

Chapter 5

Stochastic Reasoning

When reason is against a man, a man will be against reason.

T. Hobbes

5.1 Lifting Isis' Veil

Sometime during the early fifth century BC, Heraclitus famously uttered: *Φύσις κρύπτεται φιλεί*.¹ Many centuries later, Werner Heisenberg famously postulated that “Not only is Nature stranger than we think, it is stranger than we can think.” Was Heisenberg right, and what exactly he meant by “we can think”? The spirit of this book is based on the premise that the precise meaning of this sort of thoughts can attune IPS to new dimensions of human inquiry, change one’s sense of what is possible and meaningful, and guide one toward unforeseen horizons of understanding.

Metaphorically speaking, Heraclitus’ and Heisenberg’s thoughts seem to converge to a common image of Nature using some sort of a “veil” or “mask” to deceive humans and make it difficult or even impossible for them to discover the truth. History-prone readers may recall that Nature has been allegorically identified with the goddess *Isis* of ancient Egypt. The statue of Isis covered in a black veil was erected on a tomb close to Memphis. On the statue’s pedestal was engraved the inscription:

I am everything that was, everything that is, that will be, and no mortal has yet dared to lift my veil.

The ancients believed that knowledge and truth were hidden beneath Isis’ veil. The lifting of the veil represented the revelation of the truth, and to succeed in doing so is to become immortal. Accordingly, since ancient times philosophical investigations have focused on questions like: Is Isis (Nature) unknown or unknowable? Can the veil be removed from Isis (Nature) by reason, experiment, or intuition? Should the veil be removed, and what are the possible consequences?

¹ Nature loves to hide.

Perhaps, one should not be over-concerned about goddess' veil. After all, ancient Greeks expected their gods and goddesses to behave as human beings do. Humans are often masked from one another, and so do their gods. This is true in modern times, and perhaps even more so. The imaginative ways humans are masked from others, masked even from those who they love most are masterfully explored in Carolyn Parkhurst's 2003 novel *The Dogs of Babel*. Just as is the case with human behavior, all options are on the table: Nature's veil may be impenetrable, she may chose to lift the veil herself, or the veil can be finally removed using the tools of human inquiry. In the latter case, it is left to inquisitive minds to search for creative ways that could progressively, profitably, and safely lift Isis' veil, so to speak.

Resorting once more to metaphor, *stochastic reasoning*² is an attempt to lift Isis' veil using a synthesis of tools (abstract and intuitive, mathematical and physical, rational and empirical) provided by the sometimes productive-sometimes fruitless, sometimes enjoyable-sometimes agitating, sometimes exhilarating-sometimes discouraging, yet always fascinating dialectic between the human mind and Isis (Nature). The correspondence between the inner and the outer, the intellectual and the sensuous, the seer and the seen, is a daring attempt to visualize invisible Isis out of space and time. It is also an attempt to obtain a deeper understanding of the distinction between the Nature impressing itself on the mind and fashioning it, on the one hand, and the mind portraying Nature in its own creative way, on the other hand. A word of warning may be appropriate at this point. Following Niccolo Machiavelli's advice that "injuries should be inflicted all at once," this chapter exposes the readers to a good dose of mathematics.

5.2 Reasoning in a Stochastic Setting

Although many investigators would claim that they do not consciously practice formal reasoning, nevertheless, they often unwittingly practice an informal yet distinctive reasoning mode. This is true even in cases in which the investigator's reasoning begins simply with the recognition of clues. The matter is of considerable importance since it can effectively help the investigators scrutinize the main presumptions underlying their research techniques, improve their understanding of key concepts, test and reshape their intuition. It is surprising that recent debates concerning epidemiology research and its consequences in public health (Boffetta et al. 2008, 2009a, b; Blair et al. 2009) do not pay sufficient attention to the soundness of the logical reasoning that underlies each approach. Instead, the focus is on technical data analysis and empirical evidence. I will start with a review of traditional reasoning modes, and then will make the connection with uncertainty in a real-world setting.

² Already briefly introduced in Section 1.5.3.

5.2.1 Basic Reasoning Modes

It has been said that we live in a sound-bite society, in which it is the simple issues that predominantly attract people’s attention. According to this perspective, if an idea cannot be presented on a bumper sticker, it has little or no chance to succeed. But this does not mean that one has to give in to hopelessness, which is how the story of stochastic reasoning unfolds.

5.2.1.1 Elements of Reasoning

Generally speaking, reasoning is a thought process that involves arguments (sentential, syntactic, symbolic, or numerical). An argument is a mental construct that starts with specified premises or hypotheses (data, facts, observations or measurements, statements, assumptions, and physical laws), and develops certain conclusions or consequences (problem-solutions, attribute predictions across space–time, system evaluations, and new laws). There is a list of so-called indicator words, which point out which part of the argument is the premises and which the conclusions. Words like, “assuming that,” “if,” “because,” “since,” and “by virtue of” indicate the beginning of premises. On the other hand, words like “therefore,” “hence,” “so,” “consequently,” and “it follows that” indicate the beginning of conclusions. For illustration purposes, Table 5.1 gives a list of common arguments. Whatever is above the horizontal line is a premise and whatever is below the line is a conclusion. The symbol “ \therefore ” means “entails,” or “implies” in a broad sense (i.e., it is valid for any rational agent). The readers may notice that (5.1) is a commonly used argument. When the focus of the study is a physical attribute $X_p = X_{s,t}$, the premises and the conclusions may take a symbolic and/or numerical form, see argument (5.2); the X_p changes across space–time according to physical law, which means that the “premises” are causally linked to the “conclusion” (as we saw in Section 1.2.3, this is a key premise of stochastic reasoning). Measurement and prediction values in Eq. (5.2) are in suitable numerical units.

Table 5.1 Examples of arguments

<p><i>Every Thessalos is a good horserider</i> <u><i>Alkividas is not a good horserider</i></u> (5.1) \therefore <i>Alkiviadis is not Thessalos</i></p>	<p>$X_{s,t} = 0.9^t X_{0,0} + 2.1s^2t$ (Physical law) <u>$X_{0,0} = 1.3$ (Measurement)</u> (5.2) $\therefore X_{2,1} = 9.57$ (Law prediction)</p>
<p><i>All Romans spoke Latin</i> <u><i>Nero spoke Latin</i></u> (5.3) \therefore <i>Nero was Roman</i></p>	<p><i>All Romans spoke Latin</i> <u><i>Descartes spoke Latin</i></u> (5.4) \therefore <i>Descartes was Roman</i></p>
<p><i>All Italian – Americans are tall</i> <u><i>Danny DeVito is an Italian – American</i></u> (5.5) \therefore <i>Danny DeVito is tall</i></p>	<p><i>All men are mortal</i> <u><i>Danny DeVito is a man</i></u> (5.6) \therefore <i>Danny DeVito is mortal</i></p>

Table 5.2 Possible valid argument forms

Premise	Conclusion
True	True
False	False
False	True

In terms of logic, an argument may be concerned with a number of things. It could be for or against a specific thesis, suggest a solution of a problem, or lead to a novel result. In evaluating an argument one is basically interested about two items: (i) Are the premises true? (ii) Assuming that the premises are true, what kind of support do they offer to the conclusion? Although Element *i* is not the business of logic, it is of great concern in scientific investigations. Element *ii*, on the other hand, is definitely the business of logic. Valid argument is one that cannot have true premises followed by wrong conclusions (i.e., if the premises are true then one is assured that the conclusion is also true). Three classical premise-conclusion combinations associated with a valid argument are shown in Table 5.2. The word “possible” in the legend of Table 5.2 implies that a true premise and a true conclusion are not, by themselves, enough to have a valid argument; it must also hold that to assert the premise and deny the conclusion would involve a contradiction (i.e., it will be logically inconsistent). It is instructive to consider the arguments (5.3)–(5.4) in Table 5.1. Neither of these arguments is necessarily a logically valid argument (even when the premises are true, one is not assured that the conclusion is also true). As a matter of fact, it is on historical grounds that one can say that the argument (5.3) is valid (Nero was Roman), whereas the argument (5.4) is wrong (Descartes was not Roman). In fact, when the premises are indeed true (which is not the task of logic but of science, history, etc., to confirm) the argument is more than valid – it is sound. Consider the argument (5.5) in Table 5.1. This is a valid but not a sound argument (because, obviously, the premise that “All Italian-Americans are tall” is not true). Now consider the argument (5.6) in Table 5.1. This is a valid and sound argument (both premises are true in the real-world).

Rather simple arguments like the above can offer insights concerning sound reasoning that can deepen conscious awareness and improve one’s capacity for experience. As we will see later, these qualities play a key role in the development of an IPS that accounts for conditions of in situ uncertainty. Possible insights include the following: (a) A rigorous formalization may guarantee general logical validity but not substantive soundness (scientific or otherwise). An argument may be perfectly valid from a purely formal viewpoint, and yet make no sense from the viewpoint of science or even common sense. Hence, one needs more than pure logic to establish the truth of many real-world arguments. (b) It is doubtful that most real-world arguments fit the strict “premises-conclusion” formalization. Instead, there is considerable uncertainty about several aspects of the premises (e.g., physical law parameters and associated measurements are often uncertain). The “theory–evidential support” relationship is not as definite as the formalization may assume (e.g., general relativity theory is assumed valid in a wider physical domain than that covered by the available data). (c) The logical process used in Arguments (5.5)–(5.6) offers a complete

confirmation, whereas the process used in (5.3)–(5.4) only provides a partial confirmation. It is noteworthy that many everyday life arguments are based on the latter rather than the former logical process. Below, I will first examine the two traditional reasoning modes, deduction and induction, and then will briefly discuss hypothetico-deduction, which is a reasoning mode that became popular mainly during the last century. Some of the *pros* and *cons* of the three modes will be pointed out as well.

5.2.1.2 Deductive Reasoning

Deduction or deductive reasoning is reasoning from the general to the particular or less general. It evaluates the arguments on the basis of validity, i.e., it allows only valid arguments. The premises, if they were true, guarantee the truth of the conclusion, which means that deductive inferences preserve truth. For illustration purposes, let $X_{p_i}, Y_{p_j}, Z_{p_k}$ etc. denote space–time attributes with possible realizations $\chi_{p_i}, \psi_{p_j}, \zeta_{p_k}$ etc. ($p = (s, t)$). The attribute realizations χ_{p_i} and ψ_{p_j} may be linked by means of a causal relationship in a physical continuum (Fig. 5.1). The symbol “ \neg ” denotes negation (e.g., $\neg\chi_{p_i}$ means that it is not the case that the realization χ_{p_i} is true). The symbol “ \wedge ” denotes conjunction ($\chi_{p_i} \wedge \psi_{p_j}$ means that both the realizations χ_{p_i} and ψ_{p_j} are true). The symbol “ \vee ” denotes disjunction ($\chi_{p_i} \vee \psi_{p_j}$ means that either χ_{p_i} or ψ_{p_j} is true). The symbol “ \rightarrow ” denotes implication ($\chi_{p_i} \rightarrow \psi_{p_j}$ means that if χ_{p_i} is true, then ψ_{p_j} is true).³ The symbol “ \leftrightarrow ” denotes equivalence ($\chi_{p_i} \leftrightarrow \psi_{p_j}$ means that χ_{p_i} is true if and only if ψ_{p_j} is true).⁴ The symbol $\chi_{p_i} \leftrightarrow \psi_{p_j}$ also implies that $\chi_{p_i} \wedge (\neg\psi_{p_j})$ is a contradiction ($\chi_{p_i} \wedge (\neg\psi_{p_j}) \leftrightarrow \ell$), whereas $\chi_{p_i} \vee (\neg\psi_{p_j})$ is a tautology, ($\chi_{p_i} \vee (\neg\psi_{p_j}) \leftrightarrow \tau$). The symbol “ $\langle \cdot | \cdot \rangle$ ” denotes that whatever is on the right of the vertical line has the property on the left of the line. If A is a set of realizations χ_{p_i} , the $\langle \Theta | A \rangle$ and $\langle \Theta | \chi_{p_i} \rangle$ denote, respectively, that the set A or just a realization χ_{p_i} , has the property Θ ; lastly, the symbol “ \in ” means “belongs to.” Logic operators can be combined in different ways leading, to a variety of deductive reasoning results that are not always obvious a priori. In Chapter 6, we will

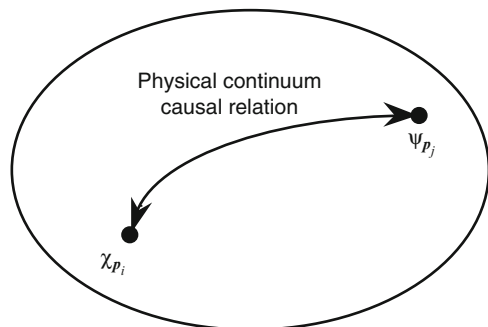


Fig. 5.1 Realizations of two different attributes linked by a physical continuum

³ What is asserted by implication is that $\neg(\chi_{p_i} \wedge (\neg\psi_{p_j}))$, i.e., it is not the case that χ_{p_i} and not ψ_{p_j} .

⁴ Equivalence is a strong logic operator that means the same as $(\chi_{p_i} \rightarrow \psi_{p_j}) \wedge (\psi_{p_j} \rightarrow \chi_{p_i})$.

Table 5.3 Deductive reasoning rules in terms of attribute realizations⁵

Modus tollens	Modus ponens	Simplification	Conjunction
$\chi_{p_i} \rightarrow \psi_{p_j}$	$\chi_{p_i} \rightarrow \psi_{p_j}$	$\chi_{p_i} \wedge \psi_{p_j}$	χ_{p_i}, ψ_{p_j}
$\neg\psi_{p_j}$	χ_{p_i}	$\therefore \chi_{p_i}$	$\therefore \chi_{p_i} \wedge \psi_{p_j}$
$\therefore \neg\chi_{p_i}$	$\therefore \psi_{p_j}$		
Absorption	Excluded middle	Disjunctive syllogism	Constructive dilemma
$\chi_{p_i} \rightarrow \psi_{p_j}$	$\chi_{p_i} \rightarrow \psi_{p_j}$	$\chi_{p_i} \vee \psi_{p_j}$	$\chi_{p_i} \vee \psi_{p_j},$
$\therefore \chi_{p_i} \rightarrow (\chi_{p_i} \wedge \psi_{p_j})$	$\neg\chi_{p_i} \rightarrow \psi_{p_j}$	$\neg\chi_{p_i}$	$(\chi_{p_i} \rightarrow \zeta_{p_k}) \wedge (\psi_{p_j} \rightarrow \omega_{p_q})$
	$\therefore \psi_{p_j}$	$\therefore \psi_{p_j}$	$\therefore \zeta_{p_k} \vee \omega_{p_q}$
Contradiction	Addition	Hypothetical syllogism	Direct generalization
$\chi_{p_i} \rightarrow \psi_{p_j}$	χ_{p_i}	$\chi_{p_i} \rightarrow \psi_{p_j}$	$\langle \Theta \mid A \rangle$
$\chi_{p_i} \rightarrow \neg\psi_{p_j}$	$\therefore \chi_{p_i} \vee \psi_{p_j}$	$\psi_{p_j} \rightarrow \zeta_{p_k}$	$\chi_{p_i} \in A$
$\therefore \neg\chi_{p_i}$		\dots	$\therefore \langle \Theta \mid \chi_{p_i} \rangle$
		$\psi_{p_q} \rightarrow \omega_{p_m}$	
		$\therefore \chi_{p_i} \rightarrow \omega_{p_m}$	

see how these logic operators (as well as the rules of Tables 5.3 and 5.4) can be considered in a stochastic logic milieu in conditions of uncertainty.

Table 5.3 provides a useful list of deductive argumentation rules in terms of attribute realizations. The same rules are valid if the attribute realizations are replaced with statements A, B, C, \dots of everyday language. Deductive reasoning is defined in a very precise way: it is the kind of reasoning in which it is logically impossible for the premises to be true and the conclusion false. According to Karl Popper (1963: 51), “The role of deductive logic reasoning remains all-important for the critical approach. . . because only by purely deductive reasoning is it possible for us to discover what our theories imply, and thus to criticize them effectively.” Mathematics is based on deductive reasoning, which is why mathematics possesses all the *pros* and *cons* of this mode of reasoning. In the case of deduction, the conclusion asserts no more information than is asserted in the premises, and generally has nothing to say about the validity of these premises per se (which is the business of science).⁶ In fact, the deductive process is so precise albeit mechanical and essentially content-free that Bertrand Russell once emphatically wrote that

Pure mathematics consists entirely of such asseverations as that, if such and such a proposition is true of anything, then such and such another proposition is true of that thing. . . It’s essential not to discuss whether the proposition is really true, and not to

⁵ Otherwise said, these are realizations of the spatiotemporal random field model (Section 5.3 below).

⁶ Whereas in induction the conclusion goes beyond, i.e. “amplifies,” the content of premises (see below).

Table 5.4 Inductive reasoning rules in terms of attribute realizations

Partial confirmation $\chi_{p_i} \rightarrow \psi_{p_j}$ ψ_{p_j} <hr/> $\therefore \chi_{p_i}$ is probable	Partial rejection $\chi_{p_i} \rightarrow \psi_{p_j}$ $\neg \chi_{p_i}$ <hr/> $\therefore \neg \psi_{p_j}$ is probable	Causal generalization ψ_{p_j} follows χ_{p_i} $\therefore \chi_{p_i}$ is probably the cause of ψ_{p_j}
Analogy $\langle \Theta_i \chi_{p_i}, \psi_{p_i} \rangle, i = 1, \dots, N - 1$ $\langle \Theta_N \chi_{p_N} \rangle$ <hr/> $\therefore \langle \Theta_N \psi_{p_N} \rangle$ is probable	Simple enumeration $\langle \Theta \chi_{p_i} \rangle \left\{ \begin{array}{l} \chi_{p_i} \in A \\ (i = 1, 2, \dots, N) \end{array} \right.$ <hr/> $\therefore \langle \Theta A \rangle$ is probable	Statistical generalization $\langle \Theta S_i \rangle \left\{ \begin{array}{l} S_i \subset \Omega \\ (i = 1, 2, \dots, N) \end{array} \right.$ <hr/> $\therefore \langle \Theta \Omega \rangle$ is probable

mention what the anything is of which it is supposed to be true. . . If our hypothesis is about anything and not about some one or more particular things, then our deductions constitute mathematics. Thus mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true.

The take-home message is that one should be aware of the *seduction-by-deduction* temptation, since in many cases a direct, uncritical implementation of deductive reasoning in real-world applications may be like using both feet to test the depth of the river.

5.2.1.3 Inductive Reasoning

Induction is reasoning from the particular to the general. It evaluates the arguments on the basis of *probability* (may allow invalid arguments that are, though, highly probable arguments on the basis of the premises). The premises, if they are true, make probable the truth of the conclusion. Accordingly, induction includes argument forms in which the conclusion does not follow necessarily from the premises (as is the case of valid deductive reasoning), but, instead, is inferred as likely. Otherwise said, inductive reasoning assures one that the conclusion is likely, but not that it is certain, and it analyzes risky arguments using probabilistic statements. There exist several classifications of inductive reasoning. One classification distinguishes between induction by enumeration and induction as inference to the best explanation or abduction. In *enumerative* induction, a conclusion is derived on the basis of a large and representative attribute sample. In *abductive* inference, a conclusion concerning one thing is obtained as the best explanation of something else. In other words, the basic difference between enumeration and abduction is that, while the former proceeds from a large and representative sample to an unrestricted conclusion, the latter proceeds from a single observed attribute or phenomenon to the explanation of another attribute or phenomenon. Abduction is frequently employed in scientific investigations; e.g., although electrons themselves cannot be seen, scientists conclude that they exist since such a conclusion provides the best possible explanation of certain observations. Table 5.4 gives a list of inductive rules. Inductive arguments are *partial confirmation* arguments to which

one can assign probability values that depend on the available knowledge, i.e., given the historical knowledge available, the probability that Nero was Roman tends to one, whereas the probability that Descartes was Roman tends to zero. In less developed fields the violation of the reasoning rules of Tables 5.3 and 5.4 frequently leads to problematic results. In clinical research, e.g., the probabilistic nature of the inductive rules in Table 5.4 is often ignored, and the rules are misinterpreted as deductive. The matter will be studied in Section 6.1, after the random field concept is introduced in Section 6.3 that follows. As in Table 5.3, the inductive rules of Table 5.4 remain valid if the attribute realizations are replaced with statements.

