Received: 28 July 2020 | Revised: 13 November 2020 | Accepted: 11 December 2020

DOI: 10.1002/qj.3964

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

SAD: Verifying the scale, anisotropy and direction of precipitation forecasts

Sebastian Buschow | **Petra Friederichs**

Institute of Geosciences, University of Bonn, Germany

Correspondence
S. Buschow, Institute of Geosciences, University of Bonn, Bonn 53113, Germany.
Email: sebastian.buschow@uni-bonn.de

Funding information
German Research Foundation (DFG), Grant/Award Number: FR 2976/2-1

Abstract

One important attribute of meteorological forecasts is their representation of spatial structures. While several existing verification methods explicitly measure a structure error, they mostly produce a single value with no simple interpretation. Extending a recently developed wavelet-based verification method, this study separately evaluates the predicted spatial scale, orientation and degree of anisotropy. The scale component has been rigorously tested in previous work and is known to assess the quality of a forecast similar to other, established methods. However, directional aspects of spatial structure are less frequently considered in the verification literature. Since important weather phenomena related to fronts, coastlines and orography have distinctly anisotropic signatures, their representation in meteorological models is clearly of interest. The ability of the new wavelet approach to accurately evaluate directional properties is demonstrated using idealized and realistic test cases from the MesoVICT project. A comparison of precipitation forecasts from several forecasting systems reveals that errors in scale and direction can occur independently and should be treated as separate aspects of forecast quality. In a final step, we use the inverse wavelet transform to define a simple post-processing algorithm that corrects the structural errors. The procedure improves visual similarity with the observations, as well as the objective scores.

KEYWORDS

MesoVICT, precipitation forecasts, structure error, verification, wavelets

1 | INTRODUCTION

The errors of modern weather forecasts can take many different forms. While everyday users may only notice that their weather app failed to predict rainfall at a specific point in space and time, such a mistake could have several possible meanings. Perhaps the coherence of a frontal precipitation band was underestimated. The likelihood of convective initiation in a certain region may have been misjudged. Alternatively, the simulated pattern was perfectly adequate but its spatial location was wrong. In some cases, the precipitating process is so small and short-lived that no present-day weather model could be expected to foresee its exact timing and placement.

In order to obtain useful diagnostic information on the merits of highly resolved simulations, many forecast verification tools aim to separate the various types of error from one another. Most prominently, spatial

displacements tend to mask all other kinds of error in a point-wise evaluation. To tackle this issue, a multitude of so-called "spatial" verification techniques have been developed throughout the last two decades. A first intercomparison of these methods was undertaken within the intercomparison project (ICP; Gilleland *et al.*, 2009), which classified the various approaches and attempted to elucidate their differences and similarities using a set of standardized test cases. The Mesoscale Verification Intercomparison over Complex Terrain (MesoVICT; Dorninger *et al.*, 2018), launched in 2014, constitutes the second phase of the ICP and focuses on the effects of uneven terrain and uncertain observations and considers forecasts of both precipitation and wind.

This study participates in MesoVICT by using both the realistic test cases of Dorninger *et al.* (2018) and the recently presented geometric tests of Gilleland *et al.* (2020). Our focus lies on isolating and understanding errors in the predicted spatial structure of quantitative precipitation forecasts. Using a two-dimensional wavelet transform, we want to separately determine whether the predicted structure was (a) too small or too large, (b) too directed or too round, and (c) oriented along the correct angle.

Several popular methods from the rich spatial verification literature have previously been used to determine a "structure" error. Using the field deformation technique of Keil and Craig (2007), Han and Szunyogh (2016) approximately corrected the forecast's location and intensity and referred to the residual error as "structural". While straightforward and intuitive, this kind of approach yields no further information on how exactly the pattern was mis-forecast. Furthermore, it should be noted that any field deformation approach which allows for a divergent optical flow will be sensitive, simultaneously, to errors in both the spatial scale and anisotropy and therefore cannot truly separate structure from location.

A more intuitive notion of structural disagreement can be obtained using object-based methods that decompose the fields into features and measure their individual properties. Such techniques are typically adapted to the special case of precipitation forecasts where well-defined discrete objects are known to exist. The popular SAL method of Wernli *et al.* (2008) defines its structure component S via the ratio between total and maximum precipitation in each object. The resulting score is related to the size and number of objects as well as the tail behaviour of the marginal distribution; directional aspects are neglected by S. Interestingly, the relative placements of the individual objects are also not included in S. It is clear that a number of small features in close proximity to one another can form a large-scale structure, perhaps driven by a single meteorological process. Such fractured pattern may result from the driving process itself, the complex terrain in which it occurs, or, in case of observational data, the measurement technique. SAL does not consider this as an element of structure, but instead includes the relative placement of the object in the location component (Wernli *et al.*, 2008). A further related drawback of this otherwise useful technique is its potential sensitivity to the details of the object identification algorithm (Weniger and Friederichs, 2016).

The popular Method for Object-based Diagnostic Evaluation (MODE; Davis *et al.*, 2006) provides a wide framework in which numerous structural properties such as feature size, aspect ratio, aspect angle and even curvature can be evaluated. Like SAL, it is mostly adapted to precipitation, can be sensitive to the object-defining algorithm and does not simply allow multiple features to form an organized super-structure on larger scales. More generally, these techniques are inherently *single-scaled*. If objects are detected by smoothing the field with a kernel of size σ and thresholding at a value T, then the pair (σ, T) defines a spatial scale: larger values of T isolate smaller intense regions, larger values of σ lead to the union of increasingly distant features into single, larger objects. If we calculate, for example, the average aspect ratio of the objects detected in this manner (following Davis *et al.*, 2006), the result is characteristic of the anisotropy on the scale defined by (σ, T). Re-arranging the objects in space (larger-scale variability) or re-arranging the pixels within an object (smaller-scale variability) leave the result of the analysis unchanged, since the aspect ratio depends only on an object's shape, not its position or internal composition.

Avoiding such pitfalls of the object-based methods, several authors have based their structural verification on indirect estimates of the spatial correlation structure. Marzban and Sandgathe (2009), Scheuerer and Hamill (2015) and Ekström (2016) rely on empirical variograms for this purpose. The latter study in particular achieves an intuitive notion of the predicted and observed spatial scale by estimating the variogram's range. Without defining a verification score, Willeit *et al.* (2015) study the climatological structure of forecast precipitation fields using Fourier transforms. Wong and Skamarock (2016) extended this work using directional information from the 2D Fourier spectra.

A similar direction was pursued in Buschow *et al.* (2019). Building on the work of Kapp *et al.* (2018), they used a two-dimensional redundant discrete wavelet transform (RDWT) to analyze the scale on which spatial variability occurs, both globally and at each grid point. The resulting scores effectively isolate scale errors in idealized (Buschow *et al.*, 2019) as well as realistic situations (Buschow and Friederichs, 2020). A first analysis of anisotropy, using the same kind of wavelet transform, was undertaken by Brune *et al.* (2018) who included a measure

of anisotropy as one component of their wavelet-based convective organization index (WOI). Following the same approach, Brune *et al.* (2020) furthermore attempted to infer the local direction in precipitation fields from the relative contributions of horizontal and vertical features to the total variability. However, as discussed in Buschow and Friederichs (2020), the classic discrete wavelet transform has inherent shortcomings in its representation of directional structures which make it especially unsuitable to the task of forecast verification. With this transform, one could rotate forecast and observation by the same angle and receive a very different score.

In the context of image processing, the same issues were long ago recognized by Kingsbury (1999), who addressed the problem using complex-valued wavelets. Their so-called dual-tree complex wavelet transform (DTCWT) forms the new basis for our verification method. A first meteorological application of this technique was recently presented by Scovell (2020) who used it to incorporate anisotropy into a stochastic noise generator for precipitation nowcasting. Nerini *et al.* (2017) pursued a similar route in their application of a localized Fourier transform to the task of reproducing non-stationary, non-isotropic rainfall variability.

After introducing the relevant datasets in Section 2, we explain in Section 3 why the original RDWT is unsuited to analyze directions and how the DTCWT solves the problem. The next step (Section 4) is to extend the idea of a central scale (Buschow *et al.*, 2019) to include anisotropy and direction. This leads to the definition of new wavelet-based structure scores in Section 5. Experiments with geometric test patterns (Section 6) and the realistic MesoVICT forecasts (Section 7) demonstrate that the new wavelets allow for the same kind of sensitive scale-verification as their predecessors. In addition, they yield valuable information on the forecast's degree of anisotropy and predominant orientation. Section 8 uses the inverse DTCWT to define a simple algorithm for correcting the structural errors detected by our approach. The algorithm is tested on individual forecasts as well as the dataset as a whole. We discuss the outcomes of all experiments in Section 9.

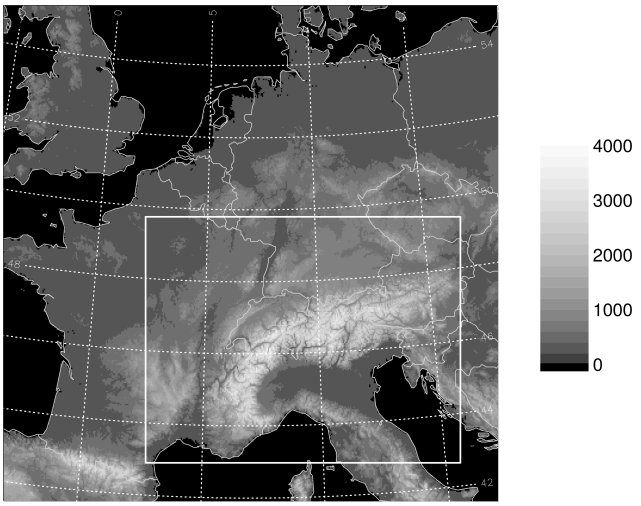


FIGURE 1 VERA orography in metres and minimum common domain for the MesoVICT dataset (white rectangle)

TABLE 1 MesoVICT cases, reproduced from Dorninger *et al.* (2018)

Case	Date	Weather event
1	20–22 June 2007	Strong convective activity north of the Alps followed by a cold front
2	18–21 July 2007	Strong convective events across an airmass boundary impinging the Alps from northwest
3	25–29 September 2007	Accelerating cold front north of the Alps and cyclogenesis in the Gulf of Genoa with heavy rains south of the Alps (Venice flood)
4	6–8 August 2007	Squall line ahead of cold front crossing the Alps causing widespread thunderstorms
5	18 September 2007	Cold front crossing the Alps causing severe thunderstorm in Slovenia
6	8–10 July 2007	Subtropical air mass advected into the Alpine region causing widespread thunderstorms

2 | DATA

In contrast to the largely homogeneous Great Plains considered in the ICP, the study area of the MesoVICT project focuses on a small, mountainous region surrounding the European Alps (Figure 1). Six case-studies on interesting weather situations in summer and autumn 2007 were selected (Table 1, reproduced from Dorninger *et al.* (2018)). Gridded analysis data of precipitation and wind are provided by the Vienna Enhanced Resolution Analysis (VERA; Bica *et al.*, 2007) which incorporates station observations and topographic information but no data from numerical models. While a variety of forecast datasets are in principle available within MesoVICT, we focus on four deterministic models which cover the entire region for all cases: the Swiss COSMO (initialized at 0000 UTC), CMH from the Canadian weather service (initialized at 0600 UTC) as well as BOLAM007 and MOLO0225 from the Institute for Environmental Protection and Research

TABLE 2 Summary of all used datasets

Abbreviation	Organization	Lead times	Δx	Citations
CMH	Environment Canada	+1 hr … +24 hr	2.5 km	McTaggart-Cowan (2009)
COSMO	MeteoSwiss	+6 hr … +24 hr	2.2 km	Ament and Arpagaus (2009)
BOLAM007	ISPRA	+12 hr … +35 hr	0.07°	Mariani and Casaioli (2018)
MOLO0225	ISPRA	+12 hr … +35 hr	0.0225°	Mariani and Casaioli (2018)
VERA	University of Vienna	—	8 km	Bica *et al.* (2007)
RADKLIM	DWD	—	1 km	Winterrath *et al.* (2018)

(ISPRA, both initialized at 1200 UTC, first 12 hr discarded). To avoid obvious artifacts of model spin-up, we consider only time steps from 0700 UTC to 2300 UTC. Further details and the references for each model are given in Table 2.

