model (Soloff et al., 1997), is widely employed in experimental setups where the assumptions of ideal imaging conditions are not fulfilled. A typical example is the presence of phase changes along the line of sight of the cameras, where the standard pinhole calibration model becomes inaccurate. However, the Soloff approach requires the use of known 3D marker positions. In general, this requirement is fulfilled with the support of a mechanical traverse. In practice, traverses can be cumbersome to operate and prone to mechanical failure, often at critical times. Moreover, they may introduce systematic errors in the calibration for several reasons:

1. the calibration plate may not be perfectly normal to the translation direction;
2. the traverse span or plate size may be insufficient to cover the entire measurement volume;
3. each camera observes the calibration plate under nearly identical viewing angles, which can bias peak detection.

To overcome these limitations, we propose a method that enables Soloff calibration without the need for a traverse, following the same principles as a free-hand calibration. Although the focus of this work is on the Soloff model, the same methodology applies to any mathematical camera model, where the coefficients are linear. The details of the formulation are discussed in Section 2.

2. Framework

In this section, the general framework is proposed. The method described in 2.1 is fully developed and supported by preliminary synthetic evidence in this extended abstract.

2.1. Free-hand calibration

The proposed calibration procedure follows a pipeline analogous to that of the pinhole model. The main objective is to determine the rotations (in Rodrigues form) and translations corresponding to all calibration poses. The method can be applied to either a single camera or a multi-camera setup. The advantage of the latter is that pose estimation is enforced to be identical across all cameras, ensuring a consistent reconstruction of the 3D volumes. In the following, we describe the method for the multi-camera case, which is the primary focus of this work; the single-camera case follows directly.

To define the global reference frame with the first pose, which is assumed to lie in $z = 0$ with the point $0, 0$ coinciding with the origin. Hence, the translation and rotation of the first pose are zero. Under these assumptions, the total number of unknowns is $2N_p N_c + 6(n_v - 1)$, where n_v is the number of poses, N_c is the number of cameras and N_p is the number of parameters of the model, $N_p = 19$ for the Soloff model. The objective function to be minimized is:

$$\min_{\mathbf{A}_k, \mathbf{R}_j, \mathbf{T}_j} J(\mathbf{A}_k, \mathbf{R}_j, \mathbf{T}_j) = \sum_{k=1}^{N_c} \left(\|f(\mathbf{X}_{pl})\mathbf{A}_k - \mathbf{x}_{k,1}\|_2^2 + \sum_{j=2}^{n_v} \|f(\mathbf{R}_j \mathbf{X}_{pl} + \mathbf{T}_j)\mathbf{A}_k - \mathbf{x}_{k,j}\|_2^2 \right), \quad (1)$$

where:

- $f(\mathbf{X})$ is the polynomial basis function; in this work we employ the Soloff model;
- $\mathbf{A}_k$ is the coefficient matrix of $f(\mathbf{X})$ (for both x and y coordinates) of the k -th camera;
- $\mathbf{R}_j$ and $\mathbf{T}_j$ are the rotation matrix and translation vector of the calibration plate at the j -th view;
- $\mathbf{x}_{k,j}$ is the set of image coordinates (x, y) observed by the k -th camera at the j -th pose;

- $\mathbf{X}_{pl}$ denotes the reference calibration plate grid markers.

Because the coefficients $\mathbf{A}_k$ appear linearly, the problem can exploit the *Variable Projection* technique (Golub & Pereyra, 1973). Given a state vector $\mathbf{SV} = (\mathbf{R}_0, \mathbf{T}_0, \dots, \mathbf{R}_{n_v}, \mathbf{T}_{n_v})$, the evaluation of the cost function proceeds as follows:

1. Compute $\mathbf{X}_j = \mathbf{R}_j \mathbf{X}_{pl} + \mathbf{T}_j$;
2. Stack all positions to form $\mathbf{X}_{tot} = (\mathbf{X}_{pl}, \dots, \mathbf{X}_{n_v})$ and $\mathbf{x}_k = (\mathbf{x}_{k,1}, \dots, \mathbf{x}_{k,n_v})$;
3. Solve for the optimal coefficients:

$$\mathbf{A}_k = (f(\mathbf{X}_{tot})^T f(\mathbf{X}_{tot}))^{-1} f(\mathbf{X}_{tot})^T \mathbf{x}_k;$$