Epicureans have held that there exist shortcuts to happiness, but induction is not one of them. As it turns out, the direct, uncritical implementation of pure or naive induction in scientific research can be problematic. David Hume (Section 2.2.9) was probably the first to put into question the legitimacy of pure induction, due to its circularity: the only grounds we have for trusting induction are circular, in the sense that inductive inferences are justified on the basis that these inferences have worked in the past. Remarkably, one of the best-known responses to Hume's challenge is one of desperation: as long as induction works, one can ignore any circularity problems. This is an inadequate argument, of course, that essentially applies to everything under the Sun. And if this is the best argument pure induction can come up with, then too bad for pure induction. Nevertheless, even this simplistic argument is not problem-free in its implementation: What is the meaning of the term "works" in the setting of the above argument? Under what special conditions the argument applies? When pure induction fails, what we learn about the source of its failure? Its inability to convincingly respond to these and similar questions has caused many scientists to seriously doubt the effectiveness of pure induction. Sir Peter Medawar (1969: 11) jokingly remarked that, "If anyone working in a laboratory professed to be trying to establish laws of Nature by induction, we should begin to think he was overdue for leave." Surprisingly, some empirical data analysts still remain in an unconscious bondage to outdated practices of pure induction that have been widely repudiated or otherwise allowed to fade away (see, also, Sections 8.2.2 and 9.4).

The above considerations by no means imply that pure induction has no place in scientific inquiry, rather its implementation makes sense in certain special cases that must be carefully considered. Most investigators would agree that in real-world studies one rather employs valid combinations of inductive and deductive elements; e.g., induction is used in the determination of premises (first stage), and the verification of conclusions (third stage), whereas deduction is used in the derivation of conclusions from premises (second stage). Due to its importance in scientific inquiry, the matter is discussed in other parts of this book.

5.2.1.4 Hypothetico-Deductive Reasoning

The hypothetico-deductive mode of reasoning is as follows: a hypothesis or theory is formulated concerning a problem, its consequences (e.g., predictions) are worked out, and then tested by means of observations and/or experiments. A test that could and does run contrary to the consequences of the hypothesis or theory is taken as a

falsification of the hypothesis or theory (Popper 1963; see also Section 1.1.2). On the other hand, a test that could but does not run contrary to the hypothesis or theory corroborates the hypothesis or theory. In hypothetico-deductive reasoning, a mental entity (hypothesis, theory, or solution) needs to be testable in some definite way, i.e. be capable of proven wrong (falsified) under certain conditions, in which case the entity is termed falsifiable. A falsifiable entity is provisionally accepted until it is falsified.

Popper claimed that for a construct to be scientific, it must satisfy the conditions of the above framework. As considered by him, falsification demands absolute specificity, in which case probabilistic statements are not directly falsifiable. The statement “It will probably rain in Paris tomorrow,” e.g., is not directly falsifiable in the above sense, because it is not a clear-cut statement. The latter is the case of mathematical statements, since they are tautological (proving mathematical theorems involves reducing them to tautologies, i.e., reducing the negative to a contradiction). The above imply some limitations of both, the conceptual framework of falsification and its practical usefulness. Imre Lakatos (1976, 1978a, b), e.g., argued that there is no falsification before the emergence of a better theory – theories and models are more often repaired than they are refuted.

5.2.2 *Transition to Stochastic Thinking*

The preceding discussion of reasoning modes provides a starting point from which to interpret as significant the conceptual gaps in standard logic between formal rules and in situ reality. Undoubtedly, the implementation of a reasoning mode in most in situ situations should involve the notions of probability and uncertainty. Given the multisourced in situ uncertainty, failing to include a suitable probability theory in the scientific field can be an obstacle to the field’s progress. In this spirit, Paul W. Glimcher (2004: 177) maintained that, “The fundamental limitations which neurobiology faces today is a failure to adequately incorporate probability theory into the approaches we use to understand the brain.”

5.2.2.1 *A Slippery Affair and Its Psychology Connections*

Having said that, it must not escape the readers’ attention that reasoning in terms of probabilities can be a slippery affair. For Charles Sanders Peirce, “This branch of mathematics is the only one, I believe, in which good writers frequently get results entirely erroneous.” In a similar vein, George N. Schlesinger (1991: 16) writes:

The susceptibility to error is caused by allowing oneself to be guided too much by intuition and common sense. In probabilistic reasoning, more often than elsewhere, things are not what they seem, and untutored innate intelligence may frequently prove an unreliable guide.

These probability features are sometimes so difficult to comprehend that practitioners armed with only a superficial knowledge of probability theory, often make nonsensical claims (see, also, Sections 6.1, 6.3 and 9.4). Jeffrey S. Rosenthal (2006) offers some insight why human intuition is often very bad in guessing probabilities. Ola Svenson (2008) gives a psychological perspective on why in many cases human intuition is completely wrong. Furthermore, a few decades ago, Amos Tversky and Daniel Kahneman (1973, 1982) published some results suggesting that people have serious difficulties with probabilistic reasoning. They claimed that much of people's thinking under conditions of uncertainty is based merely on heuristics⁷ (Workman and Reader 2004). Tversky and Kahneman attributed the poor performance of the study participants to their using heuristics: representative bias (participants are misled by what seems to be representative of the real-world), and base-rate neglect (participants failed to take prior probabilities into account). The response of the evolutionary psychology school was that, while it is true that people show rather poor intuitions when making decisions under conditions of uncertainty, however, this is due to the way things are presented to them. In many cases, e.g., people are presented with problem formulations that their minds are not evolutionary adapted to cope with. In particular, Leda Cosmides and John Tooby (1996) presented some results suggesting that when a problem is presented to a group of study participants in terms of single-case probabilities, most of them derive an incorrect solution. However, when the same problem is presented to the same group of participants in terms of frequencies, the majority of them derive the correct solution. The explanation of this apparent paradox is that while our ancestors have gained considerable benefits from evolving frequency-sensitive mechanisms, they have found little use for single-case mechanisms, in case the latter had been evolved. Two main conclusions could be drawn concerning the above views that seek to explain an agent's difficulties with probabilistic thinking: the heuristics perspective focuses on the irrationality of human reasoning, whereas the evolutionary perspective properly emphasizes its adaptive rationality. The former perspective seeks explanations in terms of proximate mechanisms, whereas the latter rather stresses ultimate explanation. A matter of significant interest is to assess how these different perspectives can affect the IPS approach that the agent chooses to use under conditions of uncertainty. This includes the solution of in situ problems in the physical and health sciences alike.

In an attempt to deal effectively with the state of affairs described above, stochastic reasoning requires from the investigator considerable levels of introspection and interpenetration, in addition to formal derivations. Unlike the mainstream paradigm, in the stochastic reasoning milieu, uncertainty characterizes not only inductive but deductive modes of argumentation too. Accordingly, logical derivations are not certain but have realistic probability values assigned to them. For reasons discussed in Sections 1.2.3, 4.3.1 and 4.3.2, uncertain attributes of a real-world system are usually linked to other uncertain attributes via physical or

⁷ Heuristics are short-cut solutions to a problem, which, while are often quick and easy (or, in some cases, "quick and dirty") to implement, do not guarantee a correct solution.

logical relations. Even when an attribute is known with certainty, in situ relations most often link it to other attributes to which they assign probability values. Only in the rare case that the strict dependency of deductive reasoning connects one attribute to another the certainty of the first can be transferred to the second. As a consequence, terms like “probable,” “causation,” “implication,” “contradiction,” and “conditional” need to be re-interpreted in the appropriate contextual settings.

5.2.2.2 The Relationship Between Logic and Psychology

Continuing our discussion of the role of psychology in human reasoning, I will start with a real-world example that is paradoxical and at the same time somehow entertaining. The statements $P = \text{In favor of family values}$ and $C = \text{In favor of assault weapons}$ logically should represent mutually inconsistent or exclusive possibilities. Said otherwise, occurrence of one of them makes the occurrence of the other highly improbable. Yet public opinion polls show a clear shifting of American attitude toward $P \wedge C$, an astonishing result that belongs to the sphere of psychology rather than logic. As far as the relationship between logic and psychology is concerned, the readers are reminded that the two contradictory viewpoints traditionally considered are: (a) logic as a tool for exploring standards of human reasoning (*philosophical* viewpoint) and (b) logic as a quarry for extracting hypotheses concerning human thought processes (*psychological* viewpoint). Viewpoint *a* has a normative structure, whereas Viewpoint *b* has a rather descriptive structure. Concerning Viewpoint *a*, it is known that mathematical (deterministic) logic assumes a closed system with controlled environment. And even within this system, logic cannot demonstrate whether a possibility expresses an objective truth or not. It can only prove the validity of a possibility relative to other possibilities that an agent already knows to be true or false. Despite the usefulness of Viewpoint *b* in certain psychological investigations, it is considered of rather limited value outside these investigations (Macnamara 1994).

In view of the above considerations, the objective of stochastic reasoning is to reshape the relationship between logic and psychology in ways that enhance the experience of the investigating agents involved: stochastic reasoning seeks a fruitful synthesis of Viewpoints *a* and *b* that accounts for the fact that standard logic does not constitute the entire thinking process, but is only part of it; the synthesis incorporates uncertainty due to multiple sources (linked to the theory of knowledge or reality itself); and also confronts the fact that an agent’s thinking in a real-world situation is a much more sophisticated process than the mechanistic scheme assumed by standard logic. In a sense, then, stochastic reasoning suggests that logic and psychology mutually constrain each other in an analogous way that mathematics and physical sciences constrain each other. Logic, e.g., could provide a rigorous language in which to express mental states and formulate these expressions in mathematical terms, which can be used in IPS under conditions of in situ uncertainty and space-time heterogeneity (Chapter 3). Human understanding and creativity are often richer than standard logic, indeed, which is

content-insensitive and ignores that agents operate in an open system rather than in the idealistic closed system of formal logic. In so far as understanding thinking changes thinking, stochastic reasoning needs to substantially enrich and even modify formal logic, if it is to incorporate in situ situations that currently elude it. Stochastic reasoning expresses a cognitively general viewpoint (where the agent can only know that there exist some entities that have a certain feature), rather than a cognitively specific viewpoint (in which the agent definitely knows the exact entities that have this feature). It is more reasonable, e.g., to claim that due to its doctors' high qualifications, most of the patients who have open-heart surgery in the *St. Therese of Liseux* hospital survive (one may even be able to provide probabilities of survival for individual patients), rather than to claim to know exactly which patients will survive. In some special cases the *cognitive general* may be reduced to the *cognitive specific*. For example, when one knows with certainty all the input parameters and coefficients of a stochastic law (Section 5.5.3 below), the associated probability distributions reduce to single values, and the solution of the law becomes deterministic. But, this is a rather unlikely scenario in the vast majority of in situ situations. Last but not least, stochastic reasoning is purposive, which means that it delivers the agent's values and principles. This is a definite advantage, since any kind of reasoning, regardless of how rigorous and sound it is, if it lacks values, is of limited use or even dangerous in human affairs.

5.2.2.3 Some Distinctions

For procedural purposes, it is important to distinguish between three key fields: probability theory, statistics, and stochastics. As described in Collani (2008), "Probability theory develops 'mathematical concepts' independently of their usefulness. Statistics develops methods for analyzing large data sets in order to detect stabilities;" whereas "Stochastics represents a conceptual and theoretical basis covering all aspects which are involved in the scientific process of making predictions." Failing to acknowledge the key differences between these fields can lead to misconceptions, such as that stochastics is merely akin to descriptive statistics, or that spatial statistics includes both stochastic modeling and geostatistics (Myers 2006).

Noteworthy limitations of mainstream statistics that have been pointed out in the literature include (e.g., Wang 1993; Sivia 1996; Christakos 2000; Hyman 2006): (i) It is dominated by symbolic thought and not the free exchange between meaning and the empirical world, or the creative thought that is open to the new and risky. (ii) Substantive inadequacy of assumptions, like statistical independency and stability, which do not account for the physics of space–time. (iii) Lack of rigorous mechanisms to incorporate important forms of core knowledge (natural laws, primitive equations, social structures, etc.). (iv) Many tests entail serious logic problems and are often irrelevant to the objectives of the study (e.g., a statistical test states the probability of the observation given that a null hypothesis is true, whereas scientific

investigation seeks the probability that the null hypothesis is true given the observation). (v) Analysis often relies on a collection of data processing recipes and number-crunching software (pattern fitting, trend projection, regression analysis, copula technology, etc.) that are introduced on the basis of mere convenience than sound reasoning and scientific insight—which is probably why Thomas Mikosch (2006b: 61) made a rather pessimistic comparison: “Living in the twenty-first century, we stand on the shoulders of giants such as Kolmogorov, Levy, Wiener and Cramer who did things not just because they could or because it was convenient.”

To avoid the above limitations, stochastic reasoning assumes a very different conceptual structure than mainstream statistics. It focuses on deep theory (founded on natural laws, phenomenological representations, and epistemic principles) that enhances its scientific content and makes it a central force in the realistic study of natural systems. This is the kind of reasoning that can incorporate, *inter alia*, the sophisticated mathematics of stochastics, which has been very successful in the study of such diverse phenomena as contaminant transport in environmental media, atmospheric turbulence, electromagnetic wave propagation through atmosphere, large-scale systems linked to disease and mortality, epidemic propagation, embryonal formative processes, and organic molecules organizing themselves into organisms of increasing complexity through random chemical processes. Stochastic reasoning is endorsed with a solid theoretical background, a sound methodology, and a useable set of tools to study complex in situ situations associated with several possible “scenarios” of how a system or attribute might change in space–time under conditions of uncertainty, rather than a single yet unrealistic “scenario.” As a matter of fact, due to its inevitably high level of sophistication, working in the field of stochastic reasoning requires a proportionally high level of intellectual effort on behalf of the investigator, who should not expect to be rewarded with a trip to the exotic Rondônia.⁸ Instead of the mouthwatering *Caruru do Pará*, pure intellectual satisfaction most probably will be the theorist’s only reward.

5.2.2.4 Interpretive Matters

In sum, nothing less is asked of an investigator today than to be at the same time within and outside things. The challenge of using stochastic reasoning in situ is often not in its formal component, but in the validity of its interpretive component in the specific application that goes beyond pure mathematics into the realms of physical knowledge and empirical observation. Interpretation issues are relevant when one needs to establish correspondence rules between natural attributes and formal mathematics that describe them, to measure and test the formal structure or to justify the methodological steps. This does not intend to imply that the two components are totally independent or merely linked by correspondence rules. Instead, the formal and the interpretive form an integrated whole. As such, the

⁸ Region in Brazil that has been the theater of NASA’s field-data acquisition campaigns.

fruitful interaction of formal and interpretive investigations plays a crucial role in the successful application of stochastic reasoning in real-world IPS. The essential connection between formal and interpretive components has been astonishingly productive, in both ways: formal techniques provide the means for understanding a phenomenon beyond sense perceptions, and interpretive investigations lead to new and more powerful formal techniques.

In short, stochastic reasoning lies at the interface of logic and empirical evidence, with strong ties to philosophy, linguistics, sociology, psychology and cognitive science. In the human inquiry milieu, stochastic reasoning acts as an intellectual catalyst that shows how different topics ran naturally into each other. Accordingly, stochastic reasoning needs to conceal any antagonistic demands of in situ observation and theory-based interpretation, which implies that the meaning of logic operators may change in the stochastic reasoning context. The strict determinism of the formal logic operators (\wedge , \vee , \neg , \rightarrow , \leftrightarrow) discussed in the previous sections is replaced by the reasonable indeterminism of stochastic reasoning. In other words, the meaning of the operators is re-interpreted to account for the uncertainty of the premises, the conclusions, and the operator-based process itself. For example, in formal logic, $\chi_p \wedge \psi_p$ denotes that both attribute realizations χ_p and ψ_p are definitely true. But in stochastic reasoning, $\chi_p \wedge \psi_p$ means: “Agent’s assertion that χ_p is true and the agent’s assertion that ψ_p is true.” These assertions are not definite but, rather epistemic, i.e., they are conditioned on the available knowledge, which means that to each assertion (or, more generally, to any combination of assertions) one can assign a probability value. Also, instead of explaining a fallacy by trying to show that a valid realization χ_p of the attribute X_p can cause an invalid realization ψ_p of another attribute Y_p (standard logic), it makes more sense to show that a probable realization χ_p can cause an improbable realization ψ_p (stochastic reasoning). This approach may involve natural laws that link X_p and Y_p , incomplete yet valuable databases, and other sources of knowledge under conditions of uncertainty. In our next example the space–time attribute X_p denotes the average daily temperature. Consider an agent’s prediction that the X_p value at $p =$ (San Diego, September 19, 2011) will be $\chi_p = 26.3^\circ$ with probability $P_{KB}[X_p = 26.3^\circ] = 0.6$.⁹ This probability refers to the agent’s assertion (based on the available knowledge base, KB) that the temperature value $\chi_p = 26.3^\circ$ has probability 0.6, rather than the standard claim that the probability of the temperature value above is 0.6. Said otherwise, since the term “probability” is used by the agent to talk about the attribute realization $\chi_p = 26.3^\circ$, it is part of the stochastic reasoning metalanguage. I will revisit the subject in Chapter 6, after we first introduce in the next section another key element of stochastic reasoning, namely, the spatiotemporal random field concept.

⁹ In mathematical terms, stochastic reasoning views the temperature attribute as a spatiotemporal random field, see [Section 5.3](#) that follows.

5.3 The Spatiotemporal Random Field Concept

Stochastic reasoning involves a variety of concepts – abstract and intuitive, formal and interpretive, epistemic and ontic, mathematical and physical. And, equally important, it involves interactions between these concepts that honor a capacity for experience, engage consciousness, and offer new ways of imagining the world. As such, the subject of stochastic reasoning is replete with theoretical issues. One of the main theoretical concepts is the *spatiotemporal random field (S/TRF)*. Let us start with the thought process that leads to the formulation of the S/TRF as currently conceived.

5.3.1 *The Possible Worlds Representation: Epicurus, Leibniz, and Voltaire*

An influential school of thought promotes the study of Nature in terms of the so-called *possible worlds representation (PWR)*. According to PWR, the agent's mental conception of Nature should involve many possible worlds (or realizations). The world that is currently observable by means of the agent's cognitive means is just one of the many possible worlds – the worlds of possibilities concerning what one may find if Isis' veil is ever lifted, metaphorically speaking.

The PWR idea can be traced back in the third and fourth centuries BC Epicurean teachings “on the plurality of worlds (*κόσμοι*).” According to Epicurus, it is possible to propose multiple explanations of a phenomenon, each of which must agree with appearances. In a famous letter to Herodotus, Epicurus writes that, “there are infinite worlds both like and unlike this world of ours” (Konstan, 1972).¹⁰ Gottfried Leibniz viewed the PWR as ideas in the God's mind who accordingly created the currently existing world to be “the best of all possible worlds.” A well-known reference to the religious relevance of this characterization is due to François-Marie Arouet, better known as Voltaire. Voltaire (2005) used the characterization in his novel *Candide* to ridicule the theologians' claim that divine justice was served by the great Lisbon earthquake on All Saints' Day, 1755 (over 30,000 lives were allegedly lost). In modal logic, the PWR is called modal actualism, which assumes that possible worlds exist as abstract entities that are distinguished from the actual world. Another PWR interpretation is modal realism, which assumes that the possible worlds exist just as surely as the actual world does. In Epibrainatics (see, also, Section 3.5.1), the PWR posits the existence of worlds within our mentally extended senses that must connect or relate with our own. PWR is inherent in the imaginative construction approach of human inquiry. An agent constructs an approach to reality (rather than reality per se), which relies on agent's coherent and creative imagination. In the context of the

¹⁰ A relevant source is Lucretius' poem on Epicureanism titled *De Rerum Natura* (“On the Nature of Things”; Book 2, ll. 1023–1089).

imaginative construction approach, one uses an instrument to see a world (say, W_2) that one can never see with the naked eye (which can only see the world W_1). The action of building and using the instrument implies that the agent assumes that there indeed exists a world to be seen. The PWR realizations in connection with the real-world situation have two main features: (a) the realizations are consistent with the physical, practical, or logical conditions of the specified problem and (b) they have different probabilities of occurrence, depending on the epistemic situation of the agent. Feature *a* may be linked to different IPS kinds (Section 2.3.4), i.e., solutions that are physically, practically, or logically possible (the reader is reminded that something that is not physical is only accessible mentally). And Feature *b* may be associated with the agent's mode of thinking, worldview, system of beliefs, cognitive means, and knowledge sources.