In Section 7.2, we move beyond the domain of the MesoVICT project and validate VERA, BOLAM007 and MOLO0225 against the gauge-adjusted radar climatology RADKLIM of Winterrath *et al.* (2018). The dates and other characteristics of the datasets remain the same, but the domain for this experiment is defined by the German national borders. Missing pixels in RADKLIM (outside Germany or due to radar failures) are set to zero in all fields to ensure comparability.

Hourly rain sums for all forecasts have been interpolated to the VERA grid at a common resolution of approximately 8 km. After cropping the data to the core regions where all datasets have non-missing values, we obtain 133×88 grid points for the Alpine domain and 99×116 for Germany. These regions are symmetrically extended to 256×256 because our implementation of the wavelet transform requires the input dimensions to be whole powers of two. Buschow and Friederichs (2020) discuss several possible ways of handling these boundary conditions and conclude that, in theory, reflective boundaries are the most elegant and appropriate solution. However, this approach may no longer be viable because we are interested not only in the scale but also in the direction and anisotropy detected by the wavelets. It is easy to imagine situations where the latter two properties are distorted, when we reflect the input image at the edges. To avoid such effects, we pad the fields with zeros instead. Following Kapp *et al.* (2018), we linearly decrease the original values to zero across ten pixels along each side in order to smooth out potential artificial edges.

Rain values below 0.1 mm are set to 0 mm, then all values x are replaced by $\log_2(x + 0.1)$ before the wavelet transform is applied. Buschow and Friederichs (2020) discuss the rationale behind this step in detail. Simply, typical plots of rain fields use logarithmic or similar colour scales in order to visualize both local extreme events and extended areas of moderate intensities. Similar ideas apply to our automatic analysis of spatial structure. The results of the wavelet transform are generally easier to understand if they are based on the same data transformation as the plots used for visual inspection.

3 | THE DUAL-TREE COMPLEX WAVELET TRANSFORM

This section introduces the basics of discrete wavelet transforms in a very concise manner. To readers who are completely new to wavelets, we recommend Torrence and Compo (1998) and Weniger and Friederichs (2016) for an introduction in a meterological context, as well the general textbook of Daubechies (1992).

Classic wavelet transforms start by selecting a function $\psi(r)$, $r \in \mathbb{R}^n$, which is localized in both space and frequency and integrates to zero over its domain of definition. From this so-called mother wavelet, a set of daughter wavelets are derived via shifting and re-scaling, i.e., $\psi_{s,u}(r) = s^{-1/2}\psi((r-u)/s)$. In a multidimensional space $\mathbb{R}^n$ with $n > 1$, the daughters can furthermore have various spatial orientations, which we will denote by the index d. A signal (a time series for $n = 1$, an image for $n = 2$) is projected onto the $\psi_{s,u(,d)}$, thereby decomposing it into components with specific scales and (in 2D) directions. This so-called *wavelet transform* is similar to the well-known Fourier transform except that the basis functions are also localized in time (or space in 2D), which allows for the correct treatment of non-stationary signals.

The bottom row of Figure 2 shows the three directional daughter wavelets associated with the two-dimensional discrete wavelet transform (DWT). Their mother is the least asymmetric Daubechies wavelet with six vanishing moments. By shifting several scaled versions of these directed, localized wave-forms across an image, one can localize and study features of various spatial scales and orientations. Looking at the diagonal daughter (45° in the bottom row of Figure 2), it becomes obvious that all attempts at deriving direction and anisotropy from these basis functions are flawed: while the vertical and horizontal daughter wavelets are rotated versions of one another,

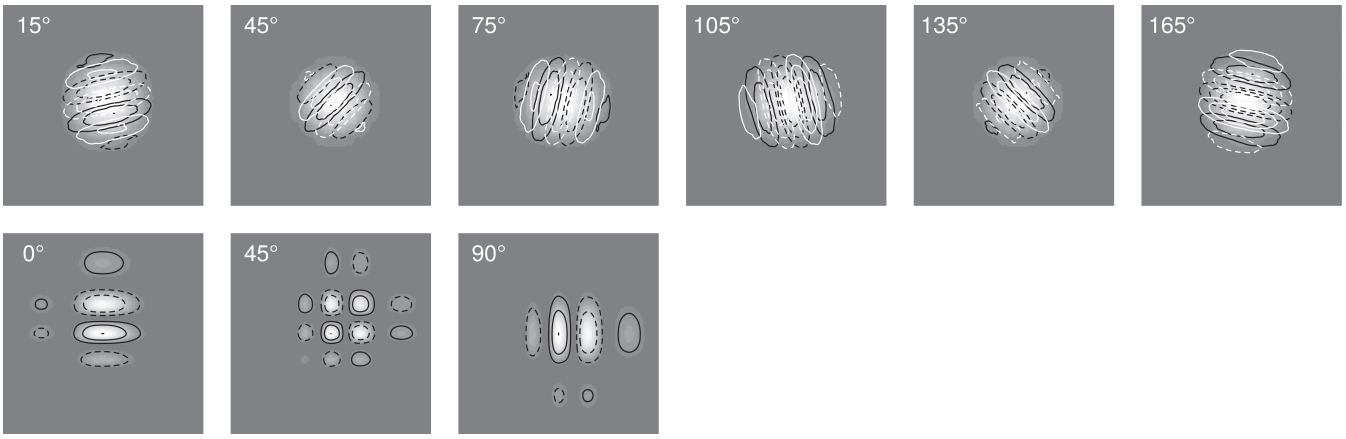


FIGURE 2 Directed daughter wavelets of the DTCWT (top row) and the regular DWT ("extremal phase" Daubechies wavelet number 6, bottom). Solid and dashed lines indicate positive and negative values, respectively, the background shading indicates the absolute amplitudes squared. White contours correspond to the imagery part of the DTCWT daughters

the daughter for 45° clearly lives on a smaller scale and cannot distinguish between the two diagonals. Using this wavelet transform, it is thus impossible to decide whether an edge is oriented at +45° or −45°. The degree to which one of the three directions dominates over the others, that is, the estimated anisotropy of a given feature, furthermore depends on the scale and the orientation of that feature.

To understand the origin of (as well as the solution to) these undesirable effects, we must briefly discuss the algorithms by which wavelet transforms are implemented. In principle, one could convolve the signal with each scaled and oriented daughter wavelet individually. This procedure is used in continuous wavelet transforms (CWTs), which allow arbitrary scales s and involve a high degree of redundancy at high computational costs. A far more efficient algorithm was introduced by Mallat (1989), paving the way for innumerable modern wavelet methods: instead of defining a continuous function ψ, the mother wavelet is represented by a finite set of filter coefficients $g_{1,\dots,n}$. Next, the so-called father wavelet $\phi(x)$ is defined by the filter coefficients $h_k = (-1)^k g_{n-1-k}$. The father wavelet is thus a reversed version of the mother, where the sign of every second coefficient has been flipped. Loosely speaking, mom performs a differentiation (high-pass) while dad is an averaging (low-pass) filter.

In one dimension, the so-called discrete wavelet transform (DWT) is then implemented by (1) convolving the signal with g and discarding every second value from the result to obtain the wavelet coefficients at the finest scale, (2) convolving the signal by h (again discarding half of the values) to obtain the input for the next coarser scale, and (3) repeating (1) and (2) until only a single value remains. By dropping every second value, we effectively shift the smallest-scaled daughter wavelet ($s = 2^0$) to every second location in the domain, the next larger one ($s = 2^1$) to every fourth, and so on, thereby removing the redundancy. The scales of this so-called "decimated" transform are no longer continuous but whole powers of two, and the larger wavelets are shifted to fewer locations. As a result, the daughter wavelets of this transformation form an orthogonal basis. The algorithm can be adapted to obtain values at all possible locations by simply not discarding any values and instead inserting zeros between the filter coefficients h_k and g_k after each level, resulting in a redundant discrete wavelet transform (henceforth RDWT).

The efficient DWT algorithm has a straightforward extension to higher dimensions which wasalso introduced by Mallat (1989) (shown in Figure 3). Given a two-dimensional matrix of input values, convolve the rows with h and then the columns with g to obtain the vertical daughter coefficients. The horizontal daughter coefficients result from applying g to the rows and h to the columns; the diagonal daughter is the product of applying g to both rows and columns. Application of h in both directions gives the input for the next coarser scale. This procedure, which can be implemented with decimation or redundancy just as in the 1D case, generates the three directional daughters seen in Figure 2. That explains the reduced scale of the diagonal wavelets (being the product of two high-pass filters), as well as the absence of a fourth filter for the other diagonal.

Recognizing the shortcomings of the classic DWT, Kingsbury (1999) introduced the so-called dual-tree complex wavelet transform (henceforth DTCWT). Instead of a single real-valued mother, they defined a complex-valued $\psi = \psi_r + i\psi_i$ with corresponding filters h_i, h_r, g_i, g_r. The two mother wavelets ψ_r and ψ_i are each other's Hilbert transform, meaning that they are 90° out of phase with each other. In two dimensions, the complex transform can be implemented by performing four regular DWT s (as in Figure 3) with all possible combinations of $h_{i,r}, g_{i,r}$

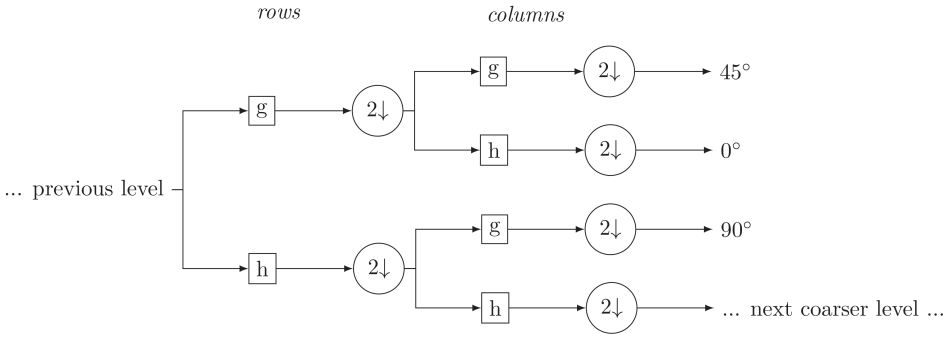


FIGURE 3 One level of the two-dimensional discrete wavelet transform (dwt). g and h denote applications of the high- or low-pass filter, respectively. "2↓" signifies a down-sampling step

applied to the rows and columns. The twelve resulting sets of coefficients are then re-combined into six complex directional daughter wavelet coefficients. A set of corresponding daughter wavelets is shown in the top row of Figure 2. Each of these functions represents one distinct direction, and the two diagonals are no longer ambiguous. Here, we have furthermore applied the method of Kingsbury (2006) to obtain optimized diagonal wavelets with very nearly the same scale as their sisters. The two issues preventing us from utilizing the directional information of the wavelet transform are thus resolved. The absolute values, shown as background shading in Figure 2, reveal a further advantage of this transform. While real and imaginary parts both constitute wave forms, the Hilbert property means that the modulus monotonously decreases from the centre of the support. Image features can thus unambiguously be located within the support of each daughter wavelet – a task which is less straightforward for the Daubechies wavelets. As a final benefit, Selesnick *et al.* (2005) report that the complex nature of the coefficients greatly reduces the shift dependence of the transform. While for the regular DWT, we must always rely on the computationally more expensive redundant transform, we can obtain robust information on the global structure of a field from the decimated DTCWT as well. We demonstrate the effective equivalence of the two transforms in Appendix B.

Regardless of the merits of the decimated DTCWT, it does not deliver fully localized information because the large-scale coefficients are only available on increasingly coarse grids due to the downsampling (cf. Figure 3). If we are interested in local characteristics at every location, a fully redundant transform is needed. In this case, an over-emphasis on very large scales, caused by their great redundancy (large overlapping areas), must be avoided. Here, we follow Kapp *et al.* (2018), Brune *et al.* (2018) and Buschow *et al.* (2019) and rely on the theory of locally stationary wavelet processes (Eckley *et al.*, 2010) to remove this large-scale bias. In a nutshell, it can be shown that the squared local wavelet coefficients have a well-defined relationship with the spatial covariances if we multiply them by a bias-correction matrix which depends on the domain size and choice of mother wavelet. This step mostly reduces the values of large-scale coefficients and re-distributes their energy to smaller scales. The theory was extended to the redundant DTCWT by Nelson *et al.* (2018). Following Buschow *et al.* (2019), any negative "energy" values introduced by the bias correction are set to zero. Based on the discussion in Buschow and Friederichs (2020), we furthermore discard the three largest scales due to their ambiguous localization (basis functions being larger than the entire domain).