4. Evaluate the cost function in Eq. (1).

This cost function can be provided to any non-linear optimizer. The optimization variables are only the $6(n_v - 1)$ extrinsic parameters, which keeps the problem low-dimensional. In this work, numerical gradients were sufficient, although future implementations could exploit Jacobian sparsity for further speedup. As a practical initialization strategy, we use the extrinsic parameters obtained from a standard pinhole calibration (OpenCV implementation (Bradski, 2000)) as the initial guess. This choice serves as a fail-safe: if convergence issues arise in the multi-camera mode, the result remains at least as accurate as the pinhole baseline. We do not exclude that for very distorted setup using a set of randomized initial state or some kind of cost function sampling might be required.

3. Results

3.1. Synthetic test case

A synthetic experiment was designed to evaluate the geometric accuracy of the proposed calibration. Four real cameras with previously determined third-order polynomial (Slooff) models were used as the ground truth. Although this setup constitutes an “inverse crime” Wirgin (2004), it isolates numerical behavior and noise sensitivity. Further experimental results are proposed in the next section to address realism 3.2. The main objective here is to examine reprojection accuracy and spatial consistency.

A synthetic dual-plane calibration target of size $[-200, 200] \times [-100, 100] \times [0, 1]$ mm³ was generated, containing $20 \times 20 \times 2$ points. Synthetic noise was introduced to emulate typical experimental uncertainties:

1. manufacturing error on the plate, $\sigma = 1$ mm (Gaussian);

2. peak-detection error, $\sigma = 1$ pixel (Gaussian);
3. radial detection bias increasing from 0 to 0.1 pixel toward the image edges, due to barrel distortion.

Random translations and rotations were applied to generate fictitious poses, with the first view aligned to the ground-truth reference frame.

The calibration was performed in two modes: (i) single-camera calibration and (ii) multi-camera simultaneous calibration. For $n_v = 13$ poses, both approaches yielded similar *apparent* reprojection errors of approximately 1.27 pixels. Using the known ground-truth models, the *true* reprojection error was 0.070 pixels for the single-camera case and 0.053 pixels for the multi-camera case, while the maximum geometrical errors in the translations of the pose are 0.05mm and 0.026mm respectively. These values are not reported to claim superior absolute accuracy, since the synthetic experiment constitutes an inverse crime, but rather to demonstrate that the true reprojection error and the geometric consistency of the multi-camera formulation are inherently superior to those of the single-camera calibration. In particular, the maximum orientation and translation errors were approximately half in the multi-camera configuration, confirming that the reconstructed 3-D space exhibits significantly less warping and is therefore geometrically more consistent.

n_v	3	8	13	18
Single-camera [pixel]	0.108	0.065	0.070	0.071
Multi-camera [pixel]	0.097	0.051	0.053	0.039

Table 1. True reprojection error for different numbers of calibration poses n_v .

A stress test was also performed by introducing an exaggerated manufacturing error of 10 mm on one target point. In this case, the true reprojection error slightly increased to 0.169 and 0.182 pixels for the single- and multi-camera cases, respectively. However, the maximum translation error was about 0.16 mm for the single-camera calibration and only 0.04 mm for the multi-camera configuration. Considering that one voxel corresponds to roughly 0.25 mm, the multi-camera solution, although slightly worse in reprojection error, remains physically consistent and free from non-linear warping. This geometric consistency ensures that the lines of sight intersect at well-defined 3-D positions, a desirable property for subsequent volumetric reconstruction.

3.2. Experimental test case

An experimental campaign was conducted using three different types of calibration plates:

- dotted target;

- multi-level dotted target;
- checkerboard target;

The calibration was performed in a water basin instrumented with four FCX-5M cameras (2448×2048 pixels) equipped with 6.5 mm lenses. This configuration was purposely chosen to include complex optical paths and refraction effects. The reprojection errors obtained with the proposed approach were compared against those from single-camera pinhole calibration using OpenCV and our implementation of the multi-camera model for pinhole.

