TOPICAL REVIEW

A network-based approach for semi-quantitative knowledge mining and its application to yield variability

Bernhard Schaubeger1, Susanne Rolinski and Christoph Müller
Potsdam Institute for Climate Impact Research, Telegraphenberg A56, D-14473 Potsdam, Germany
1 Author to whom any correspondence should be addressed.
E-mail: schauber@pik-potsdam.de, rolinski@pik-potsdam.de and cmueller@pik-potsdam.de

Keywords: yield variability, crop models, interaction network, plant process, wheat, maize, rice

Supplementary material for this article is available online

Abstract
Variability of crop yields is detrimental for food security. Under climate change its amplitude is likely to increase, thus it is essential to understand the underlying causes and mechanisms. Crop models are the primary tool to project future changes in crop yields under climate change. A systematic overview of drivers and mechanisms of crop yield variability (YV) can thus inform crop model development and facilitate improved understanding of climate change impacts on crop yields. Yet there is a vast body of literature on crop physiology and YV, which makes a prioritization of mechanisms for implementation in models challenging. Therefore this paper takes on a novel approach to systematically mine and organize existing knowledge from the literature. The aim is to identify important mechanisms lacking in models, which can help to set priorities in model improvement. We structure knowledge from the literature in a semi-quantitative network. This network consists of complex interactions between growing conditions, plant physiology and crop yield. We utilize the resulting network structure to assign relative importance to causes of YV and related plant physiological processes. As expected, our findings confirm existing knowledge, in particular on the dominant role of temperature and precipitation, but also highlight other important drivers of YV. More importantly, our method allows for identifying the relevant physiological processes that transmit variability in growing conditions to variability in yield. We can identify explicit targets for the improvement of crop models. The network can additionally guide model development by outlining complex interactions between processes and by easily retrieving quantitative information for each of the 350 interactions. We show the validity of our network method as a structured, consistent and scalable dictionary of literature. The method can easily be applied to many other research fields.

1. Introduction

Crop yields can vary strongly between years and locations. These fluctuations, or yield variability (YV), are undesirable, since they undermine food security on three dimensions (Morton 2007, Schmidhuber and Tubiello 2007, Wheeler and von Braun 2013, Thornton et al 2014). First, the amount of harvested food can be lower than necessary, second, the financial sustainability of farming systems can be challenged, and third, the access to nutritious food can be diminished by rising prices or export bans connected to variable yields (Headey and Fan 2008, Headey 2010, Coumou and Rahmstorf 2012, Chung et al 2014). Substantial fractions of historic YV can be explained by weather variability and extremes like droughts, floods, heat waves, cold spells, or combinations of them (Porter and Semenov 2005, Schlenker and Roberts 2009, Coumou and Rahmstorf 2012, Lobell et al 2013, Der- yng et al 2014, Ray et al 2015, Lesk et al 2016). Globally about one third of YV can be explained by weather variation, but with large regional differences (Ray et al 2015). Although some of the actual weather-induced variation in yields might be lost in the aggregation procedure, this leaves up to two thirds of YV to be explained (SI figure S1). Thus other environmental or management factors must cause the variation. An example for regional differences is the
influence of precipitation on YV (figure 1). Precipitation variability clearly determines wheat variability in Australia (panel (a) with inlay), while in Germany wheat does not exhibit a clear, simple response to either temperature or precipitation (panel (b); Ray et al 2015).

Variability in growing conditions is transmitted to yield levels by plant physiological processes. These form a layer of complexity that has to be accounted for when assessing future YV. A huge body of experimentally-derived knowledge describes quantitative relationships between growing conditions, plant physiological processes and yield (e.g. Barnabás et al 2008, Farooq et al 2009b, Hatfield et al 2011). Process-based crop models are frequently used to study the influence of growing conditions on crop physiology and yields apart from experiments. These models represent our current knowledge on plant interactions with their environment (Boote et al 2013, Holsworth et al 2015). They are apt to reliably reproduce spatially aggregated mean yield levels (Palosuo et al 2011, Asseng et al 2015, Martre et al 2015).

Despite the abundant knowledge about YV a consistent and comprehensive overview of its causes and mechanisms is not yet available. Apart from the study by Ray et al (2015) and similar predecessors (see references therein) other causes of YV were also researched, but focusing on subsets of possible causes only. Bakker et al (2005) decipher the contribution of soil, climate and management as important sources of spatial wheat YV in Europe. Porter and Semenov (2005) or Asseng et al (2011) consider the impacts of heat stress on crop yields, but do not consider other climatic factors like water or solar radiation, or do not discuss plant physiological processes. Other studies include Yu et al (2014), who identify temporal patterns of climate effects on wheat YV in Australia but do not consider processes, or Thornton et al (2014), who stress the importance of considering climate variability in food security assessments, and Ben-Ari and Makowski (2014), who identify the geographical distribution of crops as source of YV. At the same time crop models are deemed to lack adequate implementations of temporal YV under changing growing conditions (Rötter et al 2011, Sánchez et al 2014). In particular, extreme events like heat or drought have been found to be less well represented (Palosuo et al 2011, Rötter et al 2011, White et al 2011, Boote et al 2013, Rötter 2014, Asseng et al 2015). A comprehensive overview of the current status of crop models is provided by Boote et al (2013), who list nine cardinal points on how to improve crop models.