5.3.2 Causality–Randomness Interaction

According to many historians, Aristotle was the first to combine probability with necessity (Section 4.3.3). Aristotelian insight underlies the basic idea of stochastic reasoning concerning the co-existence and interaction of randomness with causality in the quantitative description of attributes and systems that unfold *spatially* with the course of *time*. In addition, stochastic reasoning includes a group of spatiotemporal models with attractive features that reflect the epistemic fact that agents are pattern-forming creatures: they like to, have to, or do connect things. As emphasized by Gerald M. Edelman (2006: 58), brain is a selectional system that operates *prima facie* by pattern recognition. In this respect, the S/TRF model plays an important role that aims at studying the uncertain properties of a system as a whole, and connecting them to causal relations and space–time patterns. The S/TRF model is briefly reviewed below; a detailed presentation of the mathematical theory and its various applications can be found in the literature (Christakos 1991a, b, 1992; Christakos and Hristopulos 1998).

5.3.2.1 Agents Who Are Not Mute in Their Souls

Consider an in situ attribute that varies in a composite space–time domain (e.g., air pollution concentration, water level, epidemic mortality, soil property, land use, poverty level, or commodity price). In light of the PWR, the S/TRF $X_p = X_{s,t}$ is represented as

$$X_p \rightarrow \begin{cases} \chi_p^{(1)} = (\chi_{p_1}, \dots, \chi_{p_k})^{(1)} \\ \chi_p^{(2)} = (\chi_{p_1}, \dots, \chi_{p_k})^{(2)} \\ \vdots \\ \chi_p^{(R)} = (\chi_{p_1}, \dots, \chi_{p_k})^{(R)} \end{cases}; \quad (5.7)$$

i.e., X_p is viewed as the collection of all possible space–time distributions or realizations, $\chi_p^{(j)}$ ($j=1, \dots, R$) at the space–time points $p = (p_1, \dots, p_k)$ of the attribute represented in terms of the S/TRF. In the random field setting, the attribute realizations $\chi_p^{(j)}$ have some features that are worth noticing: (i) The realizations are consistent with the physical properties, uncertainty sources, and space–time variations characterizing the attribute distribution (i.e., the multiplicity of realizations makes it possible to account for uncertainty sources and, at the same time, to adequately represent the spatiotemporal variation of an attribute). (ii) They have the epistemic quality of corresponding to ways that are consistent with the known¹¹ system properties rather than to all possible ways a system could be represented in terms of formal logic. (iii) They have different chances of occurrence, in general; each realization has a distinct probability to occur that depends on the epistemic condition of the investigator and the underlying mechanism of the in situ phenomenon. The implication of Feature iii is that Eq. (5.7) could be re-written in a more informative way as follows:

$$X_p \rightarrow \begin{cases} \chi_p^{(1)} & \text{with probability } P_{KB}^{(1)} \\ \chi_p^{(2)} & \text{with probability } P_{KB}^{(2)} \\ \vdots & \\ \chi_p^{(R)} & \text{with probability } P_{KB}^{(R)} \end{cases}, \tag{5.8}$$

where the subscript KB denotes the knowledge base used to construct the probability model in the stochastic reasoning setting of Section 5.2.2. According to the above perspective, the agent is in control of possibilities (S/TRF realizations), but not actualities. In other words, the control involves the agent’s mind in a way that the agent could predict what was *likely* to occur, but not what *will* actually occur. In this sense, the future could potentially influence the present as much as the past. That which does not exist in one realization but exists in some other realization shares certain important characteristics with that which actually exists. To understand a phenomenon and assess the actual risks linked to it, one needs to be aware not only of the favorable scenarios but also of the usually much larger number of unfavorable scenarios that did not occur this time, but could occur the next time around.¹² To put it in a literary way, *that which probably isn’t affects what is*.

It is a signature of our times that this sort of multi-thematic integrative thinking cannot be appreciated by those who promote a society with low intellectual

¹¹ “Known” is here associated with human consciousness.

¹² To use a real life scenario, since there are so many conditions that need to be satisfied simultaneously for Tiberius Finamore to be the only survivor of an airplane crash, the fact that he survived makes it tempting to believe that there was another reason that made his survival highly probable a priori (say, God likes Tiberius) rather than to admit that his chances to survive were indeed extremely small (there are many possible realizations that did not favor Tiberius’ survival but are ignored).

standards, and an eye only for the illusive “bottom line.” It is not difficult to imagine just how startling, and even frightening, the PWR idea of an infinity of possible worlds seem to those who have been programmed to think “within the box” and only believe what they can touch. On the contrary, integrative thinking under conditions of uncertainty is the kind of stuff that can be appreciated by those individuals who are not mute in their souls. Understanding, appreciating, and implementing stochastic reasoning and random field theory require a broad and penetrating imagination rather than a dry pedantic brain, and an incisiveness of the mind rather than a formulaic thinking.

5.4 Stochastic Characterization

The mathematical S/TRF theory will be presented to the extent necessary for the purposes of this book. In stochastic reasoning terms, the description of an attribute distributed across space–time concentrates on the web of possible attribute patterns across space–time and what lies beneath them. Correspondingly, the S/TRF model of such an attribute is fully characterized by its PDF, f_{KB} , which is generally defined as

$$P_{KB}[\Lambda(\chi_p, d\chi_p)] = f_{KB}(\chi_p) d\chi_p, \quad (5.9)$$

where $d\chi_p = \prod_{i=1}^k d\chi_{p_i}$ and $\Lambda(\chi_p, d\chi_p) = (\chi_{p_i} \leq X_{p_i} \leq \chi_{p_i} + d\chi_{p_i}; i = 1, \dots, k)$ for all k . When $k=1$, Eq. (5.9) reduces to the special case of a *univariate* PDF; and when $k>1$, the term *multivariate* PDF is used, instead. While the univariate PDFs define the S/TRF at a local scale, the multivariate PDF characterizes it at a global scale. Technically, $f_{KB}(\chi_p) d\chi_p$ could be replaced by the more general $dF_{KB}(\chi_p)$, where $F_{KB}(\chi_p)$ is the corresponding CDF, but at the moment continuous and differentiable CDF are assumed, in which case Eq. (5.9) is sufficient for the goals of our investigation.

5.4.1 The Holy Grail

Equation (5.9) involves both the *content* of the investigator’s thinking process and the *in situ context* of the attribute. Naturally, the construction of f_{KB} on the basis of the available KBs is a critical process with epistemic, cognitive, and psychological characteristics. It is, hence, important that the investigator uses the appropriate f_{KB} interpretation, be aware of the problem space within which the interpretation is valid, and carefully implement content-sensitive logic norms (Section 6.1) that maintain consistency among the KB elements. This being the case, it should come as no surprise that the f_{KB} is considered the *Holy Grail*, so to speak, of S/TRF analysis.

By means of Eq. (5.9), the f_{KB} assigns numerical probabilities to the X_p realizations that evolve between multiple space–time points, see Eq. (5.8). The f_{KB} describes the comparative likelihoods of the various realizations and not the certain occurrence of a specific realization. Accordingly, the PDF unit is probability per realization unit. This may be the time to remind the readers that one should not underestimate the importance of notation. The same entity may be represented using different symbols, depending on the context and the emphasis one wants to assign to the relevant variables. A probability may be denoted as P_X if it is contextually significant to denote that this is the probability function of the S/TRF X_p ; or by P_{KB} if one needs to emphasize that the probability function has been constructed on the basis of a specified KB and that underlying it is a particular methodology and worldview. Similarly, a PDF may be simply denoted as f_{KB} ; as $f_{KB}(X_p)$ if the goal is to emphasize the X_p realizations; or as $f_{X,p}, \mathbf{p} = (p_1, \dots, p_k)$ if it is necessary to indicate that the PDF is a function of the space–time domain. Also, while $f_{X,p} = f_X(p_1, \dots, p_k)$ denotes a multivariate PDF, $f_{X,p_i} = f_X(p_i)$, $i=1, \dots, k$, denotes a set of univariate PDFs.

5.4.2 Multiple Conceptual Layers

In view of Eq. (5.9), multiple realizations of the attribute under consideration are possible *before the event* (i.e., before the actual attribute distribution reveals itself). In the investigator’s mind, the attribute could exhibit one of several possible realizations, until it is observed or measured. Attribute’s probability of exhibiting any particular realization is a measure of how likely it is that the attribute exhibits this realization when it is observed or measured given the agent’s epistemic condition. Plainly speaking, the attribute realizations exist as probabilities (or potentialities), becoming certain only as they are observed or measured. Observation of an attribute realization by a conscious investigator (using the cognitive means available, natural or technical) transfers its state from one of uncertainty into one of definiteness. But this is not the whole story. Just like an onion with its many layers, each of which is attached to the one beneath it, the S/TRF has several conceptual layers, each one possessing some salient features: (a) It assumes the existence of a composite space–time manifold, i.e., space and time constitute an integrated whole than two separate entities; (b) it incorporates spatiotemporal interdependencies and cross-correlations of the attribute distribution expressed by the laws of change; (c) it is of immediate relevance to models that are mathematically rigorous and tractable and, at the same time, logically or physically plausible; and (d) it is capable of generating informative images enabling the determination of important characteristics of the attribute distribution across space–time. When representing an attribute in terms of an S/TRF, the investigator assigns to it a random character and an equally important structural character. Thus, a realization is allowed only if it is consistent with the KBs about the situ attribute and the

investigator's logical reasoning. Clearly, not all S/TRF realizations are equally probable. Depending on the underlying mechanisms, some realizations are more probable than others, and this is reflected in the PDF model of the S/TRF.

It is always enlightening to think in a literary way about a mathematical concept, and view its properties in the light of historical or cultural situations. Accordingly, let us consider two instructive yet very different examples representing extreme cases of parallel worlds. Around 250 BC, King Ptolemy Philadelphus sent 72 Hebrew scholars (six from each tribe of Israel) to translate *Septuagint* (Hebrew Scriptures) into Greek and add them to the Alexandria library. He secluded these men on the island of Phares, where each worked separately on his own translation, without consultation with one another. According to legend, when they came together to compare their work, the 72 copies proved to be identical. This is an extreme case, one must admit, in which all parallel worlds (realizations) reduce to a single world with probability of occurrence equal to one (certainty or determinism). If the *Septuagint* example represents a rather extreme case of "perfectly consistent parallel worlds," the opposite is the case of the second example drawn from contemporary European politics: politicians with radically postmodern features gained fame for their political style based on a set of parallel worlds that often contradicted each other, involved logically inconsistent accounts of events, and had no relation whatsoever to evidential truth and objective facts. These politicians are characterized by their Orwellian twists of the truth. One of their "gifts" is their ability to comfortably generate contradicting worlds (account of events): one world for their voters, another one for lobby interests, another one for activist organizations, and yet another one for foreign leaders.¹³ Unlike the parallel worlds of contemporary politics, the S/TRF worlds of scientific inquiry share a common structure that is determined by what is known about the phenomenon and by the rigorous rules of internal consistency and truth searching.

5.4.3 *Robert Frost's Moment of Choice, and the Case of Paradoxes*

In sum, because it can investigate the different forms of space–time dependency allowed by the available knowledge, the X_p model is able to generate multiple permissible realizations and provide assessments of their likelihood of occurrence. Technically, by combining $f_{X,p}$ with some kind of efficient Monte Carlo simulator, one can comfortably generate numerous X_p -realizations and look at certain of their prevalent features, thus gaining additional intuition to that obtained by studying the analytical expression of $f_{X,p}$, when available.

¹³ Europe's ruling elites (cliques would be nearer the mark) are engaged in policies that go so radically against the wishes of ordinary citizens that the rift is widening between the people and the governing elite.

5.4.3.1 The Road Not Taken

In Section 1.1.3 we suggested that just as poetry does, creative modeling feeds one's imagination with possibilities. Surely, the consideration of all possible realizations provided by the S/TRF model can be very informative and insightful in the investigator's effort to assess both how much one knows and how much one does not know about the in situ situation. But if one needs to make a choice, which realization one should select and why? This is a reasonable question that cannot be answered using technical tools alone, although it can be aided considerably using substantive means (e.g., understanding the underlying physical mechanisms, expertise with the in situ conditions, and logical considerations). A poem, on the other hand, can stir all of the senses, which means that resort to the literary way of expressing one's thinking is intriguing and often conceptually motivating. In his poem *The Road Not Taken*, Robert Frost challenges the reader's imagination with the dilemma:

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.

One road is well traveled – like a high probability random field realization. And another one is less traveled – like a low probability realization. What is then the optimal choice? This depends, the readers may comment, on the agent's grasp of the in situ situation, personal conviction, creative imagination, and well-thought objectives. Selecting the most probable realization may seem a rational decision, but is it always so? The highly improbable realization, when it occurs, can be very consequential. Indeed, as deadly worldwide epidemics and financial disasters have shown, it does not matter how rare an event is, if its occurrence is too costly to bear.

In any case, the take home message is that choices are inevitable. And just like in Frost's poem, one will not know with certainty what the specific choice of a possibility actually implies until one has lived it, until the possibility has been observed and its former probability obtains its maximum value. *Ex animo*, isn't this state of one constantly facing crucial choices and new challenges the essence of an uncertain life?

5.4.3.2 The Case of Apparent Paradoxes

Admittedly, one may get the impression that there are a few conceptual paradoxes linked to the stochastic reasoning implicit in S/TRF analysis. An apparent paradox is that the S/TRF gives answers in terms of possibilities, all considered by the investigator at the same time, which seems unreal. However, one should not forget that we already have a very good non-technical word for the mixture of possibilities that co-exist at the same time: we call it *future*, which is imperfectly known to

humans due to their incomplete knowledge about the in situ phenomenon.¹⁴ Another apparent paradox is that the S/TRF seems to imply that there is an instant awareness between the attribute values across space–time, which seems strange (especially if one is used to work with independently distributed variables). But, this awareness is in the investigator’s mind (epistemic entity), and not in the actual phenomenon (unknown ontic entity). *A posse ad esse*, from possibility to actuality, the S/TRF model allows for the observation effect: when an observation takes place at a specific space–time point, awareness is expressed by a reduced set of possibilities at all points according to the model (in technical terms, this is sometimes called conditional S/TRF simulation).

5.5 About Laws, Power Holders, and Rembrandt’s Paintings

“Laws? Like who the f*** cares?” was the attitude of the people participating in CIA programs in the early 2000s, according to a senior CIA official (Horton 2008: 50).¹⁵ However, as any rational human being (who is not blinded by the extremist dogmas and the superiority complexes of clerkdoms and power holders) knows, laws constitute an important component of organized social life and, as far as this book is concerned, scientific life too. In the latter case, the natural laws are essential ingredients of real-world IPS. As David Novak (2008: 177) puts it, “Human cultures can only avoid the question of natural law when they identify themselves alone with humankind per se and regard all outsiders as devoid of humanity.” The following is a famous passage from John Donne’s 1624 prose *Meditation XVII*:

No man is an island, entire of itself; every man is a piece of the continent, a part of the main. If a clod be washed away by the sea, Europe is the less, as well as if a promontory were, as well as if a manor of thy friend’s or of thine own were. Any man’s death diminishes me, because I am involved in mankind, and therefore never send to know for whom the bells tolls; it tolls for thee.

These lines reflect powerful ideas of the Renaissance era about the interconnectedness of human experience. This interconnectedness extends to any entity (attribute, process, phenomenon, etc.) within a system, since no entity is isolated from the other entities of the system. The breaking of the isolation and the simultaneous formation of interconnectedness is made possible by means of natural laws. For Xie Liu (2003) even literary works are created according to the natural laws of the universe. All entities depicted in these works (people, animals, trees, mountains, etc.) are in accord with the rational principle expressed in words and

¹⁴ On the contrary, determinism describes events as inevitable, effectively depriving humans of a future.

¹⁵ Remarkably, a similar attitude toward human laws was exhibited in President Lyndon Johnson’s blast to a European ambassador in the 1960s (Wittner 1982: 303): “F*** your Parliament and your Constitution . . . If your Prime Minister gives me talk about Democracy, Parliament and Constitution, he, his Parliament and his Constitution may not last very long.”

things with their specific characteristics. At this point, it would be useful to consider yet another classification of the laws used in scientific applications. In Epibraimatics, the reader is reminded, natural laws are linked to the agent's capacity of rational reason to reflect upon the conditions and content of in situ experience.

5.5.1 *Deterministic Laws*

In the case of deterministic causal laws, the value χ_p of an attribute X_p can be calculated at any space–time point $p = (s, t)$, if the relevant boundary and/or initial conditions (BIC) X_0 and coefficients a_i ($i = 1, 2, \dots, k$) are known. In a symbolic form

$$M_X(a_i, X_0, X_p) = 0, \quad (5.10)$$

where M_X denotes a model derived on the basis of a scientific theory. Law (5.10) represents persistent and reproduced features of natural phenomena, whereas X_0 represents individual, contingent, and irreproducible cases of the law's action, in general. Using Newton's laws and the necessary BIC, e.g., one could, in principle, calculate the position and velocity of an object at any time instant. Although scientific laws of the form suggested by Eq. (5.10) are assumed to be generally applicable, they may be of different levels of fundamentality (e.g., Darcy's law is of a lower level than that of Newton's laws). In addition, some of them are observable laws (macroscopic level), whereas some others are not (subatomic level). Certain laws specify the actual mechanism underlying a phenomenon, whereas other laws are purely phenomenological (Table 1.12).

5.5.2 *Statistical Laws*

In the case of a statistical law, the attribute value χ_p cannot be calculated with certainty at any space–time point p , but only its frequency of occurrence

$$F_{X:p}(\chi_p) = n_{\chi_p} N^{-1} \quad (5.11)$$

can be calculated experimentally on the basis of past data (N is the total number of data, and n_{χ_p} is the number of times the value χ_p turned out). When tossing a die, e.g., one cannot predict whether "heads" or "tails" will come up. But, on the basis of past experiments with the die, one can say that among any number of throws, it is expected that about half will turn up "heads." There is no causality present in the statistical law above (say, in the form of the Newtonian laws characterizing the motion of an object such as the die). Methodologically, statistical analysis of this

sort is based on pure induction, which means that it suffers from many of the problems associated with this sort of reasoning (Section 5.2.1).

5.5.3 Stochastic Laws

Now comes the interesting part, as far as stochastic reasoning is concerned. A stochastic law obeyed by the S/TRF X_p has the same symbolic form as Eq. (5.10),

$$M_X(a_i, X_0, X_p) = 0, \quad (5.12)$$

with one key difference: the BIC X_0 and the coefficients a_i ($i = 1, 2, \dots, k$) are now random fields.¹⁶ The (uncertain) causality between the random fields a_i , X_0 and X_p is expressed by the stochastic formulation (5.12). In this setting, a natural law can be put in the form of stochastic equations by admitting that some or all of its constituents are not perfectly known and, hence, they must be represented as random fields. The upshot is clear: while Eq. (5.10) is a law of necessity and Eq. (5.11) is a law of chance, Eq. (5.12) is a law that expresses the dialectics of randomness and necessity. In this sense, a stochastic law is closer to a natural law than to a purely statistical one. Readers have already acquainted themselves with the species growth law of Section 4.3.3, in which the attribute was represented by the stochastic differential equation

$$\frac{d}{dt}X_p - aX_p = 0, \quad (5.13)$$

where a is a known physical coefficient. The attribute's X_0 refers to the point $(s, 0)$ and is random. As a second example consider a quantum system governed by the stochastic Schrödinger law,

$$i\hbar\frac{\partial}{\partial t}\psi_p - \hat{H}\psi_p = 0, \quad (5.14)$$

where i here denotes the imaginary unit, \hbar is the reduced Planck constant, \hat{H} is the Hamiltonian operator, and ψ_p is the wave function.¹⁷ The readers may be reminded that, although Schrödinger's equation is fundamentally stochastic expressing quantum uncertainty, his thinking was not always so, as he once declared in no uncertain words that, "It has *never* happened that a woman has slept with me and did not wish, as a consequence, to live with me all her life" (Mlodinow 2001: 221).

The method that derives the equation of the corresponding PDFs from the stochastic natural law (5.12) is conceptually straightforward, but its practical

¹⁶Note that since Eq. (5.10) is deterministic, the solution $X_p = \chi_p$ has probability 1. This is not the case, however, with Eq. (5.12).