Camera	1	2	3	4
Dotted / Pinhole Single [pixel]	0.510	0.699	0.363	0.841
Dotted / Pinhole Multi-camera [pixel]	0.668	1.264	0.682	1.354
Dotted / 3 rd Single [pixel]	0.288	0.391	0.249	0.591
Dotted / 3 rd Multi-camera [pixel]	0.413	0.598	0.342	0.719
Multi-level / Pinhole Single [pixel]	0.381	0.661	0.726	0.401
Multi-level / Pinhole Multi-camera [pixel]	0.567	1.317	1.226	0.518
Multi-level / 3 rd Single [pixel]	0.361	0.504	0.660	0.293
Multi-level / 3 rd Multi-camera [pixel]	0.452	0.798	0.896	0.437
Checkerboard / Pinhole Single [pixel]	0.333	0.290	0.500	0.343
Checkerboard / Pinhole Multi-camera [pixel]	0.643	0.426	0.830	0.435
Checkerboard / 3 rd Single [pixel]	0.330	0.293	0.381	0.300
Checkerboard / 3 rd Multi-camera [pixel]	0.461	0.346	0.492	0.369

Table 2. Reprojection errors for different targets and calibration strategies.

Table 2 reports the reprojection errors for all configurations. In general, except for the dotted target, the multi-camera calibration shows higher reprojection error than the single camera models counterpart. However, as discussed previously, the reprojection error alone is not a sufficient indicator of geometric accuracy. The best-performing method in all cases is the third-order single-camera calibration, consistent with expectations since it possesses a higher number of degrees of freedom per view.

It is noteworthy that the dotted target is the only configuration in which the multi-camera calibration outperforms the pinhole model. This is not coincidental: the dotted target provides an order of magnitude more detectable points per view compared with the checkerboard and multi-level patterns, resulting in a substantially higher ratio of data points to optimization variables. It is important to emphasize that the pinhole calibration is performed in single-camera mode; therefore, the directly comparable reference is the third-order single-camera calibration, which indeed consistently exhibits the lowest reprojection error. Achieving a similar data density with a checkerboard pattern would require approximately ten times more views, which would, in turn, proportionally

increase the number of optimization variables. This indicates that the dotted target is the only configuration in which the pinhole model begins to show mild underfitting due to the much larger number of detected peaks, while the polynomial model continues to exploit the additional data to improve spatial accuracy. For this reason, the higher reprojection error observed in the multi-camera calibrations should not be interpreted as poorer performance, as the shared-geometry constraint inherently reduces the available degrees of freedom compared with both the pinhole and the third-order single-camera formulations.

3.3. Tomographic reconstruction

As a final test, we set up a simple experimental configuration using four FCX2-5M (2448×2048 pixels) cameras with 8 mm lenses arranged in a square. The flow is generated by a forced shedding jet from a pump nozzle in a basin filled with water. A 200×200 mm dotted calibration target is used.

It is worth noting that, in order to introduce an error of approximately one voxel (≈ 0.11 mm) in the z -direction using a mechanical traverse, the calibration plate would need to be misaligned by only $\alpha = \arcsin(0.11/200) \simeq 0.03^\circ$. Such an alignment tolerance is practically impossible to guarantee in a real experiment, highlighting one of the limitations of traverse-based calibration procedures.

In this experiment, we use a simple minimum Line Of Sight (minLOS) reconstruction method. This choice is intentional: minLOS provides limited suppression of ghost particles and is therefore useful to test the calibration close to the edge of functionality of the reconstruction method. Furthermore, no calibration refinement routine is applied, in order to evaluate the calibrations as directly and as raw as possible.

Two test cases are considered: one at a lower seeding density of approximately 0.03 ppp, and one at a higher seeding density of circa 0.06 ppp. The second case is intended as a stress test, since line-of-sight-based reconstruction methods such as MLOS and minLOS are generally expected to be applicable only at relatively low seeding densities, and their performance deteriorates rapidly as the number of ghost particles increases Michaelis et al. (2010); Atkinson et al. (2014).

Regarding the first experiment, minLOS is performed using all four calibrations reported in Table 3. The corresponding maximum reprojection errors are reported in Table 3.