4 | ANALYZING SCALE AND DIRECTION

In order to compactly summarize the output of the wavelet transform, Buschow *et al.* (2019) studied the central scales of the wavelet spectra. Let e_j be the (bias-corrected) squared wavelet coefficient for scale $j \in \{1, \dots, J\}$, averaged over all directions. Now consider the e_j as point-masses, located along a line at the coordinates $z_j = j$. The central scale of the wavelet spectrum is then defined as the centre of mass z_c of that arrangement. A plot of these central scales for each pixel of an input image compactly visualizes the result of the wavelet analysis by showing the dominant scale at each location. Buschow *et al.* (2019) demonstrated how this map of central scales can also serve as the basis for spatial verification.

We now extend the idea to the case of the directional wavelet spectra produced by the DTCWT. Noting that the energy of the 15° daughter wavelet should be next to those for 45° and 165°, both being a 30° rotation away (Figure 2), the natural geometry in which to arrange the $6 \times J$ coefficients is a prism with hexagonal base. Figure 4 schematically shows this arrangement. The energies for the six directions are placed along the vertices of a regular hexagon, parallel to the x–y plane. The various scales j correspond to different values of the z coordinate. Indexing

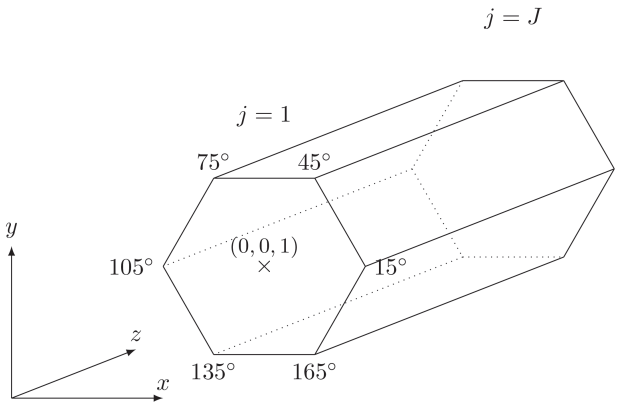


FIGURE 4 Geometry used to define the central scale, radius and angle of the DTCWT spectra

the directions by d, the coordinates for the value $e_{j,d}$ are

$$\begin{aligned} x_{j,d} &= a\cos\{60(d-1)\pi/180\}, \\ y_{j,d} &= a\sin\{60(d-1)\pi/180\}, \\ z_{j,d} &= j\,, \end{aligned} \tag{1}$$

where a denotes the arbitrarily fixed circumradius of the hexagon. Calculating the centre of mass in this geometry leads to the same central scale z_c as before. The other two central components x_c and y_c contain information on the preferred direction and degree of anisotropy. We can easily separate these two properties by transforming from the x–y plane to polar ρ, θ coordinates. The central radius $\rho_c = \sqrt{x_c^2 + y_c^2}$ then measures the total degree of anisotropy, averaged over all scales. From the central angle $\theta_c = \arctan 2(y_c, x_c)$, we can derive the angle in image space as $\varphi_c = 15° + \theta_c/2$. Note that a is merely a multiplicative factor dertermining the scale of ρ. A more detailed example of these ideas is discussed in Appendix A.

5 | DEFINITION OF SCORES

Buschow *et al.* (2019) introduced the structure score *semd*, which is given by the earth mover's distance (EMD; Rubner *et al.*, 2000) between two direction-averaged spatial mean spectra: the energy e_j of scale j is considered a point mass located at the position $z = j$ along the real line. *semd* measures the minimum cost of transporting all energy from one spectrum to another. Both spectra are normalized to unit sum, making the EMD a true metric. The EMD was preferred over other metrics because it appropriately measures shifts in the spectra as well as differences in their shape.

Based on the ideas from Section 4, the extension of *semd* to the case of directed spectra is very straightforward. Simply place the energies $e_{j,d}$ corresponding to the scales j and directions d at the corresponding vertices of the hexagon (Figure 4) and solve the transport problem to obtain the EMD. We will refer to this directed version of *semd* as $semd_d$. The radius a in Equation 1 i.e., the ratio between width and length of the prism *within which the centre resides* (cf. Figure 4), governs the relative contributions of errors in scale, direction and anisotropy to the total value of the score. For the purposes of this paper, we will set $a = (J-1)/2$ corresponding to equal weights for both components ($e_{1,15°}$ is equally far away from $e_{1,105°}$ and $e_{J,15°}$). A more in-depth explanation of this score, including the mathematical definition of the EMD, is given in Appendix A. In addition to this summary score, we introduce three helpful auxiliary quantities:

$$d\rho = \rho_c^{(for)} - \rho_c^{(obs)}, \tag{2}$$

$$d\varphi = \begin{cases} \varphi^{(for)} - \varphi^{(obs)} - 180° & \text{for } \varphi^{(for)} - \varphi^{(obs)} > 90°, \\ \varphi^{(for)} - \varphi^{(obs)} + 180° & \text{for } \varphi^{(for)} - \varphi^{(obs)} < -90°, \\ \varphi^{(for)} - \varphi^{(obs)} & \text{otherwise,} \end{cases} \tag{3}$$

$$dz = z_c^{(for)} - z_c^{(obs)}\,, \tag{4}$$

where $^{(for)}$ and $^{(obs)}$ denote quantities related to forecast and observation, respectively. The difference in central scales dz was studied in Buschow *et al.* (2019) (under the name Sp_{cd}). These authors note that dz is a lower bound on *semd* (Rubner *et al.*, 2000) which gives a rough estimate of the scale error and, crucially, determines its sign (too small or too large in scale). Analogously, we have now defined the signed anisotropy error $d\rho$ and the angular error $d\varphi$. It is important to note that $d\varphi$ is only relevant if the predicted and the observed field are both reasonably anisotropic – a circle can be rotated by any angle without actually changing. While this quantity is thus useful and intuitive for individual comparisons, it cannot simply be aggregated over many cases. We therefore define the combined anisotropy and direction error

$$dxy = \sqrt{\left(x_c^{(for)} - x_c^{(obs)}\right)^2 + \left(y_c^{(for)} - y_c^{(obs)}\right)^2}\,. \tag{5}$$

In Section 7, we show that $semd_d$ can usually be explained as a linear combination of dxy and $|dz|$, the difference in the shape of the spectra playing only a minor role.

Concluding this section, we note that the complete verification procedure has been implemented in the `sad` R-package (Buschow, 2020).

6 | GEOMETRIC TEST CASES

As a first test of our structural forecast verification based on scale, anisotropy and direction, we consider the

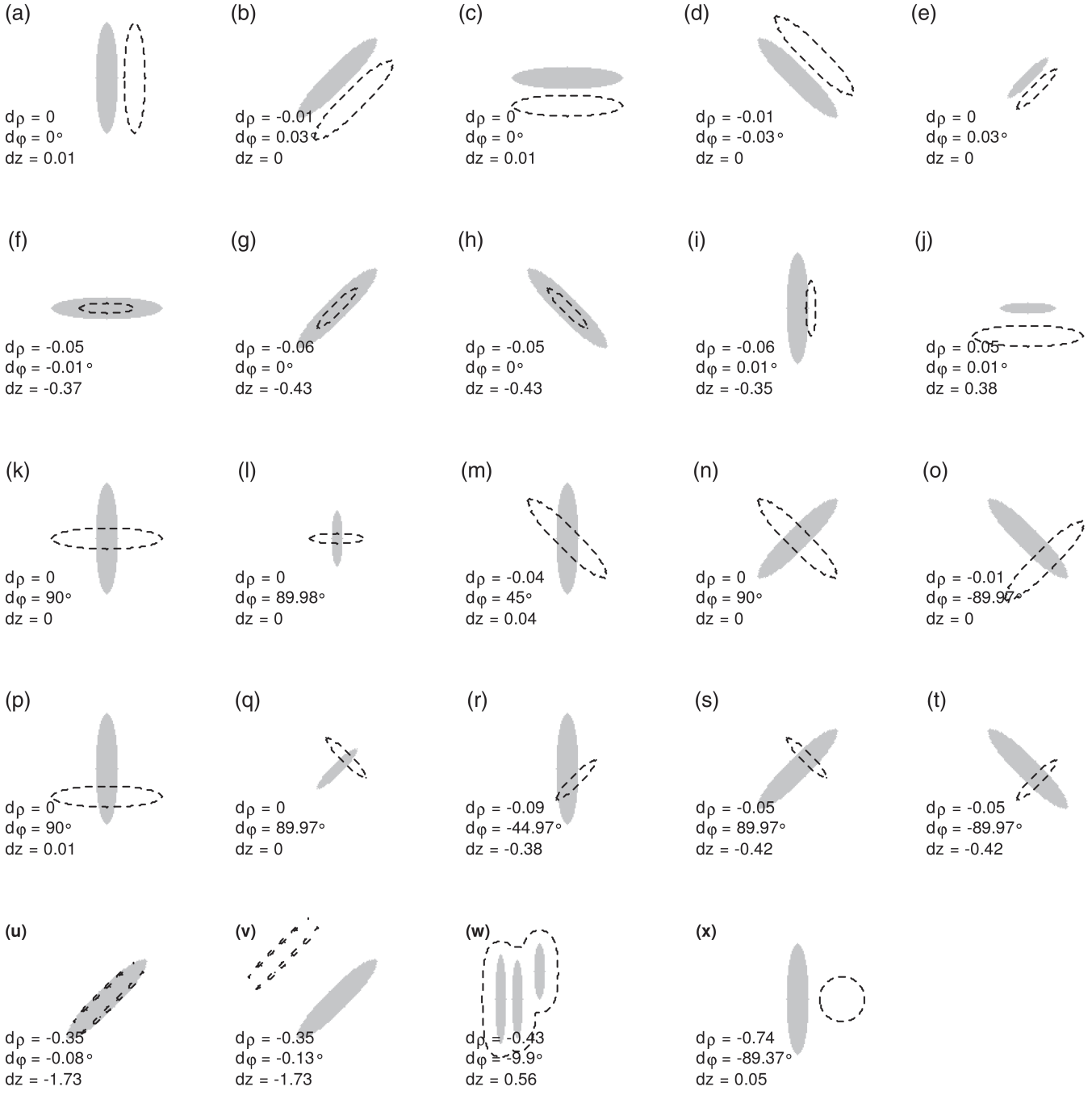


FIGURE 5 Comparisons between elliptical MesoVICT test cases suggested by Gilleland *et al.* In each panel, $d\rho$, $d\varphi$ and dz are anisotropy, angle and scale of the "forecast" (dashed contour) minus the "observations" (grey area)

geometric test cases proposed by Gilleland *et al.* (2020). These authors present a set of 50 binary images and suggest 55 pairwise comparisons between them. Here, we will discuss only the 23 comparisons between the elliptical test images because these are most relevant and interesting for our purposes.

Figure 5 shows all of the elliptical comparisons and the resulting values of $d\rho$, $d\varphi$ and dz. We have calculated the scores in this figure based on the decimated version of the DTCWT in order to test its remaining dependence on location and orientation. In Figure 5a–e, forecast and observation differ only in their location. As expected, all of our scores are close to zero with only very minor variations due to the remaining shift variance.

The situation in Figure 5f–j is more interesting. Besides possible displacement errors, the predicted area is now also too small (f–i) or too large (j). Intuitively, we expect $d\rho = d\varphi = 0$ and an identical non-zero scale error dz in all of these cases. While $d\varphi$ is indeed almost exactly zero, $d\rho$ indicates that the small ellipse appears slightly less

anisotropic than the larger one. The values of the scale error differ by less than one tenth of a whole scale and have the expected sign. Figure 5k–t introduce errors in the orientation. In all of those tests, $d\varphi$ detects the rotation with nearly perfect precision (errors <0.05°), irrespective of the presence of additional errors in location or scale.