¹⁷A detailed discussion of the underlying physics is beyond the scope of the present discussion, but the interested reader is referred to the numerous volumes on quantum mechanics (e.g., Messiah 1999).

implementation is often not an easy task. In a symbolic form, the stochastic law that $f_{X;p}$ satisfies is generally represented by

$$M_f(f_{a_i}, f_{X;0}, f_{X;p}) = 0, \quad (5.15)$$

where f_{a_i} and $f_{X;0}$ are the PDFs of the random coefficients a_i ($i = 1, 2, \dots, k$) and the BIC (X_0), respectively. In more involved physical situations, it is also possible that the stochastic law (5.15) includes the joint PDF of a_i , X_0 , and X_p . Remarkably, while the (uncertain) causality between the random fields a_i , X_0 , and X_p is expressed by the stochastic formulation (5.12), the (deterministic) causality (5.15) does not connect the attribute values themselves, but their PDFs. A more detailed analysis of the equations governing the PDFs will be given in Section 5.6.2. Epibraimatics focuses primarily on Eq. (5.12)–(5.15), since this is the formulation that best expresses the original ideas of Aristotle, Kant, Boltzmann, Schrödinger, and others concerning the fundamental connection of randomness and causality. Moreover, these equations are in agreement with the fundamental viewpoint that data are not merely numbers. Instead, they convey a message from the natural phenomenon they represent in the same way the paintings of Rembrandt convey a message from seventeenth-century Europe. From the stochastic law (5.15), one can estimate any attribute value χ_p one wishes. The term “estimate” is important here: Unlike the case of the deterministic causal law (5.10), the stochastic solution χ_p does not necessarily have probability 1. Instead, the solution has a specified probability of occurrence (between 0 and 1) with its associated estimation accuracy. To put it in different words, this sort of estimation may be seen as a process by means of which the probabilities turn into uncertain possibilities. In which case Judea Pearl (2010: 1) justifiably complains that certain fields seem to ignore the knowledge provided by science-based laws generating the data distributions: “Questions [in health, social and behavioral sciences] require some knowledge of the data-generating process, and cannot be computed from the data alone ... Remarkably, although much of the conceptual framework and algorithmic tools needed for tackling such problems are now well established, they are not known to many of the researchers who could put them into practical use.”

5.5.4 Comparative Summary

Before leaving this section, in Table 5.5 I summarize some of the salient differences between the three types of laws considered above. The readers may recall that in the case of the deterministic law, one observes the coefficients and BIC, and then derives attribute values χ_p at any p . In the case of the stochastic law the corresponding entities are random fields, which means that the law represents the co-existence of randomness and causality (the law can be conditioned by data, whenever available). Given the $f_{X;0}$ and f_{a_i} , the PDF law (5.15) derives $f_{X;p}$ across space–time. In the case of the statistical law, on the other hand, based on a certain

Table 5.5 Main characteristics of deterministic, statistical, and stochastic laws

	Entities related by law	Description of what is observed or observable	Symbolic pattern describing entities
<i>Deterministic law:</i>	Attribute values linked by causality	Law coefficients, BIC	Science-based conceptual scheme
<i>Statistical law:</i>	No causality present	Dataset	Statement about facts
<i>Stochastic law:</i>	Attribute values linked by (uncertain) causality PDFs linked by causality	Random law coefficients, BIC PDFs of coefficients, BIC Uncertain dataset	Science-based conceptual scheme

number of χ_p observations one merely derives a frequency law of the attribute that is assumed valid, in general. Having as a common origin a scientific theory, the deterministic and stochastic laws are science-based conceptual schemes, whereas the statistical law is merely a statement about observed facts.

5.6 Constructing Multivariate PDF Models

Since in stochastic reasoning one must learn to think and act in terms of PDFs (distributed across space–time and following laws of Nature) rather than single attribute values, it makes sense to develop effective ways to construct these PDF in praxis. Like Percival, the Round Table Knight who needed to grow mentally before he could locate the Holy Grail, stochastic theorists need to constantly improve their epistemic conditions in order to be able to credibly construct the complete multivariate PDFs across space–time. Since this turns out to be a difficult endeavor in real-world situations, it is not surprising that a number of “shortcuts” are often used in practice. This section examines certain well-known approaches to construct a PDF model. To paint with a broad brush, a basic classification of these approaches is as follows: (i) *Formal* PDF model construction, which includes models that are speculative, analytically tractable, or ready-made. (ii) *Substantive* PDF model construction, which includes models having a firm basis in reality (determined on the basis of scientific knowledge and evidential experience), taking into account the contentual and contextual domain of the in situ situation.

5.6.1 Formal Construction: Copulas and Factoras

First, we will study speculative PDF models that are either ready-made or possess analytically tractable properties. As to the merits of formally constructed PDFs, it must be said at the very outset that the consequences of using the wrong PDF in real-world situations can be more severe than those displayed by the PDF itself.

5.6.1.1 Ready-Made and Tractable PDF Models

There is a list of PDF models that have been derived using formal analytical techniques. The most famous example, of course, is the multivariate *Gaussian* (*normal*) PDF model

$$f_{X;p} = [(2\pi)^k |C|]^{-1/2} e^{-\frac{1}{2} \sum_{i,j=1}^k c_{ij} (\chi_{p_i} - m_i)(\chi_{p_j} - m_j)}, \quad (5.16)$$

with parameters m_i and m_j ($i, j=1, \dots, k$); $|C|$ is the determinant of a positive-definite matrix C with elements c_{ij} . The PDF (5.16) has many convenient properties that are described in the relevant literature (Tong 1990). Despite its attractive properties, many experts argue that the multivariate Gaussian PDF is unequipped to answer questions about the occurrence of rare but catastrophic events (linked to natural disasters, financial crises etc.). Beyond that, often there is no substantive justification why to prefer this type of “well behaved” PDF to those models that emphasize the possibility of “catastrophic” events.

Other well-known analytically tractable ready-made multivariate PDFs include the *Student*, *exponential*, *lognormal*, *elliptical*, *Cauchy*, *beta*, *gamma*, *logistic*, *Liouville*, and *Pareto* models (a detailed presentation of ready-made multivariate distributions can be found in Kotz et al. 2000 and Genton 2004). A basic feature of many of these multivariate PDFs is that the corresponding univariate PDF are of the same kind (e.g., if the multivariate PDF is Gaussian, so is the univariate). As the readers are aware, the inverse is often not true. That is, a univariate PDF (say, Gaussian) may be associated with a multivariate PDF of a different kind (non-Gaussian).¹⁸ These facts can cause serious problems in many applications in which one deals with non-Gaussian fields X_p that have different kinds of univariate $f_{X;p_i}$ (e.g., $f_{X;p_1,p_2}$ is non-Gaussian, whereas $f_{X;p_1}$ is Gaussian and $f_{X;p_2}$ is gamma). In such cases, a key question is how to extend the univariate PDFs that are usually available in practice to a multivariate PDF that fits the attribute of interest across space–time. This kind of problems constitutes a prime reason for the systematic development of the copula- and factor-based representations of a multivariate PDF (Section 5.6.1.2 and 5.6.1.3).

There are also PDF models that are assumed to have particularly tractable analytical forms. A rather trivial case of such a multivariate model is the PDF with full *stochastic independence*.¹⁹

$$f_{X;p} = f_X(p_1, \dots, p_k) = \prod_{i=1}^k f_{X;p_i} \quad (5.17)$$

¹⁸ In this sense, the multivariate PDF behaves like the Holy Grail of the legends of the questing king Arthur’s knights, which assumed different shapes, forms, origins, and interpretations.

¹⁹ The reader is reminded that stochastic (or probabilistic) independence is different from logical independence (Table 4.2).

for all k . This model essentially describes phenomena that do not transmit knowledge across space–time, i.e. one’s knowledge of the attribute’s state at point \mathbf{p}_i does not affect one’s knowledge of the state at point \mathbf{p}_j . Although mathematically convenient, model (5.17) is of rather limited use in real-world situations. A more interesting model is that of *partial* stochastic independence (e.g. (5.17) holds for $k=2$, but not for $k=3$, etc.). Multivariate PDFs can be derived in cases when a specific relationship is known to exist between the random-field realizations. An interesting yet rather limited model is the PDF with *spherical symmetry*, simply written as (Blokh 1960)

$$f_{X,\mathbf{p}} = f_X(\mathbf{p}_1, \dots, \mathbf{p}_k) = g(\xi), \quad (5.18)$$

where $\xi = (\sum_{i=1}^k \chi_{\mathbf{p}_i}^2)^{1/2}$. This PDF is an even function that is symmetric with respect to $\chi_{\mathbf{p}_i}$. All univariate PDFs are the same, $f_{X,\mathbf{p}_i} = f_X(\chi)$, $i=1, \dots, k$. The multivariate PDF (5.18) is determined from the univariate PDF via the integral representation

$$f_{X,\mathbf{p}_i}(\chi) = \frac{2\pi^{(k-1)/2}}{\Gamma(\frac{k-1}{2})} \int_0^\infty d\xi \xi^{k-2} g((\chi^2 + \xi^2)^{1/2}), \quad (5.19)$$

where Γ is the gamma function. The class of multivariate $f_{X,\mathbf{p}} = g(\xi)$ is obtained by assuming different univariate PDF f_{X,\mathbf{p}_i} and then inverting Eq. (5.19). In the case of stochastic independence, $f_{X,\mathbf{p}} = \prod_{i=1}^k f_{X,\mathbf{p}_i} = g(\xi)$, one finds the PDF model $f_{X,\mathbf{p}_i} = c_0 e^{c_1 \chi^2}$ (c_0, c_1 are suitable coefficients), which is the Gaussian case. In some other situations, the multivariate PDF $f_{X,\mathbf{p}}$ can be expressed in terms of its univariate PDFs f_{X,\mathbf{p}_i} ($i=1, \dots, k$) and a set of functions of $\chi_{\mathbf{p}_i}$. This could be of considerable interest, because often one has good knowledge of f_{X,\mathbf{p}_i} and seeks to construct $f_{X,\mathbf{p}}$ that is physically meaningful, and its parameters can be estimated in practice. Two noteworthy cases are considered next: Copulas and factoras.

5.6.1.2 Copula-Based PDF Models

Under certain technical assumptions, a multivariate PDF can be written in terms of the so-called *copula* (Sklar 1959; Genest and Rivest 1993; Nelsen 1999),

$$C_{X;\{\mathbf{p}_i\}} = C_X(F_{X;\mathbf{p}_1}, \dots, F_{X;\mathbf{p}_k}) = P[F_{X;\mathbf{p}_1} \leq v_{\mathbf{p}_1}, \dots, F_{X;\mathbf{p}_k} \leq v_{\mathbf{p}_k}], \quad (5.20)$$

where $F_{X;\mathbf{p}_i}$ are univariate CDF, and $v_{\mathbf{p}_i}$ are realizations of $U_{\mathbf{p}_i} = F_{X;\mathbf{p}_i}^{-1} \sim U(0, 1)$, $i=1, \dots, k$. Any distribution function with support on $[0, 1]^k$ and uniform marginals has been termed a copula (Mikosch 2006a: 5). The corresponding copula density is defined by $\varsigma_{X;\{\mathbf{p}_i\}} d\mathbf{v} = dC_{X;\{\mathbf{p}_i\}}$ (assuming copula continuity and differentiability). The $f_{X,\mathbf{p}}$ is reformulated in terms of its univariate PDF and the copula density as

$$f_{X;p} = \left[\prod_{i=1}^k f_{X;p_i} \right] \varsigma_{X;\{p_i\}}. \quad (5.21)$$

Equation (5.21) basically decomposes the multivariate PDF ($f_{X;p}$) into the product of the univariate densities ($f_{X;p_i}$) and the multivariate copula density ($\varsigma_{X;\{p_i\}}$) that expresses a certain form of interaction between univariate PDFs. Copula families with useful properties include the elliptic and the Archimedian ones (Genest and Rivest 1993).

As is the case with all technical apparatuses, the copula technology has its *pros* and *cons*. Basically, copula is yet another tool to estimate multivariate non-Gaussian PDFs, which is suitable for some applications, but not for some others (Joe 2006). Under certain conditions, copulas yield useful parametric descriptions of multivariate non-Gaussian fields (Scholzel and Friederichs 2008). According to Andras Bardossy (2006), a copula can express whether the corresponding spatial dependence changes for different attribute quantiles (high values may exhibit a strong spatial dependence, whereas low values a weak dependence) – although the situation is more difficult or even impossible for copulas to handle when multivariate (higher than second-order) copulas are considered. Copulas are scale-invariant in the sense that the copula of $Z_p = \phi(X_p)$ is equal to the copula of X_p if $\phi(\cdot)$ is a strictly monotonic function. On the other hand, one should keep in mind that the copula technology mainly applies to continuous-valued attributes so that the marginals are uniform according to the probability integral transform theorem. No general approach exists to construct the most appropriate copula for an attribute, whereas the choice of a copula family for an in situ problem is often based not on substantive reasoning, but on mathematical convenience (Mikosch 2006a, b). If construction methods are available for componentwise maxima, no unique approaches can be established for a set of attributes that are not all extremes. This is also the case of univariate analysis, where distribution functions are usually chosen on the basis of theoretical observations and goodness-of-fit criteria. Direct interpretation of the copula alone does not offer insight about the complete stochastic nature of the attribute and there is no dependence separately from the marginals. Also, copulas do not solve satisfactorily the dimensionality problem (Scholzel and Friederichs 2008). Interpretive issues concerning the copulas' in situ applications emerge too. There are many real-world attributes that are not continuous-valued but rather discrete- or mixed-valued (e.g., daily rainfall), which means that the integral transform theorem (on which the copula technology of continuous variables relies) cannot be implemented, since the $F_{X;p_i}$ are no longer uniformly distributed in the interval (0,1), thus giving rise to so-called unidentifiability issues (Genest and Nešlehová 2007). In this respect, although copulas can be used in simulation and robustness studies, they have to be used with caution because some properties do not hold in the discrete case. Applying copula models to datasets that do not satisfy the necessary assumptions, or disregarding proper inferential procedures, is like “the modeller telling Nature what to do,” which can lead to unsatisfactory results. Also, experts have linked the extensive use of Gaussian copulas in finance with the 2008 worldwide meltdown (Salmon 2009).

This is a widely publicized case in which the model (Gaussian copulas) provided a poor representation of reality (financial markets), which also showed that analysts often use copulas without a correct inferential procedure. Attracted by the possibility to select arbitrary marginals, they sometimes forget that a suitable copula should be chosen as well as marginals. In other words, assuming a priori a Gaussian copula is like assuming Gaussian marginals without any theoretical reason or empirical evidence.

5.6.1.3 Factora-Based PDF Models

The factora technology has its origins in the Gaussian tetrachoric series expansion of Karl Pearson (1901). Although the factora PDF is apparently an older concept than the copula PDF, both concepts share some common features. The class of factora PDFs extends Pearson's original insight in a non-Gaussian random field context, leading to the class of *factorable* S/TRF (Christakos 1986, 1989, 1992). Let $\theta(\chi_p)$, $\chi_p = (\chi_{p_1}, \dots, \chi_{p_k})$, be a multivariate function of $L_2(R^k, \prod_{i=1}^k f_{X;p_i})$, $r_k = \int d\chi_p \prod_{i=1}^k f_{X;p_i} \theta^2(\chi_p) < \infty$, and let $\varpi_{j_i}(\chi_{p_i})$ be sets of complete polynomials of degree $j = 0, 1, \dots$ in $L_2(R^1, f_{X;p_i})$ that are orthogonal with respect to $f_{X;p_i}$. Then, one can write

$$\theta(\chi_p) = \left[\prod_{i=1}^k \sum_{j_i=0}^{\infty} \right] (\theta_{j_1 \dots j_k} \prod_{i=1}^k \varpi_{j_i}(\chi_{p_i})) = \theta_{X;\{p_i\}}, \quad (5.22)$$

where the $\theta_{X;\{p_i\}} = \theta(\varpi_{j_i}(\chi_{p_i}), i = 1, \dots, k; j_i = 0, 1, \dots)$ is called a *factora*, and the corresponding completeness relationship is $\left[\prod_{i=1}^k \sum_{j_i=0}^{\infty} \right] \theta_{j_1 \dots j_k}^2 = r_k$ (which assures that the series expansions converge). Accordingly, the multivariate PDF is expressed as

$$f_{X;p} = \left[\prod_{i=1}^k f_{X;p_i} \right] \theta_{X;\{p_i\}}, \quad (5.23)$$

which decomposes the modeling of the multivariate PDF ($f_{X;p}$) into the product of the univariate (non-uniform, in general) densities ($f_{X;p_i}$) and the factoras ($\theta_{X;\{p_i\}}$) that express interactions between univariate functions of χ_{p_i} . This is an advantage of the way factoras are defined over that of copulas. Also, the factoras may offer a measure of the deviation of the multivariate PDF from the product of the univariate PDFs. An S/TRF X_p that satisfies Eq. (5.23) is called a *factorable* S/TRF (of order k).

For illustration purposes, in the bivariate case Eq. (5.23) can be reduced to

$$f_{X;p_1, p_2} = f_{X;p_1} f_{X;p_2} \sum_{j=0}^{\infty} \theta_j \varpi_j(\chi_{p_1}) \varpi_j(\chi_{p_2}), \quad (5.24)$$

for all p_1, p_2 . In this case, $\theta_0 = 1$, $\theta_1 = \rho_{X;p_1, p_2}$ is the correlation coefficient and $\theta_j \delta_{jj'} = \overline{\varpi_j(\chi_{p_1}) \varpi_{j'}(\chi_{p_2})}$, with $\varpi_0(\chi_{p_i}) = 1$ and $\varpi_1(\chi_{p_i}) = (\chi_{p_i} - \overline{X_{p_i}}) \sigma_{p_i}^{-1}$ for all

space–time points \mathbf{p}_i . In (5.24), knowledge of lower order statistics is linked to the first terms of the series, whereas that of higher order statistics is linked to later terms of the series. A key step is to calculate ϖ_j that are orthogonal to a univariate PDF. There exist several methods for this purpose, where ϖ_j include Hermite, Laguerre, Generalized Laguerre, Legendre, Gegenbauer, Jacobi, and Stieltjes-Wigert polynomials.²⁰ To call a spade a spade, the main challenge presented by the factora formulation is how to define factoras $\theta_{X;\{\mathbf{p}_i\}}$ with the prescribed mathematical properties and associated complete sets of orthogonal polynomials (the difficulty increases with $k > 2$). A widely applicable method is based on the formula (Jackson 1941), $\varpi_j(\chi_{\mathbf{p}_i}) = f_{X;\mathbf{p}_i}^{-1} \frac{d^j}{d\chi_{\mathbf{p}_i}^j} [v(\chi_{\mathbf{p}_i})^j f_{X;\mathbf{p}_i}]$, where $v(\chi_{\mathbf{p}_i})$ is a function that satisfies specific conditions. This formula has been used to find polynomials for a wide range of continuous univariate PDF, including the Gaussian, exponential, and Pearson (Type I). For illustration, if $f_{X;\mathbf{p}_i} = \frac{1}{\sqrt{2\pi}} e^{-\chi_{\mathbf{p}_i}^2/2}$ ($-\infty \leq \chi_{\mathbf{p}_i} \leq \infty$), the bivariate PDF is $f_{X;\mathbf{p}_1, \mathbf{p}_2} = f_{X;\mathbf{p}_1} f_{X;\mathbf{p}_2}$, $\theta_{X;\mathbf{p}_1, \mathbf{p}_2} = f_{X;\mathbf{p}_1} f_{X;\mathbf{p}_2} \sum_{j=0}^{\infty} \rho_X^{a(j)} H_{a(j)}(\chi_{\mathbf{p}_1}) H_{a(j)}(\chi_{\mathbf{p}_2})$, H are Hermite polynomials. For $a(j) = j$, the $f_{X;\mathbf{p}_1, \mathbf{p}_2}$ is bivariate Gaussian; but for $a(j) = 2j$, the $f_{X;\mathbf{p}_1, \mathbf{p}_2}$ is non-Gaussian (Christakos 1992: 162–164). This is not surprising, since to a given univariate PDF, one may associate more than one bivariate PDF. Many other examples are found in the cited literature.