Yet YV is not explicitly addressed as a topic, and stresses are only considered for heat, nitrogen and water. In Holsworth et al (2015) the authors state the effects of increased CO2, temperature extremes, pests and hydrology as inadequately represented in models. Barlow et al (2015) and Eyshi Rezaei et al (2014) describe the negative effects of frost or heat on cereals and derive modeling guidelines. We conclude that a comprehensive and systematic overview of causes and mechanisms of YV is much needed, in particular for selecting suitable process candidates for model improvement.

Therefore we systematically review the literature on YV, and provide specific recommendations on how to incorporate the findings into process-based crop
models. We adopt a novel, semi-quantitative technique for systematic reviews since the literature on plant physiology is overwhelming (more than 11 000 hits in the Web of Science² database for ‘crop variability’). The idea is to structure knowledge in a network of interactions, where management, weather and other environmental factors define crop yield via plant physiological processes (figure 3). We then rank the possible contribution of individual growing parameters to yield from their location in the network topology, independent from their frequency in the literature. Furthermore, we quantify the importance of plant physiological processes for the transmission of variability in growing conditions to YV. Finally, we use this knowledge to compile suggestions for the improved representation of YV in crop models. The method is ‘semi-quantitative’ since we do not employ quantitative relationships between growing conditions and yield. But we do quantify the impact of growing condition parameters and plant physiological processes by their contribution to the network structure.

To test the validity of the method we compare our proposals to the agenda suggested by Boote et al (2013). We consider maize (a C₄ plant), rice and wheat (both C₃ plants), representing roughly 92% of the globally harvested cereals (Ben-Ari and Makowski 2014) and planted on 41.3% of the global agricultural area (Portmann et al 2010).

This article describes a new method for mining knowledge from the literature, which is applied to review physiological mechanisms of YV. It is bound to reproduce existing knowledge to a large extent, but will check this for comprehensiveness and can thus guide future crop model development. With this review, we aim to answer three questions. First, what are key drivers of YV in wheat, maize and rice? Second, what are the central plant physiological processes involved? Third, how can the important interactions be included into crop models?

2 Materials and methods

2.1 Definitions and network terminology

Yield is an aggregate measure of crop characteristics and performance over the entire growing season. Yield can be defined as the integral of many short-term variations in growing conditions during the growing season and the plant’s reaction towards them. Temporal YV is hence the variability of this integral (equation (2.1)). We define YV as average changes in absolute yield amounts (e.g. t/ha) between growing seasons of the same crop at the same location; an example measure would be the standard deviation.

\[ \Delta \text{Yield} = \Delta \left[ \int_{\text{growing season}} \text{physiological processes (growing conditions)} \right]. \]

² http://apps.webofknowledge.com/; accessed on 11 October 2015.

We focus on the fine-grain interactions between growing conditions, plant physiological processes and yield. Spatial variability plays, next to variability over time, a decisive role (Ben-Ari and Makowski 2014). Here we assume that spatial and temporal variation share common causes like e.g. temperature variation over space or time (Blois et al 2013), such that our analysis is also valid for spatial YV. We do not consider long-term trends, including a gradual increase in yields through improved management or a shift in yield trends from changes in climatic conditions. We use the term ‘stress’ to describe any non-optimal growing condition (e.g. a heat wave). Finally, plant growth and plant development (‘phenology’) are two distinct terms: while the first is a physical accumulation of biomass over the growing season, the latter refers to advances in developmental stages, for example the transition from vegetative to reproductive growth.

A network consists of nodes (i.e. elements) and edges between these. In our case nodes refer to processes, drivers or variables and edges to interactions between them. The source/target node of an edge is its starting/end point, respectively. A path q from node A to node B, denoted as \( A \rightarrow_{q} B \), through the network follows a direction and can be direct (i.e. connecting A with B immediately) or indirect (i.e. containing intermediate nodes). The path length \(|q|\) is the number of edges it contains (illustration in SI figure S2).

2.2 Network construction

The starting point for the network construction was the basic network scheme shown in figure 3, into which subcategories and interactions were subsequently added. The interactions described in six standard physiology text books (Hay and Walker 1989, Porter and Lawlor 1991, Hall et al 1993, Larcher 1995, Hay and Porter 2006, Lamberts et al 2008) were used to add details: the network was refined with every encountered subcategory or interaction. For example, if the scanned literature stated an influence of temperature (T) on photosynthesis, these two nodes were created (if not yet existent) and an interaction arrow drawn from T to photosynthesis (if not yet existent). If an interaction edge was already present, the new reference was recorded, but no duplicate edge was included. This ensures that interactions do not gain more weight just because they are frequently stated in the literature. With more details, the categories were
subdivided. For each node and interaction it was annotated for which crop (wheat/maize/rice or all three) it is valid, thus creating crop-specific networks.