In the last three test cases (Figure 5u–w), more complicated patterns are compared. In (u) and (v), the observation is a regular ellipse, while the forecast consists of very small oblong shapes along the ellipses' boundaries. As expected, the orientation is found to be very nearly correct, whereas the scale of the forecast is deemed far too small. As for the slightly too small ellipses in previous comparisons, $d\rho$ indicates that the predicted pattern is too isotropic. In the final suggested test (Figure 5w), the observation consists of three small ellipses; the forecast shows a single large feature enveloping the three. As expected, the scale error dz is strongly positive. The anisotropy error is significantly smaller than zero which is in good agreement with our subjective judgement as well. A slight clockwise rotation, indicated by $d\varphi$, also seems reasonable since the three small ellipses are vertically oriented, while the combined pattern extends somewhat more along the diagonal. Since none of the suggested comparisons feature an obvious error in anisotropy, we have added an extra case Figure 5x which compares a circle to an ellipse. As expected, the strongly negative value of $d\rho$ correctly detects the discrepancy. Note that the apparent rotation by -90° is meaningless because one of the two images has a low value of ρ.

In summary, the geometric tests show that our structure verification overall works as intended. The angular component in particular yields almost perfect results and is very robust to changes in location, scale, overall orientation and anisotropy of the fields to be compared. We have seen that the remaining shift-dependence of the decimated DTCWT plays only a minor role for our purposes (see also Appendix B) with discrepancies on the order of ∼0.05 in all three components (compare, for example, Figures 5(r) to (s), (u) to (v), and (i) to (j)). Conversely, this also means that forecast errors smaller than 0.05 can generally be regarded as negligible.

7 | REALISTIC TEST CASES

7.1 | Verification against VERA

For a first impression of our verification technique in a realistic situation, we consider a single time step from the second MesoVICT case. Here, we focus on only two competing forecasts. Figure 6a–c show the hourly rain intensity analyzed by VERA and predicted by CMH and COSMO. At this time, precipitation was mainly induced by a quasi-stationary airmass boundary extending roughly from the German–French border to the southwesterly corner of the domain. VERA shows a relatively linear rain feature along the Rhine and a number of more amorphous cells throughout France and Switzerland. CMH overestimates the rain area slightly and the total intensity strongly by producing a nearly round rain field in the north and numerous very small convective cells across the rest of the domain. COSMO, on the other hand, simulates a single linear feature along the airmass border. According to the maps of z_c (Figure 6d–f), the spatial scales are well represented by COSMO while the structure of CMH is overall slightly too small ($dz \approx -0.3$). As expected, CMH is slightly too isotropic, whereas COSMO appears far more directed than the observations ($d\rho \approx 0.46$). In addition, a rotational error of about 14° is assigned to COSMO, which is also in good agreement with our visual impression. The slightly worse scale and much better anisotropy add up to substantially better overall rating for CMH ($semd_d \approx 0.2$) than COSMO ($semd_d \approx 0.41$).

To get an overview of the complete MesoVICT dataset, we apply the decimated DTCWT to all fields and calculate the central components ρ_c, φ_c and z_c. Figure 7 displays the distributions of the evaluated central statistics for each hourly field as well as the the total rain area and total intensity, separated by case and model. Starting with the two simple, non-wavelet quantities, we observe that all models are able to simulate approximately correct rain totals, at least as far as the average over all cases is concerned (case 5 being an exception where all models frequently predict too little rain). The rain area, on the other hand, is systematically underestimated, especially by MOLO0225. One possible interpretation would be that this model simulates variability on smaller scales than those analyzed by VERA. We can partly confirm this hypothesis with the help of z_c which shows that MOLO0225 , as well as COSMO and CMH , operate on smaller scales than VERA. BOLAM007, with its nominal resolution of approximately 7 km, produces similar, in some cases even larger, scales than VERA. The order of the five datasets, BOLAM007 being largest, followed by VERA, COSMO, CMH and the very fine-scaled MOLO0225, is consistent across all six cases.

Next, we are interested in the directional structure. Looking at the distributions of ρ_c in Figure 7, we note that the degree of anisotropy of each model depends on the weather situation. Cases 3 (featuring a Genoa cyclone), 4 (dominated by large convective cells across the Alpine region) and 6 (no organizing frontal structure) are less directed than the remaining three cases, in which cold fronts and an airmass boundary dominate the weather

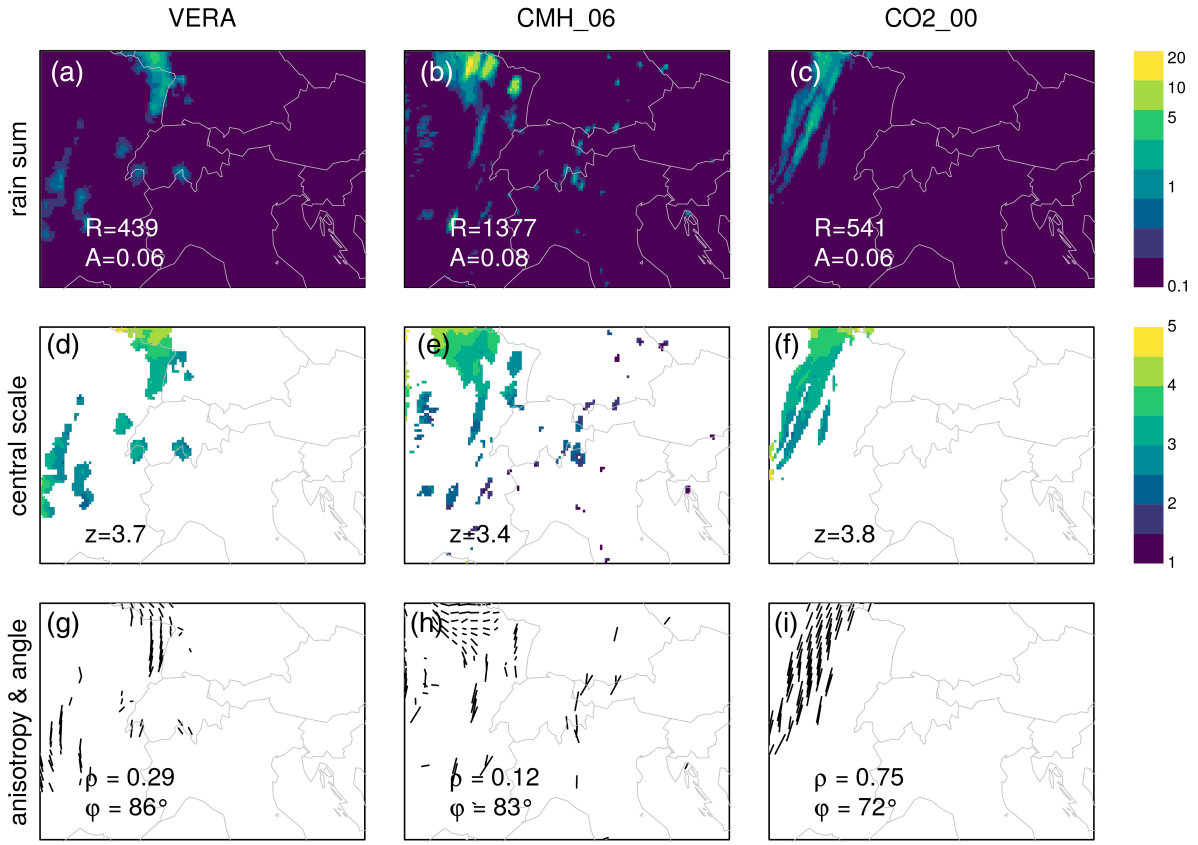


FIGURE 6 Wavelet-based analysis of observed and predicted precipitation on 20 August 2007, 1100 UTC. (a–c) Hourly rain accumulation with total rain sum in mm (R) and area fraction (A), (d–f) map of central scales with mean central scale (z) and (g–i) anisotropy (length of the arrows) and angle (direction of the arrows) with mean anisotropy (ρ) and mean angle (φ)

patterns. In comparison to VERA, CMH reproduces the average distribution of ρ_c very well. COSMO has the largest positive bias in ρ_c (is far too anisotropic), followed by BOLAM007 and MOLO0225.

The angles φ_c are not included in Figure 7 because box-plots can be misleading for a circular quantity. Figure 8 therefore shows the corresponding histograms instead. As expected, the two strongly directed cases 1 and 2 have a clear preferential direction around 45° and 60°, respectively, corresponding to the alignment of the airmass boundaries present in these cases. All models reproduce the analyzed direction reasonably well in the first case; BOLAM007 and MOLO0225 exhibit slight rotation errors in case 2. The anisotropic cases 3 and 4 feature a wide variety of directions, which are not particularly well matched by any of the models – recall that errors in the orientation are not meaningful when ρ_c is small. Case 5, which only encompasses 24 hr, has well-defined directions related to the cold front crossing the domain. All models represent the 45° orientation of this feature reasonably well. VERA's secondary peak at 90° is caused by a large rain area being cut off at the domain's eastern edge during the final time-steps of the day. Despite its relatively low anisotropy, case 6 also exhibits a well defined peak around 45°, which is present in VERA and all four forecast models. This phenomenon is likely related to the shape of the western flank of the Alps, where many of the precipitation events during this case-study were triggered.

So far, we have only assessed the modelled and observed statistics of spatial structures in each of the six cases. Figure 9 shows the corresponding distributions of the structure scores from Section 5 with respect to VERA, calculated at each time step and separated by case. COSMO and BOLAM007's systematic overestimation of ρ_c is reflected in increased values of the combined anisotropy / direction score *dxy*. Cases 2 and 5 are deemed particularly bad, while errors in the other, overall more isotropic, cases are less severe. In total, the representation of directional structures in CMH and MOLO0225 is notably better than in the other two models. The opposite result emerges for the representation of spatial scales where CMH and MOLO0225 are the worst candidates with strongly negative values of *dz*. Despite its nominally finer resolution, COSMO is only slightly too small in scale; the low-resolution BOLAM007 fares best.

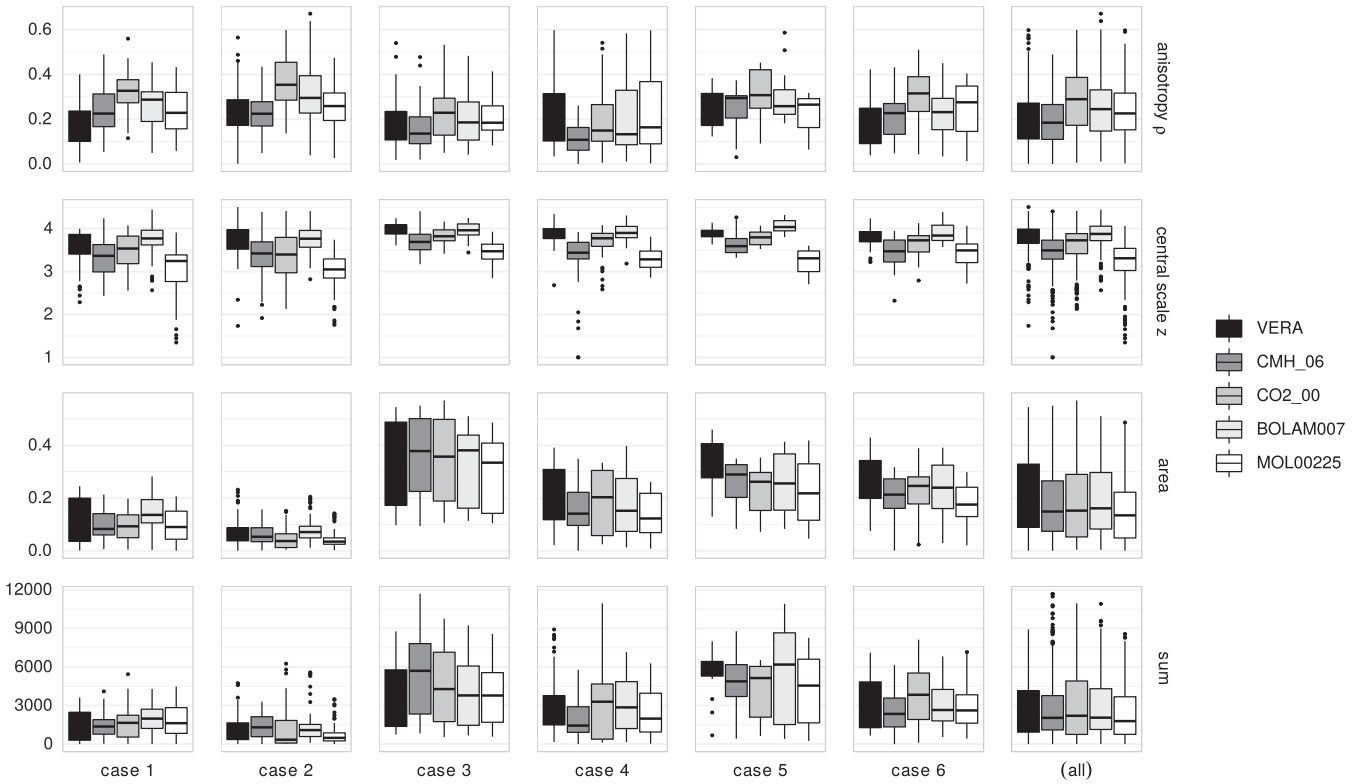


FIGURE 7 Degree of (row 1) anisotropy, (row 2) central scale, (row 3) fraction of the domain with non-zero rain and (row 4) total rain sum in mm for each of the six MesoVICT test cases and for all cases together