Some useful properties of the factorable S/TRF model may grab the readers’ attention, and so may do its limitations. If X_p is such an S/TRF field and $\phi(\cdot)$ is a strictly monotonic function, the random field $Z_p = \phi(X_p)$ is also factorable. This means that starting from the known classes of factorable fields X_p , new classes Z_p can be constructed using different kinds of $\phi(\cdot)$. In the bivariate case ($k = 2$), the PDF is written as $f_{Z;\mathbf{p}_1, \mathbf{p}_2} = f_{Z;\mathbf{p}_1} f_{Z;\mathbf{p}_2} \sum_{j=0}^{\infty} \theta_j \varpi_j[\phi^{-1}(\zeta_{\mathbf{p}_1})] \varpi_j[\phi^{-1}(\zeta_{\mathbf{p}_2})]$. Another interesting property of the factorable model is that it satisfies the relationship

$$\int d\chi_{\mathbf{p}_1} (\chi_{\mathbf{p}_1} - \overline{X_{\mathbf{p}_1}}) f_{X;\mathbf{p}_1, \mathbf{p}_2} = c_{X;\mathbf{p}_1, \mathbf{p}_2} \sigma_{X;\mathbf{p}_2}^{-2} (\chi_{\mathbf{p}_2} - \overline{X_{\mathbf{p}_2}}) f_{X;\mathbf{p}_2}, \tag{5.25}$$

for all $\mathbf{p}_1, \mathbf{p}_2$. Remarkably, (5.25) is valid for S/TRF classes other than factorable. In the special case that $\overline{X_{\mathbf{p}_i}} = \overline{X} = \mu$ and $f_{X;\mathbf{p}_i} = f_X$ (for all \mathbf{p}_i), Eq. (5.25) reduces to a more tractable form, $\int d\chi_{\mathbf{p}_1} (\chi_{\mathbf{p}_1} - \mu) f_{X;\mathbf{p}_1, \mathbf{p}_2} = \rho_{X;\mathbf{h}, \tau} (\chi_{\mathbf{p}_2} - \mu) f_X$, $\mathbf{h} = |s_1 - s_2|$ and $\tau = |t_1 - t_2|$. A direct consequence of (5.25) is $\overline{X_{\mathbf{p}_1} X_{\mathbf{p}_2}^m} = \rho_{X;\mathbf{h}, \tau} \overline{X_{\mathbf{p}_2}^{m+1}} - \mu(\rho_{X;\mathbf{h}, \tau} - 1) \overline{X_{\mathbf{p}_2}^m}$; i.e., a higher-order, two-point dependence is conveniently expressed in terms of one-point functions. Other attractive properties emerging from formulation (5.25) could be considered by the interested readers. In fact, those among the readers with an eye for unconventional results may also wish to develop the multivariate (higher than two) version of Eq. (5.25). Yet, another interesting property of the factoras is that they can generate estimators of nonlinear state-nonlinear measurement systems that are superior to those of

²⁰ For example, the Gaussian, Gamma, or Poisson univariate PDF is associated with Hermite, Laguerre, or Charlier polynomials.

the Kalman filter (Christakos. 1989). For example, the Kalman filter estimates include only linear correlation, whereas the factora estimates include linear and nonlinear correlations; also, the Kalman filter is limited to the estimation of the lower-moments (mean and variance), whereas the factora estimator can provide lower- and higher-order moments.

5.6.1.4 Comparative Comments and Pontius Pilate's Evasion

Theorists are sometimes accused of having the tendency to make otherwise simple-minded ideas and concepts look impressive, by using a sufficient string of intimidating Greek symbols.²¹ If this is indeed the case, no theorist can repeat Pontius Pilate's *αθώος του αίματος*²² evasion, and claim innocence. But if this is not the case, one cannot really see the need to misrepresent complex ideas by making them look overly simple, in the name of a misplaced and misunderstood populism in science. In a nutshell, underlying both the copula and factora technologies is the basic idea of replacing an unknown entity (original multivariate PDF) with another unknown entity (factora or copula), which is supposedly easier to infer from the available data and manipulate analytically. Whether this is actually a valid claim of practical significance depends on a number of technical and substantive issues, some of which were touched upon in the previous lines.

In technical terms, a prime advantage of the copula technology is its analytical tractability, although this is mainly valid in low dimensions (2–4). While factoras involve infinite series that have to be truncated, many copulas are available in a closed-form. This comes at the cost of some restrictive assumptions made by copulas, such as low dimensionality, uniform marginals, and the applicability of the integral transform theorem. Attempts to involve transforms of uniform marginals are rather ad hoc and can add considerable complexity to the process. Potential advantages of factoras include the elimination of the above restrictive requirements, and the rich classes of PDFs derived by taking advantage of the ϕ -property and generalization formulas (like (5.25)). The functional form of $\theta_{X;\{p_i\}}$ is explicitly given in terms of known polynomials, whereas the explicit form of $\varsigma_{X;\{p_i\}}$ is generally unknown and needs to be derived every time.

5.6.2 Substantive Construction

Though the importance of theory in real-world IPS is undeniable, explicating the relationship between theory and in situ phenomena is a perennial epistemological issue. *A cette fin*, the substantive approach of constructing multivariate PDF adopts a definite science-based viewpoint. Readers are reminded that a prime source of substantive knowledge is provided by natural laws and scientific theories. Indeed,

²¹ The situation may worsen if the theorist happens to be Greek.

²² I am pure of this blood.

investigators often have at their disposal a well-established set of natural laws to work with (e.g., one can hardly imagine a physical law free atmospheric science). Most commonly these laws have the form of algebraic or differential equations (a list of natural laws is presented in Table 1.12), whereas the study of a natural law requires some auxiliary conditions in the form of boundary and/or initial conditions (BIC), see Section 5.5.3. The crux of the matter is that if the natural laws are known, this is core knowledge that should be used in the derivation of the PDF models, which is a definite advantage of substantive model construction (e.g., prior probability problems of the so-called objective and subjective Bayesian analyses could be avoided). One may distinguish between the direct involvement of natural laws in terms of the corresponding *stochastic equations* and their indirect involvement by means of the *knowledge synthesis* framework.

5.6.2.1 The Stochastic Equations Method

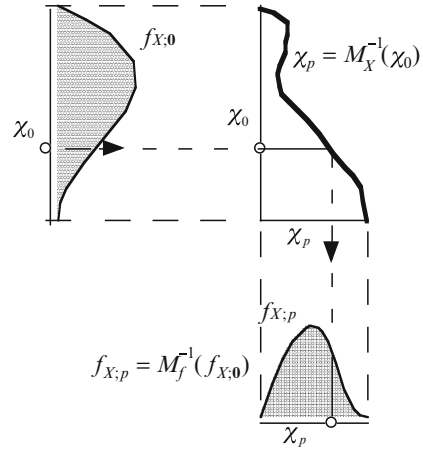
During the development of his kinetic theory of gases in the nineteenth century, Boltzmann rigorously demonstrated that reliable physical laws could be built on a stochastic foundation involving probability functions. In a similar vein, Sir Arthur Eddington remarked that, “It is impossible to trap modern physics into predicting anything with perfect determinism because it deals with probabilities from the outset” (Newman 1956). This is indeed the case of the stochastically formulated natural law (5.12) obeyed by attribute X_p . In a symbolic form, the law that the corresponding $f_{X;p}$ satisfies is Eq. (5.15). Its derivation from the stochastic law (5.12) is conceptually straightforward, but its practical implementation is often not an easy task.

When all coefficients are assumed fixed and only X_0 is random, Eqs. (5.12), (5.15) give

$$X_p = M_X^{-1}(X_0) \rightarrow f_{X;p} = M_f^{-1}(f_{X;0}). \tag{5.26}$$

A visual representation of Eq. (5.26) is attempted in Fig. 5.2, which indicates that stochastic laws have their rationale in symbolic language (in terms of mathematical attribute symbols) and visual language (probability shapes and function plots). Both languages are important in one’s effort to reproduce the laws of Nature into a coherent and comprehensive system of knowledge. Consider the Langevin-type equation $\gamma \frac{d}{dt} X_t = \sigma \zeta_t$, where X_t denotes the velocity of a particle at time t , γ is a coefficient associated with the velocity-dependent frictional term, and ζ_t is a fluctuating force term with coupling coefficients σ . The corresponding physical probability (Fokker-Planck) equation is $\frac{\partial}{\partial t} f_{X_0}(X_t) = \nabla^2 \frac{\sigma^2}{2\gamma^2} f_{X_0}(X_t)$. For illustration, examples of two different ways to construct the PDFs from the physical laws are briefly examined next. The first example is the stochastic differential Eq. (5.13) of X_p with $X_0 \sim f_{X;s,0} = e^{\mu_0 + \mu_1 \chi_{s,0} + \mu_2 \chi_{s,0}^2}$, and known coefficients μ_i ($i=0,1,2$). The analytical solution of (5.13) yields the attribute PDF (Gardiner 1990), $f_{X;p} = e^{-bt + \mu_0 + \mu_1 \chi_p} e^{-bt + \mu_2 \chi_p^2} e^{-2bt}$ as a function of p . The second example uses a quantum system governed by the stochastic Schrödinger law (5.14). The associated

Fig. 5.2 A visual representation of Eqs. (11)



probability has the form $f_{\psi;p} = |\psi_p|^2$, i.e., the PDF is determined in a straightforward manner, as soon as the solution ψ_p of (5.14) is available.²³ Several other studies can be found in the literature which focus on the derivation of useful attribute probabilities from physical laws. In subsurface flow, e.g., one notices the pioneering work of Gedeon Dagan (1982, 1989) that includes both conditional and unconditional probabilities in heterogenous porous formations, and the research efforts of Shlomo Neuman and co-workers (Neuman, 2005; Neuman and Tartakovsky, 2009) who also considered stochastic flow in fractured rocks and anomalous transport.

5.6.2.2 The Knowledge Synthesis Method

Surely, knowledge comes to people through a non-uniform network of beliefs, presumptions, self-corrections, opinions, and experiences. In the face of this, it is difficult to exactly reconstruct the process of thought. Nevertheless, there are certain important major knowledge stages that can be outlined (at the very least) and offer inspiration for IPS purposes. Chapters 6–7 present a general knowledge synthesis framework for constructing multivariate PDFs in a manner that incorporates *G*-KB (natural laws, theoretical models, scientific theories, empirical relationships) and *S*-KB (site-specific knowledge like hard data, uncertain information, secondary sources) of the in situ situation.²⁴ In Section 6.5.1, the knowledge synthesis-based PDF is compactly expressed as

$$f_{X;p} = A^{-1} \int d\chi \xi_S e^{\mu \cdot g}. \tag{5.27}$$

²³ While Newton’s laws deal with actual positions and velocities, the Schrödinger law essentially describes the evolution of probabilities.

²⁴ The core and the specificatory knowledge bases, *G*-KB and *S*-KB, respectively, were introduced in Section 1.2.3, and are discussed in detail in the Section 3.6.

where A is a normalization parameter, \mathbf{g} is a vector with elements representing the G -KB (including the natural law), $\boldsymbol{\mu}$ is a space–time vector with elements that assign proper weights to the elements of \mathbf{g} and ξ_S represents the S -KB available. When used in (5.27), the core knowledge widens horizons by abiding with site-specific data by a process of integrating S with G in a physically and logically consistent manner. Equation (5.27) accounts for local and nonlocal attribute dependencies across space–time. The above are noticeable advantages of the knowledge synthesis method of PDF building, which is why its description covers two of the following chapters.

5.6.3 *Drunkard's Search*

It is worth reviewing some technical and interpretive features of the methods used to construct a PDF. When the formal approach of Section 5.6.1 is favored, the PDF can have a variety of shapes, as long as certain conditions are fulfilled (satisfaction of the mathematical PDF admissibility requirements); and when selected from the list of models available in the literature, the PDF should not be merely a convenient choice but also a physically meaningful and internally consistent model whose parameters are obtained from the databases. When, on the other hand, the substantive methods of Section 5.6.2 are chosen, the PDF is derived directly from the in situ situation (physical laws, biological models, social constructs). Definite advantages of this method include that the derived PDFs have physical substance, and one may not need to check whether the technical conditions of Section 5.6.1 are satisfied. An obvious difficulty is that natural laws may not be available for all in situ situations. But in this case, many thinkers argue, it may be appropriate to admit that no sufficient in situ knowledge is available to pursue the task (of PDF construction) at the present time. Indeed, there is a considerable number of problems that the scientific method can solve and also a number of problems that cannot be currently solved on the basis of the existing data and current knowledge.

What the readers should take home from the discussion so far is that the search for the most adequate PDF model should be wide open and not merely a drunkard's search.²⁵ In some cases, the convenience of the formal approach may be a reasonable option, whereas in other applications a substantive approach will need to be used that accounts for physical and other kinds of knowledge sources. As already noted, a challenge faced by PDF modelers is how best to present detail-drenched in situ phenomena alongside theoretical constructions without twisting the former beyond recognition or viewing the latter as an interesting but unrealistic abstraction. On occasion, the real search begins with the recognition of clues. Like in detective novels or history books, the development of PDF rests on highlighting minuscule clues that may shed small beams of light on a hidden picture inferred by induction rather than deduced from general principles.

²⁵ The drunk looks for the lost keys only under a sidewalk lamp because this is where the best light is.

5.7 Spatiotemporal Dependence and Woody Allen’s Prose

When a thing is funny, search it for a hidden truth. In a hilarious passage, characteristic of Woody Allen’s prose, reference is made to “the bizarre experience of two brothers on opposite parts of the globe, one of whom took a bath while the other suddenly got clean” (Allen 1998: 16). Allen’s special brand of humor portrays a case of spatiotemporal dependence, in which what happens in one space–time point strongly affects what happens in another point. In sciences there exist different ways to assess the space–time change (in the sense of stochastic causality or association) of a natural attribute, and each one of them has its *pros* and *cons*. One way may be a perfect fit for an end, but not for other ends.

5.7.1 Dependence in Terms of Stochastic Expectation

Useful S/TRF tools for assessing spatiotemporal change are the *dependence functions* of the attribute X_p defined in terms of stochastic expectation. In principle, these functions, say D_{X_p} , can be calculated in terms of the PDF as follows:

$$D_{X_p} = \overline{\Theta(\chi_p)} = \int d\chi_p \Theta(\chi_p) f_{X;p}, \quad (5.28)$$

where, as usual, the bar denotes stochastic expectation, Θ is a known function. A few examples of D_{X_p} functions are given in Table 5.6. Some of these examples should look familiar to the readers, whereas some others may not. The covariance function, e.g., is obtained from the general Eq. (5.28) by letting $\Theta(\chi_{p_i}, \chi_{p_j}) = (\chi_{p_i} - \bar{m}_{p_i})(\chi_{p_j} - \bar{m}_{p_j})$, in which case $c_{X;p_i,p_j} = \iint d\chi_{p_i} d\chi_{p_j} \Theta(\chi_{p_i}, \chi_{p_j}) f_{X;p_i,p_j}$.²⁶ The reader will observe that not all dependence functions can be derived from Eq. (5.28). This is, in fact, the case of the sysketogram and contingogram functions in Table 5.6 (we will revisit these functions later in this chapter). In principle, one can assume an

Table 5.6 Examples of spatiotemporal dependence functions.

Name	Form	
Mean:	$\bar{m}_p = \overline{X_p}$	(5.29)
Covariance:	$c_{X;p_i,p_j} = (X_{p_i} - \bar{m}_{p_i})(X_{p_j} - \bar{m}_{p_j})$	(5.30)
Variogram:	$\gamma_{X;p_i,p_j} = \frac{1}{2} [(X_{p_i} - X_{p_j})^2 - (\bar{m}_{p_i} - \bar{m}_{p_j})^2]$	(5.31)
Multiple-point:	$g_{X;\{p_i\},\lambda} = \overline{\prod_{i=1}^{\lambda} (X_{p_i} - \bar{m}_{p_i})}$	(5.32)
Sysketogram:	$\beta_{X;p_i,p_j} = \log(f_{X;p_i,p_j}/f_{X;p_i}f_{X;p_j})$	(5.33)
Contingogram:	$\psi_{X;p_i,p_j} = f_{X;p_i,p_j}/f_{X;p_i}f_{X;p_j} - 1$	(5.34)

²⁶ It is easily seen that the attribute *variance*, $\sigma_{X;p_i}^2$, is obtained from $c_{X;p_i,p_i}$ if we let $p_i = p_j$.

infinite number of Θ functions, and then calculate D_{X_p} from (5.28). At this point, one may legitimately ask: Since each D_{X_p} is calculated *in terms of* $f_{X;p}$, how can D_{X_p} be used *instead of* $f_{X;p}$? This is a legitimate question. Its answer will concern us next.

5.7.1.1 Abstract and Intuitive Appraisals of Reality

Most experts agree that the usefulness of the D_{X_p} -sets consists in their anticipated ability to express important aspects of the space–time attribute distribution in a form that is more convenient to use than the multivariate PDFs, which are often difficult to obtain or may have an arbitrary and difficult to interpret general shape (Bogaert 1996; Douaik et al. 2005; Law et al. 2006; Choi et al. 2007; Coulliette et al. 2009). In theory, assuming that $f_{X;p}$ is known exactly, any D_{X_p} can be derived from Eq. (5.28). But a careful consideration of the matter shows that some important issues emerge in the real-world. If the $f_{X;p}$ shape is completely unknown, how can one select a D_{X_p} -set that enables an adequate characterization of the attribute distribution across space–time? Also, assuming that a D_{X_p} -set has been somehow selected in theory, how can it be computed reliably from the limited databases? Plainly speaking, there are no generally valid answers to these questions. The fact that $f_{X;p}$ may have an arbitrary shape creates serious difficulties concerning D_{X_p} selection. In many problems, the PDF shape is indeed completely unknown (one does not even know if the PDF is symmetric, etc.), which makes it difficult or even impossible to decide what sort of dependence functions to select. A trivial exception, of course, is the Gaussian case: the D_{X_p} -set (attribute mean and covariance) allows a formally complete characterization of $f_{X;p}$. In the vast majority of PDF cases, however, this is not possible. Even if one makes a guess concerning the general shape of $f_{X;p}$, it may be not possible, on the basis of the available datasets, to calculate those parameters that will allow an exact determination of the PDF. In a way, it is like the general solution of a differential equation that is not of much use in practice, unless realistic auxiliary conditions are available that allow the derivation of the particular solution of the in situ phenomenon.

But the situation may be not always as gloomy as seems to be implied above. Many space–time modeling experts claim that in a large number of in situ cases the D_{X_p} set consisting of the first three functions in Table 5.6, i.e.,

$$D_{X_p} = \left\{ \overline{X_p}, c_{X;p_i;p_j}, \gamma_{X;p_i;p_j} \right\}, \quad (5.35)$$

can provide a formally incomplete yet practically satisfactory description of the attribute's space–time distribution (Jones and Zhang 1997; Kyriakidis and Journel 1999; Augustinraj 2002; Ma 2003; Fernandez-Casal et al. 2003; Douaik et al. 2005; Stein 2005; Gneiting et al. 2007; Porcu et al. 2006, 2008; among others). This may be a reasonable claim, as long as practicality is not confused with mere convenience. Furthermore, the apparent success of the dependence set (5.35) in practice is one of those unexpected yet welcomed “miracles” that sometimes occur in a

scientist's life. I will call it a “disconcerting” miracle for reasons that will become clear later. It is true that the situation is much better in the exact than the non-exact sciences. In the former case, an attribute obeys a physical law or a well-tested empirical model, so that its values are causally linked across space–time. Then it makes sense to determine D_{X_p} that expresses this link in a stochastic way that accounts for the co-existence of spatiotemporal structure and chance. However, this is not necessarily valid in non-exact sciences, where the concept of dependence (correlation) may be less meaningful or even deceiving. The co-association between financial securities, e.g., is not measurable using correlation functions, because past history can never prepare one for the day when everything goes south (Salmon 2009: 112). Some experts suggest deriving mainstream dependence functions (like the covariance and variogram) in terms of copulas (Bardossy 2006; Bardossy and Li 2008). There exist other studies that favor the use of the systetogram ($\beta_{X;p_i,p_j}$) or contingogram ($\psi_{X;p_i,p_j}$) functions (Christakos 1991a, 1992). Are there any sound criteria for favoring a specific set of dependence functions over another, or is it simply a case of *fides quaerens intellectum*? The answer to this “faith seeking understanding” sort of question depends on the relation between the abstract and intuitive appraisals of reality in terms of dependence functions, the space–time structure of the underlying phenomenon, and the IPS objectives. As it will be shown in the following sections, each dependence function has its *pros* and *cons*, but some of them contain more grains of truth than others.