Afterwards a systematic search for studies in the full ISI Web of Science database was performed. A keyword list with 55 entries was created, using terms from the initial textbook-based network. General terms like ‘yield variability’ and more specific ones like ‘temperature AND wheat AND yield’ were included (full keyword list in SI table S3). Only papers after 1990 were considered to limit the number of search results. This first search for the keywords in the ‘Topic’ fields yielded 460 765 studies in total, so the results were filtered to contain only ‘Review’ papers. If this number was still large (> 200) for one search term the results were further filtered to contain the keywords in the ‘Title’ instead of ‘Topic’ (with few exceptions; SI table S3). Additionally, references to and in four large reviews (Barnabás et al 2008, Farooq et al 2009b, Hatfield et al 2011, Boote et al 2013) were searched to validate the efficacy of the keyword approach. These search criteria resulted in 8818 studies that were inspected for relevance by sequentially looking at title, abstract and full text. An article was relevant if the study included an explicit treatment of plant physiological processes, with either growing condition in- terences on them or their in- fluence on yield, and the interactions were not derived solely from modeling studies. More recent studies were selected when similar but older ones existed. Molecular details like enzyme activity or signaling molecules and genotypic or cultivar-specific differences are not considered. After this final filter step, 60 relevant papers remained from which interactions were manually included in the initial text book-based network.

Six out of 350 edges were added without explicit literature reference as they were considered obvious but have not been found in the selected literature. These are: irrigation adds to soil water content (SWC), fertilization adds to soil nutrient levels, sowing and harvesting time affect the amount of precipitation and solar radiation intercepted during the growing season, water uptake is affected by SWC, and the plant’s uptake of micronutrients influences their content in grains.

2.3. Driver and process importance

The importance of drivers as possible sources of YV was derived from the network structure. The importance of a driver d is defined by the number of different paths from d to yield amount, mediated by various plant physiological processes (equation (2.2); m denotes the maximum path length).

\[ \text{importance}_m(d) := | \{ q \mid d \xrightarrow{m} q \text{ \text{ Yield} } \land |q| \leq m \} |. \] (2.2)

A maximum path length \( m = 4 \) (i.e. at most three intermediate nodes) was chosen. This allows for possibly important indirect effects but avoids cyclic paths. Sensitivity to this assumption was tested with path lengths from 1 to 10. Each interaction was counted only once, independent of the number of studies which mentioned that specific influence. Thus a frequent occurrence of an interaction in the literature does not necessarily imply a high ranking. It is assumed that only drivers that exert substantial impact on plant physiology and are variable in nature can cause YV. Therefore each possible driver was qualitatively classified for its variability in nature and drivers with low variation were excluded. Three reduced network variants were also analyzed to search for variability drivers other than temperature (T) and precipitation (Pr). From the full network either T (air and soil, with all out-edges), or Pr along with SWC, or both T and Pr nodes (then also air humidity) were deleted; then the importance assessment was repeated.

The importance of a process for transmitting variability in growing conditions to YV was evaluated by plotting both impact values against each other (scheme in figure 2). The impact of a node v on another node w is defined by the number of paths between v and w, similar to the importance of drivers in equation (2.2). A process which is impacted by many different influences from the growing conditions (above the mean value on x-axis) and in turn substantially impacts yield levels (above mean value of y-axis) was assumed important in this respect. Processes in the other sectors of the plot did fulfill either one criterion or none at all, and were thus deemed less or not important for shaping yield amounts.

We consider this network method as ‘semi-quantitative’ since quantitative relationships between growing conditions and yield are not included, but the relative impact of growing condition parameters and plant physiological processes is evaluated by their quantitative contribution to the network structure.

The network approach explicitly integrates across physiological scales and assumes that driver or process importance is directly related to their number of network links to yield. The adequacy of these two assumptions is justified in the discussion section.

3. Results

3.1. Network structure

The crop yield interaction network contains 130 nodes and 509 edges. Of the edges 350 are interactions between nodes (functional interactions); the other edges only connect hierarchical distinctions in categories, e.g. ‘uptake’ to ‘uptake of nutrients’. Each node is connected on average by 3.92 edges (functional edges only: 2.69), the average number of studies cited

3 http://apps.webofknowledge.com/; accession dates in SI table S3.
per functional interaction is 1.53, and the nodes with the highest out-degree are temperature (49 out-edges), SWC (39) and CO2 (26). The number of edge annotations for only wheat are 105, for only maize 36 and for only rice 32; interaction references applying to all three crops summarize to 363 (SI table S1).

3.2. Relative importance of factors causing YV

SWC, with its climatic precursor precipitation (Pr), and temperature (T) are ranked as foremost influences on yield by our method. An analysis of the full and the reduced network variants suggests also the following environmental factors as physiologically important for

chilliness and heat, water logging and drought, excess and shortage of solar radiation, ozone, strong wind, nutrient shortage and excess, salt and acidity stress, pests and diseases, and toxic substances.
yield amount (but not necessarily its variability): carbon dioxide, solar radiation, soil salinity, tropospheric ozone concentration, microorganisms (e.g. mycorrhizas), soil temperature, soil pH, soil density, wind and soil nutrient levels. From the management category the following nodes are suggested as important: timing of sowing/harvesting, weed management, irrigation, soil management and drought resistance support. Figure 4 shows the relative ranking of drivers (only top 25%): the x-axis contains the four network types (full and three reduced variants) and the y-axis the number of interactions up to a path length of four (log-scale). The more interactions a factor controls, the more important it is assumed for yield formation. The results are similar for all three crops, although the relative importance can be crop-specific (SI figure S4).