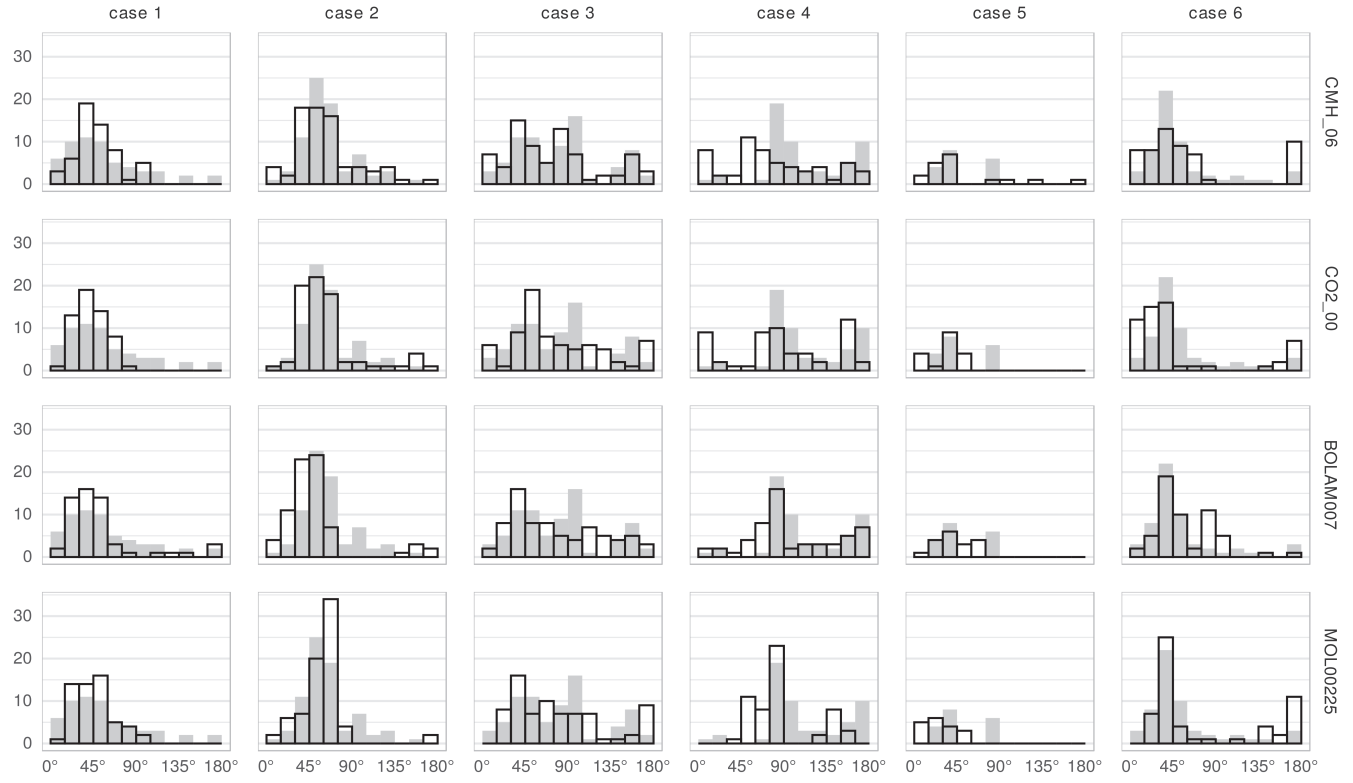


FIGURE 8 Histograms of dominant directions φ_c for all data forecasts (open bars) and VERA (grey bars), separated by case

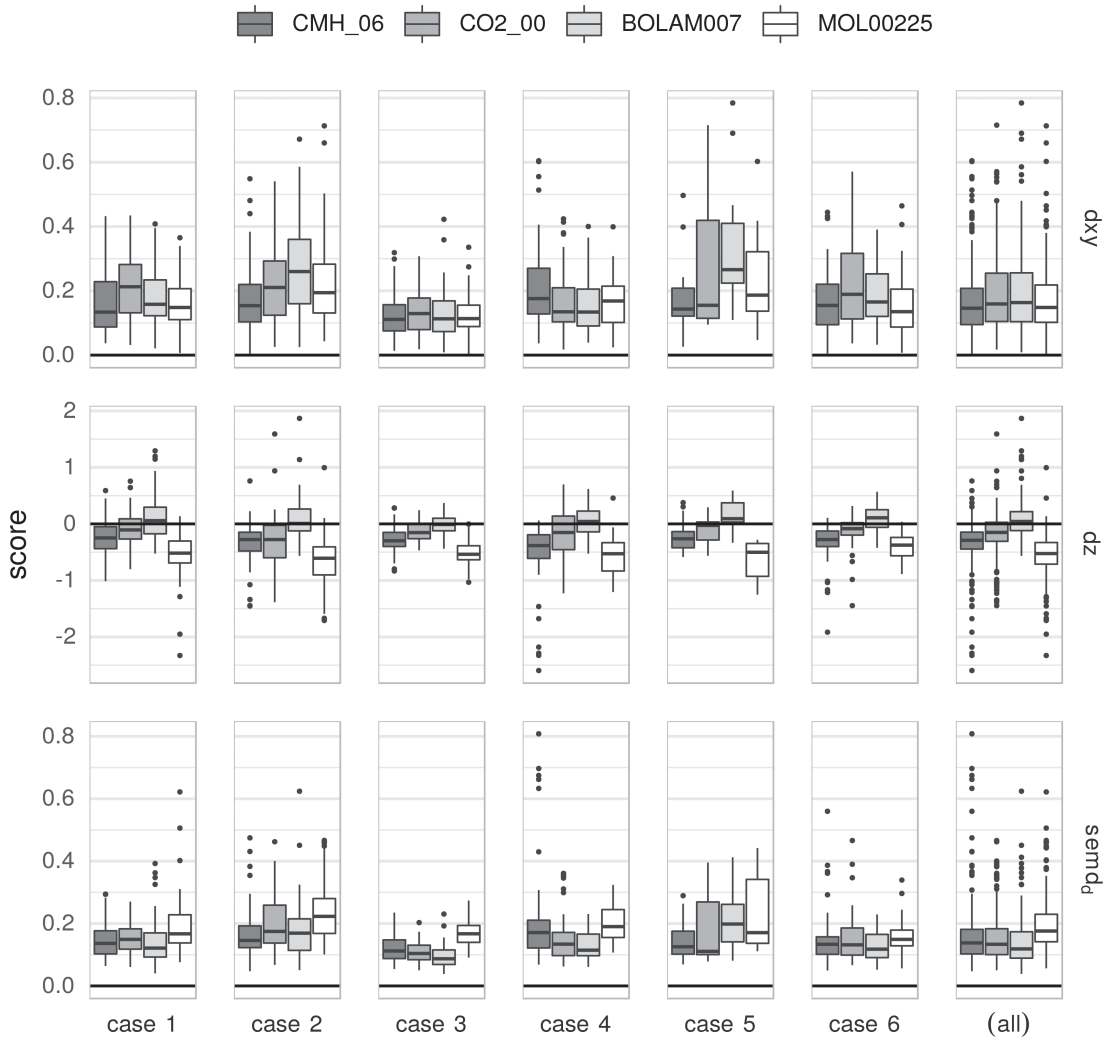


FIGURE 9 Combined (row 1) anisotropy/direction error, (row 2) difference in scale and (row 3) complete spectral EMD between VERA and the forecast models for the six cases. Positive values of dz indicate that the forecast is too large in scale

	CMH	COSMO	BOLAM007	MOLO0225	All
$\|dz\| \sim dxy$	0.23	0.04	0.11	0.05	0.05
$semd_d \sim \|dz\|$	0.82	0.58	0.51	0.69	0.69
$semd_d \sim dxy$	0.61	0.57	0.79	0.48	0.48
$semd_d \sim \|dz\| + dxy$	0.98	0.96	0.98	0.97	0.97

TABLE 3 Coefficient of determination R^2 for linear regressions of the scores on the left of $\sim$ against those on the right for each model individually and for all models

$semd_d$, shown in row 3 of Figure 9, combines both kinds of structural errors into a single score that takes into account the complete distribution of energy across directions and scales. We find that the large scale-error makes MOLO0225 the overall loser in each individual case and in total, despite its good representation of directions. For CMH and COSMO, the two kinds of error tend to average out, leading to nearly identical scores in aggregate. BOLAM007 emerges as the overall winner, largely due to having similar spatial scales as VERA.

To get a rough idea of the interrelations between our structure scores, we perform a series of linear regressions between them and consider the degree of determination measured by R^2. Table 3 confirms that the scale and directional errors are largely independent of one another. The relative contributions of dxy and dz to the overall score $semd_d$ differ from model to model, scale dominating for CMH and MOLO0225, direction for BOLAM007 and both contributing equally for COSMO. In all cases, as well as in total, $semd_d$ can almost entirely be explained as a linear

combination of the shift in scale and the shift in direction/anisotropy ($R^2 > 0.95$).

7.2 | Verification against RADKLIM

Up to this point, we have assumed VERA as a flawless representation of the true precipitation fields and interpreted all discrepancies as forecast errors. However, the fact that the coarsest model achieved the best overall rating raises some suspicions. Can VERA appropriately represent the scale structure of precipitation fields? To address this question, we shift our attention from the Alpine MesoVICT domain to Germany, where the radar-based RADKLIM dataset offers spatial observations at 1 km nominal resolution. CMH and COSMO are not available in this domain, so the other three datasets are now treated as competing forecasts to be verified against RADKLIM.

Before looking at the scores, it is again instructive to get a visual first impression from an example case. Figure 10 shows an instance of scattered convective cells across Germany during MesoVICT case 6. As expected, VERA and RADKLIM agree very well on the placement and approximate shape of the individual cells. Accordingly, the degree of anisotropy and the overall direction φ are nearly identical. However, RADKLIM reveals a much finer-scaled texture and distributes more precipitation across a smaller total area, leading to a scale error of $dz = 0.5$ for VERA. BOLAM007 and MOLO0225 both under-forecast the overall rain intensity and area and produce patterns slightly too isotropic. Variability in MOLO0225 occurs on the same small scales as in RADKLIM, and BOLAM007 is again far more similar to VERA.

The distribution of scores shown in Figure 11 reveals that our example case was in fact representative of an overall trend: while the scale errors of MOLO0225 are centred around zero, both BOLAM007 and VERA exhibit a bias towards larger scales ($dz \approx 0.3$). In terms of directional structure, VERA is by far the most similar to RADKLIM, followed by MOLO0225 and BOLAM007. With respect to the summary score $semd_d$, VERA and MOLO0225 are thus tied for first place, both performing substantially better than BOLAM007. It is worth noting that the distributions of the scores for the two forecast models have substantially heavier tails than for VERA. These outliers represent complete mis-forecasts of spatial structure, which naturally do not occur in observational datasets like VERA.