5.7.1.2 Concerning Mainstream Dependence Functions

Let us make a few observations concerning the most commonly used set of space – time dependence functions. The attribute mean function \bar{X}_p is defined at each point of the continuum, it can be calculated at a local and/or a global scale, and it gives an idea about dominant trends in the spatiotemporal variation of the attribute. The covariance $c_{X;p_i,p_j}$ and the variogram $\gamma_{X;p_i,p_j}$ measure the degree of agreement of attribute values at pairs of points p_i and p_j . In other words, the covariance and the variogram functions show how dependence between pairs of attribute values changes with space–distance and time-lag (in commonly encountered cases, e.g., the observed reduction of covariance values with spatial distance and time lag implies a reduced space–time attribute dependency). This dependence is an inherent feature of the attribute's composite variation across geographical space and during different time periods. There exist, of course, different forms of space–time dependency, which lead to distinct covariance and variogram shapes that satisfy the required mathematical *permissibility* criteria.²⁷ The $c_{X;p_i,p_j}$ and $\gamma_{X;p_i,p_j}$ models may

²⁷ A list of space–time covariance and variogram models together with their permissibility conditions can be found in Christakos (1991a, c, 1992, 2005b, 2008a, b), Christakos and Hristopulos (1998), Gneiting (2002), Kolovos et al. (2004), Christakos et al. (2000, 2005), and Porcu et al. (2006, 2008).

Table 5.7 Methods for constructing space–time dependence models

Space transforms (ST)	Probability distribution (PD)
Laws of Nature (LN)	Generalized random measure (GM)
Spectral measure (SM)	Permissible combinations of valid models (PC)

be space–time *separable* (e.g., they are expressed as the product of purely spatial and purely temporal components), whereas other models are assumed to be *non-separable* (they cannot be expressed as the above product).

Alexander Kolovos and co-workers have described a variety of mathematical and physical methods that can be used to construct valid space–time dependence models for the D_{X_p} -set of Eq. (5.35), separable and non-separable (Table 5.7; details can be found in Kolovos et al. 2004 and references therein). Using the ST method, dependence models in higher dimensionality domains are obtained from functions that are permissible in a lower dimensionality domain. The LN method expresses natural laws in a stochastic form, and the associated dependence functions are derived accordingly (see also the following Section 5.7.2). Using the SM, PD, and GM methods, valid dependence models are obtained from different kinds of measures by means of appropriate operations. In the PC method various combinations of existing permissible models can generate rich families of new space–time dependence functions. It is a matter of choice what knowledge bases and methods of analysis one should use. Each choice has its own merits and domain of applicability. However, there are cases where rationality and rigor require that some bases and methods be preferred over others. Many experts are critical of data-driven regression modeling that uses a fixed list of covariance models, independent of the underlying physics, rather than deriving them on the basis of substantive knowledge. On the other hand, the Spartan random field modeling of Dionissios T. Hristopulos²⁸ and co-workers properly uses covariance models established by means of physically or intuitively motivated interactions, instead of a purely data-driven matrix (Hristopulos 2003; Elogne et al. 2008).

5.7.1.3 The Indiscrimination Property

Studying space–time dependence functions, rather than the complete PDFs, is often a legitimate way of confronting S/TRF theory with in situ observations, and then making informative space–time predictions. But one should be always reminded of the warning of Section 5.2.2 that the uncritical implementation of probabilistic analysis can be a slippery affair. A telling example is the so-called *indiscrimination* property of lower order dependence functions: the same covariance or variogram

²⁸Beyond being a distinguished researcher, Hristopulos is an impressive performer capable of leaping across the dance floor in a series of faultless *grand jetés*. Talking about synthesizing personalities of large scope.

function can be assigned to random fields that exhibit very different space–time variation patterns (which may partially justify the characterization “disconcerting miracle” used earlier). Accordingly, the indiscrimination property raises legitimate questions regarding the validity of some theory–reality associations. The best response probably is to view the dependence function and the generated realizations in combination with other knowledge sources and analysis tools. Viewing a problem from the perspective of (intradisciplinary or interdisciplinary) knowledge synthesis is often a sound approach.

Let us consider a simple example. The two random fields X_s and Y_s are empirically related by $Y_s = v X_s$, where the field $v \sim U(0, 1)$ is independent of X_s and Y_s . Trivial calculations show that the two random fields have the same covariance, $c_{Y; s_i, s_j} \equiv c_{X; s_i, s_j}$. As one can see in Fig. 5.3, however, the two random fields can generate very different realizations representing the variation of the corresponding attribute, which means that the “black box” use of the covariance can provide a poor representation of the actual variation. In many cases, the situation can be improved significantly by conditioning the generated realizations with good quality datasets. In sum, the covariance and variogram functions should be used only when there is a deeper understanding and a valid working hypothesis about the natural system. Understanding guides one’s sensory engagement with in situ evidence, and improves one’s interpretation of the dependence functions calculated from this evidence.

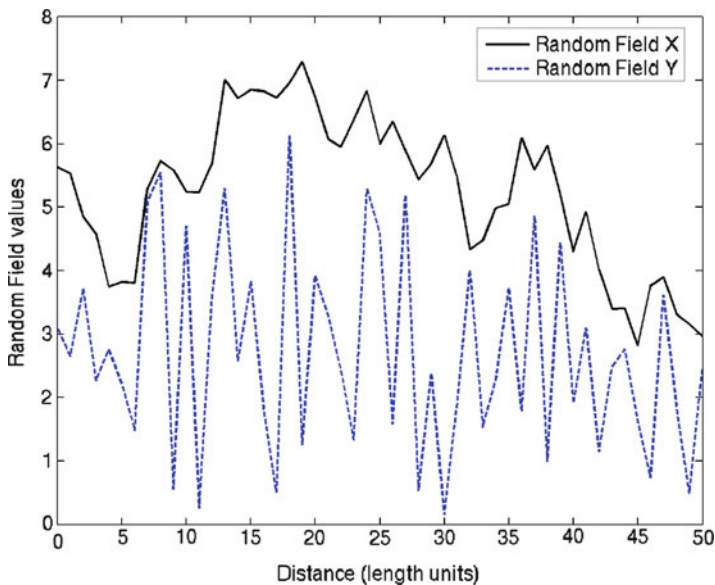


Fig. 5.3 Realizations of two random fields sharing the same covariance form

5.7.2 Dependence in Terms of Natural Laws

I hope the readers like ancient legends as much as I do. While considering dependence functions, the legend of the Irish king Fergus Mac Leda came to my mind. The king used to take long journeys in the land of Ireland, until he encountered a fierce river horse in Loch Rury. The encounter so terrified Fergus that his face became permanently distorted with fear. He kept on thinking about the incident, and since it was a matter of honor for the king to resolve the situation, he returned to Loch Rury several times to confront the monster. During the final struggle, Fergus managed to slay the monster before going down himself. But having finally resolved the matter, the king's face at last was restored and serene. It is not unusual that contemporary investigators find themselves in Fergus' position, hopefully with a few variations. During their long journeys in the land of scientific inquiry, the investigators too encounter difficult problems and serious obstacles that they cannot handle at the time, but they keep on thinking about them until they are able to finally resolve them. Let us apply this approach in the case of space–time dependence characterization, and revisit the issue of how to adequately develop the corresponding models.

Returning to the problem of space–time dependence representation, if an investigator seeks to construct a dependence model from incomplete datasets, one needs to have an understanding of what kind of space–time structure one is looking for. This understanding comes, *inter alia*, from natural laws of various kinds. This is the basic idea underlying the LN method for constructing dependence models (Table 5.7). The method has two versions. In the first version, after the solution X_p of the stochastic attribute Eq. (5.12) has been obtained, the dependence functions can be derived using Eqs. (5.29)–(5.32) in Table 5.6. In the second version of the LN method, the stochastic causality of (5.12) gives rise to the deterministic causality that connects the space–time dependence of the relevant attributes by means of the general model

$$M_D(D_{X; p_i, p_j}) = 0, \quad (5.36)$$

where $D_{X; p_i, p_j}$ denotes the dependence function set of the attribute X_p across space–time, and M_D includes the known dependence sets of a_i and X_0 . Equation (5.36) is very informative: it shows how the dependence functions propagate across space–time so that they are consistent with the attribute's natural law. It is doubtful that this kind of information can be extracted from the incomplete dataset alone. In this sense, (5.36) unites the various space–time patterns emerging from natural law in a single network. Its solution gives $D_{X; p_i, p_j}$ between all pairs of points p_i, p_j . For illustration, consider the temporal variation of the pollutant burden on a human organ, X_t , that obeys the stochastic kinematics law (Christakos and Hristopulos 1998: 284),

$$\frac{d}{dt} X_t + \lambda_t X_t - U_t = 0, \quad (5.37)$$

given $X_0 = 0$ (IC), U_t is the random uptake rate, and λ_t is the random transfer rate out of the organ. This implies, e.g., that the pollutant was first introduced in another compartment at $t = 0$ and then transferred to the compartment (5.37). The covariance of the burden fluctuation is governed by

$$D_{t_i} D_{t_j} c_{X; t_i, t_j} = \alpha e^{-\varepsilon|t_j - t_i|}, \quad (5.38)$$

where $D_{t_i} = [\frac{d}{dt_i} + \bar{\lambda}_{t_i}]$, α is the known variance of the uptake rate fluctuation, and ε^{-1} is the correlation time of the biological field. Equation (5.38) shows how the burden covariance propagates in the time domain. If one assumes that $c_{X; 0, 0} = 0$ (IC), and $\bar{\lambda} = \bar{\lambda}_t$ is constant, Eq. (5.38) can be solved for the burden covariance

$$c_{X; t_i, t_j} = \frac{\alpha}{\bar{\lambda}^2 - \varepsilon^2} [e^{-\varepsilon|t_j - t_i|} - \frac{\varepsilon}{\bar{\lambda}} e^{-\bar{\lambda}|t_j - t_i|} + \frac{\bar{\lambda} + \varepsilon}{\bar{\lambda}} e^{-\bar{\lambda}t_i - \bar{\lambda}t_j} - e^{-\varepsilon t_i - \bar{\lambda}t_j} - e^{-\bar{\lambda}t_i - \varepsilon t_j}]. \quad (5.39)$$

Eq. (5.39) is a symmetric function with respect to t_i and t_j . The burden covariance depends on the absolute time lag $|t_j - t_i|$, as well as on the disposition of both t_i and t_j with respect to uptake initiation (the burden is nonstationary, even when the uptake rate covariance is stationary). As it happens, theoretical toxicokinetics is ahead of its experimental counterpart, which means that advances in public health knowledge have to wait until the necessary experimental techniques are developed that can measure certain parameters of toxicokinetics models like Eqs. (5.38)–(5.39). Last but not least, consider a neuron morphology in which the evolution of the nerve cell potential X_p obeys the stochastic equation

$$[\frac{\partial^2}{\partial s^2} - \frac{\partial}{\partial t} - 1] X_p + a \frac{\partial^2}{\partial s \partial t} W_p + b = 0,$$

where s varies within a nerve cylinder, $t \geq 0$, a and b are constants, W_p is a white-noise field with covariance $c_{W; p_i, p_j} = \delta_{s_i, s_j} \delta_{t_i, t_j}$, and the cell is initially at rest. Then, under certain conditions (cell potential is initially zero, an infinite nerve cylinder is assumed, and smoothness requirements are met; Tuckwell 1989:69), the potential mean and variance are found from the above equation to be, $\bar{X}_p = b(1 - e^{-t})$ and $\sigma_{X, p}^2 = \frac{1}{4} a^2 [1 - \operatorname{erfc}(\sqrt{2t})]$.

5.7.3 The Predictability Power of a Model

The predictability power of the model M_D of the previous section may be considered in terms of its predictability *ranges* across space and time (ε_s^M and ε_t^M , respectively). Let $c_X^{M, S}(\mathbf{h}, \tau)$ denote the spatiotemporal covariance between the X_p values generated by M_D and those obtained from the attribute dataset S . The $(\varepsilon_s^M, \varepsilon_t^M)$ set is defined such that

$$c_X^{M, S}(\varepsilon_s^M, \varepsilon_t^M) = \eta c_X^0, \quad (5.40)$$

where $\mathbf{h} = \varepsilon_s^M$, $\tau = \varepsilon_t^M$, c_X^0 is the corresponding variance, and the value of η is selected by the investigator to represent the desired level of model predictability (usually, $0.5 \leq \eta < 1$). Equation (5.40) provides a stochastic measure of similarity between the attribute values generated by M_D and S . The longer the spatial (temporal) range is, the higher is the spatial (temporal) predictability of M_D with respect to X_p . The predictability ranges of M_D can be also compared to those of the dataset S . For example, S may include measurements Y_p via the empirical relationship $Y_p = X_p + W_p$ (W_p is the measurement error due to equipment imperfections, site conditions, etc.). Let $c_Y^S(\mathbf{h}, \tau)$ be the Y_p covariance calculated on the basis of the S data, and let $c_Y^S(\varepsilon_s, \varepsilon_t) = \eta c_Y^S$ define the corresponding correlation range set $(\varepsilon_s, \varepsilon_t)$. In this case the $c_{X,Y}^{M,S}(\mathbf{h}, \tau)$ denotes the spatiotemporal covariance between the X_p values generated by the model M_D and the Y_p values obtained from the dataset S ; and $\varepsilon_s^M, \varepsilon_t^M$ are the associated space and time ranges. If the model M_D provides an adequate representation of the attribute distribution one should expect that $\varepsilon_s^M > \varepsilon_s$ and $\varepsilon_t^M > \varepsilon_t$.

The readers may notice that in the limit when $\varepsilon_s^M = \varepsilon_s$ and $\varepsilon_t^M = \varepsilon_t$, M_D is not an improvement over the dataset S , in the predictability sense above. This is a point to be carefully investigated when using statistical regression and time series models in real-world applications (Smith et al. 2000; Hwang and Chan 2002; Martin and Roberts 2008). Since these models express X_p as a function of S data, the derivation of $c_{X,Y}^{M,S}(\mathbf{h}, \tau)$ is essentially based on the same dataset as $c_Y^S(\mathbf{h}, \tau)$, and so does the calculation of the coefficients of the statistical models. Hence, under certain circumstances it is possible that $\varepsilon_s^M \approx \varepsilon_s$ and $\varepsilon_t^M \approx \varepsilon_t$, which is a result that may doubt the validity of the models.

5.7.4 Information Theoretic and Copula Dependence Functions

Given the fundamental doctrine of scientific inquiry that one should always search for alternatives, I suggest examining the possibility of space–time dependence functions that lie outside the framework of the mainstream dependence functions. We start with the sysketogram $\beta_{X;p_1,p_2} \geq 0$, Eq. (5.33), which is also written as

$$\beta_{X;p_1,p_2} = \overline{\log f_{X;p_1}^{-1}} - \int d\chi_{p_2} f_{X;p_2} \overline{\log f_{X;p_1|p_2}^{-1}}; \tag{5.41}$$

i.e., it is a spatiotemporal dependence measure with information-theoretic properties. Eq. (5.41) may be viewed in the context of the Kullback-Leibler divergence $D(f_{X;p_1,p_2}; f_{X;p_1}, f_{X;p_2})$, where the D form is logarithmic. The sysketogram has some noticeable properties: in the case of stochastic independence (Section 5.6.1.1), it is valid that $\beta_{X;p_1,p_2} = 0$. The sysketogram depends only on the PDF and not the

X_p values.²⁹ And it is not affected if X_p is replaced by some function $\phi(X_p)$, provided that ϕ is one-to-one.³⁰ The last property means that $\beta_{X;p_1,p_2}$ is an absolute rather than a relative quantity, in the sense that the space–time correlation defined by $\beta_{X;p_1,p_2}$ is completely independent of the scale of measurement of X_p . This property is useful in physical applications in which the concepts of “scale of measurement” and “instruments window” play an important role. Similarly, when the attribute has random space–time coordinates (e.g., distribution of aerosol particles), $\beta_{X;p_1,p_2}$ is independent of the coordinate system chosen. The absoluteness property brings to mind a basic result of modern physics, according to which only absolute quantities (independent of the space–time coordinate system) can be used as essential components of a valid physical law (in which case the term “covariant” is used).

Another possible measure of space–time dependence is the contingogram $\psi_{X;p_1,p_2}$, which is based on Karl Pearson’s original idea of a discrete contingency coefficient. In the continuous space–time domain, the counterpart of Pearson’s contingency was defined in Eq. (5.34).³¹ The contingogram can be also written as

$$\psi_{X;p_1,p_2} = \int \int d\gamma_{p_1} d\gamma_{p_2} f_{X;p_1,p_2}^2 (f_{X;p_1} f_{X;p_2})^{-1} - 1, \quad (5.42)$$

which shows that in the case of stochastic independence, $\psi_{X;p_1,p_2} = 0$. As was the case with $\beta_{X;p_1,p_2}$, the $\psi_{X;p_1,p_2}$ depends only on the PDF of X_p , and is not affected if the (X_p) is replaced by the one-to-one function $\phi(X_p)$. From a stochastic reasoning viewpoint, X_p characterization provided by $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$ is cognitively general. The reader may notice a certain similarity between their definitions in Eqs. (5.33)–(5.34). Both dependence functions offer measures of the degree of X_p ’s departure from stochastic independence. It is noteworthy that if $A = f_{X;p_1,p_2}/f_{X;p_1}f_{X;p_2}$, then $\beta_{X;p_1,p_2} = \overline{\log A}$ and $\psi_{X;p_1,p_2} = \overline{A} - 1$. And by using series expansions³² (small A values), one finds $\beta_{X;p_1,p_2} \approx A - 1 = \psi_{X;p_1,p_2}$. At the moment, little is known about the in situ performance of $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$, which remains an open field of research.

There are more space–time dependence functions that do not readily fit the general formulation of Eq. (5.28). One of them is defined as follows. To an S/TRF X_p one can associate the indicator random field: $I_{X;p,\zeta} = 1$ if $X_p < \zeta$ (ζ is a cutoff), and $I_{X;p,\zeta} = 0$ otherwise. The corresponding indicator variogram of geostatistics can be written as (Bardossy 2006)

²⁹ This is easily seen in the discrete case, $\beta_{X;p_1,p_2} = \sum_i f_{1;i} \log f_{1;i}^{-1} - \sum_j f_{2;j} \log f_{1;j}^{-1}$, where only the probability values are needed and not the numerical values of the realizations.

³⁰ That is $(\gamma_{p_1}, \gamma_{p_2}) \rightarrow (\gamma'_{p_1} = \phi(\gamma_{p_1}), \gamma'_{p_2} = \phi(\gamma_{p_2}))$, in which case $f_{X;p_1,p_2} d\gamma_{p_1} d\gamma_{p_2} = f'_{X;p_1,p_2} d\gamma'_{p_1} d\gamma'_{p_2}$.

³¹ Actually, Pearson defined the discrete-valued contingency as $\varphi = [\sum_i \sum_j \frac{\eta^2(z_i, z_j)}{\eta(z_i)\eta(z_j)} - 1]^{0.5}$, where η denotes discrete probabilities. Here the idea is extended in a continuous space–time domain with $\psi = \varphi^2$.