3.2.1. Filtering drivers with low short-term variability

Only factors that are variable in nature can be drivers of variable yields. Therefore, to exclude unlikely drivers of variability, we determine the variation of each factor that is regarded as yield-influencing from our network. Table 1 lists the variability of each factor and whether it is considered in this review. The management options listed above are ‘variable’ by definition since the farmer can decide at any point in time to apply irrigation, drought support (seed priming only before the growing season), weed control or different soil management options (before and within season). Sowing times can also be highly variable between years, depending on local climatic conditions, cultivar choice, soil parameters and other factors (Craufurd and Wheeler 2009, Portmann et al 2010, Waha et al 2012, 2013). The impact of management decisions on YV is not assessed here, but should nonetheless be considered in crop models. In the following we only consider environmental variations as source of YV. Interactions between drivers and plant physiological processes are summarized in cursory depth in the next section. An extended and in-depth version with more references can be found in the SI.

3.2.2. Processes affected by water and temperature

The influence of precipitation on yield is paramount in most regions of the globe (Yu et al 2014, Ray et al 2015), and it is mediated via the SWC. SWC depends on precipitation and other factors like temperature, soil density and management (e.g. tillage) (Leakey et al 2009, Hatfield et al 2011). The fraction of SWC that is available for uptake by plant roots is further determined by soil salinity or competition (Fuhrer 2003, Tokatlidis 2014). Photosynthesis, temperature regulation, carbon allocation, nutrient uptake and reproduction strongly depend on water to function properly (Boyer and Westgate 2004, Reddy et al 2004, Barnabas et al 2008, Brouder and Volesevic 2008, Farooq et al 2009b, Gonzalez-Dugo et al 2010, Ahmed et al 2013, Jagadish et al 2014, Suzuki et al 2014). In particular, reproductive processes including anthesis and grain filling are highly sensitive to drought (Acevedo et al 2002, Boyer and Westgate 2004, Barnabas et al 2008, Lawlor and Tezara 2009, Gonzalez-Dugo et al 2010, Thitikasakul et al 2012, Powell et al 2012, Ashraf 2014, Farooq et al 2014, Jagadish et al 2014). Non-optimal water availability...
Table 1. Assessment of the natural variability of important yield-influencing factors in crop growing conditions. The first column contains the factor, the second column its short-term variability in nature (low or high), the third column lists references for the variability, the fourth column contains comments on the factor and the fifth states if the factor is included in this review.

| Factor                      | Variability | References                                           | Comment                                                                                           | Inclusion |
|-----------------------------|-------------|-----------|------------------------------------------------------|-----------|
| Soil water content and precipitation | High        | Lobell and Gourdji (2012), Donat et al (2013), Ruane et al (2015) | SWC buffers Pr variability, but eventually follows the Pr trend (Bell et al 2010) | Yes       |
| Temperature (air and soil)  | High        | Rahmstorf and Coumou (2011), Seneviratne et al (2012) | For example influences on yield see Ray et al (2015)                                             | Yes       |
| Solar radiation             | High        | Wang and Dickinson (2013)                           | Important especially when other factors are not limiting (Tolenaar and Lee 2002, de Bossor- | Yes       |
| Tropospheric Ozone          | High        | Fuhrer (2003), Martiello and Giacchi (2010), Wild et al (2012), Tai et al (2014), Hoshika et al (2015) | Ozone follows temperature, solar radiation and precursor trends nonlinearly McGrath et al (2015) | Yes       |
| Wind                        | High        | SI figure S3 for interactions                       | Aggregate effects are unclear                                                                   | No        |
| Soil nutrient pools         | High        | Fageria and Baligar (2005); Porter and Lawlor 1991 (p 173) | Nutrients are key limiting factors for yield (Boote et al 2013)                                | Yes       |
| CO₂                         | Low         | Varotsos et al (2007)                              | Though CO₂ exerts a significant ecophysiological impact on crops (Long et al 2006, Leakey et al 2009, Sakurai et al 2014, Myers et al 2014), there is only low within-season variation | No        |
| Soil salinity               | Low         | George et al (1997), Clarke et al (2002), Schofield and Kirkby (2003), Lambers et al (2008) | Could create variability in yield and production levels at spatially aggregated levels, but only low within-season variation (Ben-Ari and Makowski 2014) | No        |
| Microorganisms              | Unknown     |                                                      | May still be instrumental for understanding YV; management influences microorganisms (e.g. Gaudin et al 2015) | No        |
| Soil acidity or density     | Unknown     |                                                      | No information on its interannual variability is available                                       | No        |
also has a possibly negative influence on soil microorganism composition and on the severity of diseases (Hatfield et al. 2011, Ahmed et al. 2013).

The yield amount of wheat, maize and rice is reduced with non-optimal temperatures. Early growth, photosynthesis, carbon assimilation, stomatal conductance, plant development and root functioning strongly respond to temperature. This can diminish yields if temperature is too high or low (Schnyder 1993, Acevedo et al. 2002, Wahid et al. 2007, Barnabás et al. 2008, Craufurd and Wheeler 2009, Farooq et al. 2009a, 2011, Hatfield et al. 2011, Hasanuzzaman et al. 2013, Madhu and Hatfield 2013, Jagadish et al. 2014, Suzuki et al. 2014). Reproduction, again, is particularly sensitive to temperature extremes (Ishag and Mohamed 1996, Morison and Lawlor 1999, Dupont and Altenbach 2003, Barnabás et al. 2008, Farooq et al. 2011, Siebenmorgen et al. 2013, Jagadish et al. 2014). Many biochemical processes, like cell respiration and division, leaf senescence or membrane functionality depend on an optimal temperature range (Fuhrer 2003, Wahid et al. 2007, Farooq et al. 2009a, Mohammed and Tarpley 2009, Yadav 2010, Farooq et al. 2011, Hasanuzzaman et al. 2013, Miura and Furumoto 2013, Jagadish et al. 2014). High T can also be coupled to an increased O₃ concentration that causes damage on its own (see below).