8 | CORRECTING STRUCTURAL ERRORS

Errors related to the marginal distribution of a forecast can generally be corrected if the desired distribution is known. Such a calibration procedure may be desirable to improve the forecast or to remove marginal errors before applying further verification methods. Most spatial verification techniques do not suggest a simple way of correcting the errors they detect; the wavelet approach is an exception to this rule. As detailed in Section 3, a wavelet transform is essentially just a change of basis, which can be reversed. Similar to the well-known Fourier case, the discrete wavelet transform allows for analysis *and* synthesis. To correct the errors in the spatial mean wavelet spectrum, we can therefore (1) transform forecasts and observations, (2) multiply the forecast values at each location, scale and direction by the corresponding ratio between total observed and predicted energy and (3) reverse the transform to obtain a corrected version of the forecast. The spatial distribution of the energy of the resulting image is that of the prediction, but its distribution over scale and direction corresponds to that of the observations. The complete procedure is given by Algorithm 1. The logarithmic transform in step 2 and the limitation to scales $\leq J$ ensure that the correction is consistent with our verification. By restoring the original mean and variance of the log-transformed field in step 10, we concentrate on the spatial structure without attempting to correct the margins as well.

Algorithm 1 is applied to all forecasts in the Alpine MesoVICT dataset; the reference in each case is the VERA analysis. Figure 12 shows four examples which illustrate the effects of our structural correction. In the first case Figure 12a–c, a forecast by the MOLO0225 model was deemed too small and too anisotropic. The algorithm smooths the field, rounds the linear pattern and visibly reduces small-scale variability. The result has near-perfect scale and direction properties, while maintaining the same arrangement of features as the original.

The second example Figure 12d–f depicts a situation in which the BOLAM007 model predicted a single large-scale rain band over the Alps, whereas VERA sees a number of smaller, disjointed cells. As expected, the correction converts the continuous rain area of the forecast into several smaller objects, thereby increasing the visual similarity with VERA.

While the previous two predictions were too anisotropic, Figure 12g–i show a forecast from the CMH model that was both too small-scaled and too round. After correction, much of the small-scale variability has disappeared, while the elongated shape in the centre of the domain has been rendered more coherent and linear.

As our final example, we have included a complete mis-forecast from the CMH model, which fails to simulate the front seen in VERA and produces scattered small-scaled precipitation across the Alps. This extreme example, which is clearly related to model spin-up

Algorithm 1. Correction of structure errors

Input: forecast F, reference R, largest scale J, minimum value $R_{\min}$
Output: corrected forecast F'

1: Set values $<R_{\min}$ to zero
2: Set $F \to \log_2(F + R_{\min})$, $R \to \log_2(R + R_{\min})$
3: Standardize F and R to zero mean, unit variance
4: Forward transform $dtF = \mathrm{dtcwt}(F)$, $dtR = \mathrm{dtcwt}(R)$
5: **for all** scales $j = 1, \dots, J$, directions $d = 1, \dots, 6$ **do**
6: Calculate sum over all grid points i: $e_F = \sum_i |dtF_{i,j,d}|^2$, $e_R = \sum_i |dtR_{i,j,d}|^2$
7: set all $dtF_{i,j,d} \to dtF_{i,j,d} \cdot e_R/e_F$
8: **end for**
9: Inverse transform $F' = \mathrm{dtcwt}^{-1}(dtF)$
10: Reset mean and variance of F' to the values before step 3
11: Set $F' \to 2^{F'} - R_{\min}$
12: Set values $< R_{\min}$ to zero

in CMH, serves to demonstrate the limitations of the algorithm. The global adjustment of the wavelet spectra cannot possibly create a cold front in which no precipitation has been simulated. Instead, the two largest cells at the western domain edge are united into a smooth, elongated feature; most of the remaining small-scale variability is removed.

Having seen that the correction algorithm produces realistic-looking fields while greatly improving the visual similarity between forecast and observation, we now quantify its influence on the verification results. As expected, Figure 13 shows that both directionality and scale, measured by *dxy* and *dz*, are greatly improved. The fact that these scores are not *exactly* zero is due to the (necessary) truncation step (Algorithm 1, step 12). In addition, the inverse wavelet transform used here (following Kingsbury 2006) is not perfect due to the special treatment of the diagonal directions. The improvement of the scores is nonetheless immense, indicating that these effects play no great role – the algorithm works as intended.

While the improvement in the wavelet scores is thus almost guaranteed by design, it is interesting to see whether beneficial effects on the structural forecast skill are observed by other verification methods as well. To this end, we apply the object-based structure score S of Wernli *et al.* (2008) (using 1/15 of the observed and predicted 90% quantiles as thresholds) and the variogram score *vgs* of Scheuerer and Hamill (2015). Following Buschow and Friederichs (2020), we use the stationary, isotropic, inverse-distance-weighted version of *vgs* with $p = 2$ and scale each field by its standard deviation to concentrate on verifying the correlation structure. The bottom panels in Figure 13 confirm that both the object-based S and the variogram score *vgs* measure a significant improvement after the wavelet-based structure correction. In particular, S originally also detects the substantial scale errors of MOLO0225 and CMH; after our adjustment, these models are deemed as good as COSMO and BOLAM007, both of which see a modest improvement in S as well.

9 | DISCUSSION

The central goal of this study is to present a verification technique that evaluates the predicted spatial structure in terms of scale, anisotropy and direction. This level of detailed structural analysis is enabled by the complex dual-tree wavelet transform of Kingsbury (1999), which comprises six directional filters on a range of spatial scales. Using data from the MesoVICT project, we have demonstrated that the DTCWT can indeed replace the classic discrete wavelet transform in an analysis of spatial scales. All previous results concerning the usefulness of such an analysis for spatial forecast verification (Kapp *et al.*, 2018; Buschow *et al.*, 2019; Buschow and Friederichs, 2020) or the quantification of convective organization (Brune *et al.*, 2018; 2020) remain

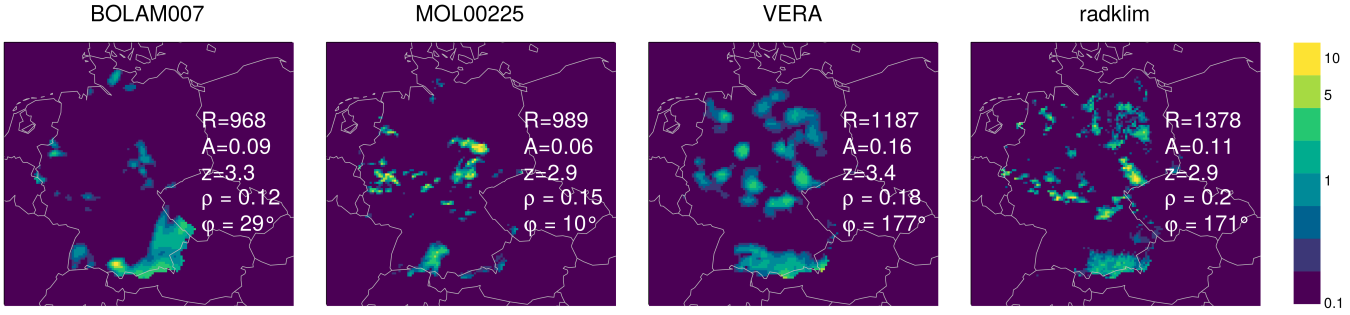


FIGURE 10 Predicted, analyzed and observed rain fields and structural characteristics on 9 July 2007 at 1900 UTC

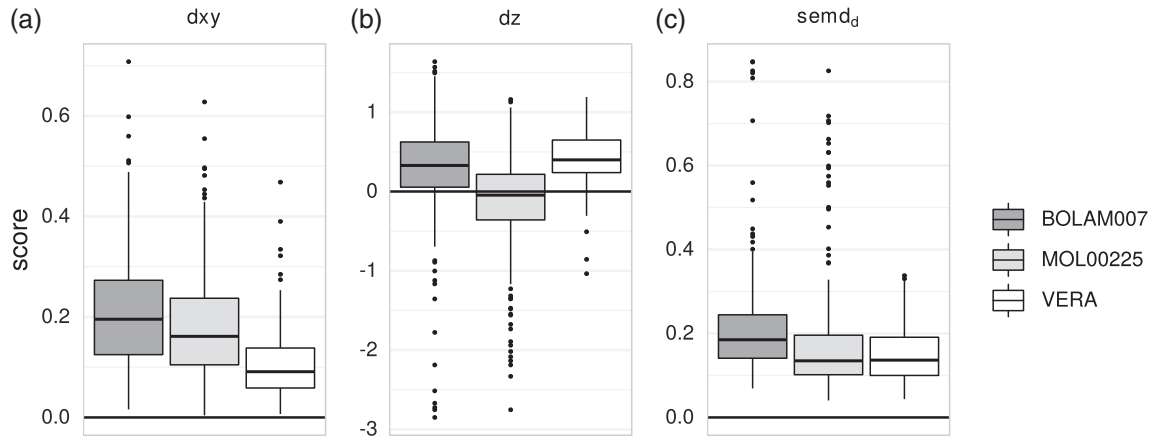


FIGURE 11 (a) Anisotropy, (b) scale and (c) total structure error of BOLAM007, MOLO0225 and VERA, verified against RADKLIM in the Germany domain. Only cases with at least 100 non-zero rain pixels in the RADKLIM image were included

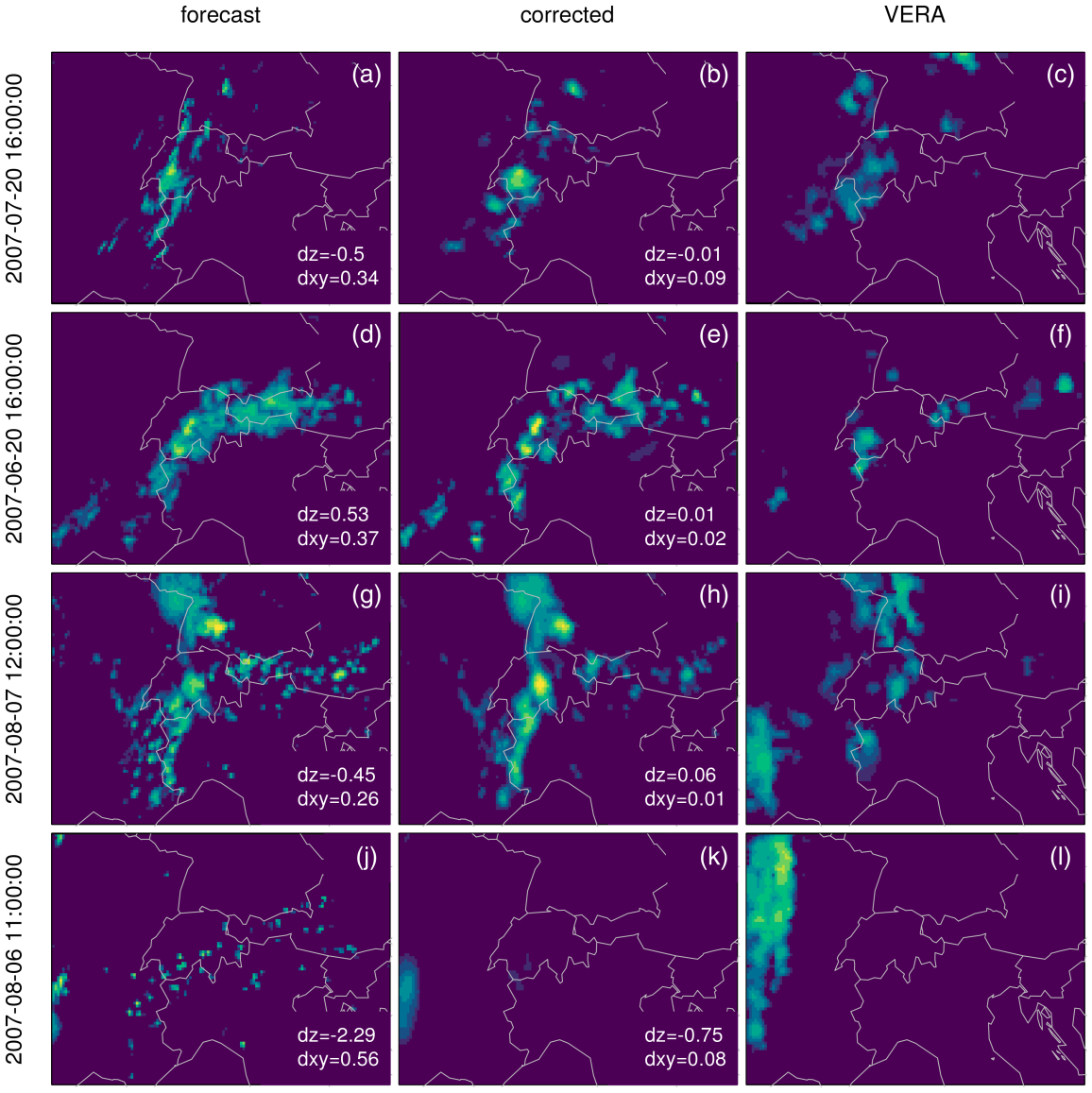


FIGURE 12 (a, d, g, j) Original forecasts, (b, e"h, k) corrected versions, and (c, f, i, l) the corresponding VERA analysis for four cases