3^{rd}	0.494
3^{rd} multi	0.615
Pinhole	0.681
Pinhole multi	1.264

Table 3. Maximum reprojection error for the calibrations used in the minLOS reconstruction.

In this case, we evaluate the averaged intensity distribution in the reconstructed voxel space and the qualitative structure of the reconstruction. Figure 1 shows the XY -averaged and normalized intensity as a function of Z . In general, the Soloff model produces a lower signal-to-noise ratio outside the illuminated volume. On the other hand, the multi-camera-based methods are able to identify the main illuminated region, whose expected size is approximately 10–13 cm.

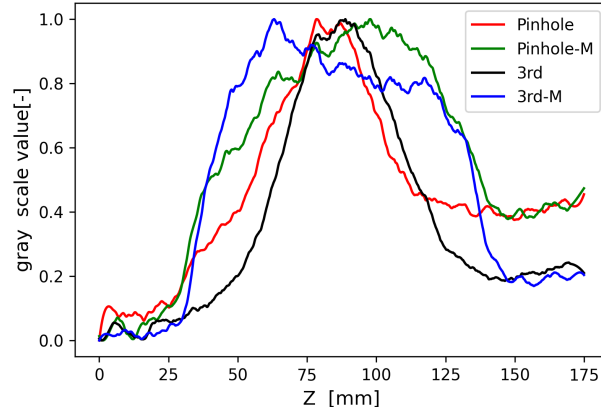


Figure 1. XY -averaged intensity.

Based only on this plot, the two multi-camera models may appear to perform similarly. However, this is not the case when considering the Y -averaged reconstruction shown in Figure 2. The multi-camera pinhole calibration produces regions at low X where the reconstructed signal is either absent or indistinguishable from the background noise level. In contrast, the third-order multi-camera calibration provides a cleaner and sharper boundary, comparable with the expected width of the light source.

It should be emphasized that Figure 2 represents only a small portion of the voxel space. If the full measurement volume were reconstructed using the multi-camera pinhole calibration, some regions would fail completely, leaving empty zones in the reconstructed volume. This behavior suggests that, although the multi-camera pinhole calibration may satisfy the imposed shared-geometry constraint, its camera model is not sufficiently expressive to preserve the physical consistency required by this optical configuration.

As a qualitative test, we perform PTV on the high-density case, corresponding to approximately 6% ppp, or about 170 thousand reconstructed particles per time step. This density is intentionally very demanding. minLOS-type reconstructions are more commonly used in a Tomographic PIV context, where ghost particles mainly contribute as uncorrelated reconstruction noise to a subsequent volumetric correlation or least-square matching. Their use for PTV is considerably more restrictive, since ghost particles can be interpreted as individual particle candidates and may therefore be linked into false trajectories. For this reason, the high-density PTV result should not be interpreted as a ground-truth validation of particle-level accuracy.