### 3.2.3. Processes affected by other important drivers

Solar radiation is the only source of energy for photosynthesis. Radiation is also the ultimate source of all weather variables like temperature. But the relation between radiation and temperature has recently become more complex (Wang and Dickinson 2013), and solar radiation affects crops in additional, distinct ways (Porter and Lawlor 1991 (p 106)). Excess radiation can damage the photosynthetic apparatus or induce oxidative stress, which both reduce the assimilation of C (Reddy et al. 2004, Lammers et al. 2008 (p 36)). Low radiation can also limit the uptake of nutrients (Lammers et al. 2008 (p 268)).

Tropospheric ozone (O₃) is known to cause substantial harm to crops in many regions (Avnery et al. 2011, McGrath et al. 2015). Its concentration in the Northern Hemisphere has risen in recent decades, with regional variation (Hoshika et al. 2015). Increased O₃ has been shown to enhance leaf senescence, to impair reproductive processes and to lower the resistance against diseases (Fuhrer 2003, 2009, Hatfield et al. 2011, Beckles and Thitsaksakul 2014). Higher O₃ can also counterbalance a fertilization effect of CO₂ (Fuhrer 2009, Hatfield et al. 2011).

Nutrients including nitrogen (N), phosphorus (P) and other micronutrients are essential determinants of crop yield. Their uptake is influenced by temperature, soil characteristics (water content, acidity, salinity), root structure, soil characteristics, weed competition and plant growth (Fuhrer 2003, Barnabás et al. 2008, Brouder and Volenc 2008, Ahmed et al. 2013, Ashraf 2014). Nutrients, especially N, and micronutrients like potassium or iron are required for photosynthesis, protein or starch synthesis, stress tolerance, turgor maintenance or ROS scavenging (Porter and Lawlor 1991 (p 13, 39, 55ff.); Hay and Porter 2006 (p 109, 188); Thitsaksakul et al. 2012, Powell et al. 2012, Suzuki et al. 2014). An excess of nutrients, in contrast, can cause misguided growth or impede grain filling (Schnyder 1993, Yang and Zhang 2006).

#### 3.2.4. Influences on yield quality

Not only yield amount, but also yield quality is variable (e.g. Larcher 1995 (p 289); Dupont and Altenbach 2003, Siebenmorgen et al. 2013). The assessment described above for yield amount has been performed for yield quality, too. It indicates that essentially the same set of drivers and plant physiological processes is important for the determination of quality (SI figure S6).

### 3.3. Selecting processes for improving crop models

The drivers of YV identified above affect yields indirectly by their influence on plant physiological processes. With our network we can also identify the relative importance of these processes to guide further development of crop models.

#### 3.3.1. Relative importance of plant physiological processes

The putatively most important processes to transmit variability in growing conditions to YV are those strongly influenced by environmental or management stimuli and at the same time exerting a considerable impact on yields. An ordering diagram is shown in figure 5, which is structured as described in figure 2. Every process is located according to its sensitivity to driver variables on the x-axis and its respective impact on yield levels on the y-axis. We identify the following plant physiological processes as important (located in the top right quadrant): plant growth (split into growth of stem, roots and leaves); the uptake of water and nutrients; photosynthesis; reproduction; stress responses including antioxidant and disease defense; (evapo)transpiration; respiration; cell-internal processes like protein synthesis, turgor maintenance or division; senescence; and stomatal opening regulation. Connections between these processes, their environmental effectors and their influence on yields are found in section 3.2.

Three physiological hierarchy levels of the processes are included in the network: cell level, tissue and whole plant. Processes on the cell level determine those on tissue level (e.g. photosynthesis is required for root growth), and these in turn determine the processes on plant level (e.g. roots define the uptake of water). Growing conditions affect crops on different levels, which is thus reflected in our network.
There are minor differences between the three crops: wheat exhibits the full set of seventeen processes mentioned above as important, while maize and rice each have three less (senescence, stem development and stomatal opening regulation); SI figure S8. Different thresholds for the maximum allowed interaction path length do not alter the results qualitatively (SI figure S5).

3.3.2. Suggestions for implementing new features in crop models
To support the implementation of new features in crop models we collocate improvement suggestions derived from network structure and importance assessment. We compare the processes identified as important in our network analysis with the status quo of current crop models, as summarized by Boote et al (2013). Table 2 compiles this information for each plant physiological process (40 in total, on different hierarchical levels): importance for YV as ranked by the network evaluation (column 2), its respective importance on yield amounts (y-axis, log scale); see methods section for details. The dashed horizontal and vertical lines represent the mean (of the log-ed values) of the x- and y-axis entries, respectively. The processes in the upper right corner are annotated in the legend on the right; these are sorted according to their descending Manhattan distance from the origin (i.e. the sum of x and y location). All other abbreviations are explained in SI table S4. Note that hierarchical levels in the network are flattened, i.e. plant physiological processes can occur both directly and indirectly. For example, stem development (‘Stmdvl’) is a subcategory of plant development and growth (‘Plndag’).