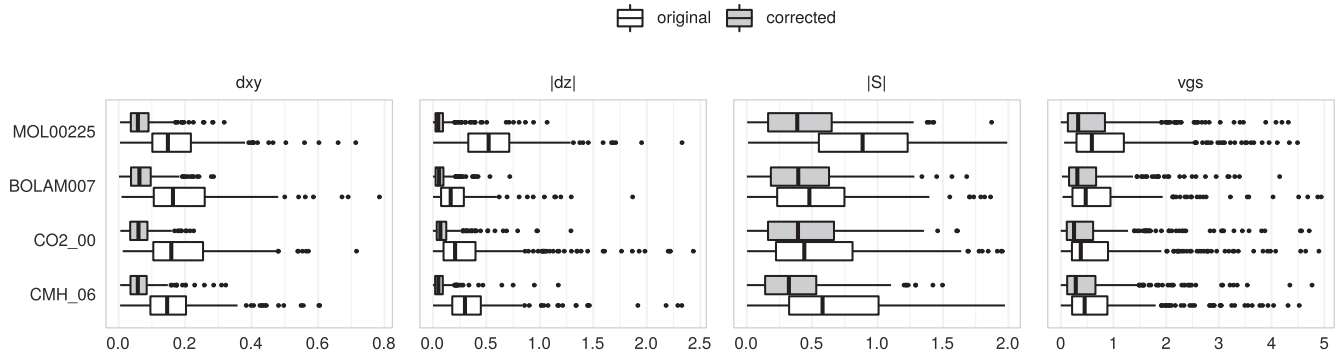


FIGURE 13 Distributions of structure scores for the original and corrected predictions of each model

valid for the dual-tree version. When only global characteristics are of interest, we can even replace the computationally expensive redundant transform by the extremely efficient decimated version. This potentially enables the use of our methods in contexts where time constraints, very large spatial domain sizes ($n_x > 1024$) or the sheer number of fields would make the redundant transform impractical. Incidentally, the decimated transform also allows for an efficient implementation of the global WOI of Brune *et al.* (2018).

However, the key innovation of the dual-tree approach lies in the analysis of the directional structure, which is impossible with the three directional filters of the classic DWT. Building on the idea of a central scale, we have introduced two further components to the centre of the wavelet spectrum, namely the degree of anisotropy ρ_c and the angle φ_c. The geometric MesoVICT test cases demonstrate that ρ_c adequately distinguishes between elongated and round patterns; the analysis of directions using φ_c is nearly flawless for simple geometric shapes.

For the purposes of verification, we advocate the use of a combined anisotropy–direction score since errors in φ_c are not meaningful when the anisotropy is low. This score, denoted here as *dxy* was used together with the signed difference in central scale *dz* and the combined structure score $semd_d$ to verify precipitation forecasts from four competing models within the MesoVICT framework. Perhaps the most important lesson from this experiment is that scale and directionality represent two independent aspects of forecast quality. Compared to VERA, CMH and COSMO achieve nearly the same average values of $semd_d$, but the composition of the structure score is very different: while COSMO simulates structures that are systematically too linear, leading to higher values of *dxy*, CMH produces excessive small-scale variability early in the day ($dz < 0$). MOLO0225 is tied for the lowest *dxy* with CMH but simulates far smaller structures than any other model. It may be interesting to note that the scale of the precipitation fields is not entirely determined by the model's nominal spatial resolution, which is nearly the same for CMH, COSMO and MOLO0225. On the other hand, BOLAM007 has a much coarser grid spacing of approximately 7 km and, somewhat expectedly, produces nearly the same distribution of spatial scales as VERA, which gives it the best overall scores in this comparison.

Should we thus conclude that the coarsest model delivers the most realistic representation of spatial structure? Doubting this, we have compared VERA, BOLAM007 and MOLO0225 to the radar-based RADKLIM dataset, which represents a realistic spatial observation of rainfall. Here MOLO0225, the overall loser in the previous comparison, is rewarded for simulating the same fine spatial scales as seen in the observations; both VERA and BOLAM007 are deemed too large. MOLO0225's representation of directionality is good but not as good as VERA's, again leaving two "forecasts" tied in terms of the overall score $semd_d$. This is a good example of independence between directional and scale-related aspects of the spatial structure: even if a comparatively coarsely resolved validation dataset hinders our analysis of spatial scales, we can nonetheless study the direction and directedness.

The degree of anisotropy and the distribution of dominant directions in a realistic precipitation field are closely tied to meteorological processes like organized or unorganized convection, moving airmass boundaries and pressure systems, as well as the interaction of these processes with the local topography. Our results demonstrate that the wavelet-based approach can meaningfully verify these directional aspects of spatial structure, and thereby indirectly the simulation of the underlying processes. Together with the analysis of scales and simple statistics such as the total rain area and accumulation, we can obtain a very detailed, objective picture of the spatial pattern in various observed and simulated fields.

An advantage of the wavelet-based approach is the existence of an inverse transform which allows us to correct the detected structural errors by a rather straightforward algorithm. The resulting post-processed fields combine the spatial placement of the predicted rain field with the global structure of the observations. We have shown that this procedure produces realistic-looking results with greatly improved wavelet and non-wavelet structure scores. The correction procedure has three main benefits. First, it enhances our intuitive understanding of the wavelet-based verification by showing us what an improved version of the forecast would have been. Second, the errors detected by the wavelet scores can be removed from the forecast before other scores are applied. In this manner, one can eliminate structure errors before verifying other aspects like the location of the predicted objects. Third, if a forecasting system exhibits strong systematic biases in its spatial structure (as was the case for MOLO0225 here), a correction to the observed climatological spectra could actually improve the value of the forecasts. How exactly such a structural post-processing could be implemented and whether it has any real-world utility is a question for future research.

ACKNOWLEDGEMENTS

Sebastian Buschow was funded by the DFG under grant no. FR 2976/2-1. We are grateful to Barbara Casati and one anonymous reviewer for their helpful criticism and encouraging remarks. Further thanks go to Eric Gilleland at NCAR and Stefano Mariani at ISPRA for providing the MesoVICT datasets. We also thank Alex Gibberd for his advice on the implementation of the bias correction for the dual-tree model and Rich Wareham for providing an open source Python implementation of the DTCWT. Open access funding enabled and organized by Projekt DEAL.

ORCID

Sebastian Buschow https://orcid.org/0000-0003-4750-361X

REFERENCES

Ament, F. and Arpagaus, M. (2009). dphase_cosmoch2: COSMO model forecasts (2.2 km) run by MeteoSwiss for the MAP D-PHASE project. Deutscher Wetterdienst, Offenbach, Germany. https://gisc.dwd.de/wisportal/showMetadata.jsp?xml=de.dkrz.wdcc.iso2150323; accessed 22 December 2020.

Bica, B., Steinacker, R., Lotteraner, C. and Suklitsch, M. (2007) A new concept for high resolution temperature analysis over complex terrain. *Theoretical and Applied Climatology*, 90, 173–183. https://doi.org/10.1007/s00704-006-0280-2

Brune, S., Kapp, F. and Friederichs, P. (2018) A wavelet-based analysis of convective organization in ICON large-eddy simulations. *Quarterly Journal of the Royal Meteorological Society*, 144, 2812–2829. https://doi.org/10.1002/qj.3409

Brune, S., Buschow, S. and Friederichs, P. (2020) Observations and high-resolution simulations of convective precipitation organization over the tropical Atlantic. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3751

Buschow, S. (2020) SAD: Verify the Scale, Anisotropy and Direction of weather forecasts. R package version 0.1.3. https://CRAN.R-project.org/package=sad

Buschow, S. and Friederichs, P. (2020) Using wavelets to verify the scale structure of precipitation forecasts. *Advances in Statistical Climatology, Meteorology and Oceanography*, 6(1), 13–30. https://doi.org/10.5194/ascmo-6-13-2020

Buschow, S., Pidstrigach, J. and Friederichs, P. (2019) Assessment of wavelet-based spatial verification by means of a stochastic precipitation model (wv_verif v0.1.0). *Geoscientific Model Development*, 12(8), 3401–3418. https://doi.org/10.5194/gmd-12-3401-2019

Daubechies, I. (1992) *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM.

Davis, C., Brown, B. and Bullock, R. (2006) Object-based verification of precipitation forecasts. Part I: methodology and application to mesoscale rain areas. *Monthly Weather Review*, 134(7), 1772–1784. https://doi.org/10.1175/MWR3145.1

Dorninger, M., Gilleland, E., Casati, B., Mittermaier, M.P., Ebert, E.E., Brown, B.G. and Wilson, L.J. (2018) The setup of the MesoVICT project. *Bulletin of the American Meteorological Society*, 99(9), 1887–1906. https://doi.org/10.1175/BAMS-D-17-0164.1

Eckley, I.A., Nason, G.P. and Treloar, R.L. (2010) Locally stationary wavelet fields with application to the modelling and analysis of image texture: modelling and analysis of image texture. *Journal of the Royal Statistical Society, Series C*, 59, 596–616. https://doi.org/10.1111/j.1467-9876.2009.00721.x

Ekström, M. (2016) Metrics to identify meaningful downscaling skill in WRF simulations of intense rainfall events. *Environmental Modelling & Software*, 79, 267–284. https://doi.org/10.1016/j.envsoft.2016.01.012

Gilleland, E., Ahijevych, D., Barbara G. Brown, Casati, B. and Ebert, E.E. (2009) Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, 24(5), 1416–1430. https://doi.org/10.1175/2009WAF2222269.1

Gilleland, E., Skok, G., Brown, B.G., Casati, B., Dorninger, M., Mittermaier, M.P., Roberts, N. and Wilson, L.J. (2020) A novel set of geometric verification test fields with application to distance measures. *Monthly Weather Review*, 148(4), 1653–1673

Han, F. and Szunyogh, I. (2016) A morphing-based technique for the verification of precipitation forecasts. *Monthly Weather Review*, 144(1), 295–313. https://doi.org/10.1175/MWR-D-15-0172.1

Kapp, F., Friederichs, P., Brune, S. and Weniger, M. (2018) Spatial verification of high-resolution ensemble precipitation forecasts using local wavelet spectra. *Meteorologische Zeitschrift*, 27(6), 467–480. ISSN 0941-2948

Keil, C. and Craig, G.C. (2007) A displacement-based error measure applied in a regional ensemble forecasting system. *Monthly Weather Review*, 135(9), 3248–3259. https://doi.org/10.1175/MWR3457.1

Kingsbury, N. (1999) Image processing with complex wavelets. *Philosophical Transactions of the Royal Society of London, Series A*, 357(1760), 2543–2560. https://doi.org/10.1098/rsta.1999.0447

Kingsbury, N. (2006). *Rotation-invariant local feature matching with complex wavelets, in 14th European Signal Processing Conference, Florence, Italy*. Piscataway, NJ: IEEE.

Mallat, S.G. (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674–693. http://dx.doi.org/10.1109/34.192463.