³² That is, $\log A = (A - 1) - \frac{1}{2}(A - 1)^2 + \frac{1}{3}(A - 1)^3 - \dots$

$$\gamma_{I_X;p_1,p_2} = F_X^{-1}(\zeta) - C_{X;p_1,p_2}(F_X^{-1}(\zeta), F_X^{-1}(\zeta)), \quad (5.43)$$

where

$$C_{X;p_1,p_2} = P[F_X(X_{p_1}) \leq v_{p_1}, F_X(X_{p_2}) \leq v_{p_2}] = C_X[F_X(X_{p_1}), F_X(X_{p_2})] \quad (5.44)$$

is the space–time dependence copula. One may, also, express $\gamma_{I_X;p_1,p_2}$ in terms of the space–time copula density, $f_{X;p_1,p_2} = (f_{X;p_1}f_{X;p_2}) \varsigma_{X;p_1,p_2}$. An interesting property of (5.43)–(5.44) is that they express space–time dependence not in terms of the bivariate probability, but as a function of the corresponding univariate probabilities. This convenience often comes at a cost, which should be taken into consideration (Section 5.6.1). Just as is the case with the direct determination of the $f_{X;p_1,p_2}$ shape, Bardossy says, the determination of the C_X form remains a difficult problem with no general solution available. In the same setting, the copulas do not contain different or more information than indicator variograms, but they allow a joint handling and a different presentation with a more parsimonious parameterization. As it turns out, the $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$ functions can be also expressed in terms of copulas: i.e.,

$$\beta_{X;p_1,p_2} = \int_0^1 \int_0^1 dv_{p_1} dv_{p_2} \varsigma_{X;p_1,p_2} \log \varsigma_{X;p_1,p_2} = \overline{\log \varsigma_{X;p_1,p_2}}; \quad (5.45)$$

and,

$$\psi_{X;p_1,p_2} = \int_0^1 \int_0^1 dv_{p_1} dv_{p_2} \varsigma_{X;p_1,p_2}^2 - 1 = \overline{\varsigma_{X;p_1,p_2}} - 1. \quad (5.46)$$

In view of Eq. (5.46), the readers may notice an analogy between $\psi_{X;p_1,p_2}$ and Kendall's rank correlation coefficient discussed by Francesco Serinaldi (2008)

$$\tau_k = 4 \int_0^1 \int_0^1 dC_{X;p_1,p_2} C_{X;p_1,p_2} - 1 = 4 \overline{C_{X;p_1,p_2}} - 1, \quad (5.47)$$

where, as usual, $dC_{X;p_1,p_2} = \varsigma_{X;p_1,p_2} dv_{p_1} dv_{p_2}$. The readers may argue that in the copula-based expressions of $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$, the arbitrarily complex $f_{X;p_1,p_2}$ is essentially replaced by what can be an equally complex $\varsigma_{X;p_1,p_2}$. Nevertheless, according to Serinaldi the advantage of expressing $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$ in terms of copulas is that these expressions involve integrals on finite supports.

Some further comparison between the mainstream space–time dependence functions and the information-theoretic dependence functions above is instructive. The $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$ have properties that may favor their use in place of the covariance $c_{X;p_1,p_2}$: (i) $\beta_{X;p_1,p_2} = \psi_{X;p_1,p_2} = 0$ in the case of stochastic independence, whereas $c_{X;p_1,p_2}$ may be 0 even when only space–time non-correlation holds; (ii) $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$ depend only on the PDF, whereas the $c_{X;p_1,p_2}$ depends on both the PDF and X_p values. (iii) $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$ are not affected if X_p is

replaced by a one-to-one function $\phi(X_p)$, which is not the case with $c_{X;p_1,p_2}$; and (iv) $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$ can be extended in a straightforward manner to the multipoint case using copulas. Property *i* implies that $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$ contain more information about space–time dependence than $c_{X;p_1,p_2}$; or that $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$ uncover dependence features that $c_{X;p_1,p_2}$ does not. Using Schwartz’s inequality, it is shown that $\rho_{X;p_1,p_2}^2 \leq \psi_{X;p_1,p_2}$. For the indicator field $I_{X;p,\zeta}$, it is valid that $\rho_{I_{X;p_1,p_2,\zeta}}^2 = \psi_{I_{X;p_1,p_2,\zeta}}$. Concerning property *iv*, the systetogram and contingogram can be expressed in terms of copulas in the multipoint case as well; i.e.,

$$\left. \begin{aligned} \beta_{X;p_1,\dots,p_k} &= \overline{\log \varsigma_{X;p_1,\dots,p_k}} \\ \psi_{X;p_1,\dots,p_k} &= \overline{\varsigma_{X;p_1,\dots,p_k}} - 1 \end{aligned} \right\}, \tag{5.48a - b}$$

respectively. Hence, as soon as the copula density $\varsigma_{X;p_1,\dots,p_k} = \varsigma_{X;p}$ is calculated using standard techniques, the multipoint systetogram $\beta_{X;p_1,\dots,p_k} = \beta_{X;p}$ and contingogram $\psi_{X;p_1,\dots,p_k} = \psi_{X;p}$ can be calculated too.

5.7.5 Spatiotemporal Homostationarity

Spatially homogeneity and temporally stationarity in the wide sense, also termed *spatiotemporal homostationarity* (STHS), assumes that the space–time mean \bar{m} of the attribute X_p is a constant throughout the space–time domain of p , and the covariance and variogram are functions only of the space–time lag $p_i - p_j = (s_i - s_j, t_i - t_j) = (\mathbf{h}, \tau)$, see Table 5.8. As happens in similar modeling cases, to perceive an STHS attribute is not to see that actual attribute; it is to see (from the perspective of one who uses) STHS. For $\beta_{X;p_1,p_2}$ and $\psi_{X;p_1,p_2}$, the STHS may be defined in the strict sense involving the corresponding PDFs (the PDFs do not change by a transformation δp of the space–time coordinates); then, $\beta_{X;p_1,p_2} = 0$ ($|p_1 - p_2| \rightarrow \infty$). The vector distance $\mathbf{h} = (r, \theta)$ consists of its magnitude $|\mathbf{h}| = r$ and its direction (angle θ). A special case of homogeneity is spatial *isotropy*: the covariance depends only on r (not on θ). Also, another way of looking at the set Θ (Fig. 4.1a) is like a set of iso-covariance contours.

The need to use simplified assumptions (such as STHS, low-order dependence functions, etc.) in real-world studies provides investigators with a perspective from which to interpret potentially significant gaps between theory and practice. Moving between theory and practice can help investigators appreciate the impact what they do has in what they think. Lastly, commonly used terms like “homogeneity,”

Table 5.8 Dependence functions for STHS fields (wide-sense)

Mean:	$\bar{X}_p = \bar{m}$	(5.49)
Covariance:	$c_{X;p_i,p_j} = c_{X;p_i-p_j}$	(5.50)
Variogram:	$\gamma_{X;p_i,p_j} = \gamma_{X;p_i-p_j}$	(5.51)

“stationarity,” “stochastic,” and “isotropy” are sometimes misunderstood, and misused. Non-stationarity, e.g., has been associated with a white-noise process, and stochasticity has been confused with spatial stationarity (Cliff and Ord 1981). As a consequence, it is suggested to use the term homogeneity instead of stationarity in the spatial case, and keep stationarity for the temporal component of the variation (Yaglom 1961).

5.8 A Generalized View of S/TRF

In real life, one is often faced with the so-called “extension” problem. Scientists, for example, constantly seek to develop mental constructs that creatively extend an existing theory in order to include new and previously unobserved phenomena, solve previously unsolvable problems, and generate new and unexpected results.

5.8.1 *Random Fields Based on Generalized Functions or Distributions*

In the early 1950s, the need emerged to extend the homogeneous spatial random field (SRF) theory to include fields with spatially nonhomogeneous features.³³ Responding to this need, the theory of generalized SRF was developed (Yaglom and Pinsky 1953; Gelfand 1955; Yaglom 1957) based on the mathematics of generalized functions (or distributions),³⁴ in the sense of Laurent Schwartz and Kiyoshi Ito (Schwartz 1950; 1951; Itô 1954). In the 1970s, parts of the generalized theory were repackaged and extended in a geostatistics context, in which case the term “intrinsic SRF” was introduced (Matheron 1973). Another extension of the generalized theory was suggested in the composite space–time domain of Section 4.2. The extension was able to study physical attributes with heterogeneous space–time variability features, and led to the development of the *heterogeneous* S/TRF theory (Section 5.3.2), which is considerably richer than the STHS class, i.e. the S/TRF theory can be linked to a larger number of in situ phenomena than the STHS one. Random fields with spatial and temporal heterogeneity orders ν and μ , respectively, and the associated support functions, satisfy physical law-based conditions of change in the composite space–time domain. In the same setting, several classes of fractal and wavelet fields were derived as special cases of the heterogeneous S/TRF theory for suitable choices of the heterogeneity orders and support functions.

³³ Again, it is preferable that “nonstationary” be associated with time series rather than spatial functions, the latter being linked to the term “nonhomogeneous.”

³⁴ Also known as SRF with homogeneous spatial increments of order ν .

5.8.2 An Operational Treatment of Space–Time Heterogeneous Attributes

Heterogeneous S/TRF theory (Christakos 1990a, 1991a, c, 1992, 2008a, b) is subject to the rules of engagement between the mind (mental construct) and its object of study (reality). In the following, for generality's sake, I assume that the space–time distribution of the attribute $X_p = X_{s,t}$ of interest is heterogeneous.³⁵ Also, for reasons that will become clear soon, the focus is on S/TRFs that satisfy the formulation

$$Q_{\nu/\mu}[X_p] = Y_p, \quad (5.52)$$

where $Q_{\nu/\mu}$ is a space–time operator, the parameters ν and μ characterize the degrees of X_p heterogeneity in space and time, respectively; and the field Y_p exhibits some specified characteristics that serve an in situ objective (e.g., it is STHS). Throughout the book, an attempt has been made to demonstrate and to advocate the fruitfulness of understanding quantitative concepts in terms of literary metaphors. Resorting to such a metaphor, the paradoxical interchange between heterogeneous and homogeneous features is perceptible in Goethe's novel *Wilhelm Meister* expressed as the unusual co-existence of liberalism and absolutism, or in Beethoven's musical composition in which dynamic expositions and regular recapitulations form a binding entity. In the continuous case, the heterogeneity parameters (ν, μ) may imply spatial derivatives of order $\nu + 1$ and time derivatives of order $\mu + 1$ operating at the point $p = (s, t)$. This is a convention, according to which a STHS field has $\nu = \mu = -1$. If the PDF of X_p is known, it is possible to readily construct the $Q_{\nu/\mu}$ operator. In fact, if the operator expresses the dynamical laws that govern the natural attribute, the X_p is fully determined. Also, it is possible that the $Q_{\nu/\mu}$ operator of Eq.(5.52) enhances one's knowledge about the original attribute represented by the S/TRF X_p . Assume that via $Q_{\nu/\mu}$ the X_p leads to Y_p ; then, the inverse operation $Q_{\nu/\mu}^{-1}$ may yield a new S/TRF representation of X_p that contains more knowledge than the original one.

There exist a large number of $Q_{\nu/\mu}$ choices in association with X_p (see examples in Table 5.9; Christakos 1992; Christakos and Hristopoulos 1998). It is instructive to consider two main groups of operators, depending on the motivation behind their construction: Group *a* includes operators linked to substantive knowledge (natural laws, scientific theories, and empirical models). Equation (5.52) takes advantage of the fact that scientists often have at their disposal a set of sound natural laws to work with. In this sense, the heterogeneous S/TRF is a scientific method rather than a purely data-driven scheme, such as the statistical regression, time series, and hierarchical techniques. These techniques are satisfied with the mere description of data across space–time, whereas the S/TRF method has an explanatory character as a result of its connection with the laws describing the mechanisms underlying the

³⁵ Heterogeneity may be interpreted, e.g., in terms of complex spatial patterns combined with varying temporal trends (at local or global scales).

Table 5.9 Examples of $Q_{v/\mu}$ operators

$$\frac{\partial^{v+\mu+2}}{\partial s_1^{v_1} \partial s_2^{v_2} \dots \partial s_n^{v_n} \partial t^{\mu+1}} [\cdot], \sum_{i=1}^n v_i = v + 1$$

$$\sum_{i=1}^n \frac{\partial^{v+\mu+2}}{\partial s_i^{v_i+1} \partial t^{\mu+1}} [\cdot]$$

$$\left(\frac{\partial^{\mu+1}}{\partial t^{\mu+1}} + \sum_{i=1}^n \frac{\partial^{v+1}}{\partial s_i^{v_i+1}} \right) [\cdot]$$

data. Knowledge produced from these natural laws is used in the definition (5.52) of the S/TRF and the derivation of the corresponding dependence models. This leads to an exact specification of attribute dependence models about which limited or no information exists in terms of other attribute models about which sufficient information is available (e.g., the hydraulic head covariance is determined from the conductivity covariance using the continuity equation and Darcy’s law).

Group *b* includes operators chosen so that they satisfy problem related requirements (e.g., they annihilate trend functions with space–time coordinates). Hence, S/TRFs defined by (5.52) are capable of handling complicated space–time patterns based on the intuitive idea that the variability of an attribute can be characterized by means of its degrees of departure from STHS. In Group *b*, more than one X_p are generated from Y_p by using different $Q_{v/\mu}$, which shows the generality of formulation (5.52). The notation S/TRF- v/μ is used to show that $Q_{v/\mu}$ eliminates composite space–time trends expressed in terms of polynomial functions $\vartheta_{v/\mu,p}$ (of degrees v in space, μ in time). When an attribute X_p has certain features, f_i ($i = 1, \dots, q$), and one or more of them are removed, it is physically possible that several of the remaining f_i cease to exist too. Which raises the question whether it is possible that by removing the heterogeneous features of X_p , one also (unintentionally and unknowingly) removes some other features. The answer to this question is twofold: when $Q_{v/\mu}$ is based on a natural law, it is very likely that $Q_{v/\mu}$ will account for all essential features of the phenomenon; and the removal of the heterogeneity features, if it happens, is not permanent, since one returns to the original X_p , after the necessary Y_p -based analysis has been performed.

In the above setting, Eq. (5.52) underscores the conceptual and methodological significance of theory in scientific inquiry. Depending on the shape of $Q_{v/\mu}$, Eq. (5.52) can be formulated and solved in the continuous- or discrete-valued domain. A solution of (5.52) in the case, e.g., of the third operator in Table 5.9 is (Christakos and Hristopoulos 1998),³⁶

$$X_p = \int d\mathbf{p}' \psi_{\mathbf{p}'} G_{0,p,p'}^{(v+1/\mu+1)} + \overline{Y_{\mathbf{p}'}} \vartheta_{v/\mu,p}. \tag{5.53}$$

³⁶ $\vartheta_{v/\mu,p} = \frac{1}{(v+1)!} \sum_{i=1}^n \theta_i^2 s_i^{v+1} + \frac{1}{(\mu+1)!} t^{\mu+1}$, $\sum_{i=1}^n \theta_i^2 = 1$, and the Green’s function satisfies $Q_{v/\mu} [G_{0,p,p'}^{(v+1/\mu+1)}] = \delta_{s-s'} \delta_{t-t'}$, where $\delta_{s-s'}$ and $\delta_{t-t'}$ are delta functions in space and time, respectively.

When the natural laws are not fully known, guidance regarding the form of $Q_{v/\mu}$ is offered by empirical relations expressed in terms, either of algebraic equations or algorithmic rules aiming to emulate physical reality. Last but not least, even if knowledge of the specific laws is not available, dependence models derived from generally applicable physical laws can be used as potential candidates for describing particular datasets, in which case the law parameters are estimated from these data. Some readers would notice that considering the mainstream paradigm that claims to reduce to simple formulas any kind of uncertainty by forcing people to think in terms of independent trials, to make bets, and to throw dices, the stochastic modeling issues discussed in this and other chapters may look as foreign to them as a Jackson Pollock expressionist creation would look to a painter of melodramatic scenes like Delaroche. Nevertheless, the complex real-world problems emerging with increasing frequency nowadays should convince one that these modeling issues do make sense and are very important indeed.

5.8.3 Spatiotemporal Dependence and Heterogeneity Parameters

The covariance $c_{X_i; p_i; p_j}$ linked with S/TRF (5.52) is generally a heterogeneous space–time dependence function. Given the $Q_{v/\mu}$ shape, a variety of $c_{X_i; p_i; p_j}$ models can be obtained, separable and non-separable. For illustration, the following covariance class is derived from a partial differential equation law (Christakos and Hristopoulos 1998):³⁷

$$c_{X_i; p_i; p_j} = \int_T d\tau' \int_V dh' G_{1, \tau-\tau'} G_{2, h-h'} c_{Y; h'; \tau'} + \vartheta_{2v+1, h} \vartheta_{2\mu+1, \tau} + \vartheta_{v/\mu, p_i} \vartheta_{v/\mu, p_j}. \tag{5.54}$$

Since the covariance (5.54) can be linked to substantive knowledge, the $c_{X_i; p_i; p_j}$ expresses a degree of stochastic space–time causation that is significantly more than the mere statistical association measured by the purely data-driven covariance or variogram. Although this is an idea logically derivable from theoretical considerations and the existing evidence, the weakness of imagination may require a wealth of carefully acquired data to make the idea psychologically possible and its potential IPS impact well-understood. The matter could be an interesting avenue of future research.

³⁷ $G_{1, \tau-\tau'} = \frac{(-1)^\mu}{(2\mu+1)!} (\tau - \tau')^{2\mu+1} \theta_{\tau-\tau'}$ ($T = (-\infty, \tau]$), $G_{1, h-h'} = \frac{(-1)^\nu}{(2\nu+1)!} (h - h')^{2\nu+1} \theta_{h-h'}$ ($V = R^1$), $G_{2, h-h'} = \frac{1}{2^{2\nu+1} \pi^{(\nu)^2}} |h - h'|^{2\nu} \log |h - h'|$ ($V = R^2$), $G_{3, h-h'} = \frac{(-1)^{\nu+1} \Gamma(\frac{1}{2}-\nu)}{2^{2\nu+2} \pi^{3/2} \nu!} |h - h'|^{2\nu-1}$ ($V = R^3$); $G_{2, h-h'} = G_{2, h'-h}$ ($V = R^2, R^3$); θ is the unit step function.

5.8.3.1 Generalized Space–Time Dependence Models

Equation (5.54) can be written in the rather more concise form, $c_{X;p_i,p_j} = \kappa_{X;p_i-p_j} + \vartheta_{v/\mu,p_i} \vartheta_{v/\mu,p_j}$, where $\kappa_{X;p_i-p_j} = \kappa_{X;h,\tau}$ depends only on the space–time distance $p_i - p_j$.³⁸ An attractive feature of the decomposition is that in certain types of spatiotemporal analysis (e.g., linear prediction), only the $\kappa_{X;h,\tau}$ part is required. This decomposition has at least one important consequence: $\kappa_{X;h,\tau}$ can be constructed first, and then $c_{X;p_i,p_j}$ is obtained by adding suitable space–time ϑ -functions. In relation to the last observation, it is valid that (Christakos and Hristopulos 1998: 148)

$$U_{Q_{v/\mu}} \kappa_{X;h,\tau} = c_{Y;h,\tau}, \quad (5.55)$$

where $U_{Q_{v/\mu}}$ is a space–time differential operator defined as the product of $Q_{v/\mu}$ and its complex conjugate operator.³⁹

In the case of STHS, a set of computational formulas provide an efficient means for calculating experimental values of low-order dependence functions – such as the covariance $c_{Y;h,\tau}$ in Eq. (5.55) – from the available database. A valid model is subsequently fitted to the experimental values using a model fitting technique. If the data are *clustered* in space, efficient algorithms exist for the practical estimation of the sample covariance (Kovitz and Christakos 2004a): a coefficient of variation of the dimensionless spatial density of the point pattern of sample locations is introduced as a metric of the degree of clusteredness of the database, and a modified covariance estimator form is used that incorporates declustering weights and proposes a scheme for estimating the declustering weights based on zones of proximity. Given the covariance $c_{Y;h,\tau}$, Eq. (5.55) provides the means for constructing the corresponding models $\kappa_{X;h,\tau}$ and $c_{X;p_i,p_j}$. For example, Equation (5.56) of Table 5.10 is linked to the simple case $c_{Y;h,\tau} = \delta_h \delta_\tau$, whereas Eq. (5.57) is linked to $c_{Y;h,\tau} = ae^{-(bh+c\tau)}$ ($n = 1$). Space transforms (Table 5.7) offer a means to produce valid covariances in $R^2 \times T$ and $R^3 \times T$ starting from the known models in $R^1 \times T$ (Kolovos et al. 2004). It seems logical that the class of $\kappa_{X;p_i-p_j}$ models is richer than that of $c_{Y;h,\tau}$, and the class of $c_{X;p_i,p_j}$ is richer than both.

5.8.3.2 On Fractal Space–Time Models

Data occasione, since fractal random fields (Feder 1988) constitute a special case of the richer class of S/TRF (5.52), several fractal covariances can be derived as

³⁸ A terminology issue emerges here. Due to mathematical associations of Eq. (5.52) with the theory of generalized functions (distributions), and in order to distinguish it from the STHS covariance $c_{Y;h,\tau}$, it seems natural to call $c_{X;p_i,p_j}$ a generalized spatiotemporal covariance, keeping in mind that the term “generalized covariance” has been used in physics (Joseph 1965) and geostatistics (Matheron 1973).