![Relative importance of processes](image)

Figure 5. Relative importance of plant physiological processes. Importance is measured by the influence that growing conditions exert on a process (x-axis, log scale) and its respective influence on yield amounts (y-axis, log scale); see methods section for details. The dashed horizontal and vertical lines represent the mean (of the log-ed values) of the x- and y-axis entries, respectively. The processes in the upper right corner are annotated in the legend on the right; these are sorted according to their descending Manhattan distance from the origin (i.e. the sum of x and y location). All other abbreviations are explained in SI table S4. Note that hierarchical levels in the network are flattened, i.e. plant physiological processes can occur both directly and indirectly. For example, stem development (‘Stmdvl’) is a subcategory of plant development and growth (‘Plndag’).

There are minor differences between the three crops: wheat exhibits the full set of seventeen processes mentioned above as important, while maize and rice each have three less (senescence, stem development and stomatal opening regulation); SI figure S8. Different thresholds for the maximum allowed interaction path length do not alter the results qualitatively (SI figure S5).

3.3.2. Suggestions for implementing new features in crop models
To support the implementation of new features in crop models we collocate improvement suggestions derived from network structure and importance assessment. We compare the processes identified as important in our network analysis with the status quo of current crop models, as summarized by Boote et al (2013). Table 2 compiles this information for each plant physiological process (40 in total, on different hierarchical levels): importance for YV as ranked by the network evaluation (column 2), its respective importance on yield amounts (y-axis, log scale); see methods section for details. The dashed horizontal and vertical lines represent the mean (of the log-ed values) of the x- and y-axis entries, respectively. The processes in the upper right corner are annotated in the legend on the right; these are sorted according to their descending Manhattan distance from the origin (i.e. the sum of x and y location). All other abbreviations are explained in SI table S4. Note that hierarchical levels in the network are flattened, i.e. plant physiological processes can occur both directly and indirectly. For example, stem development (‘Stmdvl’) is a subcategory of plant development and growth (‘Plndag’).
| Process                              | Important (Network) | Priority (Boote *et al.* 2013) | Drivers  | Improvement options for crop models                                                                 |
|-------------------------------------|--------------------|---------------------------------|----------|-----------------------------------------------------------------------------------------------------|
|                                     |                    |                                 |          |                                                                                                     |
|                                     |                    |                                 |          |                                                                                                     |
| Plant dev. & growth                 | Yes                | 1                               | x x x x x x | Inclusion of stressors in development timing; water deficit, T [B13] and salinity or O₃               |
| Roots growth                        | Yes                | 1                               | x x x x x | Growth response to edaphic [B13] and weather conditions                                             |
| Leaves growth                       | Yes                | 2                               | x x x x x | Effects of canopy architecture, plant density and supply of nutrients and assimilates [B13]         |
| Stem growth                         | Yes                | 2                               | x x x x x | Reserve accumulation and utilization under stress (T)                                               |
| Early growth                        |                    |                                 | x x x x x | Interactions of seed quality and stressors (water, T)                                               |
| Reproduction                        | Yes                | 1                               | x x x x x | Fertility effects of high T, higher mechanistic detail [B13]                                        |
| Grain filling                       |                    | 1                               | x x x x x | Stressors like high T [B13] and O₃; interactions with seed quality                                  |
| Pollen development                  |                    | 1                               | x         | Effects of high T [B13] and drought                                                                  |
| Ovary development                   |                    | 1                               | x         | Effects of heat and drought stress                                                                  |
| Silking                             |                    | x                               | x x x     | Effects of heat and drought                                                                      |
| Spikelet development                |                    | x                               | x         | Effects of high T [B13] and water lack, interactions with elev. CO₂                                   |
| Anthesis                            |                    | 2                               | x x       | Effects of high T [B13] and radiation                                                                |
| Stomata regulation                  | Yes                | 1                               | x x x x x | Regulation by T, water, CO₂ [B13] and O₃ plus interaction with canopy effects and photosynthesis [B13] |
| Uptake                              | Yes                |                                 |           |                                                                                                     |
| Uptake of water                     |                    | 2                               | x x       | Interactions with root structure and nutrient uptake                                                |
| Uptake of nutrients                 | Yes                | 1                               | x x x x x | Interactions with root structure [B13]; not only N but also P [B13]; effects of saline soils         |
| General cell processes              | Yes                |                                 | x x       |                                                                                                     |
| Division                            |                    | x                               | x         | Effects of drought and hot or cold T                                                                |
| Turgor maintenance                  | Yes                |                                 | x x       | Drought and heat effects; interactions with growth and chemical reactions like photosynthesis         |
| Process | Important (Network) | Priority (Boote et al 2013) | Drivers | Improvement options for crop models |
|---------|---------------------|-----------------------------|---------|------------------------------------|
|         |                     |                             | Pr/SWC T R₃ [O₃] Nutr SoilT |                                    |
| Protein synthesis | Yes              |                             | x x  | Impact on yield quality [B13]; effects of high T and radiation |
| Vernalization |                    |                             | x     | Possible reversal under extreme heat |
| Senescence | Yes                | 1                           | x x x | Effects of high T [B13], drought, O₃ and excess fertilizer |
| Evapotranspiration | Yes              | 1                           | x x x | Effects of elevated CO₂ and T, connection to photosynthesis [B13], effects of salinity |
| Transpiration | Yes                | 1                           | x x x | Effects of CO₂ [B13] and extreme T |
| Respiration | Yes                 | 2                           | x x  | Effects of CO₂ [B13] and extreme T |
| Photosynthesis | Yes                | 1                           | x x x | Scaling up from leaf to canopy or field [B13]; effects of non-optimal T or drought |
| Light interception |                  |                             | [no hints] |                                    |
| Production of ROS |                   |                             | x x x | Effects of extreme T and drought, and the impact of ROS on other processes like photosynthesis |
| Photospiration |                    |                             | x     | Effects of heat stress |
| Acclimation |                     |                             | [no hints] |                                    |
| Stress responses | Yes                | 2                           |         | Incorporation of diseases into crop models [B13]; interaction with energy balance |
| Antioxidant defense |                  | x x                        |         | Induction by high T or drought through increased ROS production |
| Sec. metabol. accum. |                  | x                           |         | Energy costs by high T |
| Comp. osmol. accum. |                    | x x                        |         | Energy costs by drought or high T |
| Expr. of stress prot. |                  | x                           |         | Energy costs by high T |
| Responses to diseases | 2                  | x x                        |         | Incorporation of diseases into crop models [B13]; interaction with energy balance |
| Escape or avoidance |                    | x                           | [no hints] |                                    |
| Allelopathy |                     |                             | [no hints] |                                    |
| Herbivory defenses | 2                  |                             |         | Incorporation into crop models [B13] together with pest models; interaction with energy balance |
relevant since plant responses to simultaneous changes in growing conditions often differ from the responses to individual changes (Lobell and Burke 2010, Jagadish et al 2014, Ray et al 2015). Since extreme events can induce nonlinear responses in crops, their impact on plant physiological processes is of particular importance. These influences are annotated explicitly for the network interactions where mentioned in the associated studies (full network in SI).