Mariani, S. and Casaioli, M. (2018) Effects of model domain extent and horizontal grid size on contiguous rain area (CRA) analysis: a MesoVICT study. *Meteorologische Zeitschrift*, 27(6), 481–502. https://doi.org/10.1127/metz/2018/0897. ISSN 0941-2948

Marzban, C. and Sandgathe, S. (2009) Verification with variograms. *Weather and Forecasting*, 24(4), 1102–1120. https://doi.org/10.1175/2009WAF2222122.1

McTaggart-Cowan, R. (2009). Metadata for dphase_cmcgemh: regional GEM model high resolution forecast run by CMC for the MAP D-PHASE project. World Data Center for Climate (WDCC) at DKRZ. https://doi.org/10.1594/WDCC/dphase_cmcgemh

Nelson, J.D.B., Gibberd, A.J., Nafornita, C. and Kingsbury, N. (2018) The locally stationary dual-tree complex wavelet model. *Statistics and Computing*, 28(6), 1139–1154. https://doi.org/10.1007/s11222-017-9784-0

Nerini, D., Besic, N., Sideris, I., Germann, U. and Foresti, L. (2017) A non-stationary stochastic ensemble generator for radar rainfall fields based on the short-space Fourier transform. *Hydrology and Earth System Sciences*, 21(6), 2777–2797. https://doi.org/10.5194/hess-21-2777-2017

Rubner, Y., Tomasi, C. and Guibas, L.J. (2000) The Earth Mover's Distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2), 99–121

Scheuerer, M. and Hamill, T.M. (2015) Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143(4), 1321–1334. https://doi.org/10.1175/MWR-D-14-00269.1

Scovell, R.W. (2020) Applications of directional wavelets, universal multifractals and anisotropic scaling in ensemble nowcasting: a review of methods with case studies. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3780

Selesnick, I.W., Baraniuk, R.G. and Kingsbury, N.C. (2005) The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6), 123–151. https://doi.org/10.1109/MSP.2005.1550194

Torrence, C. and Compo, G.P. (1998) A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1), 61–78

Urbanek, S. and Rubner, Y. (2012) *emdist: Earth Mover's Distance*. https://CRAN.R-project.org/package=emdist. R package version 0.3-1.

Weniger, M. and Friederichs, P. (2016) Using the SAL technique for spatial verification of cloud processes: a sensitivity analysis. *Journal of Applied Meteorology and Climatology*, 55(9), 2091–2108. https://doi.org/10.1175/JAMC-D-15-0311.1

Wernli, H., Paulat, M., Hagen, M. and Frei, C. (2008) SAL – A novel quality measure for the verification of quantitative precipitation forecasts. *Monthly Weather Review*, 136(11), 4470–4487. https://doi.org/10.1175/2008MWR2415.1

Willeit, M., Amorati, R., Montani, A., Pavan, V. and Tesini, M.S. (2015) Comparison of spectral characteristics of precipitation from radar estimates and COSMO-model predicted fields. *Meteorology and Atmospheric Physics*, 127(2), 191–203. https://doi.org/10.1007/s00703-014-0359-8

Winterrath, T., Brendel, C., Mario, H., Junghänel, T., Klameth, A., Walawender, E., Weigl, E. and Becker, A. (2018). RADKLIM version 2017.002: reprocessed gauge-adjusted radar data, one-hour precipitation sums (RW).

Wong, M. and Skamarock, W.C. (2016) Spectral characteristics of convective-scale precipitation observations and forecasts. *Monthly Weather Review*, 144(11), 4183–4196. https://doi.org/10.1175/MWR-D-16-0183.1

How to cite this article: Buschow S, Friederichs P. SAD: Verifying the scale, anisotropy and direction of precipitation forecasts. *QJR Meteorol. Soc.* 2021;1–20. https://doi.org/10.1002/qj.3964

APPENDIX A. HEXAGONAL ARRANGEMENT AND THE EMD

To further illustrate the concepts introduced in Section 4, we consider the wavelet spectra corresponding to the rain fields shown in Figure 12l,j. In Figure A1a, b, we have listed the energy values corresponding to each combination of scale j and direction d. For our calculations, each of these energies $e_{j,d}$ is treated as a point mass located at the coordinates $x_{j,d}, y_{j,d}, z_{j,d}$ (cf. Equation 1). The point masses are visualized as spheres of different volumes in Figure A1c, d. The frontal structure from Figure 12l leads to a concentration of mass at scale five and the directions around 90°. Conversely, the small, isotropic pattern of Figure 12j is reflected by a more even distribution across all directions and the three smallest scales.

ρ_c, θ_c and z_c are the barycentre of this arrangement of point masses, represented in cylindrical coordinates. Scores like $d\rho$ and dz are simply given by the difference between the central coordinates of two spectra. These simple scores are useful because they are easy to interpret, but they neglect some information on the full distribution of energy (Buschow *et al.*, 2019). To define a summary score that includes all information from the mean spectrum, we therefore use the Earth Mover's Distance (EMD) which measures the minimum total cost of transforming one arrangement of point masses into another (Rubner *et al.*, 2000). Let $m = 1, \ldots, 6J$ be an index enumerating all combinations of scale and direction (j, d). One spectrum is transformed into another by successively transferring amounts of "mass" (in our case spectral energy) $f_{m \to n} > 0$ from locations m in the first spectrum to locations n in

the second spectrum. Recalling that our "masses" are normalized such that $\sum_i e_i^{(k)} = 1$ ($k = 1, 2$ denoting the first and second spectrum), we seek a set of mass transfers which satisfy

$$\sum_m f_{m\to n} = e_n^{(2)} \quad \text{(the result of the transport is spectrum 2)},$$

$$\sum_n f_{m\to n} = e_m^{(1)} \quad \text{(all mass from spectrum 1 is transported somewhere)}.$$

Denoting the Euclidean distance between locations m and n by

$$d_{m,n} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2},$$

we can write the total work of a particular transport scheme as $\sum_{m,n} d_{m,n} f_{m\to n}$. The EMD is given by the minimum work needed, i.e.,

$$semd_d = \min\left(\sum_{m,n=1}^{6\cdot J} d_{m,n} f_{m\to n}\right), \tag{A1}$$

where the minimum is taken over all possible sets of transports $(f_{m\to n})_{m,n}$ that satisfy the requirements above. The solution of the optimization problem is found numerically via the `emdist` R-package (Urbanek and Rubner, 2012).

APPENDIX B. COMPARISON BETWEEN WAVELET TRANSFORMS

To quantify the impact of our choice of wavelet transform on the resulting structure analysis, we transform each field in the original MesoVICT dataset (Alpine domain, VERA, CMH, COSMO, BOLAM007, MOLO0225) three times: once with the decimated DTCWT which was used throughout Sections 6 and 7, once with the redundant version (used to produce Figure 6) and a third time with the sixth "Extremal Phase" Daubechies wavelet (redundant version with bias correction). Figure B1a shows the central scales z_c resulting from the spatially averaged, bias-corrected wavelet spectra. Apart from a slight linear offset, the agreement between the three analyses is close to perfect ($R^2 \approx 0.99$), and we observe no surprising outliers and no nonlinear effects. This confirms our claim that the DTCWT analyses scales in nearly exactly the same way as the usual DWT used by Buschow and Friederichs (2020). For the anisotropy ρ_c, we only compare the decimated DTCWT to its undecimated version since the DWT is not expected to agree with the dual-tree results here. Figure B1b shows almost no systematic bias; the correlation is again very high ($R^2 \approx 0.95$). We conclude that the global scale and anisotropy can be inferred from the decimated DTCWT just as well as from the undecimated version without incurring any significant, systematic double penalty.

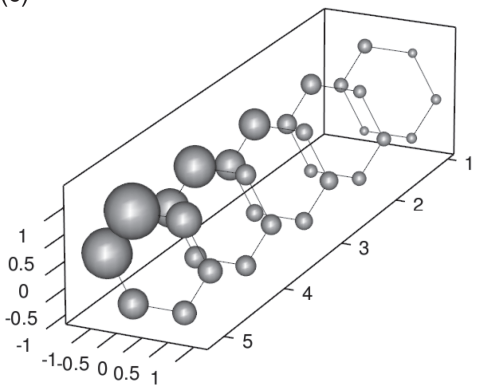
(a)

	15°	45°	75°	105°	135°	165°
$J = 1$	0	0	0	0	0	0
$J = 2$	0	0	1	1	0	0
$J = 3$	0	0	2	2	0	0
$J = 4$	1	1	5	4	1	1
$J = 5$	1	3	10	7	2	1

(b)

	15°	45°	75°	105°	135°	165°
$J = 1$	2	2	2	2	2	2
$J = 2$	2	2	3	2	2	2
$J = 3$	2	2	2	2	1	1
$J = 4$	1	1	1	1	0	0
$J = 5$	0	0	0	0	0	0

(c)



(d)

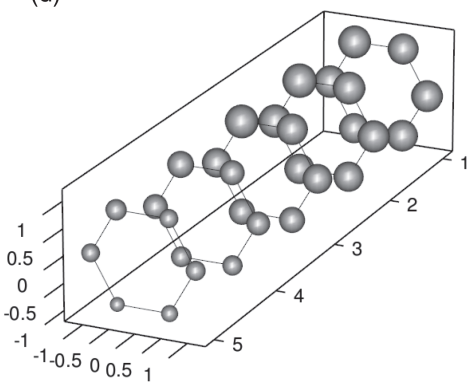


FIGURE A1 Mean wavelet spectra for the fields in Figure 12l,j. (a"b") Energy values, re-scaled and rounded to integers between 0 and 10. (c, d) Representation in a hexagonal arrangement (Figure 4); the volume of the spheres is proportional to the energy listed in the table

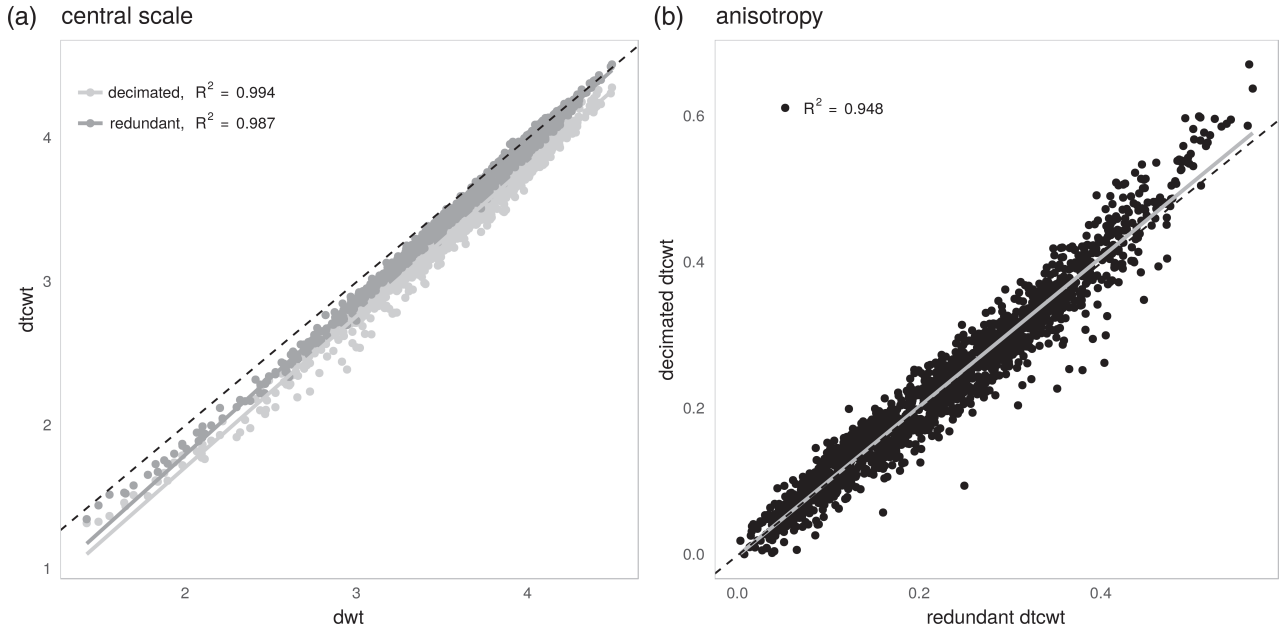


FIGURE B1 Dependence of structural characteristics on the wavelet transform. (a) central scale z_c of the DWT (DB6) against that of the decimated and redundant DTCWT. (b) Anisotropy ρ_c of the redundant and decimated DTCWT. The solid line indicates linear regression, and the dashed line marks the unit diagonal