³⁹ For example, $U_{Q_{v/\mu}} = (-1)^{v+\mu} (\partial^{2\mu+2} / \partial \tau^{2\mu+2}) (\nabla_h^2)^{v+1}$; Christakos and Hristopulos (1998: 160).

Table 5.10 Examples of $\kappa_{X;h,\tau}, r = |h|$, models⁴⁰

$$c_0 \delta_r \delta_\tau + \delta_r \sum_{\zeta=0}^{\mu} (-1)^{\zeta+1} a_\zeta \tau^{2\zeta+1} + \delta_\tau \sum_{\rho=0}^{\nu} (-1)^{\rho+1} b_\rho r^{2\rho+1} + \sum_{\rho=0}^{\nu} \sum_{\zeta=0}^{\mu} (-1)^{\rho+\zeta} a_\rho c_\zeta r^{2\rho+1} \tau^{2\zeta+1}, \quad n = 1, 2, 3, \tag{5.56}$$

$$(-1)^{\nu+\mu} [(2\nu+1)!(2\mu+1)!]^{-1} a b^{-2\nu-2} c^{-2\mu-2} e^{-br-ct} \gamma(2\nu+2, -br) \gamma(2\mu+2, -c\tau), \quad n = 1 \tag{5.57}$$

Table 5.11 Examples of fractal $\kappa_{X;h,\tau}, r = |h|$, models⁴¹

$$\kappa_{X;r,\tau} \propto r^\alpha (\tau/r^\beta)^z \tag{5.58}$$

$$\kappa_{X;r,\tau} = \sigma_X^2 \hat{f}_z(\tau/r^\beta; u_c) \hat{f}_z(r; w_c) \tag{5.59}$$

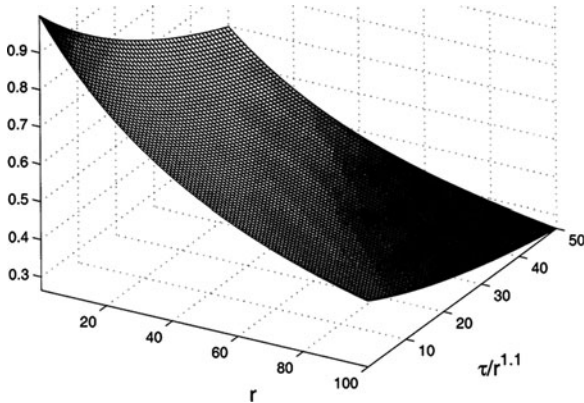


Fig. 5.4 Plot of the fractal covariance model for $\sigma_X^2 = 1, z = \alpha = -0.5, \beta = 1.1, u_c = w_c = 25$

special cases of $\kappa_{X;h,\tau}$. Examples are given in Table 5.11; the fractal model (5.59) is plotted in Fig. 5.4. The reader may notice that the function \hat{f}_z has an unusual dependence on the space and time lags through τ/r^β . For large τ , the τ/r^β is close to 0 if r is sufficiently large and \hat{f}_z is close to 1. With regard to \hat{f}_z , the equation for equidistant space–time contours is $\tau/r^\beta = c$. This dependence is physically different than that implied by, say, a Gaussian covariance model, in which case equidistant lags satisfy $r^2/\zeta_r^2 + \tau^2/\zeta_\tau^2 = c$. The difference is shown in Fig. 5.5 that plots the equidistant contours for \hat{f}_z and $r^2/\zeta_r^2 + \tau^2/\zeta_\tau^2$ as a function of space and time lags. A class of fractal S/TRFs satisfies the mathematical relationship $X_{c^\eta s, c^\zeta t} = c^H X_{s,t}$ (in the self-similarity sense), where $X_p = X_{s,t}, p = (s, t) \in R^n \times T (n = 1, 2, 3)$, and $c > 0, \eta, \zeta$, and H are the usual scaling coefficients. This class of fractal spatiotemporal random fields is denoted as *FS/TRF-H*. Under certain conditions, the corresponding expectation is written as $\overline{X_{c^\eta s, c^\zeta t}^2} = c^{2H} \overline{X_{s,t}^2}$. Consider the S/TRF (5.52), where $Q_{\nu/\mu}$ is the first operator in Table 5.9 and Y_p is a zero-mean STHS

⁴⁰ The coefficients $c_0, a_\zeta, b_\rho, a_\rho c_\zeta, a, b, c$ must satisfy certain permissibility criteria; δ_r and δ_τ are delta functions in space and time, respectively; and γ is an incomplete gamma function.

⁴¹ The $r_0 < r < r_m, \tau_0 < \tau < \tau_m$ define space-time fractal ranges; $-1 < z < 0, -0.5(n+1) < \alpha - \beta z < 0$ are permissibility conditions; σ_X^2 is variance; u_c, w_c are cutoffs.

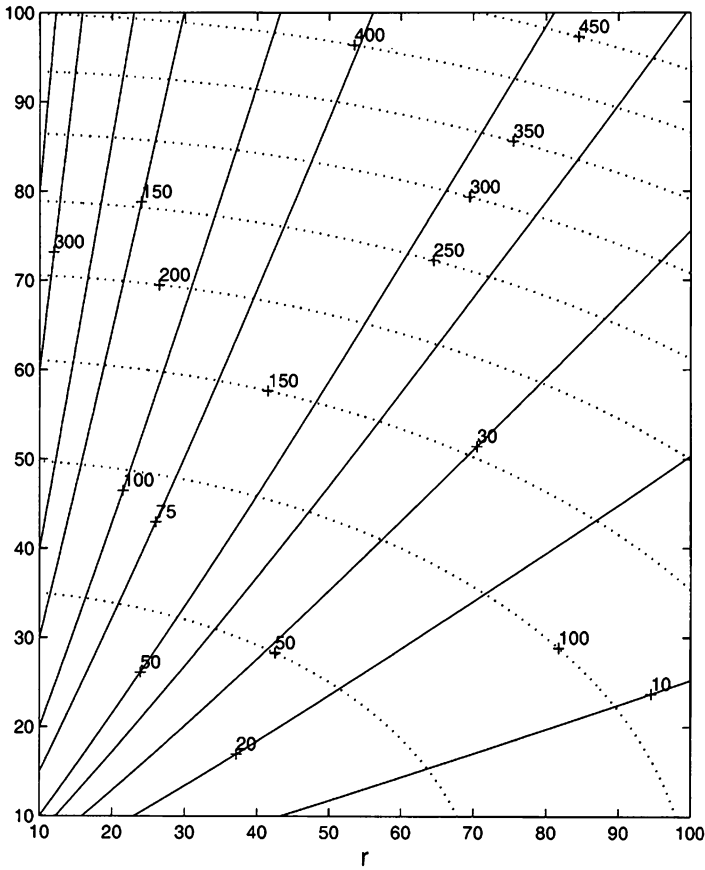


Fig. 5.5 Equidistant contours for fractal space–time dependence (*solid contours*) and for Gaussian dependence (*dotted contours*). Contour labels represent $c_0\tau/r^\beta$ values (*solid lines*) and $r^2/\zeta_r^2 + \tau^2/\zeta_\tau^2$ values (*dots*) using $c_0 = 62.95$, $\zeta_r = 10$ and $\zeta_\tau = 5$

white-noise field with covariance $c_{Y;h,\tau} = 2D\delta_h\delta_\tau$, where D is a suitable coefficient. Then, the FS/TRF- H is a member of the class of S/TRF- v/μ subject to the condition $H = n(v + \frac{1}{2})\eta + (\mu + \frac{1}{2})\xi$.

5.8.3.3 Physical Interpretation of the Heterogeneity Parameters

In addition to making mathematical considerations, one should be prepared to shift the heterogeneous S/TRF theory one layer deeper. What, when, and how a scientist can investigate in situ situations is a function of the scientist’s theoretical commitments and originality. In the case of the S/TRF theory, the heterogeneity parameters v and μ of the $Q_{v/\mu}$ operator provide a quantitative assessment of the rate of change of attribute patterns across space–time, and also offer information about

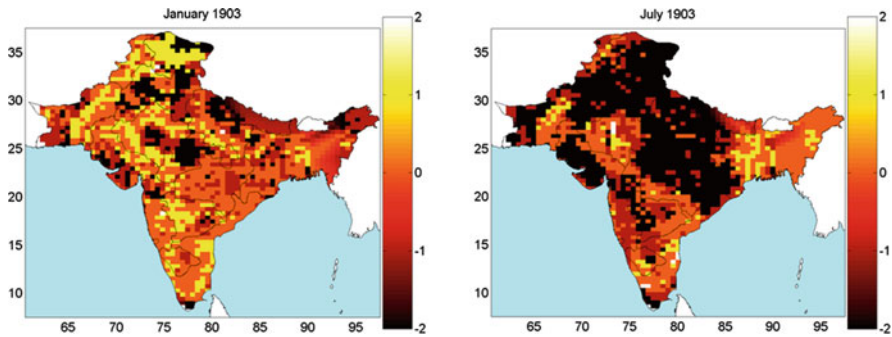


Fig. 5.6 Space–time maps of the $v - \mu$ differences associated with the Indian bubonic plague mortality distributions during different times (Yu and Christakos 2006)

the stochastic model representing the in situ system. These parameters, which are functions of space–time coordinates, determine, e.g., how “far” in space and how “deep” in time the operator searches for information about the attribute. Additional insight may be obtained by taking into consideration that the heterogeneity parameters v and μ are directly related to the fractal coefficient H .

Plots of the $v - \mu$ values associated with mortality distributions in the case of the bubonic plague in India (late nineteenth–early twentieth century) are shown in Fig. 5.6. Note that the space–time covariance model (5.56) in Table 5.10 (with $n = 2$) was used in the Indian bubonic plague study. This study shows that the S/TRF (5.52) offers a general theoretical model of the population mortality distribution that reflects the way stochastic causal influence is propagated across space–time, and gives valuable information about the attribute dynamics at the scale of interest. Generally, for natural systems that evolve within domains containing complicated boundaries and trends, the degrees of S/TRF heterogeneity should vary geographically and dynamically.

5.9 Constructive Symbiosis and Its Problems

Comments such as the following sometimes appear in the literature: “A criticism of the utility of geostatistics in agriculture is that the mathematical framework in which it is usually presented is beyond many potential users” (Nelson et al., 1999: 311). Otherwise said, a theory that can offer an improved representation of Nature but requires some extra effort on the part of its practitioners, is doomed to failure (and together with it any attempt to “lift Isis’ veil” – see Section 5.1). This is rather disappointing. Expert practitioners should appreciate the fact that theorists spend countless hours trying to create improved representations of reality, excogitating the vast physical world, and studying the significance of human existence within it. They struggle with new thoughts and imaginative mathematical constructs so that originality is not sacrificed to the Moloch of everyday pseudo-practicality. They

continuously search for shreds of evidence and meaning hidden in every aspect of the world, because they believe that an unexamined world is a world not fully appreciated, a world not explored, investigated, and discovered. In a harmonic *symbiosis* with theorists, expert practitioners should view the in situ implementation of the fruits of the theorists' labor as a basic component of scientific inquiry, rather than a chore involving "bottom line" recipes and "black-box" software (so that their users can have a carefree life, frolicking in their field of expertise without bothering to address meaning and purpose).

Along the lines of constructive symbiosis, many in situ applications of the STRF theory can be found in the relevant literature, including environmental exposure, health effects, epidemiology, earth and atmospheric sciences, forestry, ecology, geography, and history.⁴² A prime concern of these applications is not only to make possible the investigators' access to increased amounts of data, but most importantly, to present these data in a way that is consistent with the theory and improve the investigators' comprehension of the in situ phenomenon. This is made possible because, although their precise methods of inquiry differ from one discipline to the next, the investigators basically understand one another and share overlapping intellectual goals.

Naturally, the effort toward a constructive symbiosis of theory and practice is not always without difficulties. A point of friction between theorists and practitioners is often the tendency of the former to question the fundamentals of techniques that are dear to the latter. Two examples are the dependence function metrics and the popular Bochnerian criteria that are discussed below.

The readers may recall that the metric that determines space–time distance affects the mathematical *permissibility* of a dependence model; i.e., a model that is permissible for one metric may be not so for another. Moreover, when dependence models are related through a law or relationship, the permissibility of one of these models may affect that of another. For example, in light of Eq. (5.54) the permissibility of $c_{X;p_i,p_j}$ is affected by that of $c_{Y;h,\tau}$. For illustration purposes, consider a covariance of the space–time separable form, $c_{Y;h,\tau} = c_{Y;h} c_{Y;\tau}$. A general class of mathematical functions that can be associated with (Euclidean or non-Euclidean) metrics is $c_{Y;h} = e^{-N_\mu(\mathbf{h})}$, where $N_\mu(\mathbf{h}) = \sum_{i=1}^n |h_i|^\mu$ and $0 < \mu \leq 2$ (Christakos and Papanicolaou 2000). Now consider some examples of permissible models. The covariance $c_{Y;h} = e^{-|h|^2}$ is permissible for the Euclidean metric—it is not permissible for the absolute metric (Table 4.1). The covariance

⁴² The relevant literature includes (but is not limited to) the following: Serre et al. (2001, 2003a, b), Querido et al. (2007), Tuia et al. (2007), Bogaert and Fasbender (2007), Fasbender et al. (2007), Orton and Lark (2007a, b), Vyas et al. (2004), Bogaert (2002, 2004), Bogaert and Wibrin (2004), Wibrin et al. (2006), Yu et al. (2007a, b,c), Douaik et al. (2004, 2005), Serre and Christakos (1999a, b; 2003), Quilfen et al. (2004), Kolovos et al. (2002), Papantonopoulos and Modis (2006), Akita et al. (2007), Lee et al. (2008a, b, 2009), Bogaert and D'Or (2002), D'Or and Bogaert (2003), Coulliette et al. (2009), Yu and Christakos (2005, 2006), Pang et al. (2009), Puangthongthub et al. (2007), LoBuglio et al. (2007), Choi et al. (2003, 2006, 2007), Savelieva et al. (2005), Parkin et al. (2005), Augustinraj (2002), Law et al. (2006), Wang (2005), Gummer (2009), Kolovos et al. (2010), and de Nazelle et al. (2010).

$c_{Y,h} = e^{-N_1(h)}$, on the other hand, is permissible for an absolute (non-Euclidean) metric. The analysis above can be extended to include metrics of the more general form $|h| = (\sum_{i=1}^n \lambda_i |h_i|^\mu)^{\mu^{-1}}$, where $1 \leq \mu < 2$, and λ_i ($i = 1, \dots, n$) is a weight determining the “salience” of the h_i -direction. Space–time prediction and mapping depend on the metric structure assumed, since the dependence models are used as inputs in most prediction and mapping techniques. It can be shown, indeed, that the same dataset with its space–time dependence structure represented by covariance models of the same functional form can lead to different space–time predictions and maps, if prediction is performed using different metric structures (Christakos 2000).

In sum, S/TRF modeling allows the evaluation of distinct uncertainty types (conceptual and technical, ontic and epistemic); involves space–time coordinate systems to accommodate different kinds of attribute variability; makes an epistemically sound distinction between general (or core) and specificatory KBs; offers complete system characterization in terms of prediction probability laws (not necessarily Gaussian) at every mapping point (vs. a single prediction at each point); represents heterogeneous dependence patterns and landscapes (rather than artificial curve fitting, ad hoc trend surfaces, etc.); accounts for multiple-point functions representing higher-order spatiotemporal attribute dependencies; and its choice of a coordinate system and associated norm to describe a phenomenon depends on the nature of the properties being described. In fact, metric-dependent analysis of permissibility has important consequences in applications (e.g., space–time mapping, or the solution of stochastic partial differential equations) in which the investigator is concerned about the validity of space–time dependence functions associated with a physically meaningful metric (Euclidean or non-Euclidean). At this point, let me highlight that so far we have mostly been talking about theoretical space–time dependence models, rather than about their practical counterparts. Often the investigator has to make certain compromises, so to speak, concerning what an adequate and at the same time convenient representation of the theoretical model should be. It is not uncommon in practice, or “practice”, that the latter characteristic (convenience) takes precedence over the former (adequacy).

It has been noticed on various occasions that, from a mathematical standpoint, not every function can serve as a spatiotemporal dependence model. Certain formal *permissibility* criteria must be satisfied, which are based on the celebrated Bochner’s theorem of positive-definite functions. These criteria – which are valid for spatial and spatiotemporal dependence functions associated with ordinary, generalized, and fractal random fields – are discussed in detail in the relevant literature (Christakos 1984, 1992; Cassiani and Christakos 1998; Christakos and Hristopoulos 1998; Christakos et al. 2000; Kolovos et al. 2002). Numerous theoretical and applied studies derive and/or use dependence functions, the validity of which is essentially based on the criterion of Bocherian positive-definiteness (Yaglom 1986; Goodall and Mardia 1994; Jones and Zhang 1997; Ma 2003, 2008, 2009; Fernandez-Casal et al. 2003; Christakos et al. 2005; Stein 2005; Porcu et al. 2008).

It has been said that nothing in “fine print” is ever good news. And the readers should know that there is plenty in “fine print” linked to Bochner’s theorem. Certain

difficulties with the implementation of the Bochnerian permissibility criteria in practice were identified early on in the spatial analysis literature (Christakos 1984: 257). As it turns out, dependence functions that satisfy these criteria are not necessarily permissible for every random field, even if data analysis seems to associate the dependence function with this field. In the case of covariance functions, the relevant Bochnerian criterion merely guarantees that a Gaussian field exists with the corresponding positive-definite function as its covariance, but it does not necessarily imply that the covariance function is permissible for non-Gaussian fields. Spherical and cosine covariances, e.g., are compatible with the Gaussian law, but not necessarily with the Lognormal probability law. The gist of the whole business is then concentrated in the fact that the Bochnerian limitations briefly mentioned above have potentially serious consequences in real-world applications involving space–time analysis, attribute prediction, and risk assessment. Unfortunately, such facts are not always explicitly stated in the relevant literature, which makes one look like a character from Akira Kurosawa’s 1950 film *Rashomon* who prefers to live a lie than admit the truth. Truth was buried in *Rashomon* because no one could handle it.

Plato maintained that, “Serious things cannot be understood without laughable things.” Which brings us to the curious phenomenon of the so-called “Hamlets of geostatistics.” One cannot avoid noticing the unique reasoning style of certain studies characterized by their use of logically inconsistent arguments, and a profound misunderstanding of the theory’s fundamental principles. A characteristic example is the paper “To be or not to be. . . stationary? That is the question” (Myers 1989). Despite the paper’s title, its author was presumably unaware of Hamlet’s misfortunes in the Shakespearean play. *Mutatis mutandis*, the paper’s content is almost as troublesome as was Hamlet’s situation. Confusion is caused by statements like, “stationarity is not scale related” and “weakly stationary with drift,” which involve conflating concepts that need to be distinguished. Incorrect statements that are assumed to be generally applicable include, “variograms are generalized covariances (with a change in sign);” and epistemic notions are mingled with ontic ones (e.g., datasets that are samples from random field realizations vs. datasets that can be represented as a random field realization). There are several contradictory statements, e.g., “stationarity is a property of the random function, not of the data,” and a few lines later, “data sets with apparent non-stationarities.” What the author also did not know is that in Shakespeare’s original play, “when troubles come, they come not as single spies but in battalions.”⁴³ So, conspicuous inaccuracies include statements like, “in Bayesian maximum entropy, it is the posterior distribution for which the entropy is maximized,” and that spatial statistics “could be interpreted as including both geostatistics and stochastic modeling” (Myers 2006). In addition, the goal to achieve by definition what one could not achieve by logic or knowledge led to the now famous *locus classicus*: “Generalized functions, i.e., any function that is zero except at one point has a zero integral” (Myers 1993: 408). Laurent Schwartz might turn in his grave if he knew of the above

⁴³ Shakespeare’s *Hamlet* (1603, Act IV, Scene V).

Shakespearean adventures in the mathematical field he pioneered many decades ago. The reason that the above examples are mentioned in this section is instructive: to show why scientists should be prepared to be taught things they know already by “experts” who do not know them. For these self-appointed “experts,” understanding an entity is routinely based on the confusion between the name of the entity and the entity itself. This leads to nothing less than a gross perversion of technical notions, which are also irrelevant to the issues the “experts” profess to study.