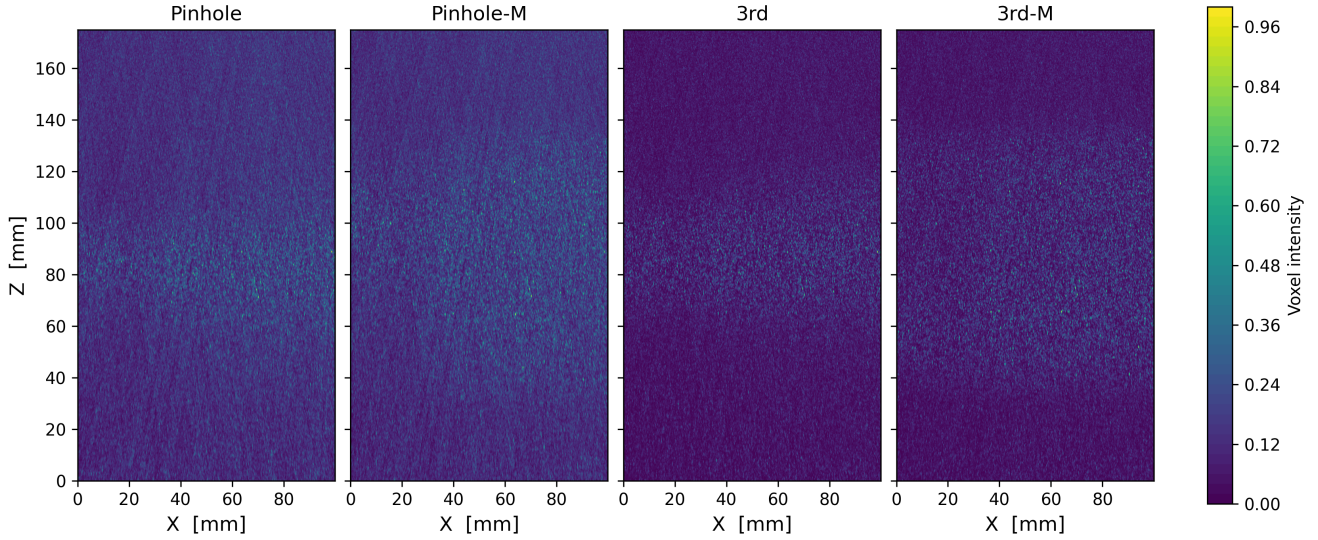


Figure 2. Y -averaged intensity.

Indeed, in the absence of ground-truth particle positions, we cannot determine how many reconstructed trajectories correspond to real particles. At this seeding density, a significant fraction of ghost tracks is expected. Tracks shorter than 9 time steps were removed; however, this criterion alone does not guarantee that the remaining trajectories correspond to physical particles. Nevertheless, the result is meaningful as a practical stress test of the calibration. At such high ppp, and using a reconstruction method with limited ghost suppression, a geometrically inconsistent calibration would be expected to produce diffuse noise, fragmented trajectories, or empty regions rather than coherent and temporally persistent track-like structures.

Therefore, the significance of this experiment is not that every reconstructed track is necessarily correct, but rather that the proposed third-order multi-camera calibration preserves sufficient multi-view geometric consistency to avoid a catastrophic collapse of the reconstruction. Even if a substantial number of tracks are ghosts, the emergence of coherent trajectories at this seeding density indicates that the calibrated camera models remain mutually consistent under conditions close to, or beyond, the practical operating limit of minLOS-based particle reconstruction.

As a final conclusive test, the tracking procedure was repeated over a subset of 51 time steps using both the multi-camera pinhole and the 3^{rd} order multi-camera calibration. In this case, all trajectories longer than 4 time steps were retained. The probability density functions (PDFs) of the track length and of a trajectory smoothness metric based on the depth-wise jerk are reported in Figure 3 and 4.

In general, the 3^{rd} order multi-camera calibration produces trajectories that are both longer and smoother than those obtained with the pinhole multi-camera model. It is important to emphasize

that the histograms are normalized by the total number of tracks in each case, allowing a direct comparison of the distribution shape.

It is also worth noting that the number of reconstructed particles per time step differs significantly between the two calibrations: approximately 3.34×10^5 particles for the 3^{rd} order model, and 7.76×10^5 for the pinhole model. While both values exceed typical expectations at this seeding density, the pinhole calibration produces a substantially larger number of particle candidates. This suggests that a significant portion of these additional particles is not associated with temporally persistent or kinematically consistent trajectories.

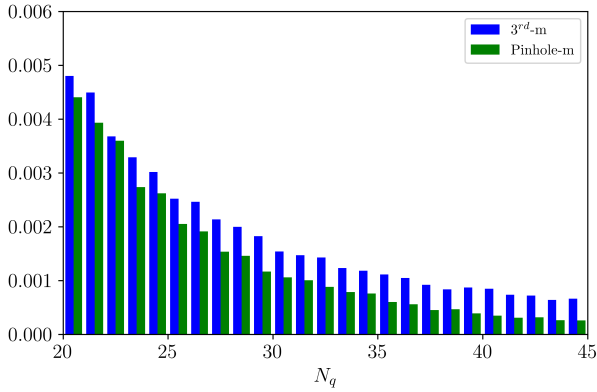


Figure 3. Probability density function of track length.