4. Discussion

4.1. Validity of the network method

A network structure, derived from literature, is employed to evaluate the importance of growing condition factors and plant physiological processes on YV. Hence it is eminent to have an unbiased knowledge base for its construction. With the systematic approach by pre-defined search terms we aim to keep a literature bias (i.e. the over-representation of aspects like temperature) at a minimum. In addition, all interaction edges have the same weight independent of how often they are confirmed (or contradicted) in the literature, which limits a potential research frequency bias. A strong representation of a process in the literature might, however, reflect its pertinency for implementation. Additionally, a broad literature coverage of aspects like heat or drought stress might stem from its agronomic importance—which further warrants their appropriate consideration in models. Therefore we argue that our findings, which are based on a large interaction network and are robust under different analysis setups, are relevant for crop models.

Our importance assessment does not consider quantitative information in the interactions. But for the relative weighting of process importance a quantitative network would not necessarily be more accurate, as it would introduce more parameters to the method. Furthermore, every quantitative parameter would depend on crop, cultivar and location—which would be beyond the scope of any single meta-study to curtail for 350 interactions. Breeding efforts have achieved higher sensitivities to selected growing conditions, e.g. N and water provision. This trend is neglected in our network, for the same reason of quantitative complexity, but we argue as above that the method would not necessarily benefit from its consideration. Another possible issue that comes with missing quantitative information is that a node with many small influences on other nodes is considered more important than a node with only few but large impacts. Yet many small impacts can also amount to large ones, and the quantity argument goes as above. But if necessary, quantitative information for any specific process can easily be retrieved from the recorded interaction references.

The network unites plant physiological processes on cell, tissue and whole plant level. Single, scale-dependent networks for each of these three would be an alternative approach that respects differences between levels. But we argue that a united approach is justified in our case for three reasons. First, small-scale processes (e.g. cell respiration) accumulate influence over the growing season and therefore can determine yields as much as large-scale processes like, for example, herbivory. Second, the network is constructed to deduce improvement suggestions for crop models, which also need to reflect yield influences on all three levels. Thus we can more easily derive these suggestions with a combined network, but endorse further differentiation in later work. Third, we explicitly aim to capture all relevant mechanisms that may or may not act synchronously to influence yields (synchrony as, for example, in vernalization).

Plant physiological processes can be grouped or aggregated in manifold ways (e.g. Boote et al 2013, Bassu et al 2014). Our network is therefore only one approach to classify these processes. It is elicited from sequential literature reading of plant physiology text books and independent articles. An assessment of how different basic network structures (differing from figure 3) would affect the results was not tested here. But the network proved flexible enough to incorporate all interactions and elements found in the literature. Some plant responses to growing conditions may not be included in the 350 interactions of our network. Yet we argue that these are likely only minor given the systematic literature mining and the robust driver ranking.

The analysis indicates temperature and precipitation as strong drivers of YV—which is well-known and thus confirms the validity of our method. But we have also identified further factors whose own variability could imply variability in yields. Drivers which our network approach labels as ‘unimportant’ are not necessarily unimportant in reality—the relative weighting applied here only assigns more weight to the others. In contrast, the drivers defined as important by the network structure have only the potential to cause variability. Yet the actual importance depends on the specific combination of the individual components of the growing conditions. Although CO2 or soil salinity are not regarded as important contributors to YV, they can strongly influence responses to other stressors via interacting effects (Jagadish et al 2014). Adequately representing yield quality is equally judged an essential target for crop model improvement, evidenced by the 13 direct influences on yield quality in the network.

The close similarity of results for wheat, maize and rice arises from two independent factors. First, the network edges are often (68%) based on publications that are valid for all three crops considered. Second, the network is qualitative only such that quantitative distinctions between crops are not accessible. Differences between crops (e.g. rice is usually irrigated, maize is a C4
crop, winter wheat requires vernalization) are not questioned. But with regard to modeling the generality of the drivers identified is beneficial, since most of the mechanistic pathways are shared between crops.

The network essentially reflects the complexity of plant regulatory systems. This entails a high ‘importance’ for drivers or processes that are involved in several regulatory pathways, i.e. importance reflects complexity. Complex systems may either be prone to abrupt state changes under disturbances (Robbirt et al 2014, Willmer 2014, Zscheischler et al 2014, Franklin et al 2016) or enhance the stability of a system (an example is resilience from biodiversity). The current network does not reveal whether a process that influences YV actually enhances or dampens it. This requires deeper inspection of each single interaction with quantitative information. Our method suggests exactly these crucial points that need further inspection. One possible inspection approach is a recursive refinement of the network into single-process subsets.

4 http://agtrials.org/; accession date 01 February 2016.

to help understanding of YV in diverse environments, and from sources other than temperature or precipitation. Regional-scale crop patterns have been studied as causes of YV by Ben-Ari and Makowski (2014), and genetic traits in YV by Mickelbart et al (2015). The focus on plant physiological process level in our analysis complements these two approaches.

5. Conclusion

We have applied a novel methodology for a systematic literature review to identify and rank the importance of drivers and plant physiological processes for crop YV. We have also derived a comprehensive list of target points for improving crop models with respect to YV. As expected, our method confirms that current modeling approaches have addressed many of the important drivers of YV. Thus our approach can be seen as a cross-validation of existing modeling concepts. However, we also show that the drivers and the mechanisms implemented are not sufficiently comprehensive, which thus can guide future model improvement. Our network is a unique structured summary of the literature and its free accessibility can support the improved representation of YV in crop models. In particular the network interaction structure and the rich quantitative literature information associated with it can serve as a starting point.

The approach could be extended by a semi-automatic text mining, extracting the most relevant information from literature databases. Text mining has successfully been applied in medical bioinformatics (Zhu et al 2013, Fluck and Hofmann-Apitius 2014, Fleuren and Alkema 2015). Our network-based review could serve as a first step towards this. We have shown its methodical validity as a structured, consistent and scalable dictionary of literature knowledge. The approach is easily applicable to many other fields of research.

Acknowledgments

The authors want to thank Frank Wechsung, Christopher Reyer, Christoph Gornott, Katharina Waha (for helpful discussions) and Deepak Ray (for data provision). BS acknowledges financial support from the German National Academic Foundation, SR from the MACSUR project (031A103B) funded by the German Federal Ministry of Education and Research (BMBF), and CM from the MACMIT project (01LN1317A) funded by BMBF.

References

Acevedo E, Silva P and Silva H 2002 Wheat growth and physiology (www.fao.org/docrep/006/y4011e/y4011e06.htm#bm06) (Accessed: 21 November 2016)

Ahmed F, Rafii M Y, Ismail M R, Juraimi A S, Rahim H A, Asfaliza R and Latif M A 2013 Waterlogging tolerance of crops: breeding, mechanism of tolerance, molecular
concentration at Mauna Loa Atmos. Chem. Phys. Discuss. 6 11957–70
Waha K, Müller C, Bondeau A, Dietrich J, Kurukulasuriya P, Heinke J and Lotze-Campen H 2013 Adaptation to climate change through the choice of cropping system and sowing date in sub-Saharan Africa Glob. Environ. Change 23 130–43
Waha K, van Bussel L G J, Müller C and Bondeau A 2012 Climate-driven simulation of global crop sowing dates Glob. Ecol. Biogeogr. 21 247–59
Wahid A, Gelani S, Ashraf M and Foolad M 2007 Heat tolerance in plants: an overview Environ. Exp. Bot. 61 199–223
Wang K and Dickinson R E 2013 Contribution of solar radiation to decadal temperature variability over land Proc. Natl Acad. Sci. USA 110 14877–82
Wheeler T and von Braun J 2013 Climate change impacts on global food security Science 341 308–13
White J W, Hoogenboom G, Kimball B A and Wall G W 2011 Methodologies for simulating impacts of climate change on crop production Field Crops Res. 124 357–68
Wild O et al 2012 Modelling future changes in surface ozone: a parameterized approach Atmos. Chem. Phys. 12 2037–54
Willmer P 2014 Climate change: bees and orchids lose touch Curr. Biol. 24 R1133–3
Yadav S K 2010 Cold stress tolerance mechanisms in plants: a review Agronomy Sustain. Dev. 30 515–27
Yang J and Zhang J 2006 Grain filling of cereals under soil drying New Phytologist 169 223–36
Yu Q, Li L, Luo Q, Earnus D, Xu S, Chen C, Wang E, Liu J and Nielsen D C 2014 Year patterns of climate impact on wheat yields Int. J. Climatol. 34 518–28
Zhu F, Patumcharoenpol P, Zhang C, Yang Y, Chan J, Meechai A, Vongsangnak W and Shen B 2013 Biomedical text mining and its applications in cancer research J. Biomed. Inform. 46 208–13
Zscheischler J et al 2014 A few extreme events dominate global interannual variability in gross primary production Environ. Res. Lett. 9 035001