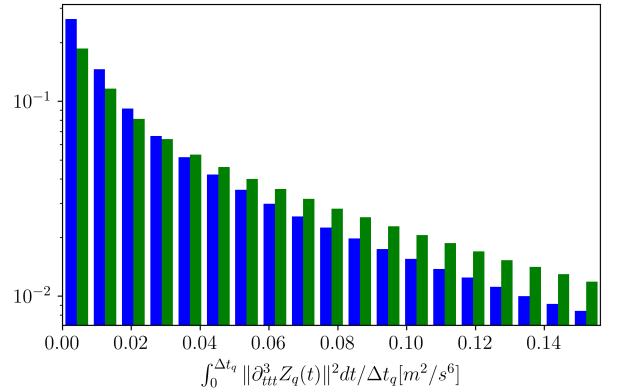


Figure 4. Probability density function of the average depth-wise jerk energy per track.

These results do not constitute a proof of correctness of the reconstructed particle trajectories. In the absence of ground truth, a significant fraction of ghost tracks is expected at such high seeding density. Rather, this analysis provides a direct comparison between two calibration strategies under identical reconstruction and tracking conditions.

The results consistently indicate that the pinhole multi-camera calibration produces shorter and less smooth trajectories, despite reconstructing a larger number of particles. In contrast, the 3^{rd} order multi-camera calibration yields trajectories that are more temporally persistent and kinematically consistent. Since the reconstruction and tracking procedures are identical in both cases, these differences can be attributed solely to the calibration. This supports the conclusion that the polynomial multi-camera model preserves a higher degree of multi-view geometric consistency in this experimental configuration.

4. Conclusions

This work introduces a generalized free-hand calibration pipeline aimed at improving the flexibility and practical usability of polynomial camera models while preserving geometric consistency

across multiple views. The proposed framework extends concepts commonly used in pinhole-based calibration to purely mathematical projection models, such as the Soloff formulation, enabling free-hand calibration without the need for a mechanical traverse.

Across the synthetic and experimental tests considered in this work, the calibration procedure remained computationally inexpensive, with calibration times below one minute in all cases. The results indicate that the proposed method can recover satisfactory calibrations and, in challenging optical configurations, can provide more physically consistent volumetric reconstructions than a pinhole-based model with comparable practical flexibility, cost, and ease of use. This is particularly relevant because the pinhole model represents the most direct competitor for free-hand calibration in terms of experimental simplicity.

A key conclusion of this work is that, when purely mathematical camera models are used for volumetric measurements, a multi-camera calibration formulation is not merely beneficial but essential. The flexibility of polynomial models allows them to represent complex optical paths, but this same flexibility can lead to non-physical solutions if each camera is calibrated independently. Enforcing shared geometry across all views provides the additional constraints required to obtain physically meaningful volumetric reconstructions.

Future work will focus on a direct comparison between the proposed free-hand polynomial calibration and classical traverse-based calibration. Such a comparison is not straightforward, since the result may depend strongly on the experimental procedure, the order in which the calibrations are performed, the mechanical accuracy of the traverse, and the operator-dependent alignment of the calibration target. A fully objective validation would ideally require an independent high-accuracy three-dimensional positioning system. Nevertheless, the present results suggest that the proposed approach is a promising alternative to traverse-based calibration, particularly in configurations where mechanical alignment errors, limited traverse accuracy, or complex optical paths may compromise the physical consistency of the reconstructed volume.

References

- Atkinson, C., Buchmann, N., Stanislas, M., & Soria, J. (2014). Adaptive mlos-smart improved accuracy tomographic piv. In *International conference on multiphase flow*.
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Golub, G. H., & Pereyra, V. (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on numerical analysis*, 10(2), 413–432.
- Michaelis, D., Novara, M., Scarano, F., & Wieneke, B. (2010). Comparison of volume reconstruction techniques at different particle densities. In *International symposium on applications of laser techniques to fluid mechanics*.
- Soloff, S. M., Adrian, R. J., & Liu, Z.-C. (1997). Distortion compensation for generalized stereoscopic particle image velocimetry. *TAM R* 855.
- Wirgin, A. (2004). The inverse crime. *arXiv preprint math-ph/0401050*.